Università degli Studi di Padova

Sede amministrativa: Padova

Dipartimento di Fisica e Astronomia "G. Galilei"

Scuola di Dottorato di Ricerca
in Fisica

Ciclo XXV

Tesi di Dottorato:

Search for the Supersymmetric Higgs boson in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel with the CMS detector at LHC.

Direttore della Scuola

Prof. Andrea Vitturi

Supervisore

Prof. Ugo Gasparini

Dottorando

Antonio Branca

Co-Supervisore

Dott. Stefano Lacaprara

Riassunto

In questa tesi è discussa l'analisi sviluppata per la ricerca del bosone di Higgs con le caratteristiche previste dall'estensione minimale del modello standard alle super-simmetrie (MSSM). La ricerca del segnale è effettuata nel canale in cui il bosone di Higgs è prodotto in associazione con uno o due b-quark e successivamente decade in due b-quark. I dati utilizzati a tal proposito sono stati acquisiti durante la presa dati del 2011 all'esperimento *Compact Muon Solenoid* (CMS) installato al *Large Hadron Collider* (LHC) del CERN di Ginevra, Svizzera. La luminosità integrata a cui corrisponde il campione di dati analizzato è pari a $\mathscr{L} = 4.8\ fb^{-1}$. Diversi trigger sono stati sviluppati per selezionare gli eventi compatibili con lo stato finale atteso prodotto dal canale considerato. Essi sono stati implementati a seconda del periodo per far fronte all'aumento della luminosità istantanea con cui sono state fornite le collisioni *pp* da LHC nel corso del 2011.
Il fondo al processo considerato è stato stimato tramite due metodi sviluppati a tal fine usando dati acquisiti dall'esperimento CMS. Le incertezze sistematiche sulla forma del fondo e sulla sua normalizzazione sono state determinate. La modelizzazione del segnale prodotto dal bosone di Higgs nel canale considerato, per diverse ipotesi di massa, è stata ottenuta tramite simulazione con metodo Monte Carlo. Anche in questo caso le incertezze sistematiche, dovute ai diversi contributi individuati, sono state determinate. La ricerca di un possibile contributo di segnale nei dati è stata effettuata eseguendo un fit ai dati del fondo stimato e del segnale simulato, tenendo conto delle rispettive incertezze statistiche e sistematiche. I fit risultano compatibili con il fondo atteso entro le incertezze. Nessun segnale viene evidenziato nel campione di dati analizzato entro i limiti di sensibilità dell'analisi.
I limiti di sensibilità dell'analisi sono stati calcolati nell'ipotesi di solo fondo nei dati. Tali limiti sono riportati in termini di valori della sezione d'urto di produzione associata dell'Higgs con una coppia di b-quark per la probabilità di decadimento in due b-quark. Questi valori sono stati inoltre interpretati in termini dei parametri del modello MSSM.
I risultati dell'analisi presentata in questo lavoro sono stati combinati con quelli ottenuti dall'analisi effettuata dal gruppo CMS di DESY (Amburgo, Germania) che, attraverso un'altra strategia, ha effettuato la ricerca del bosone di Higgs nello stesso canale. I limiti combinati delle due analisi migliorano la sensibilità dell'esperimento CMS ad un segnale prodotto da un bosone di Higgs nel canale considerato rispetto alle singole analisi.

Summary

In this thesis the analysis performed for the search of the Higgs boson predicted by the Minimal Super-symmetric Standard Model (MSSM) is presented. The process on which the search is grounded is the Higgs boson production in association with b-quarks with the subsequent decay in a b-quark pair. The analyzed data sample has been recorded by the Compact Muon Solenoid (CMS) experiment at the CERN Large Hadron Collider (LHC) in Geneva, Switzerland, during the 2011 data-taking. The data sample consists of a total integrated luminosity of $\mathscr{L} = 4.8\ fb^{-1}$, recorded with different trigger paths along the 2011 LHC operations. The trigger have been updated during the data-taking in order to cope with the increasing pp collisions instantaneous luminosity provided by the LHC. The background to the signal events has been estimated through two data-driven methods. The systematic uncertainties related to the background have been estimated for both the shape and normalization. The signal is modeled through Monte Carlo simulation technique for different hypothesis of the Higgs boson mass. The systematic uncertainties related to the signal model, from different sources, have been estimated. A possible signal contribution in the analyzed data sample is investigated through a fit to the data using the background templates estimated from the data-driven methods and the signal model from Monte Carlo simulation. The systematic uncertainties, as well as the statistics uncertainties, for both the signal and background have been taken into account in the fit. The results are compatible with the expected background within the statistics and systematic uncertainties. No signal is found in the data within the sensitivity of the analysis.
The sensitivity of the analysis has been calculated in the background only hypothesis. This is reported as upper limits on the cross section of the Higgs boson produced in association with two b-quarks times the branching ratio of the decay into a b-quark pair. This values are also interpreted as upper limits on parameters of the MSSM model.
The results of the analysis discussed in this thesis have been combined with those obtained by the DESY (Hamburg, Germany) CMS group, exploiting a different analysis strategy for searching the Higgs boson in the same channel. The combined upper limits from the two analysis improve the sensitivity of the CMS experiment to a Higgs boson signal produced through the tested channel with respect to the single analysis.

Contents

Introduction

The Large Hadron Collider (LHC) built at CERN (Geneva, Switzerland) is the biggest hadron collider available nowadays, which can provide proton-proton collisions at a nominal center of mass energy of 14 TeV and instantaneous luminosity of $\mathscr{L} = 10^{34} cm^{-2} s^{-1}$. It started its commissioning during the last months of 2009, providing proton-proton collisions at the center of mass energy of $\sqrt{s} = 900\ GeV$ by the end of November. During 2010-11 and 2012 LHC provided proton-proton collisions for physics analysis at the center of mass energy of $\sqrt{s} = 7\ TeV$ and $\sqrt{s} = 8\ TeV$, respectively.
Different physics goals justifying this big project are in the LHC agenda. Amongst them, there is the purpose of shedding light on the mechanism responsible for the Electroweak symmetry breaking (EWSB), that in the Standard Model (SM) of the elementary particles is achieved through the Higgs mechanism. It predicts the existence of a neutral scalar particle, the Higgs boson, whose mass is not predicted by the model. After about two years of LHC operations, the discovery of a boson compatible with the Higgs particle predicted by the Standard Model has been announced to the scientific community on July 2012. Nevertheless, its characteristics have still to be verified experimentally, and it might turn out that some properties of the new particle are not the one predicted by the SM.
Even if the SM is tested at the level of the per mil, thanks to the high precision measurements performed during the LEP period and at the Tevatron experiments, and the missing Higgs boson might be the particle observed by the LHC experiments, the SM is not a complete theory. From the experimental point of view, there are different observations that lead to this conclusion. For instance, the observation of the flavor oscillation of the neutrinos observed by the neutrino experiments is the evidence for a non-zero neutrino mass, which is not contemplated in the SM. Cosmological measurements show that the Universe is mostly made of components that are not described by the SM, called Dark Matter and Dark Energy. Moreover, from a theoretical point of view the SM Higgs receives radiative corrections to its mass that bring it to unphysical values.
It is for these reasons that the LHC is also searching for physics signatures beyond the SM, like those predicted by the supersymmetric theories. In the minimal supersymmetric extension of the Standard Model (MSSM), a new (fermion) boson is introduced for each SM (boson) fermion. The MSSM provides a mechanism that stabilizes the Higgs boson mass, since the radiative corrections given by the new MSSM particles cancel the SM ones. It gives also a good candidate for the Dark Matter, predicting the existence of a massive stable neutral particle with very weak interactions, known as Lightest Supersymmetric Particle (LSP). The Higgs sector of the theory is complicated by the presence of 5 scalar particles: 2 neutral CP-even, 1 neutral CP-odd and 2 charged. The particle observed at LHC is still compatible with either the light or the heavy CP-even MSSM neutral

Higgs boson.
In this thesis, the search for the neutral MSSM Higgs boson has been considered, in the mechanism where the Higgs is produced in association with b-quarks and decays into a pair of b-quarks. This production mechanism has negligible importance in the SM, while in the MSSM can have large production cross section under certain conditions. Chapter 1 is devoted to the theoretical description of the MSSM, focusing on the motivations and the Higgs phenomenology at LHC relevant for this work. Chapter 2 gives a brief introduction to the LHC with the actual parameters adopted during 2010-12 operations, and a quick overview of the experiments installed along its ring. In Chapter 3 the Compact Muon Solenoid (CMS) experiment is described in detail, giving the useful notations for the following analysis description. Chapter 4 gives an overview of the software tools used by the CMS collaboration to handle the experimental data, and of the physics objects defined to exploit the physics analysis. In Chapter 5 the analysis strategy is presented, giving the results in Chapter 6. The combination of the results with those obtained by the DESY CMS group analysis on the same channel are presented in Chapter 7. Finally, a summary and the conclusions of the work are given at the end of the thesis.

Chapter 1

Supersymmetry and minimal supersymmetric extension of the Standard Model

1.1 Introduction

The description of the fundamental components of matter and their interactions are nowadays very well described by the Standard Model of Particle Physics (SM). This is a model grounded on symmetries. As the Noether's theorem states [1], a system described by a Lagrangian with a given symmetry has a corresponding quantity whose value is conserved in time. A basic requirement for each reasonable relativistic physics theory is that the corresponding model Lagrangian must be symmetric under transformation of the Poincaré group elements. This implies the momentum, energy and angular momentum conservation laws [2]. The Lagrangian can also be constructed symmetric with respect to other symmetry transformations, which realize in an outstanding way what we observe in Nature. For instance, specifying the group of symmetry for the theory to be $U(1)$, we are requiring the Lagrangian to be invariant, meaning the corresponding equation of motion is covariant, under a transformation of the fields given by:

$$\psi^{'}(x) \;=\; e^{-ie\alpha}\psi(x) \tag{1.1}$$

Requiring the group parameter α to be dependent on the space-time coordinates ($\alpha(x)$), meaning that $U(1)$ is a Lie group, the transformations are promoted to local gauge transformations. To preserve the Lagrangian symmetry with respect to $U(1)$ local gauge transformations, a massless vector field is introduced in the model. This theory describes the electromagnetic interactions, where the new vector field introduced describes the photon [2]. The Noether conserved quantity corresponding to the $U(1)$ symmetry is the electromagnetic current.
By fixing the group of local gauge symmetry the form of interactions is uniquely specified. A complete theory which describe the experimental observations, the SM, is obtained choosing the group $SU(3) \times SU(2) \times U(1)$. In this way strong ($SU(3)$), weak ($SU(2)$), and electromagnetic ($U(1)$) interactions are described in the model. The strength of each interaction is defined by three independent coupling constants, one for each force, coming from the three factors of the SM group. Specifying the representation of the gauge group to which the particle belong, fix the three coupling constants. These would be the only parameters of

an ideal theory.
From the different experiments performed up to now, we can see that electromagnetic, strong and weak interactions of quarks and leptons are very well described by three coupling constants. Moreover, the prediction of the model, given the strength of the interactions, has been tested experimentally at the level of the per-mil [3].

1.2 Why overcome the Standard Model?

The electroweak part of the Lagrangian of the SM is [2]:

$$\begin{aligned}\mathscr{L} =&\bar{q}i\not{D}q + \bar{l}i\not{D}l - \frac{1}{4}(F^a_{\mu\nu})^2 \\ &+ |\not{D}_\mu\phi|^2 - V(\phi) \\ &- (\lambda^{ij}_u \bar{u}^i_R \phi \cdot q^j_L + \lambda^{ij}_d \bar{d}^i_R \phi^* \cdot q^j_L + \lambda^{ij}_l \bar{e}^i_R \phi^* \cdot l^j_L + h.c.)\end{aligned} \tag{1.2}$$

where quarks and leptons are described by spinor fields q and l, which are $SU(2)$ (singlets) doublets for the (right) left-handed components. u and d indicate the up and bottom component of the quarks doublet, while e is a charged lepton. The i, j indexes runs over the three number of quark and lepton families. The $F^a_{\mu\nu}$ tensor in 1.2 is the generalization of the electromagnetic tensor for the electroweak interactions and the symbol $\not{D}$ indicates covariant derivatives which implement the local gauge invariance of the Lagrangian. The first line is the pure gauge theory obtained just requiring the theory to be invariant under local gauge transformation of the SM group. This depends only on the two coupling constants that describe the electromagnetic and weak interactions for quarks and leptons.
The $SU(2) \times U(1)$ symmetry of 1.2 does not allow to introduce mass terms for quarks and leptons in a simple way[1]. The problem of the masses in the SM is overcome by means of a scalar field ϕ, the Higgs boson, introduced in the second line of 1.2, which is a doublet of weak interaction $SU(2)$. To this field is associated a potential energy $V(\phi)$, with at least two new parameters (see equation 1.4 below). The vacuum expectation value (v.e.v.) of the Higgs boson field, that is the value of the scalar field ϕ minimizing the potential, breaks down the SM symmetry [2] (known as Electro-Weak (EWK) symmetry breaking) giving masses to the W and Z bosons through the Higgs mechanism [2]. The introduction of the Higgs field allows to write down the invariant terms in the third line of 1.2 (known as Yukawa terms), which are trilinear terms linking a left and right-handed pair of quarks or leptons to the Higgs field. With the EWK breaking the Yukawa terms turns into masses and interaction with the Higgs field for the quark and leptons.
The SM symmetry allows three complex 3×3 matrices of couplings, the parameters λ^{ij} in 1.2. These become the mass matrices for quarks and leptons when the Higgs field acquires v.e.v. In the SM then there are three λ^{ij} matrices, and the parameters of the Higgs potential $V(\phi)$, which are not determined by the theory. In order to understand why the masses of the quarks, the leptons, and the W

[1] In the electromagnetic model a mass term of the type $m_\psi \bar{\psi}\psi$ invariant under $U(1)$ local gauge transformations can be written in the Lagrangian.

and Z bosons have their observed values, a deeper theory beyond the SM has to be developed.

1.2.1 The mechanism of electroweak symmetry breaking

The part of the SM group that implies electroweak interactions is $SU(2) \times U(1)$. Writing the Lagrangian 1.2 requires the choosing of a representation for the SM group, which for the EWK part is the fundamental one for $SU(2)$, given by the three components of the *weak isospin* operator, $\tau^a = \sigma^a/2$ (with σ^a the Pauli's matrices), and the *weak hypercharge* Y operator for $U(1)$. The fields of the theory, as far as the EWK part is concerned, transform with quantum numbers I, eigenvalue of the third component τ^3 of the isospin, and Y, eigenvalue of the hypercharge operator. In the SM we introduce the doublet $SU(2)$ scalar field, the Higgs doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.3}$$

transforming in $SU(2) \times U(1)$ with quantum numbers $I = \frac{1}{2}$ and $Y = \frac{1}{2}$. Taking into account the invariance under transformations of the SM group, the only terms of the potential energy associated to the Higgs field are[2]:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.4}$$

with $-\mu^2 < 0$ and $\lambda > 0$. The potential is minimized when $\phi^\dagger \phi = \mu^2/2\lambda$. This minimum is not found for a single value of ϕ, but rather for a set of values, as it is shown in Figure 1.1.

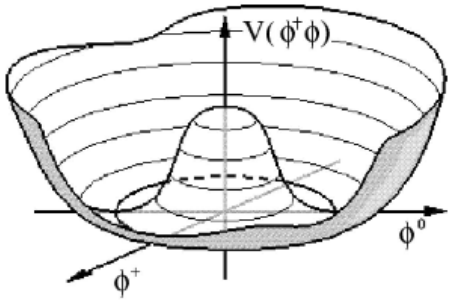


Figure 1.1: Shape of the SM Higgs potential.

One particular vacuum state is given by:

$$< \phi >= \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} v \end{pmatrix} \tag{1.5}$$

where $v^2 = \mu^2/\lambda$. This choice is arbitrary and the chosen point is not invariant under rotations in the (ϕ^+, ϕ^0) plane: this is the spontaneous symmetry breaking of the EWK symmetry.
The Higgs field can be parametrized in terms of perturbations developed around the vacuum state 1.5 as:

[2] In order to make the SM theory renormalizable, each term in the Lagrangian can have mass dimension up to 4.

$$< \phi >= e^{i\alpha(x)\cdot\tau} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + h(x)) \end{pmatrix} \tag{1.6}$$

where $\alpha(x)$ parametrize a gauge fluctuations of the Higgs field along the minimum values and $h(x)$ a gauge fluctuation perpendicular to the minimum values. The Lagrangian 1.2 can be expanded around the minimum 1.5 by using 1.6. In a particular gauge of 1.6 the three components of the Higgs field $\alpha(x)$ are cancelled and the vector boson fields W and Z acquire mass (Higgs mechanism). If g and $g^{'}$ are the $SU(2) \times U(1)$ coupling constants, the mass of W and Z have the expression:

$$m_W = g\frac{v}{2}, \qquad\qquad m_Z = \sqrt{g^2 + g'^2}\frac{v}{2} \tag{1.7}$$

The measured values of the masses and couplings then lead to:

$$v = 246\ GeV \tag{1.8}$$

which is the general scale of EWK symmetry breaking.
One component of the Higgs field remains physical, that is $h(x)$, with mass given by:

$$m_h^2 = 2\mu^2 = 2\lambda v^2 \tag{1.9}$$

The third term of 1.2 gives the mass and Higgs interactions terms of fermions. The fermion mass can be expressed as:

$$m_f = \frac{y_f}{\sqrt{2}}v \tag{1.10}$$

where y_f is the Yukawa constant appearing in the Yukawa terms after EWK symmetry breaking.
But which is the physical reason that implies the spontaneous symmetry breaking of EWK? The way the EWK breaking is realized in the SM does not give an answer to this question. In this model the spontaneous symmetry breaking is realized just because $-\mu^2 < 0$. This does not explain the physical mechanism behind the EWK breaking. Moreover, the SM shows problems in some theoretical and cosmological aspects which we briefly review in the following.

1.2.2 The problem of the Higgs mass divergence

Direct searches of the SM Higgs boson at LEP e^+e^- collider and Tevatron $p\bar{p}$ collider have led, respectively, to a lower-mass bound of $m_h > 114.4\ GeV$ [4], and to an exclusion in the range $162 - 166\ GeV$ [5], at 95% of CL. Indirect constraints from precision measurements favour the mass range $m_h < 158\ GeV$ [6, 7] at 95% of CL. More recently, searches performed at the LHC in the years 2010-11 put more stringent limits on the SM Higgs particle mass [8] and finally a boson with the signatures compatible with a SM Higgs boson has been observed with a mass around 125 GeV [9, 10]. The problem in the SM is that the Higgs boson mass receives enormous quantum corrections from the virtual effects of every particle that couples, directly or indirectly, to the Higgs field; drawing the value of the mass away from the experimental results.

For instance, the one-loop quantum correction to the Higgs mass in Figure 1.2 is due to the Lagrangian term $-\lambda_f h \bar{f}_R f_L$, coupling the Higgs to a fermion, obtained after inserting 1.6 in the third line of 1.2. The fermion can be each of the leptons and quarks of the SM, even if the largest correction comes when the fermion is the top quark. In fact, the top quark has the strongest Yukawa coupling to the Higgs boson, with $\lambda_f \sim 1$.

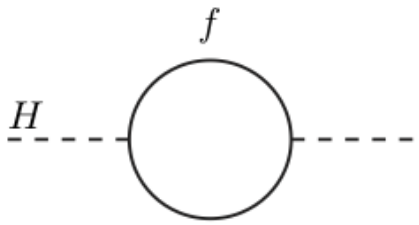


Figure 1.2: One-loop correction to the Higgs mass due to a fermion.

The Feynman diagram 1.2 yields to a Higgs mass correction factor:

$$\Delta m_h^2 = -\frac{|\lambda_f|^2}{16\pi^2}\Lambda_{UV}^2 + ... \tag{1.11}$$

where Λ_{UV} is an ultraviolet cut-off used to regulate the loop integral. The ellipses stands for terms proportional to m_f^2 growing at most logarithmically with Λ_{UV}. This cut-off has to be interpreted as the energy scale at which the SM is not reliable anymore, and new physics appears to be not negligible. Certainly new physics will be required at the Plank scale:

$$M_p = \sqrt{8\pi G_{Newton}} = 2.4 \cdot 10^{18}\ GeV \tag{1.12}$$

where quantum gravitational effects, not described by the SM theory, become important. With Λ_{UV} of the order of M_p the quantum correction to the squared Higgs mass is about 30 order of magnitude larger than the value the experiments suggest, $m_h^2 \sim 100^2\ GeV^2$. The problem of the Higgs mass divergence due to quantum corrections is also known as the *hierarchy problem* [11].

1.2.3 The evolution of couplings

The couplings of the electromagnetic, weak and strong interactions receive quantum corrections from next-to-leading order processes [2]. For instance, the scattering of two fermions through electromagnetic interaction is the sum of all possible Feynman diagrams shown in Figure 1.3.

The values of the couplings depends on the momentum transfer Q of the scattering process involving interactions. Variations of the coupling values with respect to Q are described by a set of differential equations, called the *renormalization group equation* RG [2]:

$$Q\frac{\partial}{\partial Q}\alpha(Q) = \beta_a(\alpha_a(Q)) \tag{1.13}$$

where $\alpha_a = g_a^2/4\pi\ ,\ a = 1, 2, 3$, with $g_1 \equiv g_s, g_2 \equiv g, g_3 \equiv g'$ the couplings of strong, weak and electromagnetic interactions, respectively.
In 1.13 the function $\beta_a(\alpha_a(Q))$ is computable perturbatively, leading to:

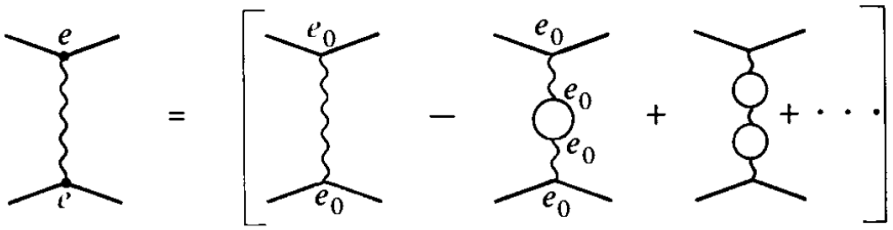


Figure 1.3: Next-to-leading order processes contributing to a fermion-fermion electromagnetic scattering. Ellipses stand for diagrams with all possible propagator modification. The charge e is the electric charge measured experimentally and is related to the electromagnetic coupling through the relation $g^{'} = e^2/4\pi$. The charge e^0 is known as the *bare* charge, referring to a vertex without loops.

$$\beta_a(\alpha_a(Q)) = \frac{b_a}{2\pi}\alpha^2(Q) + O(\alpha^3) \tag{1.14}$$

stopping the calculation at 1-loop. In 1.14 the b_a are constants which depends on the gauge group and on the particle multiplets to which the gauge bosons couple. For $SU(N)$ with particles in the fundamental representation:

$$b_{a,a\equiv 1,2} = \left(\frac{11}{3}N - \frac{1}{3}n_f - \frac{1}{6}n_s\right) \tag{1.15}$$

where n_f is the number of left-handed fermions and n_s is the number of complex scalars which couple to the gauge boson. For $U(1)$ the corresponding formula is:

$$b_3 = -\frac{2}{3}\sum_f t_f^2 - \frac{1}{3}\sum_s t_s^2 \tag{1.16}$$

where t is the particle charges. In the SM, the $U(1)$ coupling constant $g^{'}$ and the $SU(2)$ and $SU(3)$ couplings g and g_s evolve with Q according to the RG equation 1.13 with:

$$\begin{aligned} b_1 &= 11 - \frac{4}{3}n_g \\ b_2 &= \frac{22}{3} - \frac{4}{3}n_g - \frac{1}{6}n_h \\ b_3 &= -\frac{4}{3} \end{aligned} \tag{1.17}$$

In this formula, n_g is the number of quark and lepton generations and n_h is the number of Higgs doublet fields. The solution to 1.13 can be written, in terms of the measured coupling constants at $Q = m_Z$, as:

$$\alpha_a(Q) = \frac{\alpha_a(m_Z)}{1 + \frac{b_a}{8\pi} log\frac{Q}{m_Z}} \tag{1.18}$$

The evolution of couplings predicted by 1.17 and 1.18, with $n_h = 1$, is shown in Figure 1.4 [2].

As the momentum transfer Q increases, the values of the couplings converge. Specifically, the strong interaction coupling decrease, making it weaker at larger

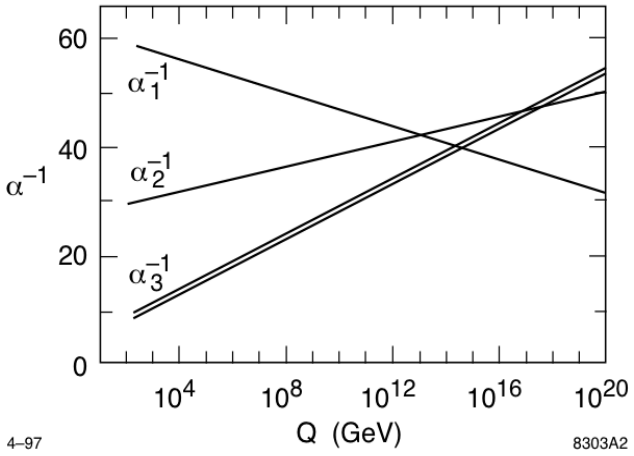


Figure 1.4: Evolution of the $SU(3)\times SU(2)\times U(1)$ gauge couplings as a function of the momentum transfer Q of the scattering process. Prediction obtained using the 1-loop RG equations of the SM. The double line for α_3 is the current experimental error in this quantity, not visible for α_1 and α_2 since it is too small.

values of Q with the possibility to compute the cross section of a strong interaction process perturbatively. This is the remarkable phenomenon of *asymptotic freedom*. Although the values of the coupling tend to converge, suggesting that they might become unified at some energy scale [12], they do not come to a unique common value at any scale. As we will see, this might be natural in theories beyond the SM.

1.2.4 Baryogenesis

From the everyday life it is evident that the world is composed only of matter and no antimatter is present. This is true in a small scale, but also moving up in larger scale in the universe. Specifically, the largest scale on which we can say that there is no antimatter goes up to the Hubble size [13]. It therefore seems that the universe is fundamentally matter-antimatter asymmetric.
A strict quantitative estimate of the baryonic matter and antimatter can be obtained from the standard cosmology. The baryon asymmetry parameter is defined as:

$$\eta = \frac{n_b - n_{\bar{b}}}{s} \tag{1.19}$$

with n_b and $n_{\bar{b}}$ the density of baryons and anti-baryons in the universe, respectively; s is the entropy density in the universe. From abundances of the light elements in the universe, H, ^3He, ^4He, D, B and ^7Li, accurately predicted by the primordial nucleosynthesis, and measurements of the Cosmic Microwave Background (CMB) fluctuations, the possible range of η at 95% of CL is [14]:

$$5.1 \times 10^{-10} \leq \eta \leq 6.5 \times 10^{-10} \tag{1.20}$$

There are reasonable arguments to think that the universe started symmetric in matter and antimatter, meaning with $\eta = 0$ [15]. The generation of the observed value of η starting from a symmetric initial condition is referred to as

baryogenesis. A theory explaining this phenomenon has to satisfy some conditions, first identified by Sakharov [16] and now known as the three *Sakharov Criteria*. They are:

1. Violation of the baryon number (B) symmetry;
2. Violation of the discrete symmetry C (charge conjugation) and CP (composition with parity P and C);
3. Departure from the thermal equilibrium;

The first of these is rather obvious. If no processes ever occur in which B is violated, then the total number of baryons in the universe must remain constant, and therefore no asymmetry can be generated from symmetric initial conditions. The second criterion is required because, if C and CP are exact symmetries, then one can prove that the total rate for any process which produces an excess of baryons is equal to the rate of the complementary process which produces an excess of anti-baryons and so no net baryon number can be created. The third criterion is required because otherwise precesses that generate an excess of baryons/anti-baryons would be compensated by inverse processes. In thermal equilibrium processes creating the excess and the inverse one would have the same velocity.
The SM has all the required criteria for the generation of the baryon asymmetry: baryon number violating processes, CP violation and non-equilibrium processes. Nevertheless, quantitatively the SM does not explain the value 1.20 of η. First of all the CP violation is too small [17–19], and second, the departure from the thermal equilibrium would require a Higgs with mass $m_h \lesssim 80\ GeV$ [20], which is not the case. Eventually, the predicted value of 1.19 within the SM is $\eta_{SM} \ll 10^{-10}$; well bellow the estimated range 1.20. The explanation of 1.20 then requires a theory beyond the SM. All the above problematic aspects of the SM find a quite natural and elegant solution in a new theory which displays an additional symmetry between fermions and bosons, called Supersymmetry.

1.3 Supersymmetry

A symmetry relating fermions and bosons, *supersymmetry* (for a comprehensive textbook on Supersymmetry, see e.g. [21]), is a transformation that transforms a bosonic state into a fermionic state, and vice versa:

$$Q|Boson> = |Ferion>, \qquad Q|Fermion> = |Boson> \tag{1.21}$$

The operator Q generating these transformation has then fermionic character: it can be chosen to have the transformation properties of a left-handed Weyl spinor, a $(\frac{1}{2}, 0)$ representation under Lorentz transformations. Since the supersymmetric operators are fermionic, carrying spin angular momentum 1/2, they must be space-time symmetry. Being fermionic the supersymmetric operator obey anticommutation relations. The Haag-Lopuszanski-Sohnius extension [22] of the Coleman-Mandula theorem [23], implies that the operator Q and its hermitian conjugate $\bar{Q}$ must satisfy an algebra of anticommutation and commutation relations with the form:

$$\begin{aligned}\{Q_\alpha, \bar{Q}_{\dot{\beta}}\} &= 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu \\ \{Q_\alpha, Q_{\dot{\beta}}\} &= \{\bar{Q}_\alpha, \bar{Q}_{\dot{\beta}}\} = 0 \\ [Q_\alpha, P^\mu] &= 0\end{aligned} \tag{1.22}$$

with P^μ the four-momentum generator of space-time translations. The states of a single-particle, called *supermultiplet*, can be realized in the irreducible representation of the algebra 1.22. Each super-multiplet contains both fermion and boson states, which are commonly known as *superpartnes* of each other. Given 1.22, the squared mass operator $-P^2$ commutes with the operators Q, $\bar{Q}$, then particles in the same supermultiplet must have equal masses. Supersymmetry operators Q and $\bar{Q}$ commute also with the generators of gauge transformations, therefore particles in the same super-multiplet must have same gauge charges (electrical charge, weak isospin, color). Finally, it can be shown that each supermultiplet contains an equal number of fermion and boson degrees of freedom.

1.3.1 Superspace and superfields

An elegant and compact description of the supersymmetric algebra representations can be obtained introducing the notion of *superspace* and *superfields*. In this formalism supersymmetric models can be constructed in such a way that supersymmetry is manifest.
A generic element of the algebra defined by 1.22 with operators P and Q is:

$$g(x, \theta, \bar{\theta}) = e^{i[x^\mu P_\mu + i(\theta^\alpha Q_\alpha + \bar{\theta}^{\dot{\alpha}} \bar{Q}_{\dot{\alpha}})]} \tag{1.23}$$

This is a Lie group where each element can be mapped by a set of coordinates:

$$\{x^\mu, \theta^\alpha, \bar{\theta}^{\dot{\alpha}}\} \tag{1.24}$$

which identifies the so called *superspace*, an extension of the space-time obtained by introducing the anti-commutating Grassmann variables θ and $\bar{\theta}$. In this superspace, supersymmetric transformations associated to the Q_α operators and generalized covariant derivatives D_α are defined (for details, see [21]).

1.3.2 Supersymmetric Lagrangian

It can be demonstrated that the highest term of a scalar and vector field expanded in powers of the Grassmann variables transforms under supersymmetry transformations as a four-divergence in the space-time. For a general superfield the highest term in the expansion is the $(\theta\theta)(\bar{\theta}\bar{\theta})$, called D-term, whereas for a chiral superfield it is $(\theta\theta)$ (or $(\bar{\theta}\bar{\theta})$), called F-term. Taking into account that a product of (chiral) superfields are still (chiral) superfields, the most general Lagrangian for a supersymmetric theory can be written as follows:

$$L = \int dx^4 [\int d\theta^4 \mathscr{L}_D + \int d\theta^2 \mathscr{L}_F] \tag{1.25}$$

where the Lagrangian densities $\mathscr{L}_D$ and $\mathscr{L}_F$ are sum of superfields and chiral-superfields, respectively. The Lagrangian L is invariant under supersymmetric transformations, since in the $d\theta^4$ integration survives only D-terms, whereas in

the $d\theta^2$ integration survives only F-terms.
Analogously with what happen in the electromagnetism, taking into account 1.25 a Lagrangian for the superfield strength W can be written as:

$$L = \int dx^4 \int d\theta^2 (W_\alpha W^\alpha + \bar{W}_{\dot{\alpha}} \bar{W}^{\dot{\alpha}}) \tag{1.26}$$

This Lagrangian leads to the equations of motion of a free spin 1 boson and a spin 1/2 gaugino.
The particles matter description, with interactions, is introduced by scalar superfields Lagrangian terms, and requiring gauge invariance under local transformations given by a group $\mathscr{G}$. A generic gauge transformation for a set of scalar chiral superfields is:

$$\begin{aligned} \Phi' &= e^{-i\Lambda}\Phi \\ \bar{\Phi}' &= \bar{\Phi} e^{-i\bar{\Lambda}} \end{aligned} \tag{1.27}$$

where:

$$\Lambda_{ij} = T^a_{ij} \Lambda_a \in \mathscr{G} \tag{1.28}$$

with T^a the generators of the gauge group $\mathscr{G}$, Φ ($\bar{\Phi}$) a vector of scalar left- (right-) handed superfields (with dimension equal to the rappresentation of the group), and Λ_a scalar chiral superfields. The most general vector superfield is:

$$V = T^a V_a \tag{1.29}$$

with V_a a generic vector superfield. The most general gauge transformations for 1.29 is:

$$e^{V'} = e^{-i\bar{\Lambda}} e^V e^{i\Lambda} \tag{1.30}$$

At this point, the most general Lagrangian invariant under supersymmetric transformations, taking into account 1.25, and under gauge transformations of the group $\mathscr{G}$ is:

$$\begin{aligned} L = \int dx^4 [& \int d\theta^2 (Tr(W_\alpha W^\alpha) + Tr(\bar{W}_{\dot{\alpha}} \bar{W}^{\dot{\alpha}})) \\ & + \int d\theta^4 \bar{\Phi} e^{n_l V} \Phi \\ & + \int d\theta^2 (m_{rs} \Phi^r \Phi^s + g_{rsk} \Phi^r \Phi^s \Phi^k + h.c.)] \end{aligned} \tag{1.31}$$

with $m_{rs}, g_{rsk} \neq 0$ only for terms invariant under gauge transformations. From this Lagrangian equations of motion for the fields contents of the superfields Φ ($\bar{\Phi}$) and V ($\bar{V}$) are obtained. These equations describe a set of spin 1/2 fermions with their supersymmetric partners, a set of spin 0 bosons (the sfermions) interacting through exchange of a set of spin 1 gauge bosons and their supersymmetric partners, a set of spin 1/2 fermions (the gauginos). All the interactions are produced by the term:

$$\mathscr{W} = m_{rs} \Phi^r \Phi^s + g_{rsk} \Phi^r \Phi^s \Phi^k \tag{1.32}$$

which is referred to as the *superpotential*. A scalar potential $\mathscr{V}$ can be derived for the Lagrangian 1.31. It can be shown that it is obtained from the superpotential 1.32 and terms involving the scalar fields of the theory [21].

1.4 The minimal supersymmetric extension of the Standard Model

The minimal extension of the SM to the supersymmetry is known as Minimal Supersymmetric Standard Model (MSSM). In the SM all left-handed fermions are $SU(2)$ doublets while right-handed fermions are $SU(2)$ singlets. This suggests to treat separately left and right handed parts of a fermion. Therefore, the description of fermions in the MSSM is achieved by introducing a scalar chiral superfield for each lepton and quark; the description of vector bosons is obtained instead by introducing a vector superfield for each of them. This brings to the first important consequence of the supersymmetric extension: in order to describe the particle content of the MSSM, we introduced a scalar boson, *sfermion*, for each fermion, and a fermion, *gaugino*, for each vector boson. A list of the supermultiplet corresponding to each superfield, describing the particle content of the MSSM, with their quantum numbers is shown in Table 1.1 and 1.2.

Table 1.1: Chiral supermultiplets content of the MSSM: spin 0 fields are complex scalars, spin 1/2 are Weyl spinors.

Names	Superfield	spin 0	spin 1/2	$SU(3)_c$, $SU(2)_L$, $U(1)_Y$
squarks, quarks (×3 families)	Q	$(\tilde{u}_L,\, \tilde{d}_L)$	$(u_L,\, d_L)$	$(\mathbf{3},\, \mathbf{2},\, \frac{1}{6})$
	$\bar{u}$	$\tilde{u}_R$	u_R	$(\bar{3},\, \mathbf{1},\, -\frac{2}{3})$
	$\bar{d}$	$\tilde{d}_R$	d_R	$(\bar{3},\, \mathbf{1},\, \frac{1}{3})$
sleptons, leptons (×3 families)	L	$(\tilde{\nu}_L,\, \tilde{e}_L)$	$(\nu_L,\, e_L)$	$(\mathbf{1},\, \mathbf{2},\, -\frac{1}{2})$
	$\bar{e}$	$\tilde{e}_R$	e_R	$(\mathbf{1},\, \mathbf{1},\, 1\ )$
Higgs, higgsinos	H_u	$(H_u^+,\, H_u^0)$	$(\tilde{H}_u^+,\, \tilde{H}_u^0)$	$(\mathbf{1},\, \mathbf{2},\, \frac{1}{2})$
	H_d	$(H_d^0,\, H_d^-)$	$(\tilde{H}_d^0,\, \tilde{H}_d^-)$	$(\mathbf{1},\, \mathbf{2},\, -\frac{1}{2})$

Table 1.2: Gauge supermultiplet content of the MSSM.

Names	spin 1/2	spin 1	$SU(3)_c$, $SU(2)_L$, $U(1)_Y$
gluinos, gluons	$\tilde{g}$	g	$(\mathbf{8},\, \mathbf{1},\, 0\ )$
winos, W bosons	$\tilde{W}^{\pm}$, $\tilde{W}^0$	$W^{\pm}$, W^0	$(\mathbf{1},\, \mathbf{3},\, 0\ )$
bino, B boson	$\tilde{B}$	B	$(\mathbf{1},\, \mathbf{1},\, 0\ )$

As it is shown in Table 1.1, the description of the Higgs doublet is naturally obtained by the scalar chiral superfield doublet H_u. However, with only one Higgs superfield, the electroweak gauge symmetry would suffer a gauge anomaly, and would be inconsistent as a quantum theory. Indeed, the anomaly cancellation condition in the electroweak gauge symmetry is:

$$Tr[T_3^2 Y] = Tr[Y^3] \tag{1.33}$$

where T_3 and Y are the third component of weak isospin and the weak hypercharge, respectively, in a normalization where the electric charge is $Q = T_3 + Y$. The trace runs over all the fermionic degrees of freedom of the theory. This condition can not be satisfied with one Higgs superfield with $Y = 1/2$ or $Y = -1/2$, and so, two Higgs superfields are required.
The interactions within the MSSM are specified by the superpotential 1.32 which, given the scalar chiral superfield in Table 1.1 and the invariance under gauge group $SU(3) \times SU(2) \times U(1)$ transformations, has the following expression:

$$\mathscr{W} = y_u^{ij} \bar{u}_i Q_j \cdot H_u - y_d^{ij} \bar{d}_i Q_j \cdot H_d - y_e^{ij} \bar{e}_i L_j \cdot H_d + \mu H_u \cdot H_d \tag{1.34}$$

where the y's are 3×3 matrices in family space, and are exactly the same Yukawa couplings as those which enter the SM, in the third line of 1.2. These couplings give masses to the quarks and leptons when the Higgs fields acquire vacuum expectation values.
The last term in 1.34, so called μ-term, is quadratic in the Higgs fields. The parameter μ that compare in this term is the only new parameter introduced by the supersymmetric extension of the SM. In the scalar potential $\mathscr{V}$ of the MSSM, the μ-term leads to the following term involving the Higgs fields:

$$\mathscr{V}(\mu) = |\mu^2|(|H_u^+|^2 + |H_d^-|^2 + |H_u^0|^2 + |H_d^0|^2) \tag{1.35}$$

This expression, similar to the one of the SM scalar potential $m^2\phi^\dagger\phi$, is non-negative with a minimum at $H_u^0 = H_d^0 = 0$. Terms analogous to the $\lambda(\phi^\dagger\phi)^2$, inducing the EWK symmetry breaking in the SM, are missing in the MSSM. Thus, the EWK symmetry breaking cannot be accomplished within the supersymmetric invariant Lagrangian of the MSSM. On the other hand, it is clear that supersymmetry must be a broken symmetry of MSSM, otherwise experiments should have observed the supersymmetric partners of fermions and bosons. In fact if supersymmetry were unbroken, particles within the same supermultiplet would have the same mass values. For instance, the *selectron* would have the same mass of its superpartner, the electron. As it will be shown, the EWK symmetry and supersymmetry breaking are related each other: terms allowing EWK symmetry breaking in MSSM are the one that breaks supersymmetry.

1.4.1 Solution of the hierarchy problem

In the MSSM Lagrangian are present terms $-\lambda_S|H|^2|S|^2$ that couples a complex scalar particle S to the Higgs boson. This term produce the one-loop quantum correction to the Higgs boson mass shown in Figure 1.5.

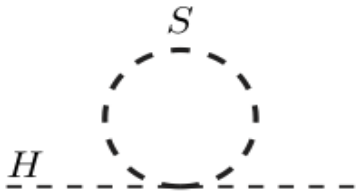


Figure 1.5: One-loop correction to the Higgs mass due to a scalar.

This Feynman diagram yields to a Higgs mass correction factor:

$$\Delta m_h^2 = \frac{\lambda_S}{16\pi^2}\Lambda_{UV}^2 + ... \tag{1.36}$$

The ellipses stands for terms proportional to m_f^2 growing at most logarithmically with Λ_{UV}. The coupling λ_S must be positive if the scalar potential is to be bounded from below.
Quadratic divergent contributions to the Higgs mass from fermions 1.11 and bosons 1.36 have opposite sign, and they will cancel since in supersymmetry $\lambda_S = |\lambda_f|^2$. This solves the hierarchy problem that occurs in the SM.

1.4.2 Unification of gauge couplings

Also for the MSSM the evolution of the $SU(3) \times SU(2) \times U(1)$ couplings are described by the RG equations 1.13. Supersymmetric extension for expressions 1.15 and 1.16 must be provided, since the particle content changed. For $SU(N)$ the gauginos give an additional contribution $-\frac{2}{3}N$:

$$b_{a,a\equiv 1,2} = \left(3N - \frac{1}{2}n_f\right) \tag{1.37}$$

where the contributions from fermions and scalars are put together into n_f, which is the number of chiral supermultiplets in the fundamental representation. For $U(1)$ the corresponding formula is:

$$b_3 = -\sum_f t_f^2 \tag{1.38}$$

where t is the fermion charge and the sum runs over chiral supermultiplets. These expressions evaluated for $SU(3) \times SU(2) \times U(1)$ with n_g quark and lepton generators and n_h Higgs fields, gives:

$$\begin{aligned} b_1 &= 9 - 2n_g \\ b_2 &= 6 - 2n_g - \frac{1}{2}n_h \\ b_3 &= -2n_g - \frac{3}{10}n_h \end{aligned} \tag{1.39}$$

The evolution of couplings predicted by 1.39 and 1.18, with $n_h = 2$, is shown in Figure 1.6 [2]: the gauge couplings unify at a scale $m_U \sim 2 \times 10^{16}\ GeV$, close to the Plank scale.

Supersymmetry seems to overcome the difficulty that the Standard Model has in linking in a simple way to grand unification.

1.4.3 A possible scenario for baryogenesis

In the MSSM new sources of CP violation are present [24, 25], which can overcome the smallness of CP violation of SM. Moreover, it as been shown [26, 27] that in the presence of a light top squark (stop), with mass lower than about 120 GeV, and a Higgs with mass below about 127 GeV, the departure from the thermal equilibrium can be realized.

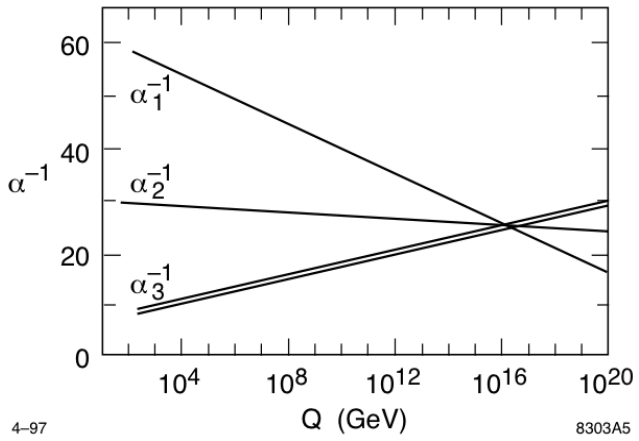


Figure 1.6: Evolution of the $SU(3) \times SU(2) \times U(1)$ gauge couplings as a function of the momentum transfer Q of the scattering process in the MSSM. Prediction obtained using the 1-loop RG equations of the SM. The double line for α_3 is the current experimental error in this quantity, not visible for α_1 and α_2 since it is too small.

1.4.4 Soft supersymmetry breaking

There are two ways in which a symmetry can be broken: either by introducing explicit symmetry-breaking terms in the Lagrangian, or by spontaneous symmetry breaking. The last means that the model should have a Lagrangian density that is invariant under a symmetry, but a vacuum state that is not. This is what actually happens for electroweak symmetry in the SM, and is preferred theoretically since otherwise the introduction of explicit mass terms for the fermions and massive gauge bosons would spoil renormalizability. Thus, by definition the supersymmetry is spontaneously broken if the vacuum state $|0\rangle$ is not invariant under supersymmetry transformations $Q_\alpha|0\rangle \neq 0$, $\bar{Q}_{\dot{\alpha}}|0\rangle \neq 0$. Many models of spontaneous symmetry breaking have been proposed. These always involve extending the MSSM to include new particles and interactions at very high mass scales, and there is no consensus on exactly how this should be done. From a practical point of view, it is extremely useful to follow the first way and just introduce extra terms that break supersymmetry explicitly in the effective MSSM Lagrangian. By introducing these terms, we must preserve the solution to the hierarchy problem, and avoid terms which would introduce divergent corrections to the Higgs mass. The complete set of terms breaking supersymmetry without compromising the hierarchy problem solution, called soft supersymmetric breaking terms, have been classified rigorously by Girardello and Grisaru [28]. They are the following:

$$\mathscr{L}_{soft} = -(\frac{1}{2}M_a\lambda^a\lambda^a + \frac{1}{6}a^{ijk}A_iA_jA_k + \frac{1}{2}b^{ij}A_iA_j + t^iA_i) + c.c. - (m^2)^i_j A^{j*}A_i \quad (1.40)$$

They consist of gaugino masses M_a for each gauge group, scalar squared-mass terms $(m^2)^{ji}$, bilinear and trilinear scalar coupling b^{ij} and a^{ijk}, respectively, and tadpole couplings t^i. For the MSSM 1.40 has the following expression:

$$
\begin{aligned}
\mathscr{L}_{soft}^{MSSM} = &-\frac{1}{2}(M_3\tilde{g}\tilde{g} + M_2\tilde{W}\tilde{W} + M_1\tilde{B}\tilde{B} + c.c.) \\
&- (\tilde{\bar{u}}\mathbf{a_u}\tilde{Q}H_u + \tilde{\bar{d}}\mathbf{a_d}\tilde{Q}H_d + \tilde{\bar{e}}\mathbf{a_e}\tilde{L}H_d + c.c.) \\
&- \tilde{Q}^\dagger\mathbf{m_Q^2}\tilde{Q} + \tilde{L}^\dagger\mathbf{m_L^2}\tilde{L} + \tilde{\bar{u}}\mathbf{m_{\bar{u}}^2}\tilde{\bar{u}}^\dagger + \tilde{\bar{d}}\mathbf{m_{\bar{d}}^2}\tilde{\bar{d}}^\dagger + \tilde{\bar{e}}\mathbf{m_{\bar{e}}^2}\tilde{\bar{e}}^\dagger \\
&+ m_{H_u}^2 H_u^* H_u + m_{H_d}^2 H_d^* H_d + (bH_uH_d + c.c.)
\end{aligned}
\tag{1.41}
$$

In the first line M_3, M_2 and M_1 are the gluino, wino and bino mass terms. The second line contains the trilinear a^{ijk} scalar couplings. Each of $\mathbf{a_u}$, $\mathbf{a_d}$, $\mathbf{a_e}$ is a complex 3×3 matrix in family space, with dimension of mass. They are in one-to-one correspondence with the Yukawa couplings of the superpotential. The third line consists of squark and slepton mass terms of the $(m^2)_i^j$ type. Each of $\mathbf{m_Q^2}$, $\mathbf{m_L^2}$, $\mathbf{m_{\bar{u}}^2}$, $\mathbf{m_{\bar{d}}^2}$, $\mathbf{m_{\bar{e}}^2}$ is a 3×3 matrix in family space that can have complex entries, but they must be hermitian, so that the Lagrangian is real. In the last line $m_{H_u}^2$ and $m_{H_d}^2$ are squared-mass terms of the $(m^2)_i^j$ type, while b is the only squared-mass term of the type b^{ij} that can occur in the MSSM. The expression 1.41 is the most general soft supersymmetry breaking Lagrangian of the form 1.40 that can be written for MSSM. It is important to note that 1.41 preserve the $SU(3) \times SU(2) \times U(1)$ gauge invariance. An explicit breaking of this symmetry would generate unitarity violation.
Unlike the supersymmetry-preserving part of the Lagrangian, 1.41 introduces many new parameters that were not present in the SM. A careful count [29] reveals that there are 105 masses, phases and mixing angles in the MSSM Lagrangian that cannot be rotated away by redefining the phases and flavor basis for the quark and lepton supermultiplets, and that have no counterpart in the SM. Thus, in principle, supersymmetry breaking appears to introduce a tremendous arbitrariness in the Lagrangian. The reason for having to deal with so many parameters is that although the way to parametrize broken supersymmetric theories is known very well, it is not known how the symmetry is actually broken. So supersymmetry is not just a model, it is rather a class of models, each scenario differing from the others by the way the parameters are related among themselves. Once we subscribe to any given supersymmetry breaking mechanism, e.g. supergravity, anomaly mediation, gauge mediation, gaugino mediation, and so on, the number of independent parameters gets drastically reduced.

1.4.5 Electroweak symmetry breaking and the Higgs boson

The scalar potential involving only the Higgs scalar fields in the MSSM is obtained using the superpotential expression 1.34 completed with the soft supersymmetry breaking terms 1.41:

$$
\begin{aligned}
\mathscr{V}(H_{u,d}^0, H_{u,d}^\pm) = &(|\mu|^2 + m_{H_u}^2)(|H_u^0|^2 + |H_u^+|^2) \\
&+ (|\mu|^2 + m_{H_d}^2)(|H_d^0|^2 + |H_d^-|^2) \\
&+ [b(H_u^+ H_d^- - H_u^0 H_d^0) + c.c.] \\
&+ \frac{1}{8}(g^2 + g'^2)(|H_u^0|^2 + |H_u^+|^2 - |H_d^0|^2 - |H_d^-|^2)^2 \\
&+ \frac{1}{2}g^2|H_u^+ H_d^{0*} + H_u^0 H_d^{-*}|
\end{aligned}
\tag{1.42}
$$

The terms proportional to $m^2_{H_u}$, $m^2_{H_d}$ and b are the contributions from the $\mathscr{L}^{MSSM}_{soft}$, the ones allowing the EWK symmetry breaking.
The minimum of 1.42 should break the EWK symmetry. The analysis can be simplified by a $SU(2)$ gauge transformations that rotate away a possible v.e.v for one of the weak isospin components of one of the scalar fields: this gives $H^+_u = 0$ at the minimum of 1.42. At the minimum must be verified $\partial\mathscr{V}/\partial H^+_u\big|_{H^+_u=0} = 0$, which implies also $H^-_d = 0$. This means that at the minimum of the potential electromagnetism is not broken, which is good. Setting $H^+_u = 0$ and $H^-_d = 0$ the expression 1.42 become:

$$\begin{aligned}\mathscr{V}(H^0_{u,d}) =&(|\mu|^2 + m^2_{H_u})|H^0_u|^2 + (|\mu|^2 + m^2_{H_d})|H^0_d|^2 - b(H^0_u H^0_d + c.c.) \\ &+ \frac{1}{8}(g^2 + g'^2)(|H^0_u|^2 - |H^0_d|^2)^2\end{aligned} \tag{1.43}$$

The coefficient b can be taken real and positive, by absorbing any possible phase with a redefinition of the phase of H_u or H_d. For a minimum of 1.43 the product $H^0_u H^0_d$ must also be real and positive, so the v.e.v.'s of H^0_u and H^0_d must have equal and opposite phases. To do this a $U(1)$ transformation can be used, since H^0_u and H^0_d have opposite weak hypercharges ($\pm 1/2$). Since v.e.v.'s and b can be simultaneously chosen real and positive, as a convention, it follows that CP cannot be spontaneously broken by the Higgs scalar potential. This means that Higgs scalar mass eigenstates can also be chosen as eigenstates, with a well-defined eigenvalue, of CP; at least at tree level. In order to have a minimum in 1.43, the potential has to be bounded from below for arbitrarily large values of the scalar fields. The scalar quartic interactions in 1.43 stabilize the potential for almost all arbitrarily large values of the fields. However, for the special directions in field space $|H^0_u| = |H^0_d|$, the quartic contributions are identically zero. In order for the potential to be bounded from below, the quadratic part of the scalar potential has to be positive along those directions. This condition is satisfied if:

$$2b < 2|\mu|^2 + m^2_{H_u} + m^2_{H_d} \tag{1.44}$$

which implies also that $|H^0_u| = |H^0_d| = 0$ cannot be a maximum of the potential. It could be a minimum, which would mean that EWK symmetry does not breaks. A study of the Hessian matrix of the scalar potential gives the condition for $|H^0_u| = |H^0_d| = 0$ to be a saddle point:

$$(|\mu|^2 + m^2_{H_u})(|\mu|^2 + m^2_{H_d}) < b^2 \tag{1.45}$$

this is automatically satisfied if either $(|\mu|^2 + m^2_{H_u})$ or $(|\mu|^2 + m^2_{H_d})$ is negative, or rather that either $m^2_{H_u}$ or $m^2_{H_d}$ is negative.

The conditions 1.44 and 1.45 have then to be satisfied in order to get non-zero v.e.v.'s for H^0_u and H^0_d. A last requirement is that these conditions have to be compatible with the observed phenomenology of EWK symmetry breaking. Writing the v.e.v.'s of the fields as $v_u \equiv \langle H_u \rangle$ and $v_d \equiv \langle H_d \rangle$, the relationship between these and the mass of the Z boson and the EWK gauge couplings is:

$$v^2_u + v^2_d = v^2 = 2m^2_Z/(g^2 + g'^2) \simeq (174\ GeV) \tag{1.46}$$

The v.e.v.'s ratio is traditionally written as:

$$\tan\beta \equiv \frac{v_u}{v_d} \tag{1.47}$$

Being v_u and v_d real and positive, then $0 < \beta < \pi/2$.
Now the equations determining the Higgs v.e.v.'s, which follow from imposing the stationary conditions, $\partial\mathscr{V}/\partial H_u^0\big|_{H_u^0=v_u} = \partial\mathscr{V}/\partial H_d^0\big|_{H_d^0=v_d} = 0$, such that 1.46 and 1.47 are satisfied are:

$$\begin{aligned} m_{H_u}^2 + |\mu|^2 - cot\beta - (m_Z^2/2)cos(2\beta) = 0 \\ m_{H_d}^2 + |\mu|^2 - tan\beta - (m_Z^2/2)cos(2\beta) = 0 \end{aligned} \tag{1.48}$$

It can be easily shown that these equations satisfy the necessary conditions 1.44 and 1.45.

1.4.6 A physical explanation of electroweak symmetry breaking

The parameter $m_{H_u}^2$ in 1.45, is driven at negative values at the electroweak scale by the RG evolution. The most important one loop terms in the RG equation for H_u mass parameter are the following:

$$\frac{d}{dlogQ} M_{H_u}^2 = \frac{1}{(4\pi)^2}[3\lambda_t^2(M_{H_u}^2 + M_{\tilde{t}_L}^2 + M_{\tilde{t}_R}^2) + ...] \tag{1.49}$$

with contributions from the Yukawa top quark λ_t and the scalar squarks masses $M_{\tilde{t}_L}$ and $M_{\tilde{t}_R}$. RG equations for the masses of $\tilde{t}_{L,R}$ show that their values increases at low energies. Thanks to the heaviness of the top quark, the equation 1.49 predicts a negative value of $m_{H_u}^2$ at the EWK scale [30]. This behavior is shown in Figure 1.7, together with the evolution of the $\tilde{t}_{L,R}$ masses. So in the MSSM EWK symmetry breaking is actually driven by quantum corrections due to the top quark.

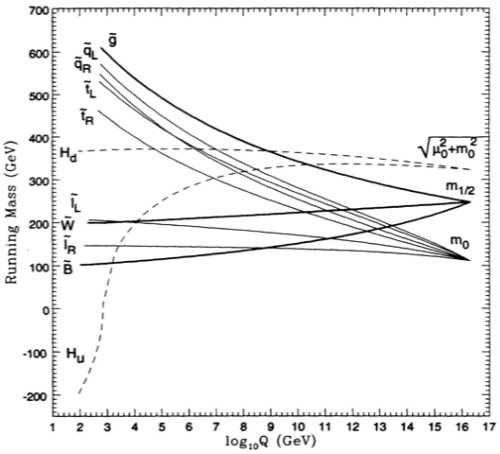


Figure 1.7: Running of the soft supersymmetry breaking mass terms from the grand unification scale to the electroweak scale.

1.4.7 The tree-level Higgs scalar fields masses

As it is shown in Table 1.1, there are two complex $SU(2)$ doublets scalar Higgs fields. These correspond to eight real degree of freedom. With EWK symmetry breaking three of them are the would-be-Nambu-Goldstone boson G^0, $G^{\pm}$, which become the longitudinal modes of Z^0, $W^{\pm}$ massive vector bosons. The remaining five Higgs scalar mass eigenstates consist of two CP-even neutral scalars h^0 and H^0, one CP-odd neutral scalar A^0, and two charged scalars $H^{\pm}$. By convention h^0 identify the lightest Higgs neutral scalar. The gauge eigenstate fields can be expressed in terms of the mass eigenstate fields as:

$$\begin{pmatrix} H_u^0 \\ H_d^0 \end{pmatrix} = \begin{pmatrix} v_u \\ v_d \end{pmatrix} + \frac{1}{\sqrt{2}} R_\alpha \begin{pmatrix} h^0 \\ H^0 \end{pmatrix} + \frac{i}{\sqrt{2}} R_{\beta_0} \begin{pmatrix} G^0 \\ A^0 \end{pmatrix} \tag{1.50}$$

$$\begin{pmatrix} H_u^+ \\ H_d^{-*} \end{pmatrix} = R_{\beta_\pm} \begin{pmatrix} G^+ \\ H^+ \end{pmatrix} \tag{1.51}$$

where R_α, R_{β_0}, $R_{\beta_\pm}$ are orthogonal rotation matrices:

$$R_{\theta=\alpha,\beta_0,\beta\pm} = \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \tag{1.52}$$

such that the quadratic part of the potential has diagonal squared masses:

$$\begin{aligned} \mathscr{V} =& \frac{1}{2} m_{h^0}^2 (h^0)^2 + \frac{1}{2} m_{H^0}^2 (H^0)^2 + \frac{1}{2} m_{G^0}^2 (G^0)^2 + \frac{1}{2} m_{A^0}^2 (A^0)^2 \\ &+ m_{G^\pm}^2 |G^+|^2 + m_{H^\pm}^2 |H^+|^2 \end{aligned} \tag{1.53}$$

The condition that the Higgs scalar potential is minimized in v_u and v_d, gives $\beta_0 = \beta_\pm = \beta$, $m_{G^0} = m_{G^\pm} = 0$, and the masses of the Higgs scalar fields at tree-level:

$$\begin{aligned} m_{h^0,H^0}^2 &= \frac{1}{2}(m_{A^0}^2 + m_Z^2 \mp \sqrt{(m_{A^0}^2 - m_Z^2)^2 + 4 m_Z^2 m_{A^0}^2 sin^2(2\beta)}) \\ m_{A^0}^2 &= 2b/sin(2\beta) \\ m_{H^\pm}^2 &= m_{A^0}^2 + m_W^2 \end{aligned} \tag{1.54}$$

all the Higgs scalar masses are determined once are known m_{A^0} and $tan\beta$. The masses of A^0, H^0 and $H^{\pm}$ can in principle be arbitrarily large since they all grow with $b/sin(2\beta)$. Conversely, the mass of h^0 is bounded from above. From the equation of the h^0 mass it follows, at tree-level:

$$m_{h^0} < m_Z |cos(2\beta)| \tag{1.55}$$

If this was the exact upper bound, the lightest Higgs scalar would have already been excluded by LEPII. Actually, the tree level formulae 1.54 receive significant quantum corrections, especially in the case of the h^0 [31–34]. The main corrections come from the $t - \tilde{t}$ sector, and for large value of $tan\beta$ also from the $b - \tilde{b}$ sector, with values depending on additional MSSM parameters entering in the loops. The most important parameters for the corrections to m_{h^0} are the top mass m_t, the mass M_{SUSY} (by convention is $M_{SUSY} \equiv M_{\tilde{t}_L} = M_{\tilde{b}_L} = M_{\tilde{t}_R} = M_{\tilde{b}_R}$),

stop and sbottom mixing parameters[3] X_t, X_b, which in turn depends on the parameter μ and the trilinear Higgs-stop and Higgs-sbottom couplings $A_{t,b}$. The lightest Higgs scalar depends furthermore on the gaugino mass parameter M_2. At the two level loop also the gluino mass, $m_{\tilde{g}}$, enters the prediction of m_{h^0}. A part for the top mass, which is fixed experimentally, the other parameters are set by hand and define the so called benchmark scenario. Different scenarios are defined, among which the $m_{h^0}^{max}$, where the parameters are chosen such that the maximum possible Higgs boson mass as a function of $tan\beta$ is obtained [35]. It has been shown [36] that in the $m_{h^0}^{max}$ scenario the upper bound on the lightest Higgs scalar is $m_{h^0} < 140\ GeV$. Thus it is a fairly robust prediction of supersymmetry at the electroweak scale that at least one of the Higgs scalar bosons must be light. The masses of the scalar Higgs, including quantum corrections, are shown in Figure 1.8.

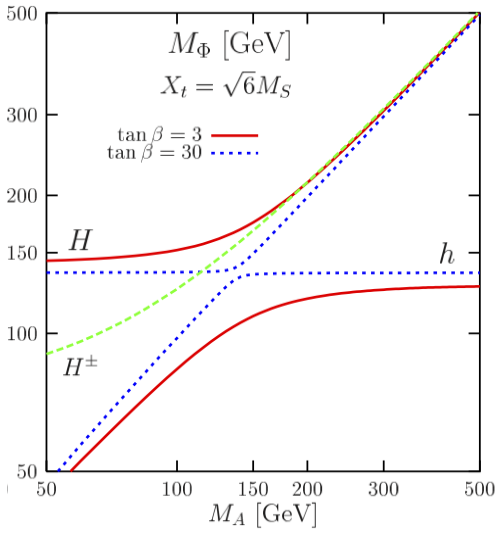


Figure 1.8: Masses of h^0, H^0 and $H^\pm$ Higgs bosons as a function of the A mass for $tan\beta = 3$ and $tan\beta = 30$ in the $m_{h^0}^{max}$ scenario. Quantum corrections are included.

As can be seen, the lightest Higgs reaches is maximum value $M_{h^0}^{max} \sim 135\ GeV$ for large values of M_A. Also, for large M_A values the A, H and $H^\pm$ bosons become degenerate in mass. Whereas, for light pseudoscalar Higgs boson, $M_A \lesssim M_{h^0}^{max}$, the h^0 mass is very close to the A mass. The difference is particularly small for large $tan\beta$ values. These mass degeneracy could be helpful experimentally since the signal strength can be enhanced for those mass values.

1.4.8 Tree level couplings of the neutral scalars Higgs to SM particles.

The couplings of the MSSM Higgs scalars to the gauge bosons are obtained from the kinetic terms of the fields H_u, H_d, in the Lagrangian:

[3]The left and right handed component of sfermions can mix. Indeed, the mass matrix has non vanishing off diagonal parameters. These elements cannot be cancelled by a unitary transformations of the sfermionic fields, like it is done for the fermions. This is because of the trilinear soft supersymmetry breaking terms for sfermions.

$$\mathscr{L}_{kin.} = (D^{\mu}H_u)^{\dagger}(D_{\mu}H_u) + (D^{\mu}H_d)^{\dagger}(D_{\mu}H_d) \tag{1.56}$$

where the covariante derivative D_{μ} is:

$$D_{\mu} = \partial_{\mu} + ig\frac{\tau}{2}\cdot W_{\mu} + i\frac{g'}{2}YB_{\mu} \tag{1.57}$$

Taking only the trilinear interaction terms and using 1.50 to obtain the physical scalar Higgs fields:

$$(\partial_{\mu} + ig\frac{\tau}{2}\cdot W_{\mu} + i\frac{g'}{2}YB_{\mu})^2 v(sin(\beta-\alpha)h^0 + cos(\beta-\alpha)H) \tag{1.58}$$

where the rotation angle α is defined in 1.52.
Performing the transformations on the gauge fields, W_{μ}, B_{μ}, to obtain the physical gauge boson fields, in 1.58 one can read the couplings for the interaction terms WWh^0, WWH, ZZh^0, ZZH. These couplings turns out to be the one of the SM modified by $sin(\beta-\alpha)$ and $cos(\beta-\alpha)$ factors. They are summarized in Table 1.3. One can also note that A boson does not couple to gauge bosons at tree level.
The couplings of the MSSM Higgs scalars to fermions can be obtained from the first three terms of the superpotential 1.34. Retaining only the scalar Higgs fields terms and expanding them in terms of the physical fields using 1.50, for the first family follows:

$$\begin{aligned}\mathscr{L}_{Yuk} = &-\frac{gm_u}{2M_W sin\beta}[\bar{u}u(H^0 sin\alpha + h^0 cos\alpha) - i\bar{u}\gamma_5 u A^0 cos\beta] \\ &-\frac{gm_d}{2M_W cos\beta}[\bar{d}d(H^0 cos\alpha - h^0 sin\alpha) - i\bar{d}\gamma_5 d A^0 sin\beta]\end{aligned} \tag{1.59}$$

where:

$$y_u = \frac{gm_u}{2M_W sin\beta} \qquad\qquad y_d = \frac{gm_d}{2M_W cos\beta} \tag{1.60}$$

are the Yukawa couplings of the superpotential. From the up-type fermion Yukawa coupling expression, y_u, it is clear that if $sin\beta$ is to small, then the top coupling y_t will be non-perturbatively large. On the other hand, from the down-type fermion Yukawa coupling expression, y_d, if $cos\beta$ is to small, then the bottom and tau couplings, $y_{b,\tau}$, will be non-perturbatively large. So, requiring that $y_{t,b,\tau}$ do not blow up above the EWK scale implies roughly lower and upper bounds on $tan\beta$, which are $tan\beta \gtrsim 1.2$ and $tan\beta \lesssim 65$, respectively. There are good theoretical motivations that prefer large value of the $tan\beta$ parameter. Grand unification theories (GUT) models based on $SO(10)$ can unify the running y_t, y_b and y_τ, at the unification scale; this requires $tan\beta \approx m_t/m_b \simeq 40$.
In 1.59 the γ_5 coupling for the A^0 shows that this is a pseudoscalar ($CP = -1$), whereas the h^0 and H^0 are scalars. The neutral Higgs boson couplings to fermions can be read from 1.59 and are summarized in Table 1.3. One can notice that the couplings are the same as SM one modified by $sin\beta$ and $cos\beta$ factors.

It can be shown that large values of m_{A^0} implies $sin\alpha \approx -cos\beta$ and $cos\alpha \approx sin\beta$. It follows that in this limit the couplings of the lightest neutral Higgs

Table 1.3: Neutral Higgs boson couplings to fermions and gauge bosons in the MSSM normalized to the SM Higgs boson couplings.

Φ	$g_{\Phi\bar{u}u}$	$g_{\Phi\bar{d}d}$	$g_{\Phi VV}$
H_{SM}	1	1	1
h^0	$\frac{cos\alpha}{sin\beta}$	$-\frac{sin\alpha}{cos\beta}$	$sin(\beta-\alpha)$
H^0	$\frac{sin\alpha}{sin\beta}$	$\frac{cos\alpha}{cos\beta}$	$cos(\beta-\alpha)$
A^0	$cot\beta$	$tan\beta$	0

h^0 become those of the SM Higgs, while the couplings of H^0 are the same as those of the A^0. For values of m_{A^0} not too large and high $tan\beta$ the couplings of h^0,H^0 and A^0 are enhanced for the bottom-type fermions and suppressed for the up-type fermions. Since the Yukawa couplings are proportional to the mass of the interacting fermion, the enhancement is stronger for bottom quark and tau lepton, whereas the suppression is stronger for top quark.
Specifically, five different regimes can be identified, each one leading to a given behavior of the MSSM Higgs decays, depending on the values of M_A and also $tan\beta$:

- the *decoupling regime* occurs at large values of M_A, and more precisely at $M_A \gtrsim 300\ GeV$ for low $tan\beta$ values and $M_A \gtrsim M_{h^0}^{max}$ for $tan\beta \gtrsim 10$. In this case, the CP-even h^0 boson reaches its maximal value $M_{h^0}^{max}$ and its coupling to fermions and gauge bosons become SM-like. The heavier H boson has approximately the same mass as the A boson and its interactions are similar: its couplings to gauge bosons almost vanish and the couplings to bottom (top) quarks and τ leptons fermions are (inversely) proportional to $tan\beta$. Hence, in this regime there is a SM-like Higgs boson, and two pseudo scalars like;

- the *anti-decoupling regime* occurs when $M_A \lesssim M_{h^0}^{max}$ and is exactly opposite to the decoupling one. The roles of the h and H bosons are reversed and at large $tan\beta$ values, the h boson is degenerate in mass with the pseudo-scalar A, $M_{h^0} \simeq M_A$, while the H boson has a mass close to its minimum which is in fact $M_{h^0}^{max}$. So all the Higgs particles are light. In this regime, it is the h^0 boson which has couplings close to those of A, while the H boson couplings are SM-like;

- the *intense-coupling regime* occurs when the mass of the pseudo-scalar A boson is close to $M_{h^0}^{max}$. The three neutral scalar Higgs have comparable masses in this regime, $M_{h^0} \simeq M_H \simeq M_A \simeq M_{h^0}^{max}$. The mass degeneracy is more effective for $tan\beta$ large. In this regime, both the h and H bosons have still enhanced couplings to b-quarks and τ leptons and suppressed couplings to gauge bosons and top quarks, as is the pseudo-scalar A. Hence, one approximately has three pseudo-scalar like Higgs particles, with mass differences of the order of $10-20\ GeV$;

- the *intermediate-coupling regime* occurs for low $tan\beta$ values, $tan\beta \lesssim 5-10$, and a not to heavy A boson, $M_A \lesssim 300-500\ GeV$. We are not yet in the decoupling regime: both CP-even Higgs bosons have nonzero couplings to gauge bosons and their couplings to down-type (up-type) fermions, as for

the A boson, are not strongly enhanced (suppressed), since $tan\beta$ is not too large;

- the *vanishing-coupling regime* occurs for relatively large values of $tan\beta$ and intermediate to large M_A values, as well as for specific values of the other MSSM parameters. These parameters, when entering the radiative corrections, could lead to a strong suppression of the couplings of one of the CP-even Higgs bosons to fermions or gauge bosons, as a result of the cancellation between tree-level terms and radiative corrections.

1.4.9 Neutral Higgs bosons production and decays at LHC

The production of the neutral MSSM Higgs bosons at LHC involves the same processes that occur for the SM Higgs. Specifically, the production mechanism are: associated production with W/Z bosons, vector boson fusion, gluon-gluon fusion and the associated production with heavy quarks. The pseudoscalar A cannot be produced in association or via fusion of vector bosons at tree level, since direct coupling of A with vector bosons are forbidden in the MSSM by CP invariance. As it as been pointed out in the previous paragraph, if m_{A^0} is not too large and $tan\beta$ is high, the couplings of the neutral Higgs bosons to bottom quarks are enhanced, and the production in association with bottom quarks, with the following decay of the Higgs boson into bottom quark pairs, become important at LHC. Contrarily at what happens for the SM Higgs, where the Yukawa coupling to the bottom quark is tiny. So in the MSSM these channels, shown in Figures 1.9 and 1.10, are experimentally interesting and are the topic of this thesis.

Figure 1.9: Diagram contribution to the production of the Higgs boson in association with a b-quark pair, $pp \to b\bar{b} + \Phi$ with $\Phi = h, H, A$, at lowest order. Two production mechanism are possible: two gluons in the initial state split into two b-quarks with the subsequent Higgs production through $b\bar{b}$ fusion (left diagram) or a quark anti-quark in the initial state produce a gluon subsequently splitting in a b-quark pair, where one of the b-quarks radiate a Higgs boson (right diagram).

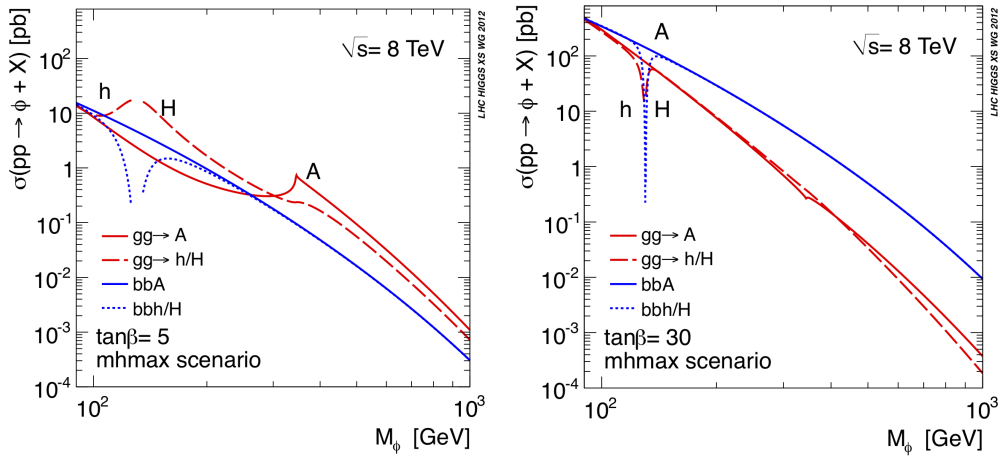
The cross section for a neutral scalar MSSM Higgs $\Phi = h, H, A$ for production in association with b-quarks, is that of the SM Higgs multiplied by the squared MSSM couplings to fermions normalized to the SM:

$$\sigma(pp \to b\bar{b}\Phi) = g^2_{\Phi b\bar{b}}\sigma_{SM}(pp \to b\bar{b}\Phi) \tag{1.61}$$

This is shown in Figure 1.11 as a function of the Φ mass and compared to the cross section of other production mechanism. Important QCD corrections to the tree level calculation have to be taken into account. These cross section distributions clearly show that the production in association with bottom quarks become relevant at high $tan\beta$.

The partial decay width of a neutral Higgs boson $\Phi = h, H, A$ into fermion pairs is given in the Born approximation by:

Figure 1.10: Diagram contribution to the production of the Higgs boson in association with one b-quark, $pp \to b + \Phi$ with $\Phi = h, H, A$, at lowest order. The production mechanism involves a b-quark from the sea and a gluon in the initial state. The Higgs boson is produced through t-channel (left diagram) or s-channel (right diagram).



(a) h^0, H^0 and A^0 production for $tan\beta = 5$. (b) h^0, H^0 and A^0 production for $tan\beta = 30$.

Figure 1.11: Neutral MSSM Higgs production cross sections at the LHC at the center of mass energy of 8 TeV for gluon fusion and the associated production with bottom quarks, including QCD corrections. The $m_{h^0}^{max}$ scenario is considered.

$$\Gamma(\Phi \to f\bar{f}) = N_c \frac{G_\mu m_f^2}{4\sqrt{2}\pi} g_{\Phi f\bar{f}}^2 M_\Phi \beta_f^p \tag{1.62}$$

where N_c is the color number, G_μ the Fermi constant, m_f the fermion mass, $g_{\Phi f\bar{f}}$ the coupling of the Higgs to the fermion normalized to the SM Higgs coupling, M_Φ the mass of the Higgs boson, $\beta_f^p = (1 - 4m_f^2/M_\Phi^2)^{1/2}$ and $p = 3(1)$ for the CP even (odd) Higgs boson.
For quark final states, important QCD correction have to be included and for the light quarks, the running masses defined at the scale of the Higgs masses have to be adopted to absorb the bulk of these corrections. For the bottom quark, specifically, and for $M_\Phi \sim 100 - 1000\ GeV$ this results in a decrease of the partial decay width by roughly a factor of two, as it happens also in the SM case. The values of the branching ratios (BR) for the various decays in SM particles are shown in Figure 1.12 as a function of the decaying particle mass. In the various regime identified in the previous paragraph, the decay patterns of the neutral Higgs bosons can be summarized in the following way:

- in the two $tan\beta$ scenarios shown in 1.12, the *decoupling regime* is realized for $M_A \gtrsim 150\ GeV$ for $tan\beta = 30$ and $M_A \gtrsim 400 - 500\ GeV$ for $tan\beta = 3$. The lightest Higgs boson reaches its maximum value $M_{h^0}^{max}$ and is SM-like and, thus, decays as the SM Higgs. Since $M_{h^0}^{max} \lesssim 140\ GeV$, dominant modes are the decays into $b\bar{b}$ pairs and into WW^* final states. The decays into $\tau^+\tau^-$, gg, $c\bar{c}$ and also ZZ^* final states are at the level of few percent and the loop induced decays into $\gamma\gamma$ and $Z\gamma$ at the level of a few per mille. The total decay width of the h^0 boson is small, $\Gamma(h) \lesssim O(10\ MeV)$. For the heavier Higgs bosons, the decay pattern depends on $tan\beta$. For $tan\beta \gg 1$, because of the strong enhancement of the Higgs couplings to down-type fermions, H and A will decay almost exclusively into $b\bar{b}$ ($\sim 90\%$) and $\tau^+\tau^-$ ($\sim 10\%$) pairs. The $t\bar{t}$, when kinematically allowed, and all other decays are strongly suppressed for $tan\beta \sim 30$;

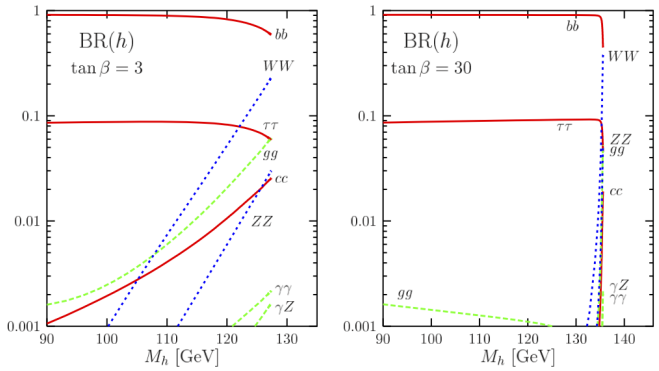
- in the *anti-decoupling regime* corresponds to $tan\beta = 30$ and $M_A \lesssim 130\,GeV$. The lighter CP-even h^0 and the CP-odd A bosons will mainly decay into $b\bar{b}$ ($\sim 90\%$) and $\tau^+\tau^-$ ($\sim 10\%$) pairs. All other modes are suppressed at the level below the per mille, except for the gluonic decays of the h^0 and A bosons. Although their masses are small, h^0 and H have relatively large total widths, $\Gamma(h, A) \sim O(1\ GeV)$ for $tan\beta = 30$. The heavier CP-even Higgs boson will have a mass $M_H \sim M_{h^0}^{max}$ and will be SM-like;

- the *intense-coupling regime* corresponds, in the two scenarios in 1.12, to $M_A \sim 120 - 140\ GeV$ and $tan\beta = 30$. The couplings of both CP-even h and H bosons to gauge bosons and up-type fermions are suppressed, while the couplings to down-type fermions, in particular b-quarks and τ leptons, are strongly enhanced. Then, the branching ratios of h and H to $b\bar{b}$ and $\tau^+\tau^-$ final states are dominant ones, as those of A boson, with values of $\sim 90\%$ and $\sim 10\%$, respectively;

- the *intermediate-coupling regime*, which in 1.12 corresponds to $tan\beta = 3$ and $M_A \lesssim 350\ GeV$, the Higgs couplings to bottom quarks and τ leptons are not strongly enhanced;

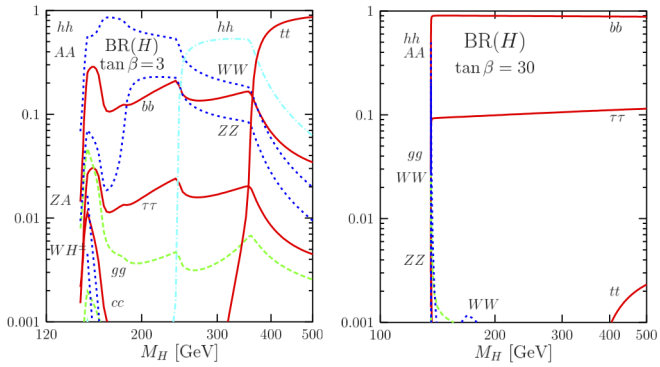
- the *vanishing-coupling regime* can occur at large $tan\beta$, moderate or high M_A values, and specific MSSM parameters. This regime leads to suppression of one of the two CP-even Higgs couplings to specific fermion or gauge boson.

From the discussion above, it is clear that the MSSM predicts a sizable production of neutral Higgs scalar and/or pseudo-scalar particles, of the order of tenth or even hundreds of pico-barns, produced in association with b-quarks. Their most important decay mode is by far the decay into a pair of b-quarks, leading to a final state of at least three energetic b-jets to be detected in the detector acceptance.
This thesis will discuss the physics analysis developed for the search of this signature and the main results obtained in the production cross-section of a narrow resonance $\Phi \to b\bar{b}$ in the mass range $90 - 350\ GeV$.



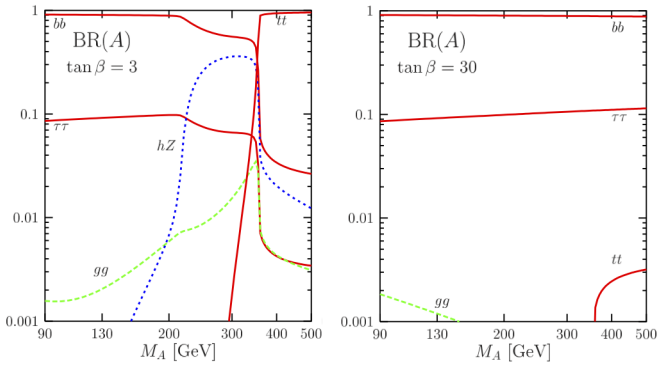
(a) h branching ratios for $tan\beta = 3$.

(b) h branching ratios for $tan\beta = 30$.



(c) H branching ratios for $tan\beta = 3$.

(d) H branching ratios for $tan\beta = 30$.



(e) A branching ratios for $tan\beta = 3$.

(f) A branching ratios for $tan\beta = 30$.

Figure 1.12: The decay branching ratios of the neutral Higgs bosons as a function of their masses for two values of the $tan\beta$ parameter. The values are obtained in the $m^{max}_{m_{h^0}}$ scenario.

Chapter 2

The CERN Large Hadron Collider

2.1 The Large Hadron Collider

The CERN Large Hadron Collider (LHC) is the most recent particle accelerator designed and built for frontier research in fundamental interactions of elementary particles [37]. It is installed in the 26.7 km tunnel which formerly hosted the LEP electron-positron collider. Two beams of protons are injected in the accelerator, which is a double superconducting proton storage ring, with an initial energy of 450 GeV. The beams circulate in opposite directions inside the ring and collide in four different points along the accelerator, where are installed four experiments: ALICE, ATLAS, CMS and LHCb. In the design project the protons are accelerated up to a corresponding energy of 7 TeV for each beam, reaching the highest center of mass energy obtainable in laboratory nowadays of $\sqrt{s} = 14\ TeV$. The proton injection is provided by the CERN accelerator chain already used by LEP, with some important upgrades. This, together with the use of the preexisting tunnel, represented a considerable gain in the total cost of the LHC machine. The schemes of the LHC and the accelerator chain, made by the Linac (Linear accelerator), PSB (Proton Synchrotron Booster), PS (Proton Synchrotron) and SPS (Super Proton Synchrotron), are shown in Figure 2.1 and Figure 2.2 respectively.

2.2 Main characteristic of LHC

Dipolar superconducting magnets [37] are used to generate a magnetic field along the ring and perpendicular to the plan containing LHC. This gives the bending power needed to keep the protons inside the pipes which are surrounded by the magnets. Since the LHC was designed as a *pp* collider, this required two paired rings and a special magnet design, twin-bore magnets, to allow same-charge particles rotate in opposite directions inside the pipes.
The magnets, shown in Figure 2.3, are made of NbTi cables cooled down to 1.9 K using superfluid helium. This allows currents which provide a bending field larger than 8 T. Other large superconducting accelerator (Tevatron-FNAL, HERA-DESY and RHIC-BNL) all used classic NbTi superconductors, cooled by supercritical helium at temperatures slightly above 4.2 K, with fields below or around 5 T.

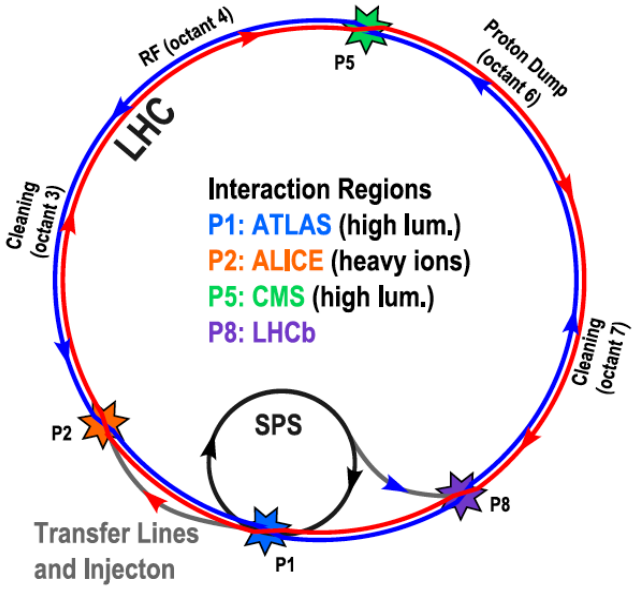


Figure 2.1: Layout of the CERN Large Hadron Collider, showing the interaction regions of the four main LHC experiments and the twin-ring design of the collider, and the accelerator chain. The CMS experiment cave is located beyond access point 5, in Cessy (France).

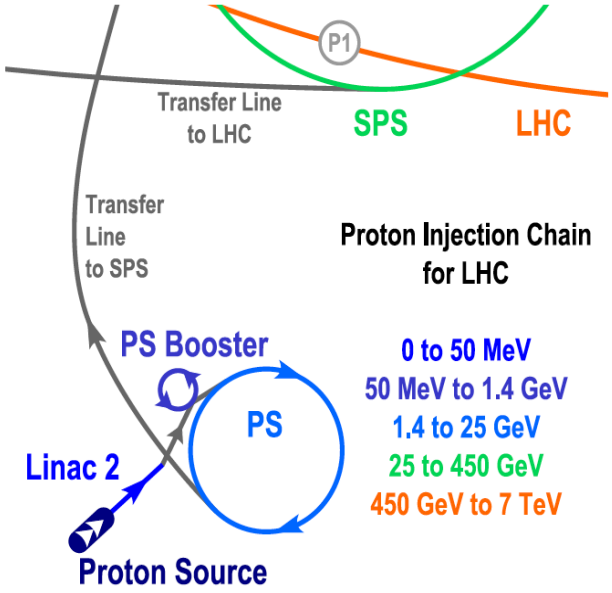


Figure 2.2: Layout of the CERN Large Hadron Collider Injection chain, showing the accelerators involved from proton source to the delivery of 7 TeV beams. All accelerators belonging to the CERN complex that are dedicated to other purposes are omitted.

The acceleration system is made of eight radio frequency cavities (RF) per beam, each delivering 2 MV (an accelerating field of 5 MV/m) at 400 MHz. The cavities, operating at 4.5 K, are grouped in four cryomodules, with two cryomodules per beam, and installed in a long straight section of the machine. A

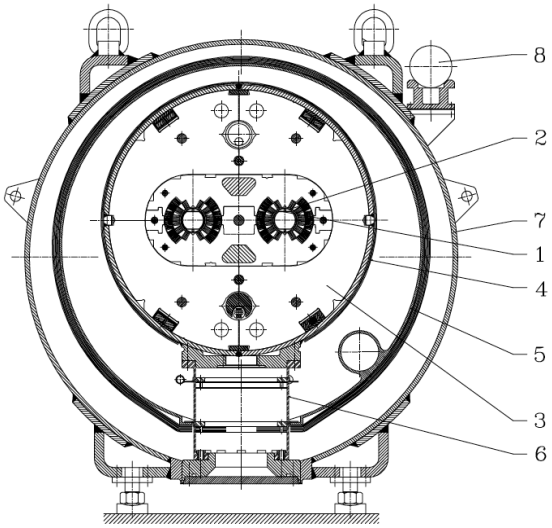


Figure 2.3: Cross Section of Dipole Magnet and Cryostat. 1. Beam screen, 2. Cold bore, 3. Cold mass at 1.9 K, 4. Radiative insulation, 5. Thermal shield at 55 to 75 K, 6. Support post, 7. Vacuum vessel, 8. Alignment target

RF accelerating system implies that the beams are made of bunches of protons, instead of being continuous. The distance between two consecutive bunches is constant, and such that the protons inside each bunch entering a RF cavity are subject to an electric field in the same direction of the proton's velocity. This way the beam is accelerated.
One of the important parameters of LHC is the energy at which the protons can be pushed. The following relation holds for circular accelerator:

$$P \ = \ 0.3Br \tag{2.1}$$

where P is the maximum momentum (in GeV) that a particle can reach given the magnetic field B (in T) generated by the magnet system and the radius r (in m) of the ring. The latter being fixed for LHC by the LEP tunnel, this formula shows the importance to achieve the highest possible value of the magnetic field, in order to reach high energies for the accelerated particles.
Another important parameter for LHC is the frequency (R) of a given process that take place in the interaction between the protons:

$$R \ = L\sigma \tag{2.2}$$

which is proportional to the cross section (σ) of the process through a constant, L, known as the instantaneous luminosity. The constant L depends only on machine parameters, that can be tuned to make L as higher as possible to get the maximum rate R:

$$L \ = \ \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n\beta^*}F \tag{2.3}$$

where γ is the Lorentz factor and f is the revolution frequency of the protons, k_B is the number of bunches, N_p is the number of protons in a bunch, ϵ_n is the transverse emittance normalized, β^* is the betatron function at the interaction point (IP) and F is the reduction factor taking into account the beams crossing

angle in the IP.
The protons inside the ring oscillate in the transverse plan to the beams around an equilibrium position (betatron oscillations). The maximum displacement of protons from the equilibrium position is determined by the betatron function, β, and the emittance, ϵ, which are then related to the transversal area of the beams. Specifically, the factor $4\pi\epsilon_n\beta^*$ in the denominator of the relation 2.3 is the transversal area of the beams at the IP. Reducing the transversal area of the beams at the IPs gives higher values of L. For this reason near the IPs are installed quadrupole magnets that have high focalizing power and are used to squeeze the beams.
Table 2.1 summarizes the goal of the machine parameters value at the nominal CM energy of 14 TeV. Whereas, table 2.2 summarizes the achieved values of the machine parameters at the CM energy of 7 TeV and 8 TeV, the operational energies during 2010-2011 and 2012 LHC operations respectively. In the tables, the luminosity values indicate the maximum instantaneous luminosity achieved at the beginning of each fill of the machine. The decision to run at lower energies than the nominal one was taken after the accident that took place on September 2009. A malfunction was caused by a faulty electrical connection between two magnets. This resulted in mechanical damage and release of helium from the magnet cold mass into the tunnel. Detailed studies of the malfunction allowed to identify means of preventing a similar incident from occurring in the future, and to design new protection system for the machine. With the new protection system the safest energy reachable is approximately half of the design energy. After the experience accumulated in the first two years of running at 7 TeV and specific additional studies performed, it was decided to go up to 8 TeV during 2012 running. A first long shutdown for LHC is foreseen for 2013-2014, during which important upgrades will make possible for the machine to reach the nominal energy of 14 TeV for the runs foreseen in the years 2015-18.

Table 2.1: Nominal values of the principal parameters of LHC

Parameter		Value	Unit
Energy per proton	E_p	7	TeV
Number of protons per bunch	N_p	1.5×10^{11}	
Number of bunches	k_B	2808	
Interval between two bunches		25	ns
Bunch Length (r.m.s.)		53	mm
Bunch radius at IP (r.m.s.)		15	μm
Beams crossing angle		300	μrad
Instantaneous Luminosity	L	10^{34}	$cm^{-2}s^{-1}$

2.3 A high intensity hadron collider

The main difference of the LHC with respect to the previous CERN accelerator LEP is the kind of particle that are used for collision: the LHC is a *pp* collider. An electron-positron collider has the important drawback of the particle energy loss because of the synchrotron radiation, which makes this machines more expensive during operations. The amount of energy dissipated in synchrotron radiation is

Table 2.2: Real values of the principal parameters of LHC

Parameter		Value (7 TeV)	Value (8 TeV)	Unit
Energy per proton	E_p	3.5	4	TeV
Number of protons per bunch	N_p	1.5×10^{11}	1.66×10^{11}	
Number of bunches	k_B	1380	1380	
Interval between two bunches		50	50	ns
Beams crossing angle		120	145	μrad
Instantaneous Luminosity	L	4.02×10^{33}	7.54×10^{33}	$cm^{-2}s^{-1}$

proportional to γ^4 (where $\gamma = \frac{E}{m}$, with E the energy and m the mass of the accelerated particle), and for protons is a factor $(m_p/m_e)^4 \sim 10^{13}$ lower than electrons/positrons.
At the high energy of the LHC the interaction during pp collisions occurs between partons, the proton constituents, quarks and gluons, that bring a fraction x of the total momentum of the beam. The CM energy of a partons collision is $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$, where x_1 and x_2 are the fractions of momentum of the two partons respectively. It can vary in a wide energy interval up to 14 TeV (for the nominal energy) since the fractions x_1 and x_2 have values between 0 and 1 with probability given by the so called parton density functions.
The use of antiproton for one of the LHC beams, like in the previous hadronic colliders, $Sp\bar{p}S$ at CERN and Tevatron at Fermilab, would have simplified the magnets layout. Unfortunately, the antiproton production process has very low efficiency. High luminosities needed to search for new, rare physics processes at a CM energy of $\sqrt{s} = 14\ TeV$ require high intensity beams, that makes the production of antiproton beams unpractical and costly.
The total inelastic cross section for pp collisions at nominal CM energy of 14 TeV, extrapolated from measurements at lower energies, is $\sigma = 80\ mb$ [38, 39]. Taking into account relation 2.2, the frequency of pp interaction is then of the order of 1 GHz. A luminosity $L = 10^{34}\ cm^{-2}s^{-1}$ with a bunch spacing of $T_b = 25\ ns$ and a bunch fill ratio[1] $\simeq 0.8$ lead to approximately:

$$N_{ev} = L \times \sigma \times T_b \times \frac{N_b^{tot}}{N_b^{fill}} \simeq 25 \tag{2.4}$$

superimposed events per bunch crossing, also called "pile-up" events, in each of the high-luminosity collision points where the ATLAS and CMS experiments are housed. High collision rate and pile-up impose several constraints on the experiments: high granularity and fast response of the detectors, a fast data acquisition system and high resistance to the radiation damage of the read out electronics system to cope with the high particle flux produced during interactions. The accuracy and rapidity of the trigger system which selects events online is a fundamental requirement to be achieved. Specifically, the trigger system is very important since it allows to discard events that are physically not relevant, lowering by 5-6 order of magnitudes the high frequency rate of interaction dominated by background events. The cross section as a function of the CM energy and the frequency of interesting events as a function of the mass of the produced

[1]The bunch fill ratio is the ratio between the number of filled buckets at $T_b = 25ns$, $N_b^{fill} = 2808$, and the available ones, $N_b^{tot} = 3564$.

particles at LHC is shown in Figure 2.4. For instance, the production of a Standard Model (SM) Higgs boson with mass in the interval $114\ GeV \leq m_H \leq 1\ TeV$, where the lower limit is from previous LEP experiments and the higher limit is from theoretical arguments, is dominated by the gluon-gluon fusion. As it is shown in Figure 2.4 the frequency of production for this channel is between the mHz and the Hz at the nominal luminosity. This is about 10 order of magnitudes lower than the total inelastic pp cross section.

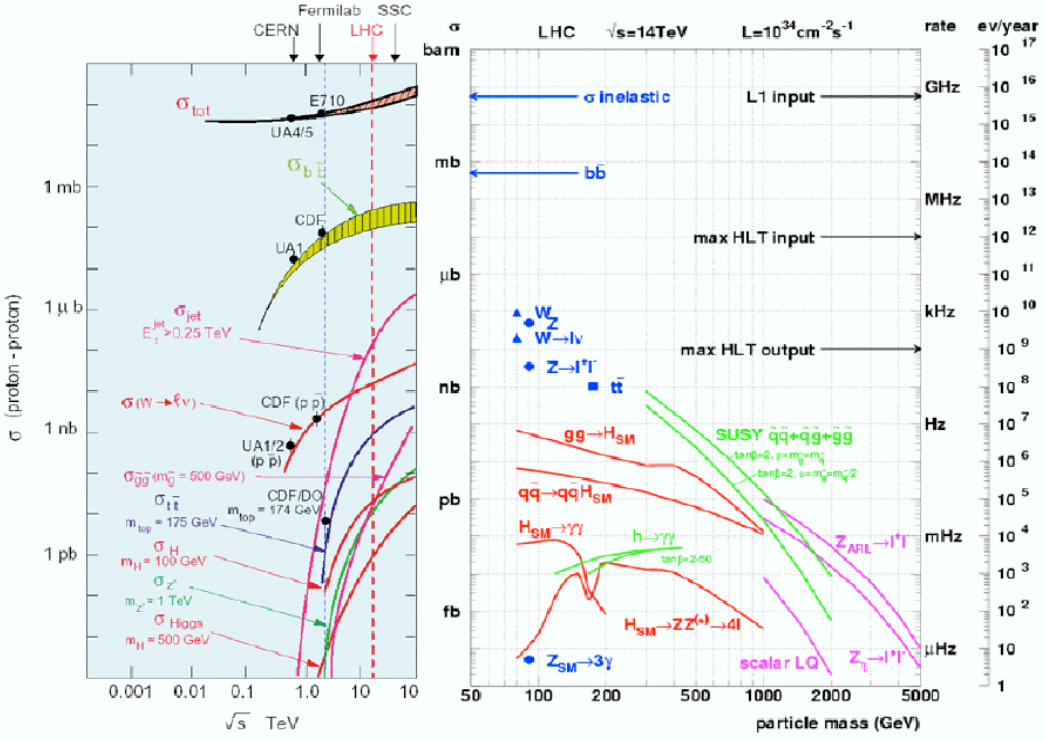


Figure 2.4: Cross sections as a function of the CM energy (left) and events frequency as a function of the mass of the produced particle at LHC (right) for interesting processes.

2.4 The LHC experiments

Four main experiments are installed along the LHC ring, as it is shown in Figure 2.1: ALICE (A Large Ion Collider Experiment) [40], ATLAS (A Toroidal LHC ApparatuS) [41], CMS (Compact Muon Solenoid) [42], LHCb (LHC Beauty experiment) [43]. In addition, there are also two experiments devoted to pp cross section measurements, TOTEM [44] and LHCf [45].
LHCb and ALICE are designed for studies of CP violation in the b quark physics and quark-gluon plasma produced in the heavy ion collisions, respectively. Whereas, ATLAS and CMS are designed for studies of the pp collisions with the goal of the discovery of the SM Higgs boson and possibly new physics signatures at the TeV energy scale.
Since the purpose of ATLAS and CMS is the study of a wide range of physics processes, they are also known as general purpose experiments. They then share some characteristics. The detectors they are made of have to be able to detect

final state particles in nearly all the solid angle around the interaction point (IP): both experiments are made of a cylindrical central part, the barrel, and two endcaps. This makes the experiments almost hermetic, except for the holes due to the proton pipes, the quadrupoles near the IPs focalizing the beams, the wiring of the read out electronics and the tubes of the cooling system. The choice of the magnet system, adopted to generate the magnetic field needed to bend the charged particle for the momentum measurement, is a fundamental issue. In fact, the choice of the magnetic system constrains the dimension of the experimental apparatus and the characteristics of the sub detectors. For CMS a single superconductor solenoid has been employed, whereas for ATLAS an external toroid coupled together with an internal solenoid have been used. A tracker system is installed in the region closer to the IPs of both experiments.
The energy measurements of the particles produced in collisions are made by the hadronic and electromagnetic calorimetric systems. A good measure of the total particle energy is extremely important to detect indirectly particles weakly interacting, like neutrinos or particle related to new physics (that might be candidate to explain the so called dark matter, that is the non barionic matter highlighted by the astrophysics observations). The calorimetric system is installed just outside the tracker system. The most external area of the two experiments is covered by the muon system.
In chapter 3 we give a detailed description of the CMS experiment.

Chapter 3

The Compact Muon Solenoid Experiment

3.1 General idea

The Compact Muon Solenoid (CMS) [42] experiment is built almost hermetically around one of the IPs of the LHC, in order to provide great efficiency in the individuation and measurement of a wide range of final states emerging from pp interactions. The detector has been built to match the goals of the LHC physics programme. The typical requirements are: a good muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution, the ability to determine unambiguously the muon charge up to $p_T \leq 1TeV$. Good charged-particle momentum resolution and reconstruction efficiency in the inner tracker. A pixel detector near the IP is needed to provide triggering and offline tagging of τ's and b-jets. Hadronic and electromagnetic calorimeters as much hermetic as possible and providing energy measurements of neutral particle with high resolution. Specifically, a good electromagnetic energy resolution and high spatial granularity is needed to ensure a good diphoton and dielectron mass resolution. Also, a good hadronic energy resolution has to be provided to ensure a good dijet mass resolution and good measurement of particles weakly interacting, through the determination of the missing transverse energy in the event. The design of CMS, described in this chapter for each sub-detector, meets these requirements.

3.2 The CMS detector

The structure of the CMS detector is shown in Figure 3.1. The coordinate system adopted by CMS has the origin centered at the nominal collision point inside the experiment, the y-axis pointing vertically upward, the x-axis pointing radially towards the center of the LHC and the z-axis parallel to the beam with is direction fixed by the right-hand rule. The azimuthal angle ϕ is measured from the x-axis in the x-y plane and the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the z-axis. In this reference frame, the quadri-momentum of a particle is $p_\mu = (E, p_x, p_y, p_z)$. Usually the CMS reference does not correspond with the CM frame of a collision between partons, since the total partons momentum could be different from zero in the CMS frame. The following quantity:

$$y \equiv \frac{1}{2} \cdot ln\left(\frac{E + p_z}{E - p_z}\right) \tag{3.1}$$

called rapidity, is invariant under a boost of the CM frame along the z axis and is then used to describe angular distributions in an event. The momentum of particles produced in collisions is such that they are ultra-relativistic ($p \gg m$), then their rapidity can be approximated with the pseudo-rapidity η, defined as:

$$\eta \equiv -ln\left(\frac{\theta}{2}\right) \tag{3.2}$$

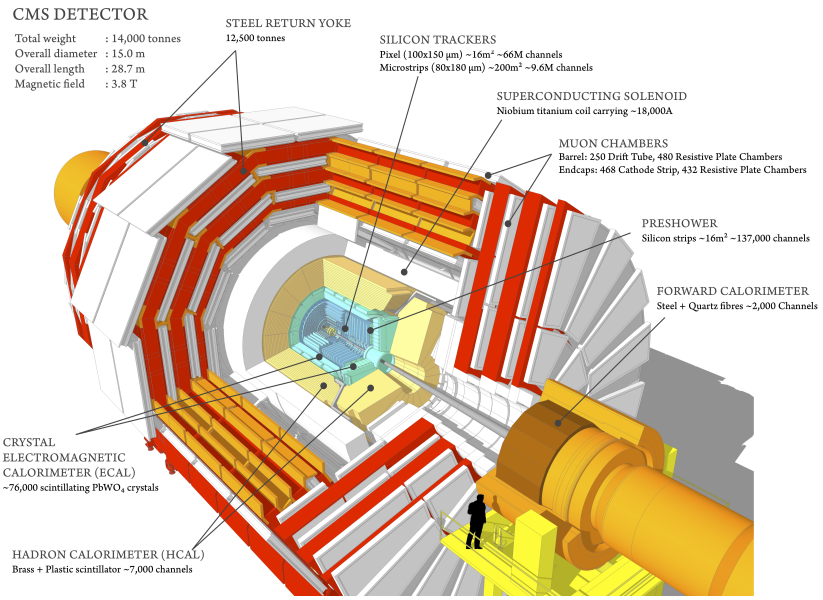


Figure 3.1: Schematic view of the CMS detector. The shells structure, from the interaction point toward the outside, is shown.

The CMS detector is composed of different detector shells surrounding the IP, with cylindrical symmetry in r and z axis along the beams. Radial sub-detector rings, centered in the beam direction, form the end-cap of the cylinder. In the innermost volume is installed the tracker, with a $5.8m$ length and $2.6m$ diameter. Ten layers of silicon microstrip detectors, which provide the required granularity and precision needed to deal with the high track multiplicity, make the sensitive part of the tracker. To improve the measurement of the impact parameter of charged-particle tracks, as well as the position of secondary vertices, three layers of silicon pixel detectors are placed close to the IP. Surrounding the tracker there is the electromagnetic calorimeter (ECAL) made of lead tungstate ($PbWO_4$) crystals with coverage in pseudo-rapidity up to $\mid \eta \mid \leq 3.0$. Silicon avalanche photodiodes (APDs) in the barrel region and vacuum phototriodes (VPTs) in the end-cap region are employed to detect the scintillator light. In front of the end-cap ECAL is installed a pre-shower system which helps to reject neutral pions. The ECAL is surrounded by the hadron calorimeter (HCAL), with a coverage up to $\mid \eta \mid \leq 3.0$. HCAL is made of calorimetric towers built with layers of brass, as absorbing material, superimposed to layers of plastic scintillator. The scintillator light is detected by hybrid photodiodes (HPDs).

The ECAL thickness, in radiation lengths, is larger than $25X_0$, while the HCAL thickness, in interaction lengths, varies in the range $7-11\ \lambda_I$, depending on η. An outer hadron calorimeter (HO), referred to as "tail catcher", is placed outside the barrel calorimeter. In the end-caps, coverage up to $\mid \eta \mid \leq 5.0$ is provided by an iron/quartz-fibre calorimeter. The Cerenkov light emitted in the quartz fibers is detected by photomultipliers. Tracker and calorimeter systems fill up the volume inside the superconducting solenoid. The superconducting solenoid provides a uniform magnetic filed with lines parallel to the beam direction and values up to $4T$. CMS is designed as a compact detector (as it is stated in its acronym), then a large bending power is needed to measure precisely the momentum of high energy charged particles, forcing the use of a superconducting magnet. The return magnetic field saturate the $1.5m$ of iron, the return yoke. Organized in 5 wheels in the barrel and 3 disks in each end-cap, the return yoke holds the muon detector system: 4 layers of muon stations cover the barrel and the end-caps. In the barrel region each muon station consists of several layers of aluminium drift tubes (DT), in the end-caps cathode strip chambers (CSC) are installed. In both regions resistive plate chambers (RPC) are also installed, to ensure good trigger redundancy. Overall, the CMS detector is $21.6m$ and has a diameter of $14.6m$. It has a total weight of $12500t$.

3.3 Tracker system

In order to exploit the high and uniform magnetic field, the tracker system is installed in the innermost central region of CMS. High particle flux comes from the IP: at the nominal luminosity, about 10^3 particles for each bunch crossing pass through the tracker, produced by about 25 pile-up pp interactions. At the design luminosity, this happens every 25 ns, that is with a frequency of 40 MHz[1]. This makes needful a tracker system with high granularity and fast response, in order to obtain reliable trajectory identification and attribution to the correct bunch crossing. These features imply a high power density of the on-detector electronics which in turn requires efficient cooling. At the same time the amount of material making the tracker system has to be the minimum possible, to limit multiple scattering, bremsstrahlung, photon conversion and nuclear interactions. The intense particle flux also may cause severe radiation damage to the tracking system, then material with high radiation hardness is needed to guarantee a detector operation of about 10 years. These requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology.
Figure 3.2 shows a scheme of the tracker system. The detector is made of two components: the Pixel Tracker and the Silicon Strip Tracker.

3.3.1 The Pixel Tracker

The particle flux goes from 1 MHz/mm^2 at $r = 4\ cm$ close to IP, to 60 kHz/mm^2 at $r = 22\ cm$. A pixel technology have been adopted to cope with this high particle flux in the region $r \leq 10\ cm$. Each pixel size is $100 \times 150\ \mu m^2$ in (r, ϕ) and z respectively.
The pixel detector consists of $\sim 66 \cdot 10^6$ pixels, disposed in ~ 1500 modular detec-

[1] In the 2010-12 runs, the bunch crossing time spacing was 50 ns.

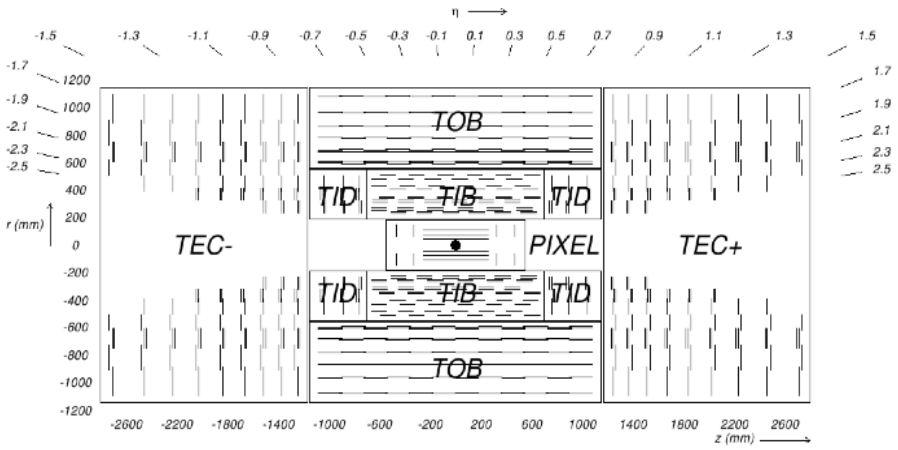


Figure 3.2: Schematic section of the Inner Tracking System in the (r, z) plane. The η ranges of the different subsystems are also shown.

tor units, called modules, installed in the barrel and end-caps. The pixel barrel consists of 3 layers with a total length of 53.3 cm, installed at $r = 4.4, 7.3, 10.2$ cm. Each end-cap is made of two disks at $\mid z \mid = 34.5, 46.5$ cm and radii ranging from 6 to 15 cm. Electrons from the ionization of the sensors by the charged particle, are collected to form the signal giving the position of the particle. Since the electrons drift in the $4T$ magnetic field, the resulting Lorentz force leads to charge spreading of the collected signal charge over more than one pixel in the barrel. The forward detectors are tilted at 20^o in a turbine-like geometry to induce charge-sharing, due to the geometric effect of particles entering the detector at an average angle of 20^o away from normal incidence. Signals from the different pixels are then interpolated to obtain the position of the ionizing particle. The pixel system provides in this way a measure of three high precision points on each charged particle trajectory up to $\eta \leq 2.5$, with a spatial resolution of 10 μm in the (r, ϕ) plane and 20 μm in z. A small impact parameter resolution is then reached, which is important for efficient secondary vertex reconstruction coming from b quark and tau decays.

3.3.2 The Silicon Strip Tracker

In the region between radii 20 $cm \leq r \leq 110$ cm the particle flux lowers and a silicon micro-strip detector technology can be adopted. Each sensor is a silicon micro-strip. The silicon strip tracker is divided into 4 sub-systems: Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB), where the silicon sensors are disposed parallel to the beamline on cylindrical layers; Tracker Inner Disk (TID) and Tracker End-Caps (TEC), where the silicon sensors are disposed radially in rings mounted over disk layers, installed at the ends of TIB and TOB.
The TIB and TID are made of 4 barrel layers installed between 20 $cm \leq r \leq$ 55 cm, $|z| \leq 65$ cm, supplemented by 3 disks at each end. TIB/TID provide 4 (r, ϕ) measurement on a trajectory using silicon sensors with 320 μm thick. In the TIB are used sensors with 10 cm length and 80 μm pitch in the two innermost layers, giving an hit resolution of 23 μm, and 120 μm pitch in the two outermost layers, giving an hit resolution of 35 μm. In the TID the mean pitch varies between 100 μm and 141 μm.

The TOB surrounds TIB/TID and is made of 6 layers installed between 55 $cm \leq r \leq 116$ cm, $|z| \leq 118$ cm. Its silicon sensors are 500 μm thick, 25 cm length and with a pitch of 183 μm on the first 4 layers and 122 μm in the remaining two layers. It provides other 6 $r - \phi$ measurements with single point resolution of 53 μm and 35 μm, respectively. TEC enclose the previous sub-detectors at 124 $cm < |z| < 282$ cm with 9 disks carrying 7 rings of micro-strips, 4 of them are 320 μm thick while the remaining 3 are 500 μm thick. TEC strips are radially oriented and their pitch varies from 97 μm to 184 μm.
The first two layers and rings of TIB, TID and TOB, as well as three out of the TEC rings, carry strips on both sides with a stereo angle of 100 milliradians to measure the other coordinate: z in barrels and r in rings. This layout ensures 9 hits per track in the silicon Strip Tracker in the full acceptance range $|\eta| \leq 2.4$, and at least four of them are two-dimensional.

The performance of the tracking system are shown in Figure 3.3, where the resolution of the transverse momentum of muons coming from J/Ψ decay, measured as explained in [46], is shown as a function of η of the track. Due to the effect of the multiple scattering, the momentum resolution of reconstructed muons (see Chapter 4) in the typical momentum range relevant for muons originating from J/Ψ decays ($3 - 30$ GeV) is completely dominated by the performance of the tracking system.

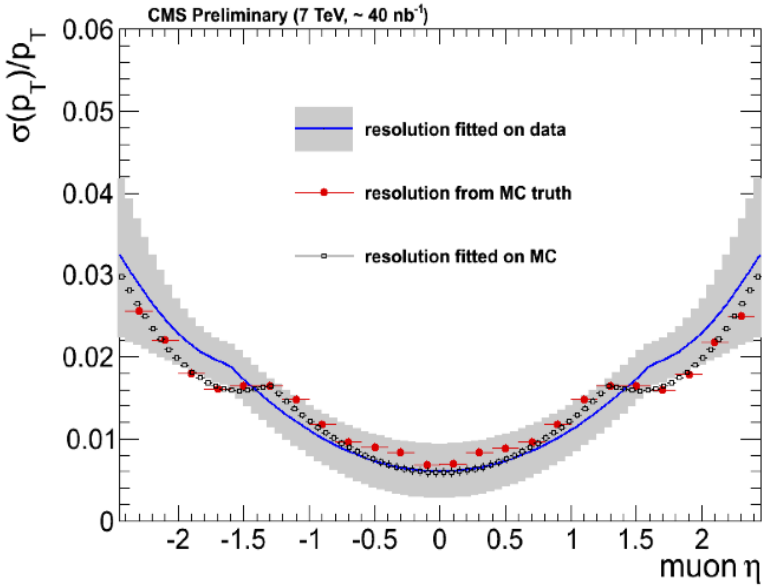


Figure 3.3: Resolution on the transverse momentum of muons coming from J/ψ as a function of η of the track. The results fitted on data (blue line) compared to Monte Carlo resolution computed from Monte Carlo truth (red points) and fit (black squares) are shown.

3.4 The Calorimiter system

The Calorimiter system is used to measure the energy of electrons, photons and hadrons, by stopping them in the material of the detector. The detector is composed of Electromagnetic Calorimeter (ECAL) [47], which measures energy of

electrons and photons, and Hadron Calorimiter (HCAL) [48], which measures the energy of hadrons.

3.4.1 Electromagnetic Calorimeter

The use of high density crystals allowed the design of a fast, fine granularity and radiation resistant electromagnetic calorimeter. Figure 3.4 shows ECAL layout. A number of 61200 lead tungstate ($PbWO_4$) crystals are mounted in the barrel part, covering the pseudorapidity range $|\eta| < 1.479$, closed by 7324 crystals in each of the two end-caps, covering the rapidity range $1.479 < |\eta| < 3.0$. A preshower detector is placed in front of the end-cap crystals. The high density ($8.28\ g/cm^3$), short radiation length ($0.89\ cm$) and small Molière radius ($2.2\ cm$) of the crystals result in a fine granularity and a compact calorimeter. Electrons and photons passing through the crystals develop an electromagnetic shower, in which the electrons produce scintillation light. The scintillation decay time of the crystals is of the same order of magnitude as the LHC bunch crossing time: about 80% of the light is emitted in 25 ns. The amount of scintillation light produced is proportional to the energy of the initial electron or photon entering the calorimeter. The light output is relatively low and varies with temperature: at the nominal operating temperature of 18^oC about 4.5 photoelectrons per MeV, with a wavelength distribution showing a broad maximum at $420 - 430\ nm$, are collected by the photodetectors. As photodetectors are used Avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the end-caps. These photodetectors are suitable because of their high gain in amplifying the scintillator light and the ability to operate in the $4T$ magnetic field.

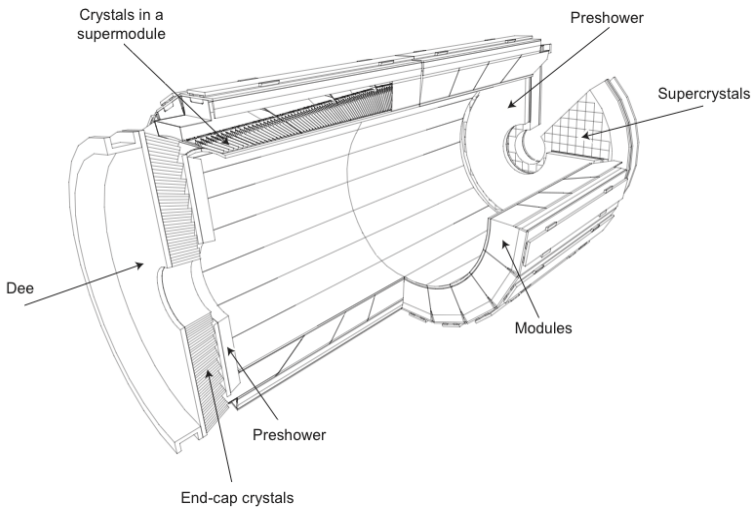


Figure 3.4: Layout of the CMS electromagnetic calorimeter showing the different components.

The ECAL energy resolution is parametrized as a function of the energy:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.3}$$

where S is the stochastic term, due to fluctuation in the lateral shower containment and in the energy released in the preshower, N the noise term, due to

Table 3.1: Constant terms entering the ECAL energy resolution parametrization.

Constant	Barrel ($\eta = 0$)	End-Cap ($\eta = 2$)
Stochastic term	2.7%	5.7%
Constant term	0.55%	0.55%
Noise(low luminosity)	0.55 GeV	0.155 GeV
Noise(high luminosity)	0.210 GeV	0.245 GeV

electronics, digitization and pile-up, and C the constant term, due to intercalibration errors, energy leakage from the back of the crystal and non-uniformity in light collection. These terms have been measured for different crystal modules during test beams with electrons [49] and the values are reported in Table 3.1.

3.4.2 Hadron Calorimeter

The HCAL detector is segmented in different parts, called towers, along η and ϕ, as it is shown in Figure 3.5 in the r,z plane. Each tower is made of superimposed layers of brass, a material with small interaction length used as absorber, and plastic scintillator, which is the active material.
The HCAL is constrained in the space left by the tracker and ECAL systems inside the solenoid volume. It was penalized in favor of good energy resolution ECAL, to increase as much as possible the sensitivity to a Higgs decaying into two photons, and a good tracker system. The absorber material has been maximized, at the expense of the active plastic scintillator, in order to contain as much as possible the hadronic component of the collision products. The resulting total interaction length is $\sim 10\ \lambda_I$. In order to improve the resolution on the missing E_T measurement by the overall calorimetric system, a calorimeter for each end-cap near the beam pipe, the Hadron Forward (HF), and a layer of scintillator in the barrel region just outside the solenoid, the Hadron Outer (HO), have been installed. Moreover, the HO increase the total interaction length over the $\sim 10\ \lambda_I$, minimizing further on the probability of punch through, that is the hadronic particles that overstep the HCAL and the solenoid reaching the muon system.
Studies on the HCAL energy resolution have been done and results are reported in [49]. An energy resolution of $\sigma_E/E = 100\%/\sqrt{E_T} \oplus 5\%$ is achieved in the central region of the detector.

3.5 The Muon system

The CMS Muon system [50] consists of sub-detectors (Drift Tube Chambers, DT, and Resistive Plate Chambers, RPC) installed in cylindrical shells with axis along the beam line in the barrel and sub-detectors (Cathode Strip Chambers, CSC, and RPC) installed along rings in the end-caps. A scheme of the Muon system, both in the transversal and longitudinal view, is shown in Figures 3.6 and 3.7 showing the four layers ("Stations") in which the barrel and end-cap chambers are organized, respectively.

The geometrical scheme used is such that the system covers, without cracks, the pseudo-rapidity interval up to $|\eta| < 2.4$. This allows muon identification,

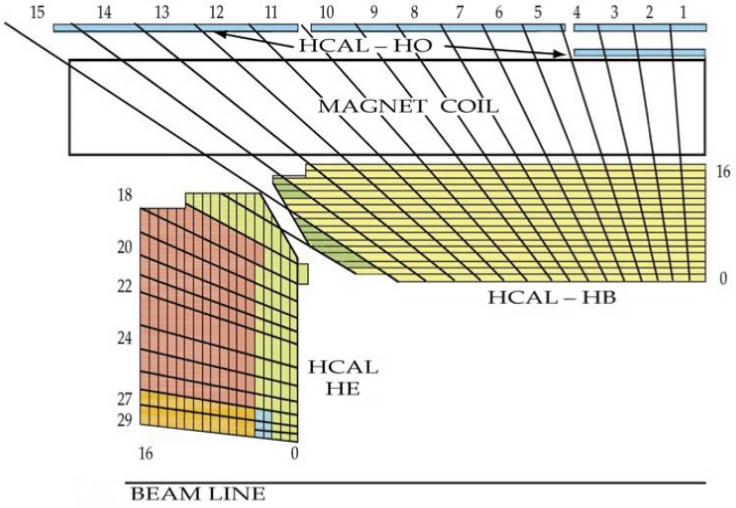


Figure 3.5: Layout of the CMS hadron calorimeter showing the different components. The tower segmentation in the r,z plane is also shown.

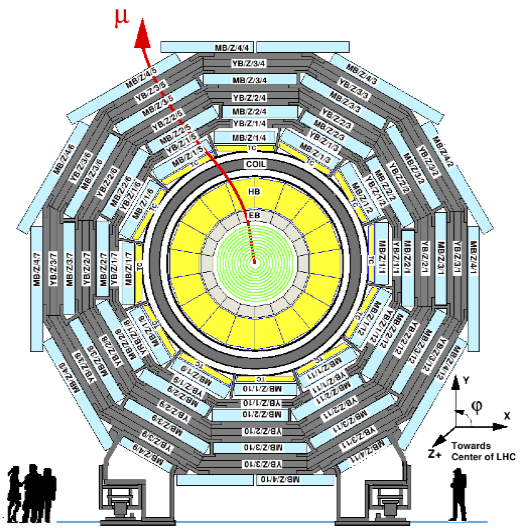


Figure 3.6: Transversal view of the CMS detector. In the outer layers, outside the solenoid, are installed the Drift Tube Chambers and Resistive Plate Chambers.

momentum measurement and event trigger in the wide interval $10^o < \theta < 170^o$. The high magnetic field in the saturated return yokes ($\sim 1.8\ T$) is such that a good momentum resolution is achieved also without the tracker system, which is of fundamental importance in the first level of trigger. In Figure 3.8 is shown the muon momentum resolution for muons produced from Z boson decays as a function of the η of the track [51], as reconstructed in the tracker and muon system. In the momentum range above $\sim 100\ GeV$, typical for muons originated by the decay of boosted Z bosons, the contribution of the muon system to the momentum resolution of the muons globally reconstructed in CMS (see Chapter 4 for details) starts to be important.

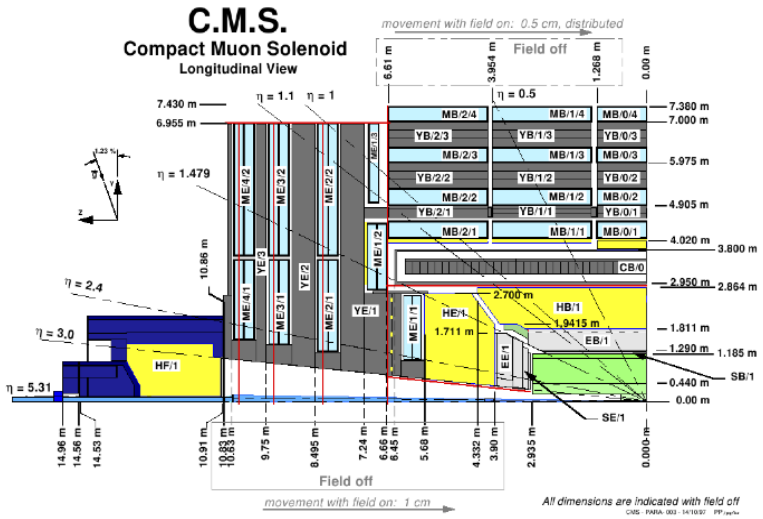


Figure 3.7: Longitudinal view of one-fourth of the CMS detector. In the outer part of the detector are installed the Drift Tube Chambers in the barrel and Cathode Strip Chamber in the end-caps. Resistive Plate Chambers are installed in both barrel and end-caps.

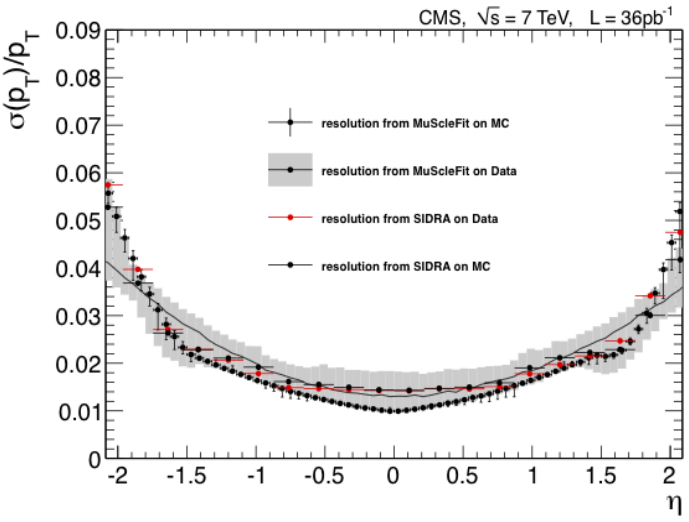


Figure 3.8: Transverse momentum resolution for muons from Z boson decay on real and Monte Carlo simulated data, obtained with two fit methods [46].

3.5.1 Drift Tube Chambers

In the barrel region, small neutron-induced background, low muon flux and relatively low and uniform magnetic field contained in the return yokes allowed the use of DT chambers. The return yoke is made of 5 wheels, each divided in 12 wedges, called sectors. There are 4 chambers housed among the yoke layers in each sector. The pseudo-rapidity range covered by the DT is $|\eta| < 1.2$.
The basic constituent of a DT is a cell, shown in Figure 3.9, with rectangular transverse section of $42 \times 13\ mm^2$. Cells are made from two parallel aluminium planes and I-shaped aluminium beams (I-beams). An electric field is produced inside the cell by an anode, a 50 μm diameter gold-plated stainless-steel wire

placed in the center of the cell, and cathodes, 50 μm thick and 11.5 mm wide aluminium tape, placed on both sides of I-beams. Field electrodes are positioned at the top and bottom of the drift cell, in correspondence of the anode, to give a proper shape to the electric field. Cathodes and electrodes are insulated with respect to the aluminium planes and I-beams, which are set to ground, by means of mylar tapes.

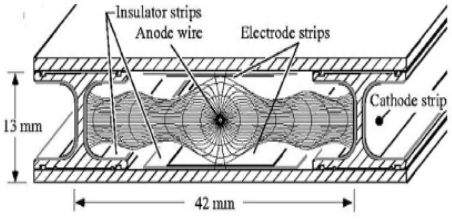


Figure 3.9: Sketch of a DT cell showing drift lines and isochrones.

Cells are filled by a mixture of 85% Ar and 15% CO_2 gas. A muon crossing a cell ionises the gas, producing electrons that drift toward the anode. Since the gas mixture and the electric field are such that the electron drift velocity is constant, with a value of $v_{drift} \simeq 54\ \mu m/ns$, the drift time can be used to measure the position of the muon inside the cell. The drift time is obtained from the time measured using a high performance Time to Digital Converter (TDC), after subtraction of a time pedestal. The time pedestal contains contributions from the latency of the trigger and the propagation time of the signal, within the detector and the data acquisition chain. Each TDC time is associated to a given cell (Digi) and is read by the data acquisition. The hit position, that is the distance of the muon track with respect to the anode wire, is reconstructed as:

$$x_{hit} = t_{drift} \cdot v_{drift} = (t_{TDC} - t_{ped}) \cdot v_{drift} \tag{3.4}$$

where t_{TDC} is the measured time and t_{ped} is the time pedestal.
Since the distance between the anode and a cathode is 21 mm, the maximum drift time is $\sim 390\ ns$ corresponding to ~ 15 bunch-crossing. This time is small enough to face the muon flux coming from pp interaction, using a suitable buffer read-out electronic. The drift time resolution for each cell is $\sim 3\ ns$, which implies a spatial resolution for a single muon hit of $\sim 150\ \mu m$.
A DT chamber is made of 3 super-layers (SL), each built from 4 layer of cells, shown in Figure 3.10. These 4 layer of cells are glued together with the cells parallel each other and staggered by half a cell between two consecutive planes. This is a technique that solves the left-right ambiguity due to the fact that a single cell alone would not be able to distinguish the crossing side of a muon with respect to the anode [52].
The two outer SL in a DT have the anode parallel to the beam direction, allowing the measure of the muon track on the (r,ϕ) plane, where it is bend by the magnetic field. The measurement in the r-z plane is done by the remaining SL, with anodes perpendicular to the beam direction. This SL is not installed in the outermost DT chambers.
The mean-timer technique [53] applied to 3 consecutive layer of cells allows the identification of the bunch-crossing of the pp collision that produced the muon. This is of great importance in the muon trigger system of CMS.

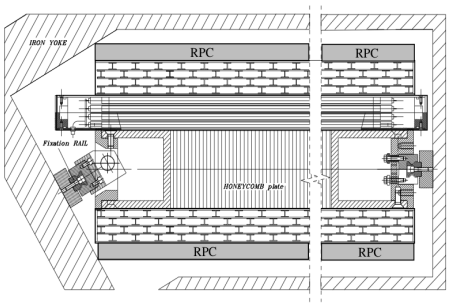


Figure 3.10: Sketch of the (r,ϕ) section of a DT chamber installed in the iron yoke. The two SLs with anode along the beam direction and the one with anode perpendicular to it are shown.

3.5.2 Cathode Strip Chambers

In the end-caps region background, muon flux and magnetic field are high. Moreover the magnetic field is not uniform. This draws to the use of Cathode Strip Chambers (CSC) in this region, suitable for their fast response, fine granularity and radiation hardness. The CSC are mounted in concentric rings on the end-cap yokes (YE$\pm$1,$\pm$2,$\pm$3), covering the pseudo-rapidity range $0.9 < |\eta| < 2.4$. All the rings, but the external one in YE$\pm$1, have the chambers overlapped in ϕ to avoid gaps in the geometrical acceptance.
As the Figure 3.11 shows, a CSC has trapezoidal shape in the r-*phi* view covering a region of 10^o or 20^o in ϕ.

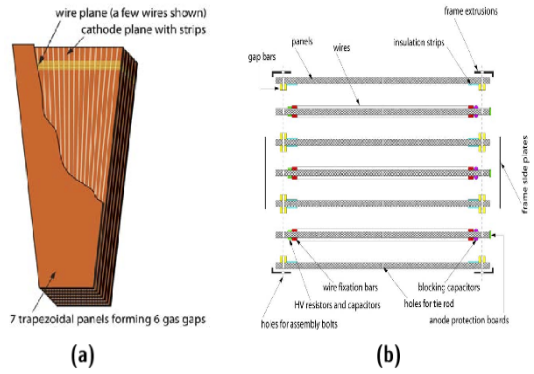


Figure 3.11: (a) Scheme of a CSC. (b) Detailed view of the 6 layer mechanical structure of a CSC.

The technology used to develop CSC is the multi-wire proportional chamber: 6 layers of anode wires are superimposed to 7 cathode plane. The volume between two cathode planes is filled with a gas mixture of $Ar/CO_2/CF_4$ in a percentage 40%, 50% and 10%. The anode wires are perpendicular to the beam direction. On the cathode plane are built radial strips at values of $\Delta\phi$ constant.
A muon crossing the volume between two cathode planes of a CSC produces an electron avalanche in one of the anode wires, giving the radial coordinate of the muon track. The charge collected on the anode induces a charge on the closest cathode strips, producing a signal which interpolated gives a precise measure of the ϕ coordinate. The resolution of a measured hit is 100 μm in the radial and $10 mrad$ in the ϕ coordinates.

3.5.3 Resistive Plate Chambers

Both the DT and CSC provide trigger on muons with a p_T resolution of 15% and 25% respectively at the trigger Level-1 (see section 3.7.1). An additional, dedicated muon trigger device based on RPC has been installed in the muon system. These detector combine adequate spatial resolution with a time resolution comparable to that of scintillators. The last feature makes RPCs able to tag the time of an ionizing event in a much shorter time than the nominal 25 ns between 2 consecutive bunch crossings, identifying unambiguously the bunch crossing to which a muon track is associated. With the given spatial resolution, RPCs are able to provide an estimate of the transverse momenta suitable for the muon trigger, independent from that provided by the other muon systems.
In each barrel sector are installed 6 RPC: two in each one of the two inner muon stations chamber and one in the remaining ones. In the end-caps there is a RPC layer in each ring. As it is shown in Figure 3.12, RPC is made of two chambers superimposed. Each of the two chambers is built from two layers of Bakelite, material with high resistivity ($\sim 10^{10}\ \Omega cm$), filled with a gas mixture of $C_2H_2F_4/iC_4H_{10}/SF_6$ in a percentage 96.2%, 3.5% and 0.3%. The external surface of the Bakelite layers is covered with graphite, making the chamber a capacitor able to produce a strong electric field inside the volume. Between the two chamber it is inserted a layer of aluminium strips insulated from the graphite. A muon crossing a RPC produce a charge avalanche in each of the two chambers which, collected on the Bakelite layers, induce a signal on the strips identifying the time and position of the muon.

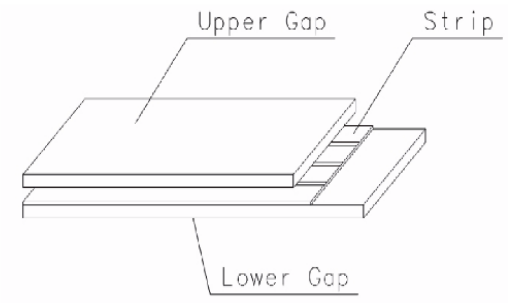


Figure 3.12: Sketch of a RPC. The two chambers (Upper Gap and Lower Gap) and the aluminium strips are shown.

3.6 Forward detectors

3.6.1 CASTOR

The CASTOR [42] (Centauro And Strange Object Research) detector is a quartz-tungsten sampling calorimeter installed at 14.38 m from the interaction point, covering the pseudorapidity range $5.2 < |\eta| < 6.6$. The main advantages of quartz calorimeters are radiation hardness, fast response and compact detector dimensions, making them suitable for the experimental conditions encountered in the very forward region at the LHC. CASTOR is used in both pp and heavy-ion physics in general, but specifically it is designed to search for exotic events that are likely to appear in the very forward region of the CMS detector. It is made of

tungsten absorber and quartz plates sandwich, with 45^o inclination with respect to the beam axis, azimuthally divided in 16 sectors, longitudinally segmented in 14 reading units (2 electromagnetic and 12 hadronic). The detector shape is cylindrical with an inner radius of $r_{in} = 3.7\ cm$, an outer radius of $r_{out} = 14\ cm$ and a depth of 10.5 λ_I. An energy resolution better than 1% is achieved in the acceptance range.

3.6.2 The Zero Degree Calorimeter

The Zero Degree Calorimeters [42] (ZDC) is located downstream of the first beam dipole magnets between the two beam pipes at $\sim 140m$ on each side of the CMS IP. It complements the very forward region of CMS, covering the region $|\eta| > 8.3$, useful for heavy ion and pp diffractive studies. Sampling calorimeters using tungsten and quartz fibers have been chosen for the detection of the energy in the ZDC with a design similar to HF and CASTOR. Each ZDC has two independent parts: the electromagnetic (EM) and hadronic (HAD) sections. The total depth of the combined system is $\sim 7.5\ \lambda_I$.

3.7 Trigger and data acquisition systems

At the nominal bunch crossing interval of 25 ns, the pp collisions rate provided by LHC is 40 MHz. Whereas the rate at which the events can be written on tape is technically limited to $\sim 100\ Hz$, given the average size of a single event data of $\sim 1\ MByte$. Moreover, even if all the events could be stored on tape, it would be impossible to process and analyze all the resulting large amount of data. These considerations require physics selections of the events during data taking, known as online selections, which implies a drastic reduction of the rate of events that have to be stored. This task is performed by the trigger system, divided into two steps called Level-1 (L1) Trigger [54] and High Level Trigger (HLT) [55], respectively. The trigger system is designed in such a way that the combined L1 and HLT selections reduce the event rate by about 6 orders of magnitudes.

3.7.1 Level-1 Trigger

The L1 Trigger consists of custom-designed, largely programmable electronics in part installed on the sub-detectors and designed to reduce the rate up to 100 kHz. The event selection at this level is performed on rough data from the calorimetric and muon systems, while the full information of a bunch crossing from the sub-detectors is stored in a *First In-First out* (FIFO) memory. The time within which L1 has to perform the selection is 3.2 μs at the nominal LHC bunch crossing time spacing. This time is fixed by the total amount of bunch crossing that can be temporarily stored in the FIFO memory, which is 128. For the events passing the L1 selections the information stored in the FIFO are read by the *Data Acquisition System* (DAQ) and processed by the HLT.
The L1 Trigger has local, regional and global components, has it is sketched in the Figure 3.13. For the muon system all the sub-detectors, DT, CSC and RPC, contribute to the trigger. Each DT and CSC chamber generates a local trigger. Local trigger objects are segment of tracks in the r-ϕ plane and sample of hits in the r-z plane from DTs and tridimensional segments of track from CSC. RPCs generate regional trigger objects which are muon track candidates. Local trigger

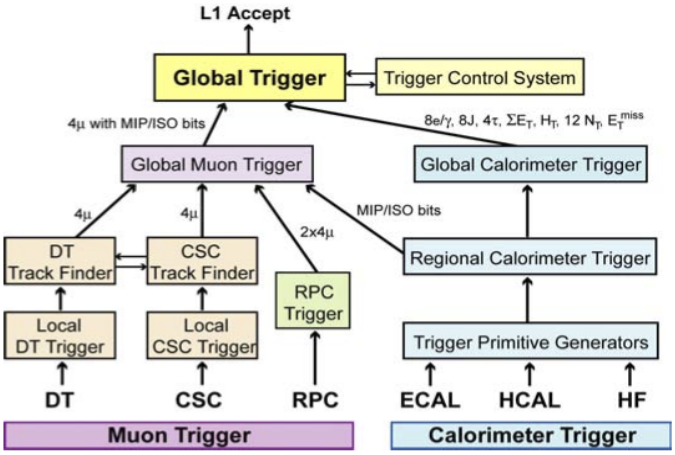


Figure 3.13: Architecture of the L1 Trigger.

objects from DTs and CSCs that pass selections are sent to the *Regional Muon Trigger*, consisting of DT and CSC *Track Finder*. Here muon track candidates are built in different regions of the muon system. The regional muon candidates are classified according to their momentum and quality, then the 4 best candidates in each region for DT, CSC and RPC, are sent to the *Global Muon Trigger* (GMT). The GMT combine the *Track Finders* and RPC regional trigger muons, making muon candidates with refined momentum and quality in the whole CMS volume.
The calorimeter system is divided in trigger towers each with $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$ coverage in the η-ϕ plane up to $\eta = 1.74$. Beyond this region the towers are larger. In each trigger tower the energy deposits in ECAL crystals and independently energy deposits in HCAL towers are summed to obtain the energy in the trigger tower. This information is sent to the *Regional Calorimeter Trigger* (RPG), able to determine electron/photon candidates, transverse energy sum (defined as $E_T = Esin\theta$), informations for the identification of jets from τ decays and relevant informations concerning muons in a calorimeter region. A calorimeter region consists of a 4×4 trigger tower in all the calorimeter system but the HF, where a region coincides with one trigger tower. The *Global Calorimeter Trigger* (GCT), using the RPG outputs, determines in the whole calorimeter system electron/photons, jets, the total transverse energy E_T, the missing transverse energy $\not{E}_T$, the scalar transverse energy sum of all jets above a programmable threshold (denoted H_T).
The particle candidates come out from GMT and GCT with measured transverse momenta or energy, η-ϕ coordinates and quality. These, together with global quantities like E_T, $\not{E}_T$, H_T, are analyzed by the *Global Trigger* (GT) which take the final decision to accept or reject an event at L1. The GT has different stages, among which the most important is the Global Trigger Logic (GTL), where algorithms calculations are performed. Basic algorithms apply only transverse momentum, energy and jets multiplicity cuts on the GMT and GCT candidates and global variables. Since additional information is also available, like the position of particles and their quality, more complex algorithms can also be implemented. Up to 128 algorithms can be executed in parallel. The algorithm output is one bit, sent to the Final Decision Logic (FDL), which takes the decision on the event accordingly to the trigger mask applied. The selected

events are sent to the data acquisition for read-out.

3.7.2 High Level Trigger

The HLT is designed to reduce the $\sim$ 100 kHz of the L1 output down to $\sim$ 100 Hz. A first step of HLT consists in the identification and measurements of particle candidates and global variables using only the information coming from calorimeters and muon system (the step is denoted L2). It must be stressed that at this stage the full information from the detector systems (the same available at the offline reconstruction stage) is used. A selection is performed on the basis of the information provided by the L2, thus resulting in improved energy and momentum resolution in triggering objects live electrons and muons with respect to L1. The tracker system information is used in the second HLT step, denoted as L3. This implies the analysis of a larger amount of data, resulting in a more time consuming process. At this level the particle candidates are reconstructed with a momentum or energy resolution close to the one achieved with the offline analysis. Suitable thresholds on the measured quantities of particle candidates and the usage of correlations between variables related to different physical objects are able to reduce the output rate of 3 orders of magnitude with respect to L1.

3.7.3 Data Acquisition System

The Data Acquisition System (DAQ) [55] reads data coming from sub-detectors and carry them to a farm of processors built with $\sim$ 1000 commercial CPU. This is the *Filter Farm* (FF), where the HLT algorithms run performing the events selection. The complexity of the event reconstruction at the HLT level is limited only by the available computing power, which is proportional to the number of CPUs. To achieve a rejection factor of 10^3 a number of 1000 CPU are needed. A scheme of DAQ is shown in Figure 3.14.

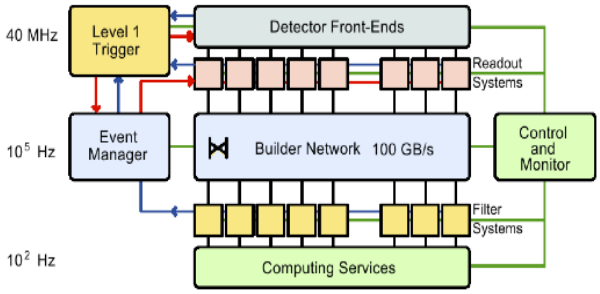


Figure 3.14: Architecture of the Data Acquisition System.

The data read by each sub-detector are stored continuously in 40 MHz pipelined buffers by Front-End Systems (FES). If the event is accepted by the L1, a signal is sent by the Timing, Trigger and Control (TTC) system and corresponding data are extracted from the buffers and pushed into the DAQ system by the Front-End Drivers (FEDs). Front-end Read out Links (FRLs) read data from FEDs and are able to merge data from two FEDs. The event builder assembles the data belonging to the same L1 Trigger from all FEDs into a complete event and transmits it to one Filter Unit (FU) in the FF of the Event Filter for further processing. The main technological challenge in building the DAQ was to develop a system able to provide to the FF assembled events with a dimension of

$\sim 1\ MByte$ at 100 kHz. Back-pressure from the down-stream event-processing, or variations in the size and rate of events, may give rise to buffer overflows in the sub-detectors front-end electronics, which would result in data corruption and loss of synchronization. The Trigger-Throttling System (TTS) protects against these buffer overflows. It provides fast feedback from any of the sub-detector front-ends to the Global Trigger Processor (GTP) so that the trigger can be throttled before buffers overflow.

3.8 Data Quality Monitoring

The purpose of the Data Quality Monitoring (DQM) [49] system is to guarantee the quality of physics data collected by the CMS experiment. The system permits to detect problems as early as possible in order to maximize the amount of high quality data recorded. An important step after the data are recorded and reconstructed with the CMS software is the "data certification" process, which determines the quality of the data, discarding those events that are not suitable for physics analysis, exploiting the DQM tools. The data used for the analysis presented in this thesis have been judged has good data for physics analysis by this certification process.
The DQM infrastructure provides an interface for the production of monitoring objects (histograms showing interesting monitoring variables), and a tool for their visualization. The visualization tool, *Graphical User Interface* (GUI), is a web-based user interface built in *Python*, $C++$ and *JavaScript* code. The monitoring objects are produced in two steps. At the first step (source processes) event data are processed and interesting informations are produced for monitoring purposes. A second step (client processes) elaborates the information from the sources, performing tests that end up in histograms containing the information about the status of the CMS system. These two steps are integrated in the CMS software. The monitoring is performed online during the data taking (Online DQM) and offline after the full event reconstruction (Offline DQM). The online checks provide a fast evaluation of problems that can affect data taking, so that a prompt reaction can avoid loosing good data. Also, a first evaluation on data quality is performed online. The offline checks use the information available from full event reconstruction with more refined calibration and specific datasets for each sub-detector to provide a more detailed evaluation of the data. Offline DQM is used to certify the quality of reconstructed data and validate calibration results, software releases, and simulated data. The data certification leads to a list of good data to be used for the analysis. The information is stored in a *JavaScript Object Notation* (JSON) file, containing a time interval list of the periods of data-taking suited for final physics analyses.

The Online and Offline DQM workflows, depicted in Figure 3.15, can be summarized as follows:

3.8.1 Online DQM

A subset of the data-stream acquired and recorded by the CMS detector are reconstructed and monitored directly during the data-taking. This allows to give immediate feedback about the detector status. Events are delivered to data quality monitoring applications at about $5 - 10\ Hz$, using the workflow depicted in

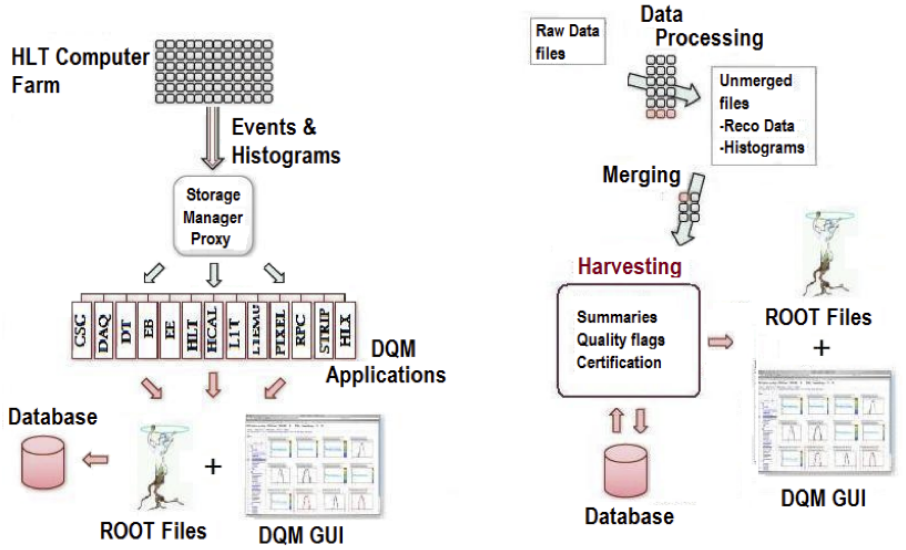


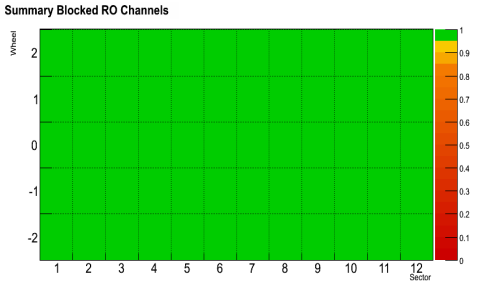
Figure 3.15: DQM workflows for online (left) and offline (right).

Figure 3.15 (left). This stream of data (DQM monitoring event stream) is filtered from the main data stream[2] by applying specific trigger path selections. Moreover, each DQM application selects data from the DQM monitoring stream by trigger selection specific for the needs of the sub-detector (e.g. the DT system use muon HLT paths). Each DQM application runs its own algorithms, source and client steps, producing the monitoring objects. All the resulting monitoring objects (in form of histograms) are made available to the DQM GUI for visualization in real time and also saved for future analysis.
At the Online DQM stage, checks on readout errors, occupancy of the detector channels, rate of the trigger primitives and local reconstructed objects (e.g. track segments in the muon chambers) are performed in order to test the status of the sub-detectors and the goodness of recorded data. At this level, data are declared good if a large fraction of the detector channels register hits. A small number of readout errors is tolerated.
As an example, Figure 3.16 shows some distributions that are checked during data-taking for the DT sub-system. The status of the read-out, the digitized hits occupancy, the local reconstruction of segments and the local trigger status are summarized in the distributions shown. A color code from red to green is used to easily spot the presence of a possible problem and its severity. If a problem is spotted in one of these summary plots, a much larger number of detailed plots are available to find out its origin.

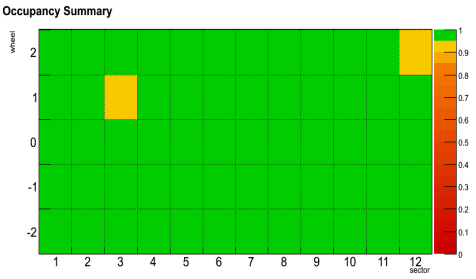
3.8.2 Offline DQM

Offline DQM, schematically represented in Figure 3.15 (right), runs as part of the reconstruction process at the central CERN data storage system (Tier-0), and of the re-reconstruction process with updated detector conditions at the other

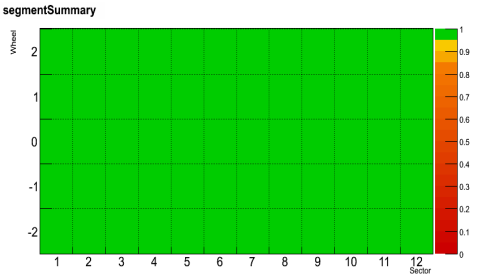
[2]The flux of data acquired by the CMS detector is divided in different main streams, depending on which way those data will be used for. For instance, a stream (stream A) collecting all the data that will be used for physics analysis is defined.



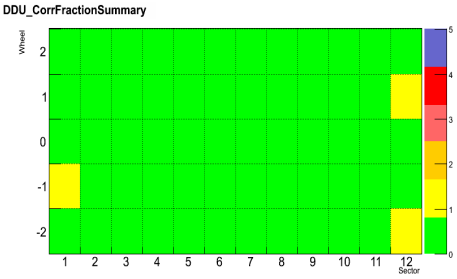
(a) Summary of the Read-Out (RO) channels status.



(b) Summary of the digi occupancy.



(c) Summary of the status of the reconstruction of segments at DT chamber level



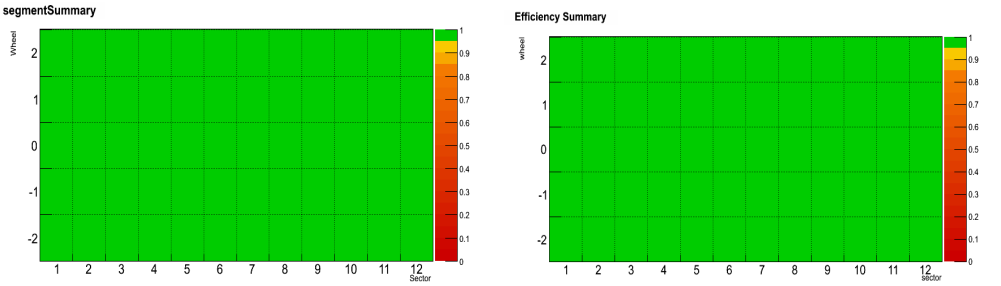
(d) Summary of the status of the DT local trigger.

Figure 3.16: Online monitoring objects showing the summary of the DT subsystem status for a reference run with high statistics taken during 2012 data-taking period. They are 2D distributions with the wheel number in the vertical axis and the sector number in the horizontal axis. Each bin shows the behavior of a single sector.

storage systems distributed on the Wide Area Network around the world (Tier-1s). It also runs at the stage of validation of the CMS software releases, simulated data, and alignment and calibration results. Two main steps define the Offline DQM workflow. As for the online DQM, in the first step the histogram monitor objects are created and filled with information from the CMS event data. In a second step, known as *harvesting*, the histograms are extracted from the event data files and summed together across the entire run to yield full event statistics on the entire dataset. All the final monitoring objects are saved and uploaded to the DQM GUI web server.
At the Offline DQM stage the quality of the reconstructed physics object can be checked in data from the different datasets[3]. Again as an example, Figure 3.17 shows the distributions that are checked after the full event reconstruction offline to certify the recorded data at the DT sub-system level. For DTs, the Offline DQM applications runs over the *SingleMu* dataset, which is rich of events with muons. The two 2D distributions show a map for the whole DT sub-system of the reconstructed segments quality, which is good if the bins are green. The two 1D distributions display the mean and width of the segment hit residual distributions for each chamber. These two variables are sensible to problems like changes in

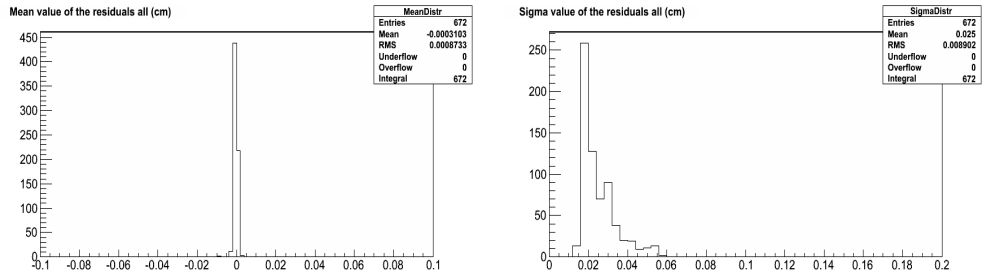
[3]Each dataset is made by events selected by a given group of HLT trigger paths. These paths have roughly the same trigger logic. For instance, the SingleMu dataset is defined by all the HLT paths that require a single muon in the event.

calibration constants conditions, gas mixture and pressure.
Like in the Online DQM, also for the offline if a problem is spotted in the main summary plots shown in Figure 3.17, many additional detailed plots are available to track down its origin. At this level, the data lost because of problems might be recoverable, with a re-reconstruction starting from raw data, or not recoverable anymore. This depends on the cause of the problem: possible data lost because of bad calibration constants can be recovered with a new set of calibration constants in the re-reconstruction.



(a) Summary of the status of the reconstruction of segments.

(b) Summary of the efficiency of the association of segments to standalone muon tracks.



(c) Distribution of the mean value of the residuals on the distance from the wire for the 1D RecHits compared to segments.

(d) Distribution of the sigma (from Gaussian fit) of the residuals on the distance from the wire for the 1D RecHits compared to segments.

Figure 3.17: Offline monitoring objects showing the summary of the DT subsystem status for a reference run with high statistics taken during 2012 data-taking period. They are two 2D distributions, with the wheel number in the vertical axis and the sector number in the horizontal axis, and two 1D distributions.

Chapter 4

CMS event reconstruction

4.1 CMS framework

The final goal of the CMS Software is to process and select events in the HLT and the off-line reconstruction stage, delivering the results to experimenters within the CMS Collaboration and providing tools to analyze the processed information in order to produce physics results. The overall collection of software packages, referred as CMSSW [49], is built around a Framework, an Event Data Model (EDM), Services, and reconstruction modules that process event data so that physicists can perform analysis. The physics and utility modules are written by detector groups. The modules can be plugged into the application framework at run time, independently of the computing environment. The software is developed keeping in mind not only performance but also modularity, flexibility, maintainability, quality assurance and documentation. CMS has adopted an object-oriented development methodology, based primarily on the $C++$ programming language.
The primary goal of the CMS Framework and EDM is to facilitate the development and deployment of reconstruction and analysis software. The event data model is centered around the "Event". The Event holds all data that were taken during a triggered physics event. Events are processed by passing the Event through a sequence of *modules*. The purpose of a *module* is to allow independent development and verification of distinct elements of triggering, simulation, reconstruction, and analysis. The exact sequence of *modules* is specified by the user. When an Event is passed to a *module*, that *module* can get data from the Event, process them, and put data back into the Event. The sequence of modules is executed by the CMS Framework executable, *cmsRun*. A user-written configuration file for the executable, in *Python* code language, defines the sequence of modules and all the information needed to run them. Each *module* can be configured with a set of parameters, which are set in the configuration file.
To be able to fully process an Event, for instance in an analysis, additional information outside of the Event itself (like magnetic field measurements, calibration and alignment constants) are required. Those informations are provided by Services, that deliver all non-Event data.

4.2 Event reconstruction and relevant physics objects for the present analysis

Event reconstruction is the operation of constructing physics quantities from the raw data, the output from the sub-detectors, collected by the experiment. At the high level reconstruction step (the global reconstruction) these physics quantities correspond to physics characteristics of a particle, so at this level reconstruction corresponds to the identification and measurement of a particle produced in the LHC collisions. The reconstruction process is a collection of independent units, each one providing a set of corresponding reconstructed objects as output. Each reconstruction unit is implemented in the CMS framework as a *module*, in which algorithms, that are able to process data from the Event, are implemented. The reconstructed quantities are then stored again in the Event for usage in high level analysis.
A widely used Event data reconstruction technique is the so called *particle flow* [56], which enhance the performance of the detector. Particle flow attempts to reconstruct all stable particles in an event by combining information from all sub-detectors. This way, an optimal determination of particle direction, energy and type is obtained. The algorithm categorizes all particles into the following five types: muons, electrons, photons, charged and neutral hadrons. The resulting list of particles can then be used to construct a variety of higher-level objects and observables such as jets, missing transverse energy (E_T^{miss}), taus, lepton and photon isolation variables, b-jet tagging discriminators, etc.
In the following, the reconstruction of the main physics quantities used in the analysis presented in this thesis are briefly described.

4.2.1 Tracker tracks

The default reconstruction of charged particles tracks in the silicon tracker system is performed by the combinatorial track finder (CTF) [57]. Triplets or pairs of hits with an additional constraint from the beamspot or a vertex[1] are used as initial estimates, or seeds, of tracks. The seeds are then propagated outward in a search for compatible hits. As hits are found, they are added to the seed trajectory and the track parameters and uncertainties are updated. This search continues until either the limit of the tracker is reached or no more compatible hits can be found, yielding the collection of hits that belong to the track. In the final step, this collection of hits is fit to obtain the best estimate of the track parameters.

4.2.2 Primary interaction vertex

The primary vertex identification [58] starts from the reconstructed tracks, selected on the basis of their compatibility with the beam spot, number of hits and normalized track χ^2. The tracks are clustered into several primary vertex candidates, according to the z-coordinate of the point of closest approach of the tracks to the z-axis. Several primary interaction vertices can be found in the same bunch crossing due to the occurrence of more than one pp collision, the pile-up. A vertex fit is performed in (x, y, z) coordinates for each primary vertex candidate

[1] This is done recursively: since the vertex is defined using the reconstructed tracks, then at the beginning the beam spot is used as constraint, updating the track information as soon as primary vertex candidates are available.

using the corresponding tracks. The primary vertex candidates compatible with the beam line are retained.

4.2.3 Jets

Quarks or gluons produced in the LHC collisions cannot be observed directly, but they fragment into stable hadrons, which can be detected in the tracking and calorimeter systems. These hadrons generated from a parton produce a roughly collimated flow of particles called *jet*.
The identification and measurement of physics quantities related to jets [59] is performed by different kind of algorithms, which assembles a collection of objects (calorimetric towers, particles) to obtain the final jet. These algorithms provide a good measurement of the initial parton energy and direction. Beside this, they should be *collinear safe*, such that the result is unchanged if e.g. the energy carried by a single particle is instead distributed among two collinear particles, and *infrared safe*, such that the result of the jet finding is stable against the addition of soft particles. Jet algorithms which do not satisfy one of the two conditions yield ambiguous results and lead to unnecessary uncertainties when applied to calculations in perturbative theory.
The following jet algorithms are implemented in the CMS framework:

- *Iterative Cone* (IC): is a simple cone-based algorithm used since the first CMS online operations in the HLT, thanks to its short and predictable execution time. Calorimeters towers or particles with $E_T > 1\ GeV$, sorted in descending order, are considered as starting points (seeds) for an iterative search for stable cones such that all inputs with $\sqrt{\Delta\eta^2 + \Delta\phi^2} \leq R$ from the cone axis are associated with the jet, with R the cone radius. A cone is considered stable if its geometric center agrees with the (η, ϕ) location of the sum of the constituent four vectors within a certain tolerance. However, this jet algorithm is neither collinear- nor infrared- safe, and then is not suitable for physics analysis;

- *Midpoint Cone* (MP): as the IC this algorithm is based on an iterative procedure to find stable cones, but the infrared-safety is addressed. This improvement is obtained by introducing a second iteration of the list of stable jets found in the first iteration. In the second iteration for every pair of jets that are closer than the cone diameter, a *midpoint* is calculated as the direction of the combined momentum. These midpoints are then used as additional seeds to find the final jets. Since each input object to the algorithm can be associated with several jets, splitting/merging algorithms are applied afterwards to ensure each input belongs to a single jet. This algorithm has proved to be infrared safe for Leading and Next-to-Leading order perturbative QCD, but not beyond, and not collinear safe;

- *Seedless Infrared-Safe Cone* (SISCone): in contrast to iterative cone clustering algorithms, which look for stable cones by starting only at the particles above a threshold (the seed), this algorithm searches for all possible stable cones. The SISCone algorithm exploits the fact that a circle enclosing a set of particles can be moved around such that two of the particles lie on its circumference. Conversely, all possible stable circles of radius R can be determined by testing the circles defined by a pair of particles and radius

R. The algorithm first finds all the stable cones. Then, these stable cones are split/merged using the same procedure as the Midpoint Algorithm. It is collinear- and infrared-safe to all orders of perturbative QCD and demands only slightly higher execution time compared to the Midpoint Cone algorithm;

- *Inclusive* k_T: this is a cluster-based jet algorithm. The cluster procedure starts with a list of input objects, stable particles or calorimeter cells. For each object i and each pair (i, j) the following distances are calculated:

$$\begin{aligned} d_i &= p_{T,i}^2 \\ d_{ij} &= min(p_{T,i}^2, p_{T,j}^2)[(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2]/R^2 \end{aligned} \tag{4.1}$$

  where R^2 is a dimensionless parameter normally set to unit. The algorithm search for the smallest d_i and d_{ij}, and call it d_{min}. If d_{min} is a d_{ij}, the objects i, j are merged into a new object with momentum the sum of the momentum of i, j. If a distance of type d_i is the smallest, then the corresponding object i is removed from the list of input objects and filled into the list of final jets. The procedure is repeated until all objects are included in jets. This algorithm is both collinear- and infrared-safe to all orders of perturbative QCD;

- *anti-*k_T: is a variation of the k_T algorithm. It uses 4.1 with $1/p_{T,i}^2$ instead of $p_{T,i}^2$, everything else is the same.

The energy measurement obtained by this jet algorithms does not match the energy of the initial parton originating the jet. The jets at this level are also called *raw* jets. The bias in jet energy reconstruction is caused by different reasons, among which the most important are: non-linear response of the calorimeters, detector segmentation, presence of material in front of calorimeters, electronic noise, noise due to physics (pile-up, PU, of interactions from same bunch crossing). The following levels of correction are applied to the raw jets in order to obtain the energy value that is closer to the true energy of the initial parton:

1 *Offset* (L1): the PU and electronic noise effects are removed. This correction can be estimated using events collected by a random trigger, without any preconditions except a beam crossing, and referred to as zero bias events. The offset contribution is evaluated as the average calorimeter energy deposited inside a cone of radius R, and depends on the η of its axis;

2 *Relative* (η) (L2): the variation in jet response with η is flattened, using the *dijet* imbalance method applied on collision data. This method is based on the principle of transverse momentum conservation: in the case of a two parton final state, the resulting particle jets have equal transverse momentum, in an ideal case. The unbalance between the jets transverse momentum that is observed on average, is due to the variation of the jet response across the detector versus η;

3 *Absolute* (p_T) (L3): the calorimetric energy response to a particle level jet is smaller than unity and varies as a function of the jet p_T. The absolute

correction removes these variations and make the response equal to unity at all p_T. This corrections is obtained from simulation using the Monte Carlo truth, and from real data exploiting $\gamma + jets$ and $Z + jets$ events;

4 *Electromagnetic Fraction* (L4): detector response is different for electromagnetic particles and hadrons, and since the ratio between the electromagnetic and hadron energy, $R_{E_{em}/E_{had}}$, depends on the energy, then the calorimetric response is not flat as a function of this ratio. A correction is applied to flatten the response versus $R_{E_{em}/E_{had}}$. This correction is obtained from simulation using the Monte Carlo truth, and from real data exploiting $\gamma + jets$ events;

5 *Flavor* (L5): the jet p_T is corrected for the specific parton flavor. L3 correction is for QCD mixture of quarks and gluons, but they differ each other. Moreover the jet shape depends on the quark flavor. This implies that the calorimetric response depends also on the parton type. These corrections can be developed from simulated events using Monte Carlo truth and data-driven methods using $t\bar{t}$ events. In the latter case, the invariant mass of the top quark reconstructed exploiting the $t \to W^+\bar{b} \to jj\bar{b}$ is constrained to the world average measurement, adjusting the light and b-quark jet corrections. The light jet corrections are cross checked by calculating the invariant mass of the two jets assigned to the W decay;

6 *Underlying Event (UE)* (L6): the hard interaction produced in a *pp* collision is surrounded by the remaining particles in the *pp* fragmentation (underlying event) that affect the measured energy of partons from hard interaction. This effect can be removed, even if it is difficult since the underlying event depends on the details of the hard interaction;

7 *Parton* (L7): the aim is to correct the jet back to the originating parton. This correction is obtained exploiting the Monte Carlo truth;

L1 to L3 corrections are mandatory, whereas the remaining correction levels are of second order of importance; in particular, the last four may or may not be included in the final analysis depending on their relevance and impact on the final systematics uncertainty in the specific physics process under study.
The jet p_T resolution is shown in Figure 4.1 for jet reconstructed with the particle-flow tecnique using the anti-k_T algorithm. The results are obtained with the asymmetry method for QCD simulated events and data [60]. This method exploits momentum conservation in the transverse plane of dijet events and is based almost exclusively on the measured kinematics of the dijet events. An asymmetry variable is defined as the ratio between the vectorial p_T differences over the sum of the dijet system. The variance of the distribution of this variable is related to the jet p_T resolution.

4.2.4 b-quark Jets

The jets produced by b-quark hadronization are characterized by quantities that allow to identify them (b-tagging), discriminating jets from gluons (g), light flavor quarks (u, d, s) and c-quark [61]. Especially the b-quark long lifetime with respect to the other quarks and its semileptonic decay properties can be exploited for the b-tagging.

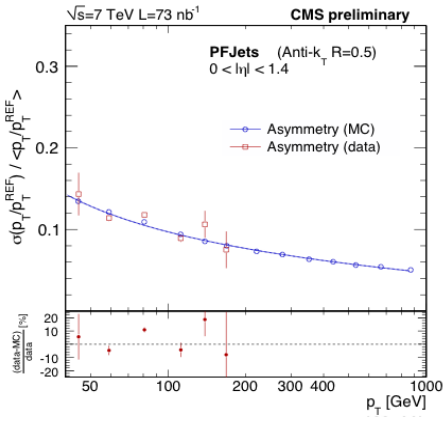


Figure 4.1: Particle-flow jet p_T resolution for jet reconstructed using the anti-k_T algorithm in the region $0 \lesssim \eta \lesssim 1.4$. The asymmetry method is used, for both data (red) and MC (blue).

The impact parameter of a track with respect to the primary vertex can be used to distinguish the decay products of a b hadron from other tracks. A sign is associated to the impact parameter, which is given by the scalar product of the vector pointing from the primary vertex to the point of closest approach of the track with the jet direction. Tracks originating from the decay of particles traveling along the jet axis will tend to have positive impact parameter values. In contrast, the impact parameters of prompt tracks can have positive or negative values. A good discrimination variable for b-tagging is the impact parameter significance, S_{IP}, defined as the ratio of the signed impact parameter to its estimated uncertainty. The distribution of S_{IP} is shown in Figure 4.2 [62].

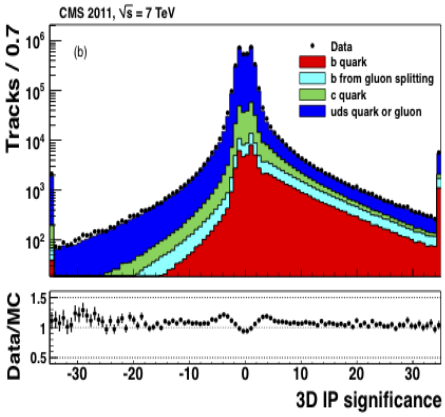


Figure 4.2: Distribution of the impact parameter significance as measured in the (x, y, z) CMS reference frame. The data are compared with the Monte Carlo simulated events split for the different flavor component.

In CMS an algorithm based on S_{IP} is used for b-tagging, called *Track Counting* (TC). It sorts the tracks in a jet by decreasing values of impact parameter significance. The probability to have several tracks with high positive value of the S_{IP} is low for light-flavor jets. Based on this, two versions of the TC algorithm have been implemented in the CMS framework:

- *Track Counting High Efficiency* (TCHE): the S_{IP} value of the second

ranked track is used as discriminator value;

- *Track Counting High Purity* (TCHP): the S_{IP} value of the third ranked track is used as discriminator value;

The extension of the TC algorithm is obtained combining the information of the impact parameter of different tracks in a jet. Two discriminators are obtained by two additional algorithms:

- *Jet Probability* (JP): uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex;
- *Jet B Probability* (JBP): is the same as JP, but the algorithm gives more weight to the tracks with the highest S_{IP} (up to a maximum of four tracks are used, which is the average number of charged particles from b hadron decay);

Another powerful way to discriminate b-jets from other jets is to look for the presence of a secondary vertex inside the jet. Variables associated to a secondary vertex that are useful in b-tagging are: the flight distance and direction, defined by the vector between the primary and secondary vertex, track multiplicity and invariant mass associated to the secondary vertex. The *Simple Secondary Vertex* (SSV) algorithm uses the significance of the flight distance as discriminating variable. There are two versions of the SSV algorithm:

- *Simple Secondary Vertex High Efficiency* (SSVHE): secondary vertices with at least two associated tracks are considered;
- *Simple Secondary Vertex High Purity* (SSVHP): secondary vertices with at least three associated tracks are considered;

The efficiency of SSV algorithms is limited by the secondary vertex reconstruction efficiency to about 65%. By using additional variables, a discrimination can be provided even in cases where the secondary vertex is not reconstructed, increasing the efficiency with respect to the SSV algorithms. This is a more complex approach and involves the use of secondary vertices (when reconstructed) together with track-based lifetime variables, leading to the *Combined Secondary Vertex* (CSV) algorithm. A distribution of the discriminating variable obtained with CSV is shown in Figure 4.3 [62]. Two likelihood ratios are built from these variables. They are used to discriminate between b and c jets and between b and light-flavor jets.

The performances of the b-tagging algorithms are summarized in Figure 4.4 [62], where the predictions of the simulation for the misidentification probabilities (the efficiency to tag as b-jet a non-b-jet) are shown as a function of the efficiency to tag a b-jet.

Loose selections corresponding to 10% of misidentification probability for light flavor jets lead to a b-jet tagging efficiency of $80 - 85\%$, whereas tight selections corresponding to 0.1% of misidentification probability for light flavor jets lead to a b-jet tagging efficiency of $45 - 55\%$. For medium and tight selections the best performances are achieved with the CSV algorithm. The separation of c from b jets is naturally more challenging since the lifetime of particles originating in c-quarks fragmentation processes is not negligible. For the CSV algorithm an explicit tuning for light-flavor and c-jet rejection is performed, which gives the best c-jet rejection values for medium to tighter selections.

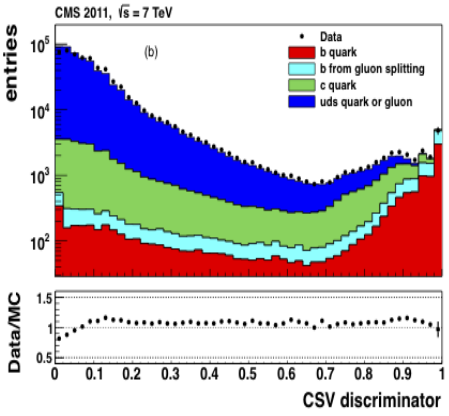


Figure 4.3: Distribution of the CSV b-tagging discrimination variable. The data are compared with the Monte Carlo simulated events split for the different flavor component.

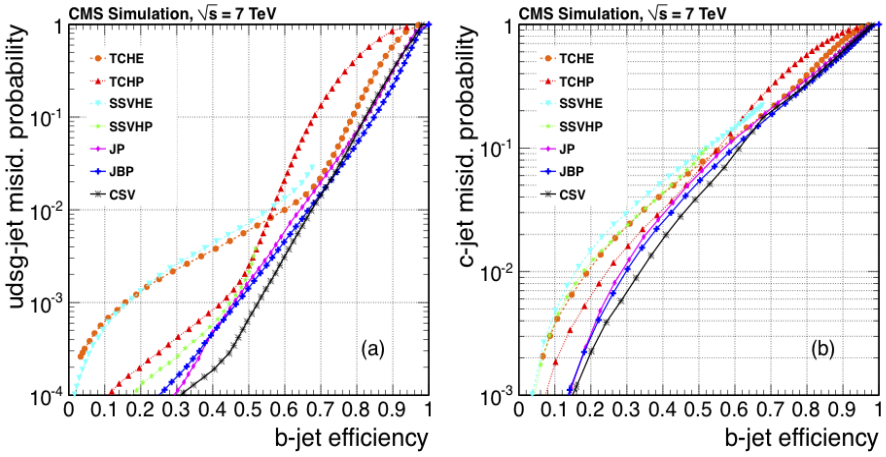


Figure 4.4: b-tagging performance curves from simulation for the different algorithms. (a) light-flavor jet and (b) c-jet efficiencies as a function of the b-jet efficiency.

4.2.5 Muons

The reconstruction of muons originated by pp collisions starts independently in the silicon tracker (*tracker track*) and in the muon spectrometer (*standalone-muon track*) [51]. A complete muon reconstruction in the whole CMS volume, matching the informations of the tracker and muon system, can be achieved following two different approaches:

- *Global Muon reconstruction* (*outside in*): starting from a standalone muon, a matching tracker track is found and a *global-muon track* is fitted combining hits from tracker and standalone muon tracks. The greater is the muon p_T, the more the global fit can improve the momentum resolution with respect only the tracker track, since the level arm is bigger;

- *Tracker Muon reconstruction* (*inside-out*): all tracker track with $p_T > 0.5\ GeV$ and $p > 2.5\ GeV$ are considered as possible muon candidates. They are extrapolated toward the muon system, taking into account the

energy loss in the material, the magnetic field and the uncertainty due to multiple scattering. If at least a muon segment (inside a DT or CSC, depending on which part of the CMS volume is involved in the extrapolation) matches the extrapolated track, the corresponding tracker track is classified as a *tracker-muon track*. The matching is performed looking at the spatial distance between the extrapolation and the segment inside the muon chamber.

At low momentum the tracker-muon is more efficient than the global muon reconstruction. In fact, as the muon p_T drops, the penetration inside the muon system is less deep. This results in higher probability for low p_T muons to produce just one segments in the first muon station layer. For about 1% of the muons from collisions, it happens that only a *standalone-muon track* is reconstructed, and both of the above approaches fail. The failing rate is very low thanks to the high tracker track efficiency.
In the CMS framework, the results of these algorithms are merged in a single collection of muon candidates. A given physics analysis can achieve the desired balance between identification efficiency and purity by applying a selection based on the muon identification variables. Three basic selections are: *soft-muon selection*, where the muon is a tracker muon, with additional requirements on the spatial matching of the segment in the muon chambers; *tight-muon selection*, where the muon is global, with additional requirements, such as cut on the normalized χ^2 of the track fit, at least one muon chamber hit included in the final track fit, segments matched in at least two muon stations, the corresponding tracker track with more than 10 silicon tracker hits (including at least one pixel hit) and a small distance between the closest point of the track to the primary vertex (impact parameter); *particle-flow muon selection*, a selection is performed on all the muon candidates reconstructed with the standard algorithms. This selection has been optimized to identify muons in jets with high efficiency, keeping low the fake rate from misidentified charged hadrons.
The muon identification efficiency for global muons with tight selections, obtained with the "Tag-and-probe" method for efficiency measurement is shown in Figure 4.5 [51]. Global muons with tight selection coming from Z boson decay have a fake rate probability of the order of 3.6% and 4%, in the forward and central region of the detector.

A powerful variable used to determine whether a muon (or in general a lepton) is inside or outside a jet is the isolation. This quantifies the amount of energy of the particles detected in a region around the track of the reconstructed muon. The region where the isolation variable is computed is a cone in the (η, ϕ) plane:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = \sqrt{(\eta - \eta_0)^2 + (\phi - \phi_0)^2} \tag{4.2}$$

where η_0 and ϕ_0 identify the cone axis, which is the direction of the inner muon track at the vertex position. The vertex is defined as the point of closest approach of the track to the beam line on a plane perpendicular to the beam.
Four muon identification variables are defined:

- *Tracker-Iso*: is the sum of the transverse momenta of tracker tracks within the isolation cone. Only tracks with $p_T > 1\ GeV$, $\Delta z < 0.2\ cm$ and $\Delta r < 0.1\ cm$ are considered, where Δz and Δr are the minimum distances from the tracker track to the cone vertex;

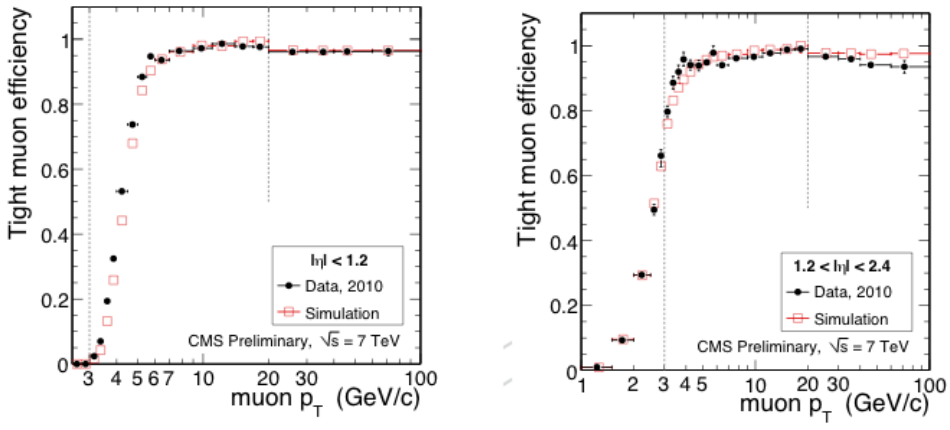


Figure 4.5: Muon identification efficiency in data compared to simulation as a function of the muon p_T for global muons with tight selections. The measurement is done using $J/\psi \to \mu^+\mu^-$ for $p_T < 20\ GeV/c$ and $Z \to \mu^+\mu^-$ for $p_T \geq 20\ GeV$. The results are shown for the barrel and overlap region (left) and for the end-caps (right).

- *Ecal-Iso*: is the sum of the transverse energies deposited in the electromagnetic calorimeter by particles passing through it. Only crystals with $E > 0.25\ GeV$ are considered;
- *HCAL-Iso*: is the sum of the transverse energies deposited in the hadron calorimeter by particles passing through it; no cuts are applied on the calorimetric towers;
- *Combined-Iso*: is the linear sum of the Tracker, Ecal and HCAL Isolation variables;

The efficiency to select an isolated muon, defined as the fraction of the muons that pass a given combined isolation threshold over all the muons in a sample of muon candidates originating from Z decays, is shown in Figure 4.6. Two methods are used for this purpose: the Tag-and-Probe and the Lepton Kinematic Template described in [51].

4.3 Simulation

Detailed simulation of the events in the CMS detector is of primary importance to perform physics analysis and detector studies. This concerns both the simulation of the physics processes at the event generator level and the particle interactions with the detector, including the response of the detector electronics. The simulation study may have different goals: feasibility and/or improvement of a given physics analysis, control and optimization of detector performance, etc.
The simulation workflow at CMS is as follows:

- a physics group configures an appropriate Monte Carlo event generator to produce the data samples of interest;

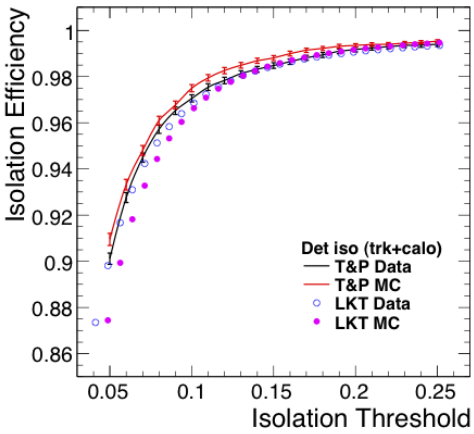


Figure 4.6: Efficiency of the combined isolation algorithm on muons from Z decays as a function of the isolation variable threshold. The results are shown for both data and Monte Carlo using two methods: the Tag-and-Probe and Lepton Kinematic Template [51].

- the production team/system runs the generator software to produce generator event data files;
- the physics group validates the generator data samples and selects a configuration for the detector and physics simulation (detector configuration, physics cuts, etc.);
- the production team/system runs the simulation of CMS, with generator events as input, to produce (using the standard CMS framework) simulation hits in the sensitive detectors;
- the physics group validates these hit data which are then used as input to the subsequent digitization step, which emulates the response of the detector, including the trigger and electronic readout and introducing the effects due to the pile-up of different events in the same and/or nearby bunch crossings. This step converts hits into digitizations (also known as *digis*) which correspond to the output of the CMS electronics. From this point, the remaining part of the reconstruction and analysis workflow is exactly the same as for the real data;

The event generators that provide the collision events as input for the detector simulation are written in $FORTRAN$ or $C++$ code. Among the most important are $PYTHIA$ [63], $MadGraph$ [64] and $HERWIG$ [65]. In particular, $PYTHIA$ and $MadGraph$ generators were used to produce the simulated samples used in the work of this thesis, as described in the following. The simulations take into account the leading order (LO) process for the samples generated with $PYTHIA$ and next to leading order (NLO) processes for the samples generated with $MADGRAPH$. The detailed CMS detector and physics simulation is based on the $GEANT4$ [66] simulation toolkit and the CMS object-oriented framework and event model [49]. The simulation is implemented for all CMS detectors in both the central region (Tracker, Calorimeters and Muon Systems) and in the

forward regions (CASTOR calorimeter, TOTEM telescopes, Roman Pot detectors and the Zero Degree Calorimeter, ZDC), including the field map from the 3.8 T solenoid. The full simulation program implements the sensitive detector behavior, hit collection and digitization (i.e. detector response), track selection mechanisms. The digitization step, following the hit creation step, constitutes the simulation of the electronic readout used to acquire data by the detector and DAQ systems. It starts from the hit positions and simulated energy losses in the sensitive detectors, and produces an output that needs to be as close as possible to real data coming from CMS. Information from the generation stage (e.g. particle type and momentum) is preserved in the digitization step. The last step is to use the simulated sub-detectors output as input for the reconstruction software, that provides the physics objects for analysis purposes.
For the purpose of this analysis, several physics processes have been considered. They are briefly described in the next sections.

4.3.1 Signal samples

- $b\bar{b} + \Phi \rightarrow 2b2\bar{b}$

  a total of 11 samples were produced with the $PYTHIA$ generator, one sample for each Higgs mass hypothesis considered for the analysis: $m_{\Phi} = 90, 100, 120, 130, 140, 160, 180, 200, 250, 300, 350\ GeV$. The simulated processes are the $gg \rightarrow b\bar{b} + \Phi, \Phi \rightarrow b\bar{b}$ with two gluons in the initial state and $qq \rightarrow b\bar{b} + \Phi, \Phi \rightarrow b\bar{b}$ with two quark in the initial state, as described in section 1.4.9. The cross sections are obtained at next-to-next leading order (NNLO) for each Higgs mass hypothesis and for a given set of the MSSM parameters through *bbh@nnlo* [67] and $FeynHiggs$ [36, 68–70].

4.3.2 Background samples

- QCD sample

  this sample of $\sim 25\ M$ events was produced with the $PYTHIA$ generator, requiring $\hat{p}_T \geq 20\ GeV$ for the hard scattering[2]. The generated events are filtered, retaining only those events containing at least one muon generated with $p_T \geq 15\ GeV$. The corresponding cross section of the process is $2.97 \cdot 10^8\ pb$, with a filter efficiency[3] of $\epsilon_{eff} = 2.86 \cdot 10^{-4}$

- $t\bar{t} + jets$

  this sample of $\sim 3.7\ M$ events was produced with the $MADGRAPH$ generator. For the final states into taus, the tau-lepton decays are simulated using the $TAUOLA$ package [71]. The total cross section, from CMS measurements, is $150 \pm 9(stat.) \pm 17(syst.) \pm 6(lumi)\ pb$ [72].

- $Z \rightarrow bb$

  this sample of $\sim 2.2\ M$ events was produced with the $PYTHIA$ generator, requiring $M_Z \geq 50\ GeV$ for the virtual Z boson in the simulated process. The total cross section is 3270 pb with a filter efficiency of $\epsilon_{eff} = 0.695$.

[2]The $\hat{p}_T$ is the exchanged momentum between the two initial interacting partons.
[3]The filter efficiency is the efficiency of the filters applied to the generated events.

- $Z \rightarrow l^+l^- + jets$

  this sample of $\sim 36\ M$ events was produced with $MADGRAPH$ generator, requiring $M_Z \geq 50\ GeV$ for the virtual Z boson in the simulated process. For the final states into taus, the tau-lepton decays are simulated using the $TAUOLA$ package. The total cross section, from CMS measurements, is $974 \pm 7(stat.) \pm 7(syst.) \pm 180(th.) \pm 390(lumi.)\ pb$ [73].

- Dibosons: ZZ, WW and WZ

  these samples, each one with $\sim 4.2\ M$ events, were produced with $PYTHIA$ generator. All the leptonic final states were simulated for the Z/W boson decays. For the final states into taus, the tau-lepton decays are simulated using the $TAUOLA$ package. The total cross sections, from CMS measurements, are $\sigma_{ZZ} = 3.8^{+1.5}_{-1.2}(stat.) \pm 0.2(syst.) \pm 0.2(lumi.)\ pb$, $\sigma_{WW} = 55.3 \pm 3.3(stat.) \pm 6.9(syst.) \pm 3.3(lumi.)\ pb$, $\sigma_{WZ} = 17.0 \pm 2.4(stat.) \pm 1.1(syst.) \pm 1.0(lumi.)\ pb$ [74].

Chapter 5

Analysis strategy for the Minimal Supersymmetric Standard Model Higgs search in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel

5.1 Expected signature and signal characterization

The production of the MSSM Higgs boson in association with b-quarks, with the subsequent decay in a b-quark pair, leads to an experimental signature made by up to four jets. These jets can be identified with the b-tagging algorithms that have been described in section 4.2.4.
In the analysis presented in this thesis, the signal is modeled through Monte Carlo simulation for each assumed mass m_Φ of the Higgs boson. The simulated events have been re-weighted to guarantee that the simulated PU distribution matches the one observed in data in the different data taking periods[1]. This is shown as an example in Figure 5.1 for the first part of the 2011 run (RunA). In Figure 5.2 the p_T and η spectra of the four b-quarks at the generation level of the signal events are shown for $m_\Phi = 120\ GeV/c^2$. The two leading jets originate from the Higgs decay in about 79% of the events. At this mass point, for about 58% of the events two jets from the Higgs decay are reconstructed by the anti-k_T jet algorithm exploiting the Particle Flow technique, as explained in section 4.2.3, with a cone size of $R = 0.5$ in the (η, ϕ) plane and $p_T \geq 30\ GeV$. As it can be seen in Figure 5.2, the dynamic of the production mechanism is such that frequently the fourth b-quark jet is very forward, outside the pixel and tracker region $\eta < 2.5$, where the b-tagging algorithms do not work, since no tracks are reconstructed. At least one of the b-jets often contains a muon inside, due to the 10% branching ratio in muons of the b-quark. Figure 5.3 shows the spectrum of such muons at generation level for signal events at $m_\Phi = 120\ GeV/c^2$. As it can be seen, the spectrum is soft, and the softer muons are more likely to come from the b-jets produced in association with the Higgs boson. About 20% of the events have a muon with $p_T \geq 5\ GeV/c$. The fraction of these events decrease as the p_T threshold increase: for $p_T \geq 12, 15, 17\ GeV/c$ the fraction of events containing a muon is $7, 4.5, 3.5\%$, respectively.

[1]The PU distribution varies with the increasing of the LHC luminosity.

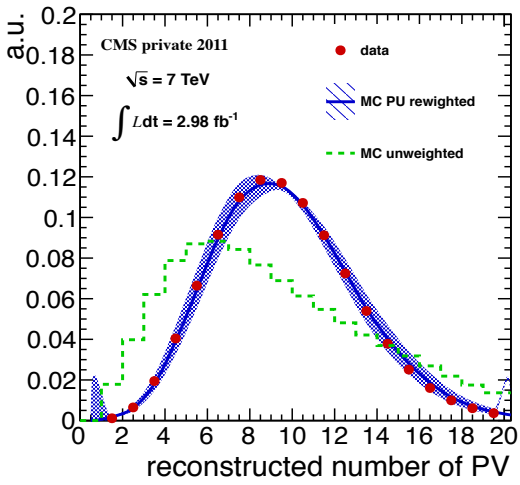
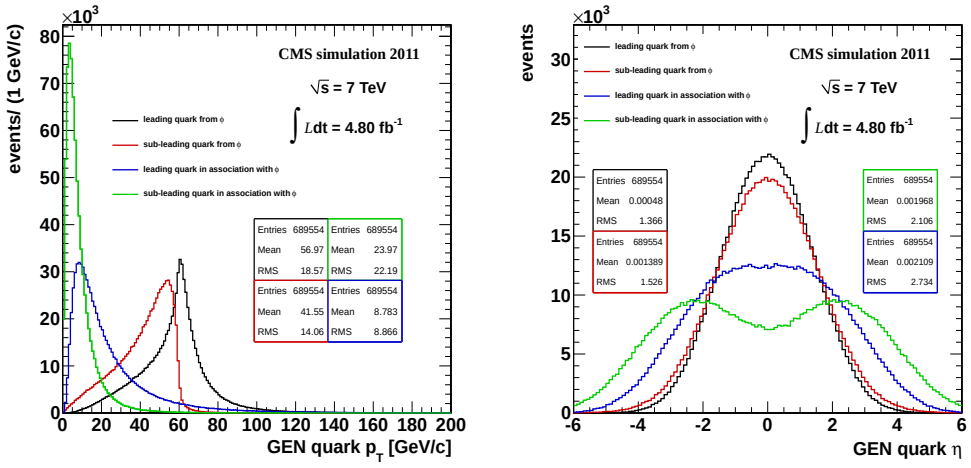


Figure 5.1: Reconstructed number of primary vertices in signal Monte Carlo simulation compared to data collected in the first part of the 2011 run (RunA). The original distribution is compared to the re-weighted one.



(a) (b)

Figure 5.2: Distributions of b-jet p_T in signal Monte Carlo simulation for $m_\Phi = 120\ GeV/c^2$ at generation level. The leading and sub-leading jets from the Higgs boson decay and the pair produced in association are shown.

The expected signal for events with three b-tagged jets can be seen in Figure 5.4, where the invariant mass distribution for the two leading b-tagged jets for a sample of simulated events corresponding to an integrated luminosity of 4.8 fb^{-1} is shown. The trigger and offline selections, that will be discussed afterwords, are applied. In the figure, is also shown the distribution of the invariant mass of the reconstructed jets correctly associated to the b-quarks originating from the Higgs decay. Choosing the two leading jets as those jets generated by the Higgs decay, we are reconstructing the correct Higgs mass in about 34% (73%) of the cases for $m_\Phi = 120$ ($m_\Phi = 250$) GeV/c^2. Fake invariant mass reconstruction from combinatorial background lowers as the Higgs boson mass increases, due to

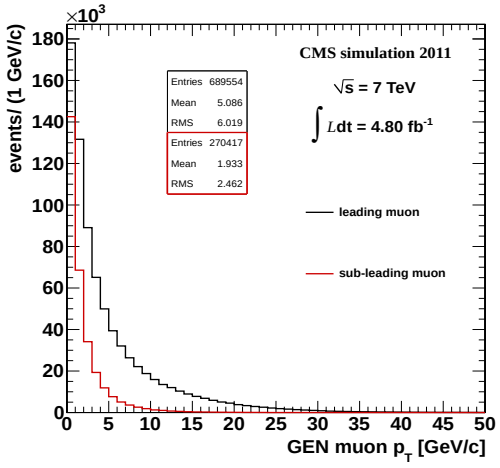


Figure 5.3: Distribution of the p_T for the leading and sub-leading muon from b-quark semi-leptonic decay at generation level.

the hardening of the p_T spectrum of the jets coming from Higgs decay. The mass resolution is 11% and 17% for the two considered masses, respectively.

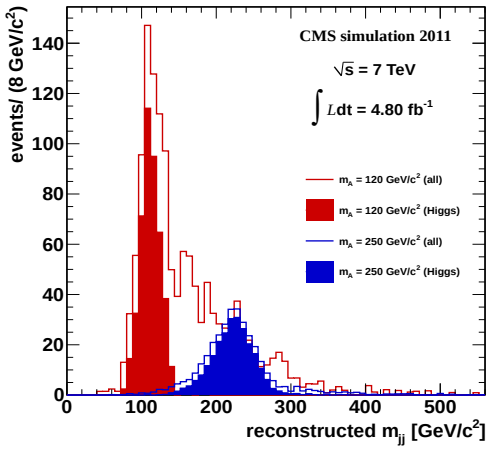


Figure 5.4: Reconstructed invariant mass for a simulated supersymmetric Higgs signal with $m_\Phi = 120\ GeV/c^2$ (red) and $200\ GeV/c^2$ (blue). The distribution reconstructed with the two leading jets (line) is shown together with the distribution obtained when the pair of jets originate from the Higgs decay (full histogram). Events yields are normalized to 4.8 fb^{-1}, assuming $\Phi = h, A, H$ cross section predicted by MSSM at $tan\beta = 30$.

The main background sources come from multi-jets events containing b-jets from QCD. This background is largely irreducible and an accurate prediction of its yield and shape is required to preserve the sensitivity of the analysis. Monte Carlo simulation is used for studies and checks, but the prediction used for the signal search relies on data-driven methods, as will be discussed afterwards. The simulation is not reliable for a quantitative QCD background determination because of the uncertainties of multiple b-jets production in gluon splitting and flavour excitation processes. Other backgrounds are $t\bar{t} + jets$, $Z \to b\bar{b} + jets$

and di-boson (ZZ, WZ, WW) event production, in which true b-jets can be generated and/or energetic non b-jets can be mis-tagged as b-jets. Monte Carlo simulation is used to estimate these contributions.

5.2 Data samples and trigger selection

The analysis is performed on a data sample collected during the 2011 data taking period. The dataset named $MuHad$ containing the events selected by the HLT triggers specifically implemented for the analysis is used for the MSSM Higgs search, whereas the $SingleMu$ dataset triggered by a single muon exceeding various pre-defined p_T thresholds is used to perform trigger efficiency studies[2]. The signal candidate events are selected online by triggers requiring two high p_T jets in the central part of the detector, at least one of them b-tagged with the TCHE algorithm implemented in the HLT algorithm. The jets are reconstructed online using anti-k_T algorithm exploiting only the calorimetric information. In order to further reduce the background coming from multi-jet QCD and to take the trigger rate at $\sim 5 - 10\ Hz$, the semi-leptonic decay of the b-quark is exploited, requiring a moderately high p_T globally reconstructed muon.
The trigger requirements start at L1 by requiring the presence of a muon candidate, which, in the second part of the data-taking, was required to be in coincidence with energetic L1 jet candidates, to cope with the higher instantaneous luminosity conditions. A list of the trigger paths implemented for the analysis for different data taking periods[3], with their main characteristics, is given below:

- **Path 1**: HLT_Mu12_CentralJet30_BTagIP
  1. global muon with $p_T > 12\ GeV$;
  2. at least one jet with $p_T > 30\ GeV$ and $|\eta| < 2.6$;
  3. one jet b-tagged online with TCHE discriminator greater than 3.5, exploiting the primary vertex reconstructed in the plane perpendicular to the beam-line (2D vertex);
- **Path 2**: HLT_Mu12_DiCentralJet30_BTagIP3D
  1. global muon with $p_T > 12\ GeV$;
  2. two jets with $p_T > 30\ GeV$ and $|\eta| < 2.6$;
  3. one jet b-tagged online with TCHE discriminator greater than 3.5, exploiting the primary vertex reconstructed in 3D;
- **Path 3**: HLT_Mu12_DiCentralJet20_DiBTagIP3D1stTrack
  1. global muon with $p_T > 12\ GeV$;
  2. two jets with $p_T > 20\ GeV$ and $|\eta| < 2.6$;
  3. two jet b-tagged online with TCHE discriminator greater than 3.5, exploiting the primary vertex reconstructed in 3D;

[2]These control triggers are generally largely pre-scaled, to keep the rate at a few Hertz level.
[3]It must be stressed that in the 2011 run the LHC instantaneous luminosity increased by more than one order of magnitude from the beginning to the end of the run (from $\sim 2 \cdot 10^{32}\ cm^{-2}s^{-1}$ to $\sim 6 \cdot 10^{33}\ cm^{-2}s^{-1}$)

- **Path 4**: HLT_Mu12_eta2p1_DiCentralJet20_DiBTagIP3D1stTrack

    1. global muon with $p_T > 12\ GeV$ and additional L1 requirement of $|\eta| \leq 2.1$;
    2. two jets with $p_T > 20\ GeV$ and $|\eta| < 2.6$;
    3. two jet b-tagged online with TCHE discriminator greater than 3.5, exploiting the primary vertex reconstructed in 3D.

The triggers are listed in the chronological order in which they were used. They present tighter requirements on some of the physics objects used to perform the online selection due to the increasing instantaneous luminosity during the 2011 data-taking period. The amount of data collected with the different trigger paths and the corresponding signal efficiencies for a Higgs boson with mass $M_\Phi = 120\ GeV$ are reported in Table 5.1.

Table 5.1: HLT paths, corresponding data sample collected and signal efficiency for a Higgs boson with mass $M_\Phi = 120\ GeV$. The reported luminosity is obtained considering data that satisfy the DQM certification requirements, as explained in section 3.8.

HLT paths	run range	$\int \mathscr{L} dt$ $[pb^{-1}]$	signal efficiency [%]
Path 1	163738-165633	180.9	3.64 ± 0.19
Path 2	165970-166967	537.1	2.28 ± 0.11
Path 3	167039-173198	1108.6	1.66 ± 0.07
Path 4	173236-180252	2979	1.65 ± 0.06

5.3 The event selection

The data collected using the different trigger paths are further filtered by off-line selections. These further offline requirements are listed below:

- a global muon with $p_T > 15\ GeV$;
- at least three jets, with $p_T > 30\ GeV$ for the two leading and $p_T > 20\ GeV$ for the third and following ones. All of them have to be within $|\eta| < 2.6$;
- distance between any pair of jets $\Delta R_{ij} > 1$ in the (η, ϕ) plane, to prevent ambiguities in b-tagging algorithms and to remove contributions from gluon splitting in background processes;
- the two leading jets must have a CSV b-tagging discriminator $CSV > 0.8$ (corresponding to a misidentification rate of about 0.3% for jets originating from light quarks);

This way, a sample of two b-tagged jets plus a third jet is selected. This sample is used for the background modeling, as it will be discussed in section 5.5. For the final signal search the following requirement is added:

- the third jet, in p_T order, must have a CSV b-tagging discriminator $CSV > 0.7$ (corresponding to a misidentification rate of about 1% for jets originating from light quarks).

The corresponding data reduction after each selection criteria is reported in Table 5.2 for the different trigger paths. In Table 5.3 the yields predicted from the Monte Carlo simulation for the different background contributions is shown. Each process is normalized to the integrated luminosity collected in 2011 and the trigger efficiency (discussed afterwards) is taken into account[4]. It can be noticed that the requirement of three b-jets suppresses all the electroweak contributions to a negligible size, with the exception of $t\bar{t} + jets$ and $Z \to b\bar{b} + jets$. For the last process only leading-order processes are considered in the simulation, so that the jet multiplicity may be underestimated. A comparison to the expected jet multiplicity in $Z \to \mu\mu$ simulated sample generated including next to leading order (NLO) precesses shows that the accompanying jets multiplicity is larger by almost a factor of two. The expected background yield then must be multiplied by the same factor, with a large systematic uncertainty. Comparing Table 5.3 with Table 5.2 can be noted that the expected contribution from the main QCD background is only in qualitative agreement with data. This is expected since the Monte Carlo simulation is not suitable for the multi-jet contribution estimation. The corresponding signal efficiency of the offline selections is shown in Tables 5.4 and 5.5 for the Higgs mass points considered in the analysis, divided in a *Low* and *High* mass region respectively, as will be discussed in section 5.5.2.

Table 5.2: Data reduction after each selection cut for the various trigger paths.

Cut	Path1	Path2	Path3	Path4	Path4[a]	All
All	3027717	4532555	2244550	1237147	5690304	16732273
$p_T^\mu > 15\ GeV/c$	1757902	2678935	1337394	742231	3222677	9739139
$\#jets >= 3$	665962	1245655	639616	404082	1556012	4511327
$\Delta R_{ij} >= 1$	513981	957884	498996	315284	1219439	3505584
$CSV(1^{st} - jet) > 0.8$	242982	492734	297838	184075	714506	1932135
$CSV(2^{nd} - jet) > 0.8$	52345	112428	162029	99175	387708	813685
μ in 1^{st} or 2^{nd} jet	50708	108551	156147	95760	374774	785940
$CSV(3^{rd} - jet) > 0.7$	3245	7323	12796	7623	29208	60195
$\int \mathscr{L} dt\ [pb^{-1}]$	180.9	537.1	1108.6	652.2	2326.816	4805.7

[a]After September 2011 new jet energy corrections have been implemented in the online trigger. The path running in that period, affected by this change, is Path4.

A possible issue that could raise in the data analysis is related to the different trigger paths used to collect the sample of events to analyze. In fact, the shape of the key distributions could be different for each trigger. In this case, the analysis should be performed for each trigger path separately. A check has been made in order to evaluate this possibility. In Figure 5.5, as an example, the distributions for the p_T of the second jet and the invariant mass of the two leading jets are shown for the events divided by the different trigger paths and requiring only

[4]As already mentioned, the QCD prediction is purely indicative and will not be used in the analysis.

Table 5.3: Events reduction for MC simulated background samples at the different stage of event selection. The trigger efficiency is included only starting from the cut $\#jets >= 3$ by applying the data driven turn-on curves discussed in section 5.4. For $Z \to bb$ sample, a scale factor of about ≈ 1.9 must be applied, due to the LO generation used (see text for details). The MC scale factor is the ratio between the 2011 integrated luminosity used in this analysis (4.8 fb^{-1}) and the equivalent integrated luminosity available for the Montecarlo sample. The quoted errors are statistical only.

Cut	QCD	TTJet	Z→bb	ZZ	WZ	WW
σ $[pb]$	2.97E+08	150	3270	3.8	17.0	55.3
MC ϵ_{filter}	2.86E-04	1.	0.695	1.	1.	1.
MC lumi scale factor	16.2	0.011	4.87	0.005	0.012	0.033
All	406884036	730360	10920083	20599	50308	133723
$p_T^{\mu} > 15\ GeV$	362703147	193972	212716	1848	6448	23448
$\#jets >= 3$	32221594	104007	24643	251	655	1855
$\Delta R_{ij} >= 1$	26817799	77939	20890	191	495	1371
$CSV(1^{st} - jet) > 0.8$	7318818	29146	7872	35	54	54
$CSV(2^{nd} - jet) > 0.8$	1189598	8859	2238	8	8	3
μ in 1^{st} or 2^{nd} jet	1165417	3878	2114	6	4	1
$CSV(3^{rd} - jet) > 0.7$	72043±747	303±5.8	284±36	1.84±0.07	0.44±0.05	0.03±0.03

Table 5.4: Events reduction for simulated signal samples, normalized to a cross section for $\tan\beta = 30$, for mass points in the *Low* mass region (see section 5.5.2) $M_H < 200\ GeV/c^2$. The trigger efficiency is included only starting from the cut $\#jets >= 3$ by applying the data driven turn-on curves.

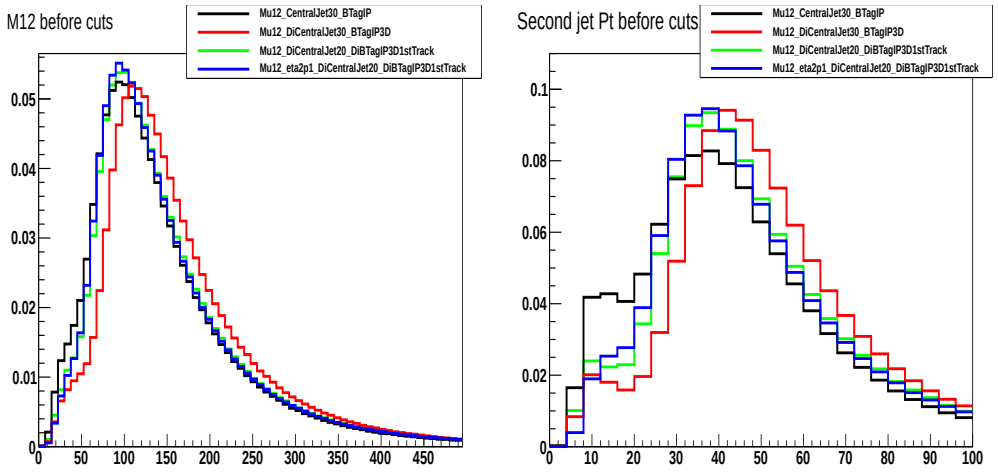
	M_H $[GeV/c^2]$						
Cut	90	100	120	130	140	160	180
All	3181487	2282039	1225179	942164	664435	425194	271531
$p_T^{\mu} > 15\ GeV$	79166	71272	53760	47703	37522	29513	21972
$\#jets >= 3$	13416	13012	11321	10649	9028	8009	6546
$\Delta R_{ij} >= 1$	11709	11423	9796	9161	7689	6729	5386
$CSV(1^{st} - jet) > 0.8$	7452	7339	6416	5970	5133	4440	3545
$CSV(2^{nd} - jet) > 0.8$	4202	4078	3752	3405	3003	2662	2172
μ in 1^{st} or 2^{nd} jet	3617	3508	3222	2937	2611	2321	1873
$CSV(3^{rd} - jet) > 0.7$	1338	1324	1305	1160	1079	953	744
ϵ [%]	0.042	0.058	0.107	0.123	0.162	0.224	0.274

good quality of the reconstructed jets as offline selection. As can be expected, the differences in shape are quite evident, since the triggers differs each other. In Figure 5.6 the same distributions are shown, for the events where the full offline selection is applied. They are compatible between the different trigger

Table 5.5: Events reduction for simulated signal samples, normalized to a cross section for $\tan\beta = 30$, for mass points in the *High* mass region (see section 5.5.2) $M_H \geq 200\ GeV/c^2$. The trigger efficiency is included only starting from the cut $\#jets >= 3$ by applying the data driven turn-on curves.

	$M_H\ [GeV/c^2]$			
Cut	200	250	300	350
All	178458	69624	29935	13310
$p_T^{\mu} > 15\ GeV$	16306	8040	4014	1999
$\#jets >= 3$	5290	3012	1686	901
$\Delta R_{ij} >= 1$	4267	2381	1303	685
$CSV(1^{st} - jet) > 0.8$	2834	1584	850	441
$CSV(2^{nd} - jet) > 0.8$	1747	986	531	271
μ in 1^{st} or 2^{nd} jet	1509	860	465	239
$CSV(3^{rd} - jet) > 0.7$	591	340	181	92
ϵ [%]	0.331	0.488	0.605	0.691

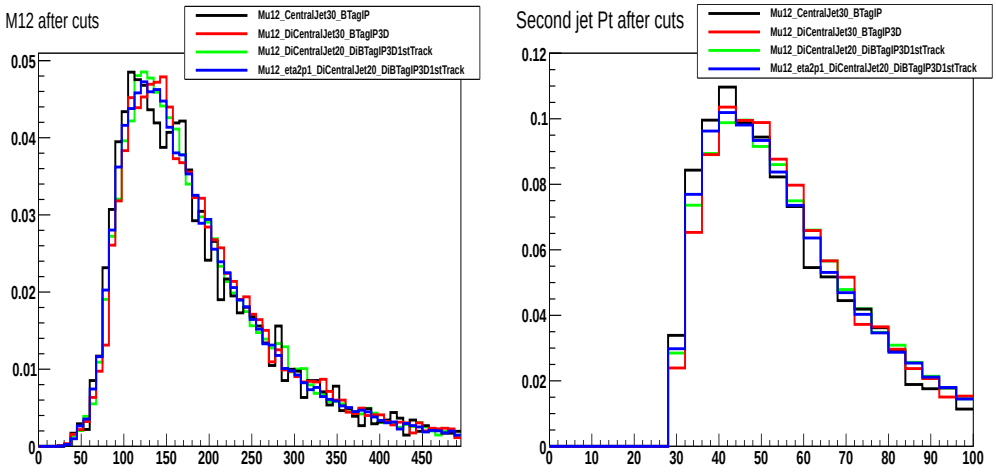
paths within the statistical uncertainties. Other kinematical parameters have been checked, confirming the behavior seen in Figure 5.6. As a conclusion, all the data can be analyzed at once, regardless of the trigger path.



(a) Invariant mass distribution of the two leading jets.

(b) p_T spectrum of the sub-leading jet.

Figure 5.5: Distributions of the sub-leading jet p_T and invariant mass of the two leading jets for each trigger path used in the analysis, before the full offline selection.



(a) Invariant mass distribution of the two leading jets.

(b) p_T spectrum of the sub-leading jet.

Figure 5.6: Distributions of the sub-leading jet p_T and invariant mass of the two leading jets for each trigger path used in the analysis, after the full offline selection.

5.4 Trigger efficiencies

To measure the trigger efficiencies of the different paths used in this analysis, the following factorization, which is justified below, is exploited:

$$\epsilon_{HLT} = \epsilon_{\mu} \cdot \epsilon_{hadr} \tag{5.1}$$

where ϵ_{μ} is the efficiency to trigger a muon with $p_T = 12\ GeV/c$ at the HLT level and ϵ_{hadr} is the efficiency to trigger an event that offline contains the hadronic and b-tagging properties described by the offline selection in section 5.3. The ϵ_{μ} is obtained from single muon triggered events with lower thresholds, and it is shown in Figure 5.7. The remaining contribution to the trigger efficiency, ϵ_{hadr}, is obtained from a sample of data collected by single muon trigger (requiring a muon with $p_T = 12\ GeV/c$) and filtered with the offline selection of section 5.3 as:

$$\epsilon_{hadr} = \frac{N^{evts}_{triggered}}{N^{evts}_{All}} \tag{5.2}$$

where N^{evts}_{All} is the number of all the events triggered by the single muon trigger and passing the analysis offline criteria, and $N^{evts}_{triggered}$ is the number of these events that in addition fired the trigger path for which the efficiency is measured. The threshold used for the muon p_T offline selection, 15 GeV, corresponds to a trigger efficiency on the muon in the plateau region (see Figure 5.7). This justifies the factorization in equation 5.1. On the contrary, if the threshold used for the muon p_T offline selection had not been in the plateau region, the ϵ_{hadr} would have been dependent on ϵ_{μ} in a non-trivial way. The measured ϵ_{hadr} is shown in Figure 5.8 for one of the trigger paths listed in section 5.2. For the other trigger paths similar distributions are obtained.

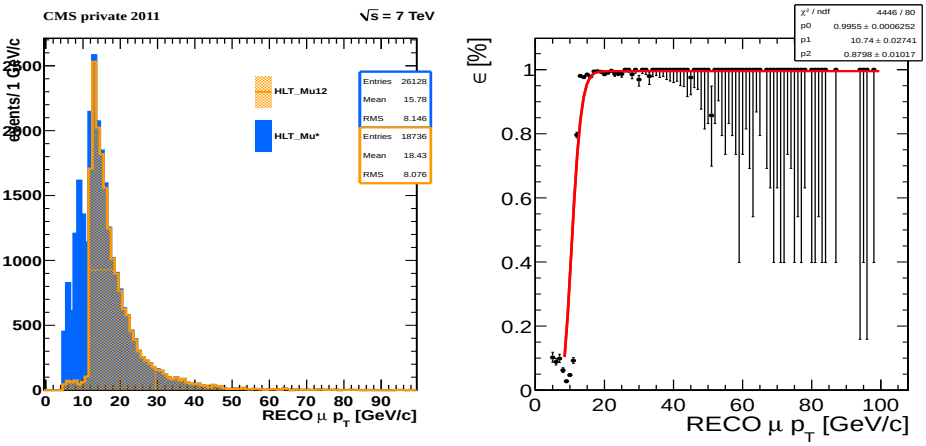


Figure 5.7: Trigger efficiency of a single muon trigger path with $p_T = 12\ GeV/c$ threshold measured on events triggered by single muon paths with lower p_T thresholds, as a function of the muon p_T (right plot). The data (black dots) are fitted with an error function (red line). The distributions of the muon p_T spectrum for the events used to measure the efficiency, superimposed to the distribution of the muon p_T spectrum of the events passing the single muon trigger with $p_T = 12\ GeV/c$ threshold, is shown on the left.

A one-parameter fit using an error function is performed on the trigger efficiency distributions obtained from data. The turn-on efficiency curves obtained from the fit are used to determine the overall efficiency of the signal selection. This will be used in the signal search analysis and in the computation of the observed limits on the production cross section. The error on the fitted parameter parametrizing the turn-on curves will be used to compute the systematic uncertainty due to the trigger efficiency and affecting the signal yield, as it will be discussed in section 6.1.1. Moreover, they are used for the background evaluation from Monte Carlo simulation reported in Table 5.3.

5.5 Background determination

As already discussed, the background determination cannot rely on the Monte Carlo simulation. Thus, two data driven methods have been developed to perform the background shape and normalization estimation: the *B-Tagging Matrices* and the *Hyperball* methods, which are described in the following.

5.5.1 The B-Tagging Matrices method

The prediction for the distribution of the relevant kinematic variables that are used for the signal search is performed starting from a sample of events selected by the criteria listed in section 5.3, except for the last one on the third jet. This way a sample of two b-tagged jets plus a third jet is obtained (this is called bbj). The estimation of the distribution for a given variable x, that will be measured in the sample selected with the full events selection presented in section 5.3, $F(x; bbb)$, is obtained from the distribution of the same variable x measured in the bbj sample, weighting each event with the probability for the third jet to be identified as a b-jet, $P_{b-tag}^{3rd-j}(j)$:

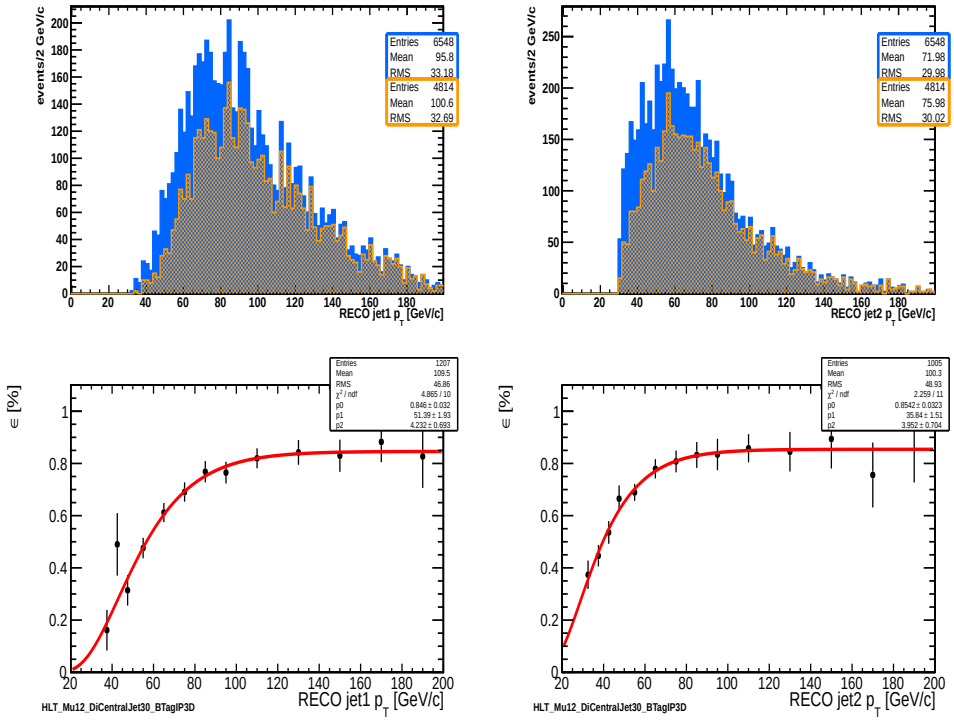


Figure 5.8: Trigger efficiency for the hadronic and b-tagging part, ϵ_{hadr}, for the trigger **Path 2**, as listed in section 5.2, shown in the bottom plots as a function of the leading (left) and sub-leading (right) jet p_T. The data (black dots) are fitted with an error function (red line). The top plots show the distributions of the leading (left) and sub-leading (right) jet p_T spectrum for the events used to measure the efficiency, superimposed to the corresponding distribution of the events passing the trigger.

$$F(x;bbb) = F(x;bbj) \times P_{b-tag}^{3rd-j}(j). \tag{5.3}$$

The probability $P_{b-tag}^{3rd-j}(j)$ can be written as follows:

$$P_{b-tag}^{3rd-j}(j) = \epsilon_b \cdot f_b + \epsilon_c \cdot f_c + \epsilon_{light} \cdot f_{light} \tag{5.4}$$

where ϵ_b, ϵ_c and ϵ_{light} are the probabilities to tag a b, c or light jet as a b-quark jet, respectively. The f_b, f_c and f_{light} are the fractions of b, c and light quarks that originate the third jet in the bbj sample, respectively.
The b-tagging probabilities and the quark fractions depends on the third jet and event characteristics. A Monte Carlo study as been done in order to determine the best parametrization. The best choice results in the parametrization of the b-tagging probabilities, $\epsilon_{b,c,light}$, as a function of the third jet p_T, $|\eta|$ and jet charge multiplicity (that is the number of charged tracks in the jet). For the flavor fractions $f_{b,c,light}$ a more complex parametrization is used:

$$f_{b,c,light} = f\left(E_T^{3rd-jet}, |\eta^{3rd-jet}|\right) \times f(\Delta R_{1,2}, \Delta R_{\Phi,3}) \tag{5.5}$$

The equation 5.5 shows that the flavor fractions are factorized in two terms: the first one is parametrized, like the b-tagging probabilities, as a function of

the E_T and $|\eta|$ of the third jet. The second term is parametrized as a function of topological variables of the event: the distance $\Delta R_{1,2}$ between the two leading jets, and the distance $\Delta R_{\Phi,3}$ between the momentum resulting by the vector sum of the two leading jets momenta, i.e. the Higgs candidate Φ, and the third jet. These ΔR distances are measured in the (η, ϕ) space, as usual with $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The second factor in 5.5 is used only to model the shape of the predicted distributions through 5.3, so that it is normalized to unity.
The b-tagging probabilities are determined from Monte Carlo simulation, as will be discussed in section 5.5.3. The flavor fractions are determined from data in the *bbj* sample, as discussed in detail in section 5.5.4. A bias on the measure of the flavor fraction f_b is given if a Higgs signal is present. In fact, the sub-sample of three real jets in the *bbj* sample would contain the signal and background contributions, so that the f_b fraction would estimate the sum of the two's, partially hiding the signal search. To avoid the bias that would spoil the Higgs search in this channel, a control region in the *bbj* sample, where no signal or a negligible contribution of it is present, has to be defined.

5.5.2 Control region definition

The discriminating power of different kinematic variables have been studied in Monte Carlo simulated samples of QCD and signal events, after the full offline selection cuts are applied. It turned out that there is not a single variable for which a good separation between QCD background and signal is optimal. The best variables with higher discrimination power have been combined to obtain a likelihood ratio discriminator which has the following expression:

$$Discr = \frac{\Pi_i p_i^{Signal}(x_i)}{\Pi_i p_i^{Signal}(x_i) + \Pi_i p_i^{QCD}(x_i)} \tag{5.6}$$

where p_i is the normalized distribution, for signal and QCD, for a given kinematic variable x_i, which in fact is the probability density function for that kinematic variable.
The choice of the best discrimination variables depends on the Higgs mass, which in principal makes the analysis rather complex, given the many considered mass points. However, close Higgs mass points have similar kinematical characteristics, allowing to simplify the problem. The considered Higgs mass range is divided into two intervals, for which a discriminator can be defined by a unique set of variables for both regions. The two regions are identified by $M_\Phi < 200\ GeV/c^2$ and $M_\Phi \geq 200\ GeV/c^2$, which we call *Low* and *High* mass region, respectively. Only three mass points, with equal weights, for each region have been used to compute the discriminator 5.6: $M_\Phi = 120, 130, 140\ GeV/c^2$ for the *Low* and $M_\Phi = 250, 300, 350\ GeV/c^2$ for the *High* mass region. The variable used to compute equation 5.6 are listed in Table 5.6. The distributions of these variables for signal and background (normalized to unity area) are shown in Figure 5.9 and 5.10 for low and high mass regions, respectively. The distributions of the resulting discriminator likelihood ratios for signal and background (normalized to unity area) are shown in Figure 5.11 for low and high mass. The *control region* is defined as the sample of the events having a discriminator value $Discr < 0.4$.

Table 5.6: List of variables used to compute the discriminator as defined by 5.6.

Name	Description
$p_T^{1,2,3}$	transverse momentum of the 1^{st}, 2^{nd} and 3^{rd} leading b-jet
$\Delta\Phi_{i,j;i,j=1,2,3}$	azimuthal angle between each of the three b-jet pairs
$\Delta\eta_{1,2}$	pseudo-rapidity difference between the 1^{st} and 2^{nd} leading b-jets
$\Delta\eta_{i,j}^{min,max}$	minimum and maximum pseudo-rapidity difference between the three b-jet pairs
$\Delta R_{1,2}$	distance in the (η,ϕ) plane between the 1^{st} and 2^{nd} leading b-jets
$\Delta R_{i,j}^{min}$	minimum distance in the (η,ϕ) plane between the three b-jet pairs
$\Delta\phi_{\Phi,3}$	azimuthal angle between the Higgs candidate and the 3^{rd} leading b-jet
$\Delta\eta_{\Phi,3}$	pseudo-rapidity distance between the Higgs candidate and the 3^{rd} leading b-jet
$\alpha_{\Phi,3}$	angle between the Higgs candidate and the 3^{rd} leading b-jet
$cos\theta^*$	cosine of the angle between the Higgs decay direction and one of the b-jets from the Higgs decay, in the Higgs candidate rest frame
N_{jets}	number of jets passing the offline selection

5.5.3 Determination of the offline b-tagging efficiencies

The efficiency to tag a jet originated from b-quark, c-quark or light parton $(u, d, s, gluon)$ as a b-jet has been obtained from Monte Carlo multijet simulation. The sample used for this purpose is selected with the offline cuts, except for the last one on the third jet. This sample is divided in categories depending on the flavor of the parton which originates the third jet. The identification of the parton originating the third jet is made by associating in the (η,ϕ) space the closest parton to the reconstructed jet, before the hadronization of the parton is simulated. The parton has to reside in a cone of radius $\Delta R < 0.3$ with respect to the jet direction. The efficiency is defined as the number of jets of a given flavor category that pass the b-tagging requirement $CVS > 0.7$ over the number of all the jets in that category. The variables used to parametrize the efficiencies are the jet transverse momentum p_T, the jet pseudo-rapidity η and the charge multiplicity (two ranges of this variable have been used $N_{charged\ tracks} \leq 10$ and $N_{charged\ tracks} > 10$). The results are shown in Figures 5.12-5.14 for b, c and light jets, respectively.
A detailed study of the b-tagging efficiency has been performed, using several data driven methods to validate the Monte Carlo simulation, as described in [75]. This includes checks on dedicated $t\bar{t} \to b\bar{b}jjX$ selected data, which provide samples of high purity reconstructed b-jets and light jets from hadronic W decays. The study shows that a scale factor to the efficiency found in the Monte Carlo

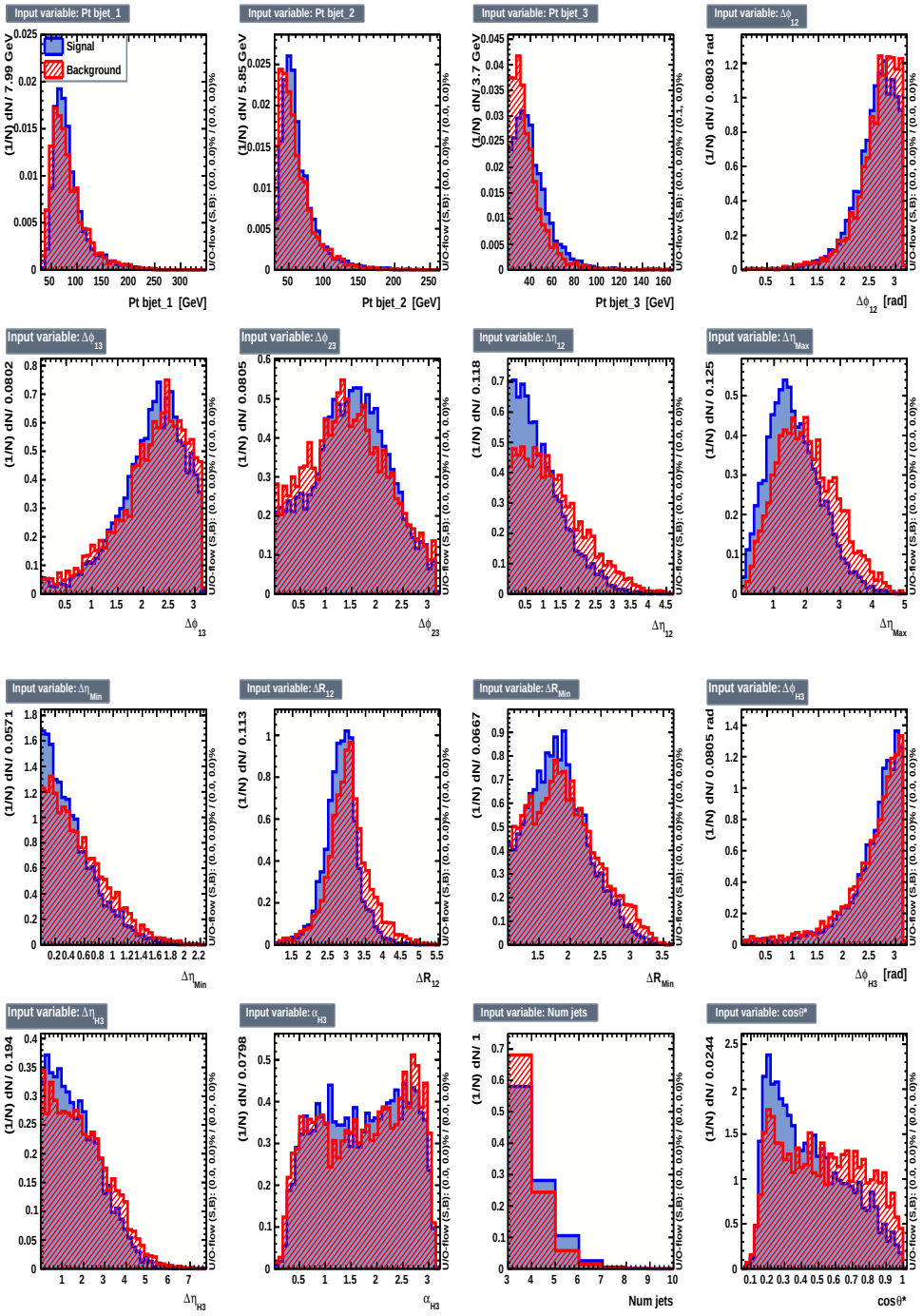


Figure 5.9: Normalized distributions of the kinematic variables used in the discriminator definition for the low mass region (see text). The QCD (red) and signal (blue) distributions are superimposed. The signal is the combination of the three mass points: $M_\Phi = 120, 130, 140\ GeV/c^2$.

simulation must be applied in order to reproduce the efficiency measured in the data. For cut on the b-tagging variable $CSV > 0.679$, which is very close to the one used in this analysis for the third jet ($CSV > 0.7$), the scale factor is 0.95 ± 0.03, and it is fairly independent on the b-jet p_T for the momentum range

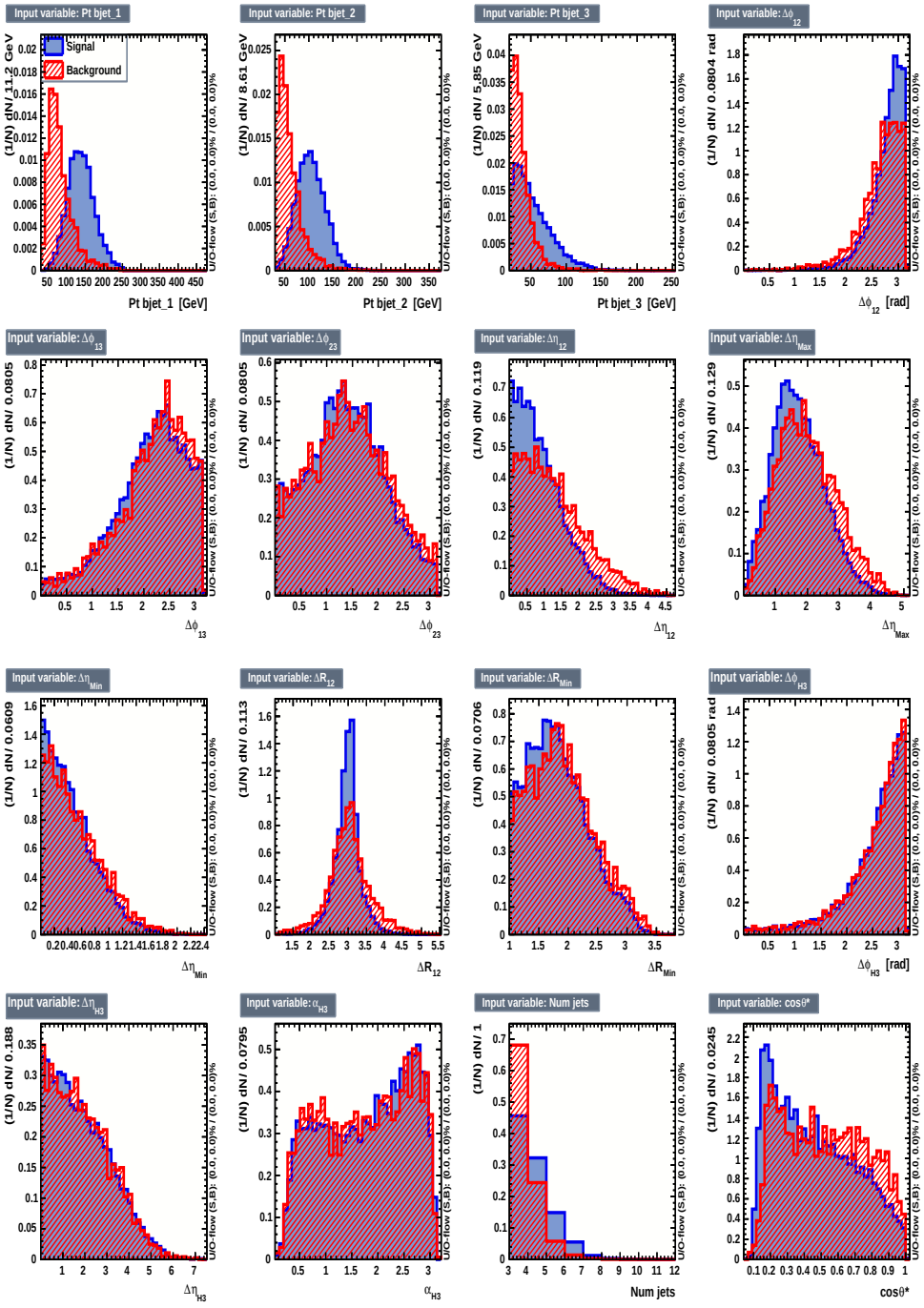
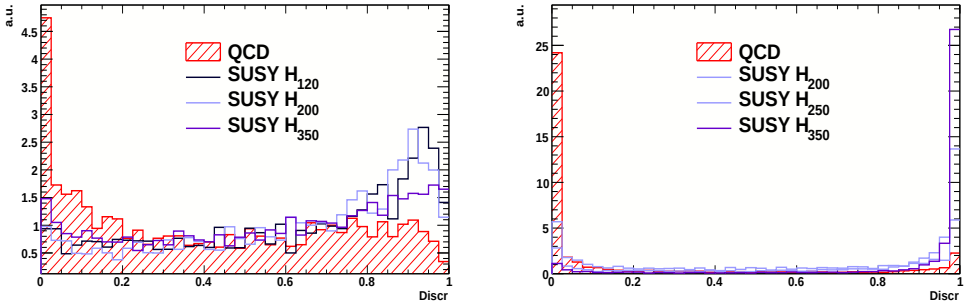


Figure 5.10: Normalized distributions of the kinematic variables used in the discriminator definition for the high mass region (see text). The QCD (red) and signal (blue) distributions are superimposed. The signal is the combination of the three mass points: $M_\Phi = 250, 300, 350\ GeV/c^2$.

relevant in this analysis. This scale factor is applied to the b-tagging efficiency.



(a) Discriminator at low mass region.

(b) Discriminator at high mass region.

Figure 5.11: Distributions of the discriminator for the low and high mass regions. The distributions of the QCD and the corresponding signals used for the discriminator computation in each mass region are superimposed.

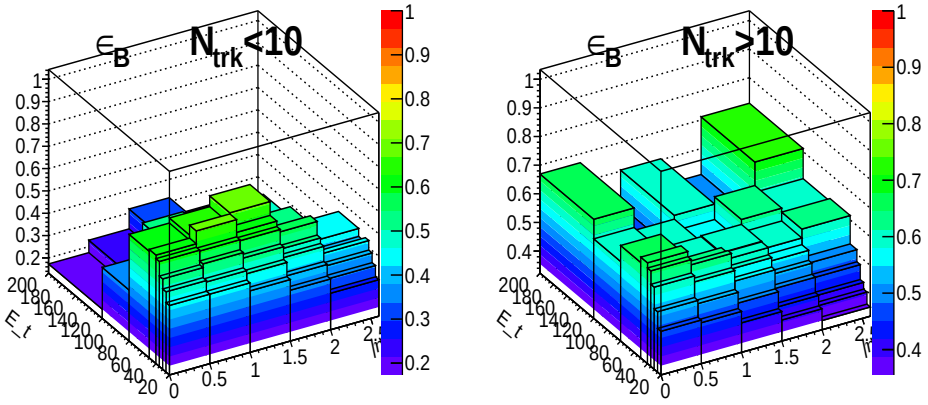


Figure 5.12: b-tagging efficiency as a function of p_T, η of the b-tagged ($CSV > 0.7$) third jet originated from b-quarks. The distribution is split for low (left) and high (right) charged tracks multiplicity.

5.5.4 Measurement of the third jet flavor fractions

As was already stated before, the flavor fractions $f_{b,c,light}$ for the third jet are obtained from the data exploiting the control region of the *bbj* sample. Two variables sensitive to b-jet characteristics are used to extract these fractions: the invariant mass of all the tracks, assumed to be charged pions, associated to the secondary vertex found in the third jet, $TagMass$, and the $JetBProbability$ of the third jet, defined in section 4.2.4. The *bbj* sample is divided in different intervals of the four variables used in equation 5.5 for the parametrization of the third jet flavour content. For each interval, a fit of the two variables is performed in order to obtain the fractions, using templates for the $TagMass$ and $JetBProbability$ distributions of b, c and light-partons obtained from Monte Carlo simulated events. The interval division in p_T and η of the third jet is performed in a way to have a good estimation of the flavor fractions in the whole phase space, with enough statistics for each interval to perform the fit.
The sample of simulated events used to determine the templates for the two variables, $TagMass$ and $JetBProbability$, for the different flavor components

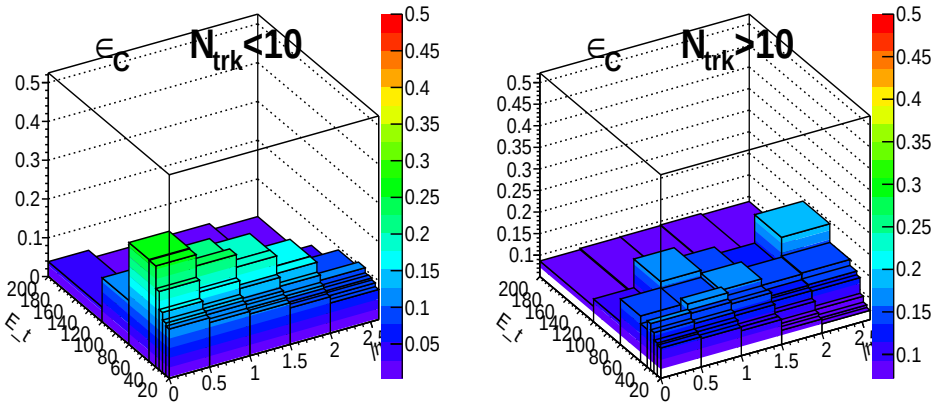


Figure 5.13: b-tagging efficiency as a function of p_T, η of the b-tagged ($CSV > 0.7$) third jet originated from c-quarks. The distribution is split for low (left) and high (right) charged tracks multiplicity.

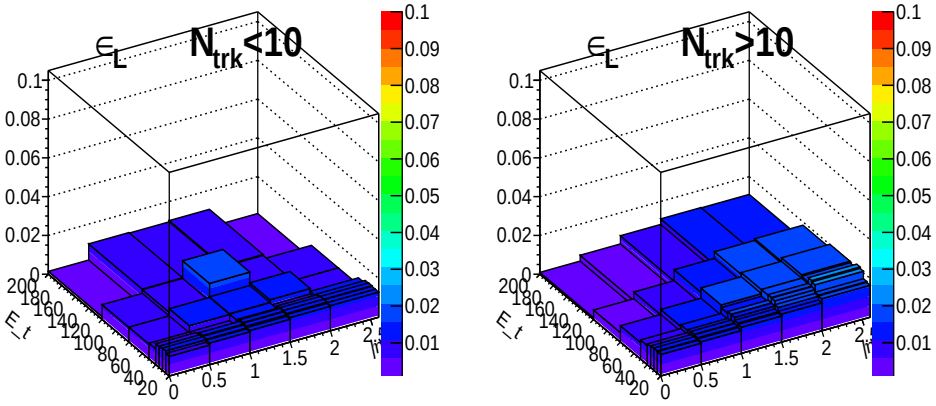


Figure 5.14: b-tagging efficiency as a function of p_T, η of the b-tagged ($CSV > 0.7$) third jet originated from light partons. The distribution is split for low (left) and high (right) charged tracks multiplicity.

of the third jet has been selected applying the offline selections, except the last one requiring the b-tagging on the third jet. To increase the available Monte Carlo statistics, the b-tagging requirements on the two leading jets have been removed. This selection, in fact, should not affect the physics of the third jet, thus introducing biases. In Figure 5.15 are shown the results of the Kolmogorov-Smirnov tests between the different templates obtained with and without the b-tagging requirements on the two leading jets, demonstrating the compatibility between the templates. The templates obtained are normalized to unity and used to perform the fits in the intervals of the variables used in equation 5.5. A two dimensional likelihood fit in the $TagMass$ and $JetBProbability$ variables space is adopted for such purpose. It can happen that no secondary vertex is reconstructed in the third jet, so that the $TagMass$ variable is not defined. In this case only a one dimensional fit of the $JetBProbability$ variable is performed. Since the bbj sample is made up of the b, c-quarks and light-partons categories, the condition:

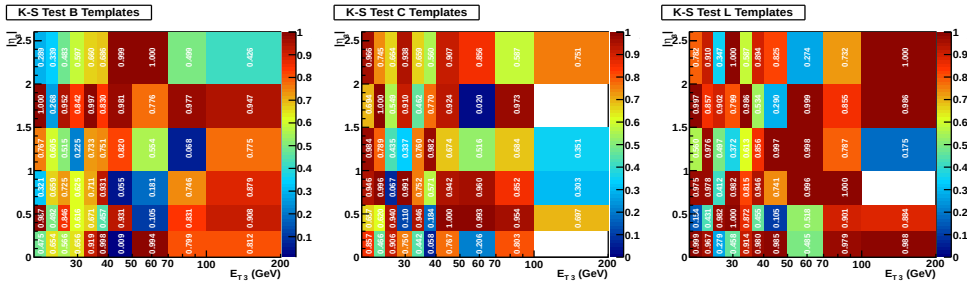


Figure 5.15: Kolmogorov-Smirnov test between templates with and without the b-tagging requirement on the two leading jets for the b-jet (left), c-jet (center) and light-parton (right) templates. The figure shows the parametrization in p_T and η only.

$$f_b + f_c + f_{light} = 1 \tag{5.7}$$

must be satisfied. In order to achieve the correct treatment of parameter errors near the boundaries, the fractions $f_{b,c,light}$ are not directly fitted, but they are replaced with the following parameter set:

$$\begin{aligned} \epsilon_1 &= \frac{1}{2}(1 + sin f_b); \\ \epsilon_2 &= \frac{1}{2}(1 - \epsilon_1)(1 + sin f_c); \\ \epsilon_3 &= 1 - \epsilon_1 - \epsilon_2; \end{aligned} \tag{5.8}$$

The likelihood function to be maximized in the fit is:

$$\begin{aligned} \mathscr{L}^{2D} &= \Pi \mathscr{P}_{TagMass}(n_i, \mu_i) \mathscr{P}_{jetBProb}(n_i, \mu_i); \\ \mathscr{P}(n_i, \mu_i) &= \frac{e^{\mu_i} \mu_i^{n_i}}{n_i!} \end{aligned} \tag{5.9}$$

where n_i and μ_i are the number of events and the expected number of events in each interval. The expected number of events μ_i is a combination of the b, c-quarks and light-flavor events:

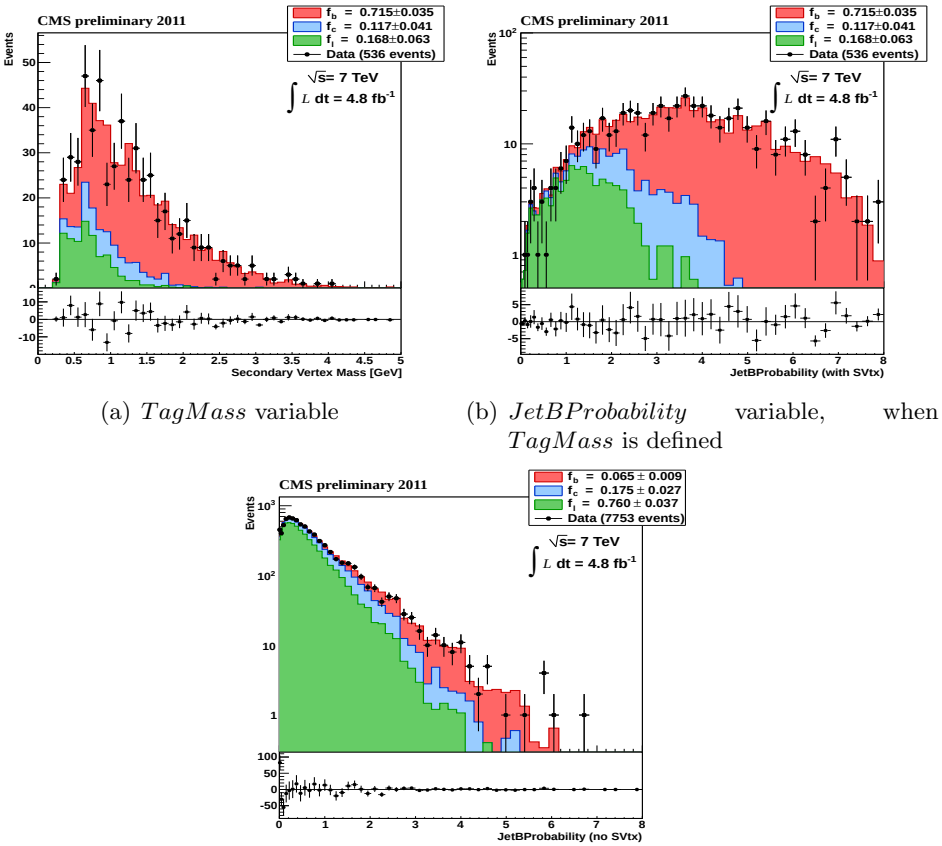
$$\mu_i = [\epsilon_1 n_i^b + \epsilon_2 n_i^c + \epsilon_3 n_i^{light}] \cdot \sum_i^{bins} n_i \tag{5.10}$$

If the $TagMass$ variable is not defined, only the second term of the likelihood 5.9 is considered for the maximization.
From the maximization, the values of $\epsilon_{1,2,3}$ are obtained, from which the flavor fractions $f_{b,c,light}$ are extracted for events in which the variable $TagMass$ is or is not defined. These values are combined through the weighted average expression:

$$f_{b,c,light} = \frac{f^{2D}_{b,c,light} \cdot N^{2D} + f^{1D}_{b,c,light} \cdot N^{1D}}{N^{2D} + N^{1D}} \tag{5.11}$$

where the appendix $2D$ and $1D$ refers to the events fitted with the two and one dimensional likelihood, respectively.

As an example, the template fits to the data for a given considered interval of the third jet variables are shown in Figure 5.16. The $TagMass$ and $JetBProbability$ variables are correlated, as shown in Figure 5.17. This is not an issue for the analysis, since the fit errors are not used in the analysis. As a closure test of the procedure, the results for the fitted flavor fractions in the simulated data in the control region are compared in Figure 5.18 with the true flavour fractions. The comparison with the Monte Carlo true values shows a small bias of the order of 6% averaged on the whole parameters space. As it will be shown in the next section, this has a negligible effect on the final prediction of the background shape and normalization. The results of the extracted fractions in data are shown in Figure 5.19, divided in two sub-samples: the one collected with single online b-tag trigger paths used in the first part of the data-taking period and the one collected with double online b-tag trigger paths, used in the second part of the data-taking. As expected, the f_b is higher in the last data sample.



(a) $TagMass$ variable

(b) $JetBProbability$ variable, when $TagMass$ is defined

(c) $JetBProbability$ variable, when $TagMass$ is not defined

Figure 5.16: Flavor fractions template fits to data (black dots) for one of the intervals in p_T, η of the third jet. The red, green and blue histograms are the b, c and light-partons, respectively. (a) and (b) are the 2D fit to $TagMass$ and $JetBProbability$. (c) is the 1D fit to the $JetBProbability$, when the $TagMass$ is not defined.

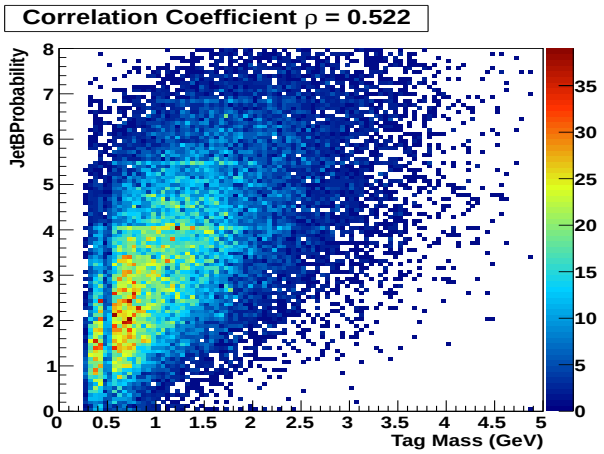
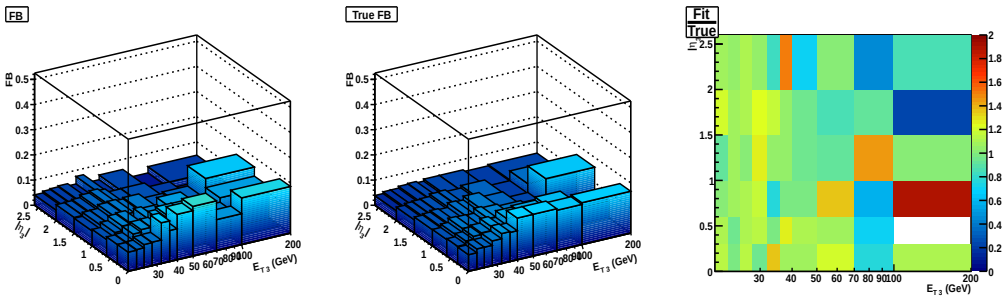
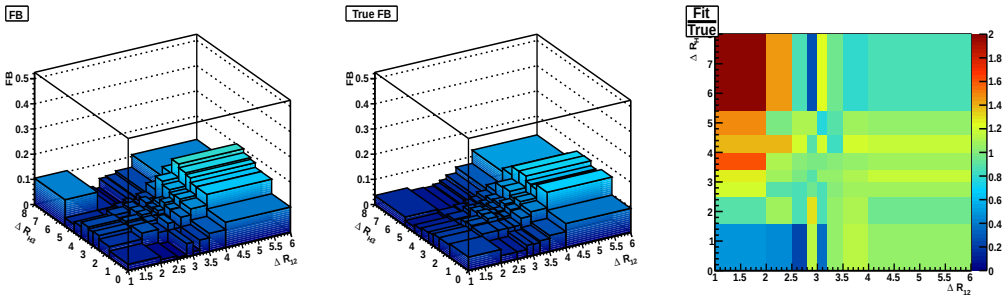


Figure 5.17: Two dimensional distribution in the $TagMass$ and $JetBProbability$, when a secondary vertex is reconstructed.



(a) $(p_T, \eta)_{3^{rd}\,jet}$ *parametrization*

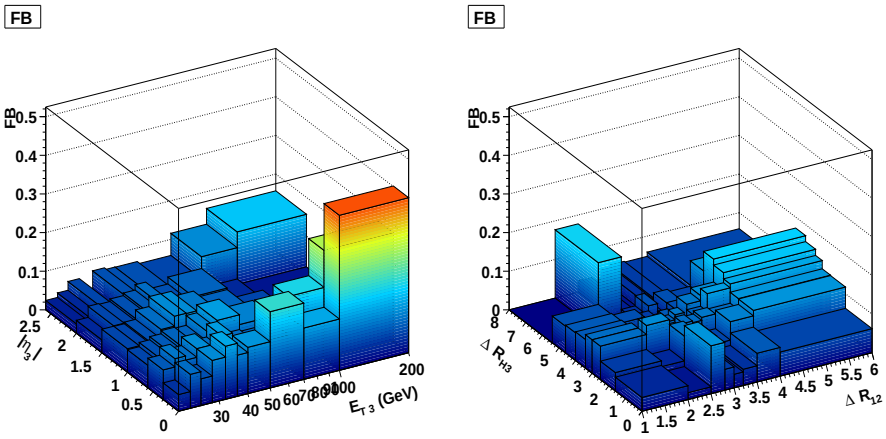


(b) $(\Delta R_{1,2}, \Delta R_{\Phi,3})$ parametrization

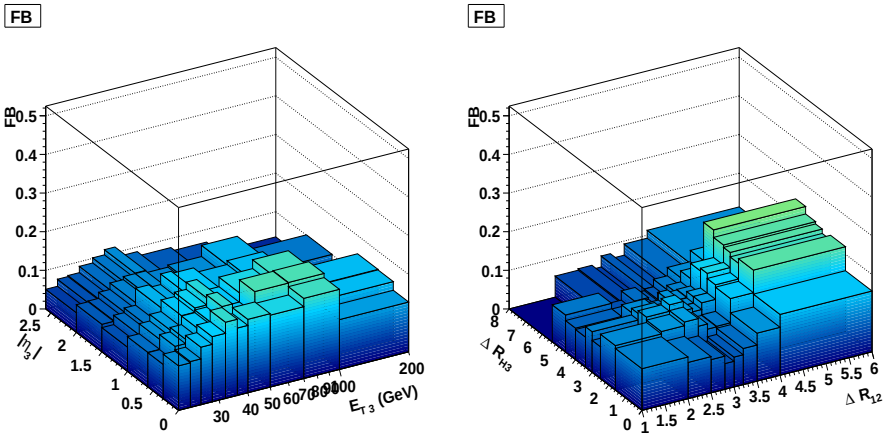
Figure 5.18: Comparison between fitted (left) and true (center) b flavor fraction, f_b, on Monte Carlo simulated QCD events, shown for the (p_T, η) parametrization (upper plots) and ΔR parametrization (lower plots). The ratio between fitted and true f_b is also shown (right).

5.5.5 Prediction of kinematic variables in the *bbb* sample

By the usage of the b-tagging efficiencies and flavor fractions, the interesting kinematic variables for this analysis can be predicted in the *bbb* sample, starting from the *bbj* sample, both in the control region and in the signal region, where we look for the signal presence. The predictions on the distribution of some key variables, together with the observed distributions, are shown in Figures 5.20 through 5.23, for the p_T and η of the third jet, the invariant mass of the two



(a) Results for the data sample collected with the single b-tag trigger paths.



(b) Results for the data sample collected with the double b-tag trigger paths.

Figure 5.19: Results of the fitted b-jet flavor fraction on data for the $(p_T, \eta)_{3^{rd}-jet}$ (left) and $(\Delta R_{1,2}, \Delta R_{\Phi,3})$ (right) parametrization.

leading jets $M_{1,2}$ and the discriminator likelihood ratio variable. All these distributions are shown for the low mass region search. The corresponding ones for the high mass region are shown in Figures 5.24 through 5.27. As a reference, the distributions for a Higgs signal of $M_\Phi = 120\ GeV$, for the low mass distributions, and $M_\Phi = 250\ GeV$, for the high mass distributions, are superimposed. The expected Higgs signal yield corresponds to the available integrated luminosity of 4.8 fb^{-1} times the expected production cross-section in the MSSM for $tan\beta = 30$. The distributions are shown for the control and signal regions both for the Monte Carlo QCD simulation, as a closure test, and for the data. In all the plots, the predictions are normalized to the observed events. The ratio between the observed and predicted events is also shown in the lower part of each plot, together with the result of the Kolmogorov-Smirnov test on their compatibility. It can be seen that in the Monte Carlo control region the shape and normalization are in very good agreement, showing a full closure of the technique. For the Monte Carlo signal region, one can notice a very good agreement in shape. Also the normalization shows a nice agreement within the statistical uncertainties of the simulated sample available, which however has almost a factor of ten less statis-

tics than the real data. It is worth to stress that the signal region is a completely separate sample with respect the one where the b-tagging efficiencies and flavor fractions are obtained. For what concerns the data, the shape agreement is also very good in the control region. However for the predictions the overall normalization factor shows an overestimation of about 10%.
The same distributions have also been split in the five different trigger paths, to check if any bias is present. The results for the invariant mass of the two leading jets and the discriminating variable are shown in Figures 5.28 and 5.29. Within the available statistics the shapes are very well predicted for all the triggers. A normalization scale factor similar to that found on the full data sample is observed, with the exception of the **Path 2** trigger, which is slightly lower.

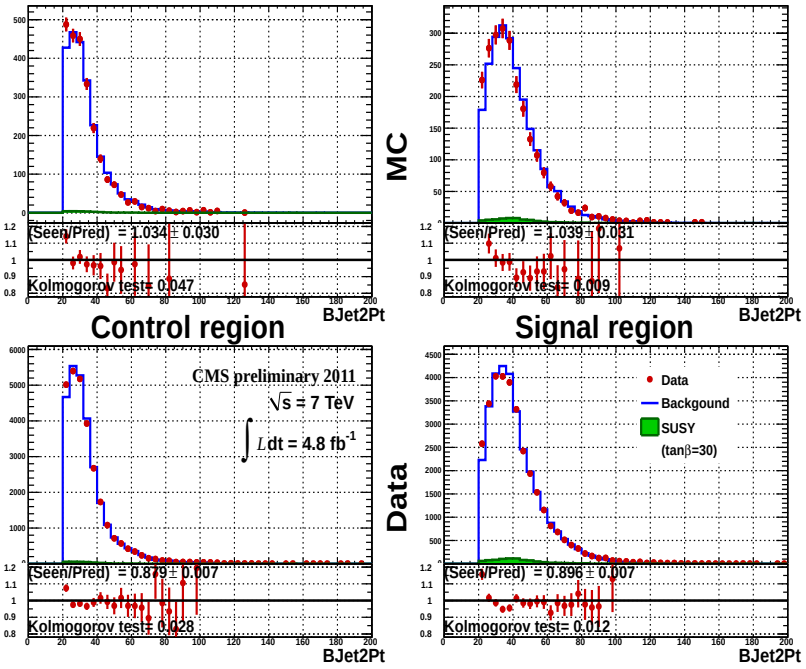


Figure 5.20: p_T of the third b-jet, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the low mass region. The predicted distribution (blue histogram) from bbj sample is normalized to the observed distribution (red dots) in the bbb sample. The ratio between observed and predicted distributions is also shown below each plot. The expected Higgs signal in MSSM, shown by the green histograms, corresponds to $M_\Phi = 120\ GeV$ and $tan\beta = 30$.

5.5.6 The hyperball method

An estimation of shape and normalization of the background is possible with an alternative and independent method to the one presented in the previous section. This method, known as the Hyperball method, is also data-driven. The prediction of the relevant kinematic variables that are used for the signal search is performed starting from a sample of events selected using the criteria listed in section 5.3, except for the b-tagging requirements, that here are applied only on the leading jet. This way a sample of one b-tagged jets plus a second and third jets is obtained

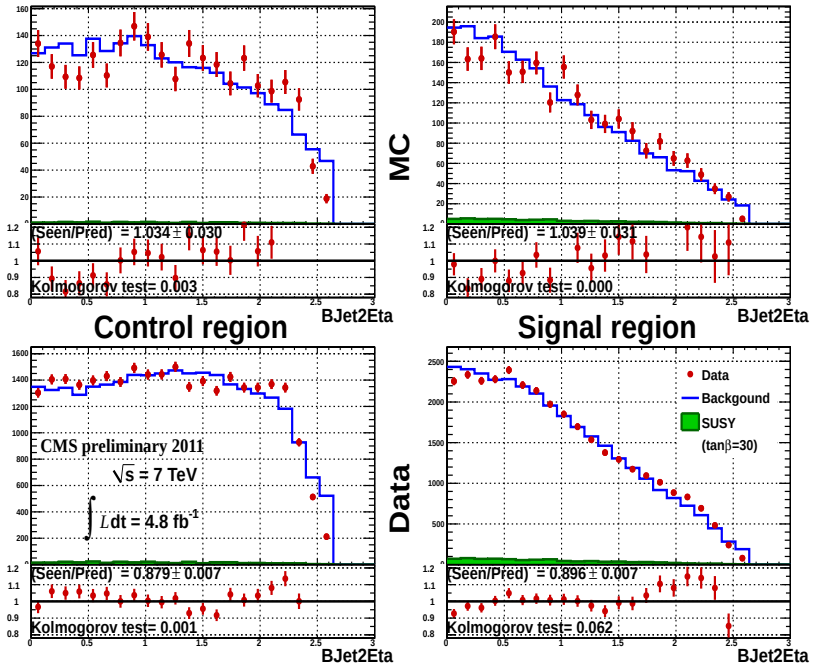


Figure 5.21: η of the third b-jet, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the low mass region. The predicted distribution (blue histogram) from bbj sample is normalized to the observed distribution (red dots) in the bbb sample. The ratio between observed and predicted distributions is also shown below each plot. The expected Higgs signal in MSSM, shown by the green histograms, corresponds to $M_\Phi = 120\ GeV$ and $tan\beta = 30$.

(this is called bjj). In order to keep the hyperball algorithm results completely orthogonal from those of the B-matrix method described in the previous section, we exclude events where the two leading jets are both b-tagged, since those are the events which are the basis of the background prediction in the B-matrix approach. The prediction of the variables for the bbb sample is obtained applying a weight to each event belonging to the bjj sample, which is the probability that all the three jets are b-tagged in the event. The bjj sample is divided into a control and signal region, using the same discriminator exploited in the $B-Tag$ Matrix method. In the control region a subset of N_H events are selected as *similar* events (see below) to those in the signal region. The probability that all the three jets are b-tagged is obtained from the ratio between the bbb over the bjj events in the *similar* sample of events.

5.5.7 Sample of *similar* signal events from control region sample

The criterion used to identify the sample of *similar* events to those in the signal region from the control region sample is based on the multi-dimensional distance defined by the following expression:

$$D^2 = \sum_{i=1}^{n_V} w_i^2 (x_i - y_{\alpha i})^2 \tag{5.12}$$

where the x_i are the variables that characterize the signal region events, the

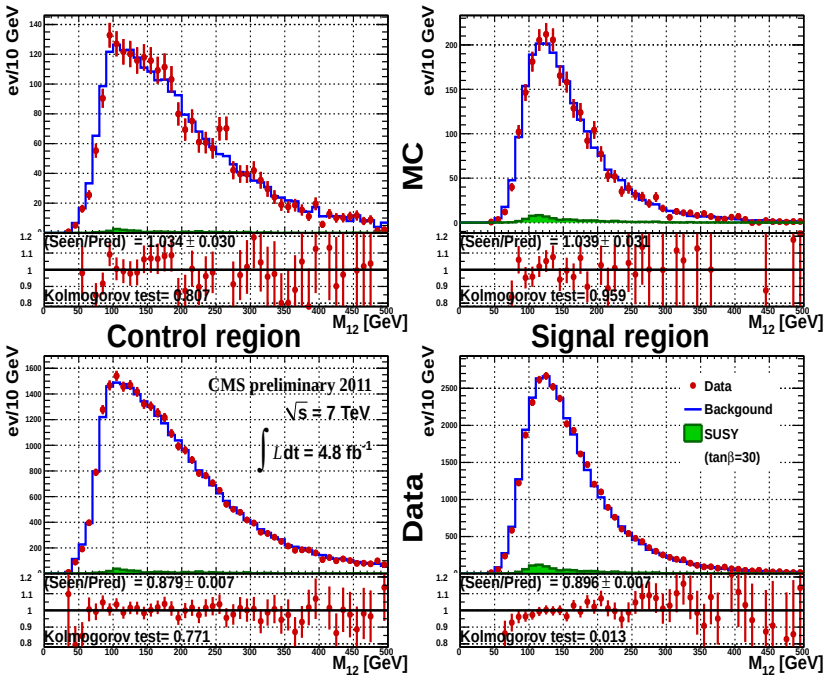


Figure 5.22: Invariant mass of the two leading b-jets, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the low mass region. The predicted distribution (blue histogram) from *bbj* sample is normalized to the observed distribution (red dots) in the *bbb* sample. The ratio between observed and predicted distributions is also shown below each plot. The expected Higgs signal in MSSM, shown by the green histograms, corresponds to $M_\Phi = 120\ GeV$ and $tan\beta = 30$.

$y_{\alpha i}$ are the corresponding variables for the α event in the control region. The weights w_i account for the different dispersions of the variables and they are computed by mean of the procedure explained below. The n_V are the number of the variables chosen for the signal event characterization, that are listed in Table 5.7. The subset of *similar* events is selected as the set of events which have the minimum distance 5.12 to the signal region events.
To compute the weights w_i, the range of each variable x_i is divided into 10 intervals, with different sizes chosen to have all intervals equally populated; for each interval the fraction of three b-tagged events, f_i, is computed. For any value of the variable x_i, the weight w_i is given by choosing the two intervals with $\bar{x}_i$ averages immediately smaller ($\bar{x}_{iL}$) and bigger ($\bar{x}_{iR}$) than x_i, and computing the ratio of the differences between the *bbb* events fraction and the x_i average distance:

$$w_i = \frac{f_{iR} - f_{iL}}{\bar{x}_{iR} - \bar{x}_{iL}} \tag{5.13}$$

To reduce the bias introduced by assuming that the average *bbb* events fraction does correspond to the fraction at the average $\bar{x}_i$, a least-square fit of the fraction itself to a quadratic function of x_i is performed, and its value at $\bar{x}_i$ is taken. The above procedure estimates the gradient of the b-tagging probability in each of the directions of space as if this were independent on the particular location in the space where we need to estimate it.

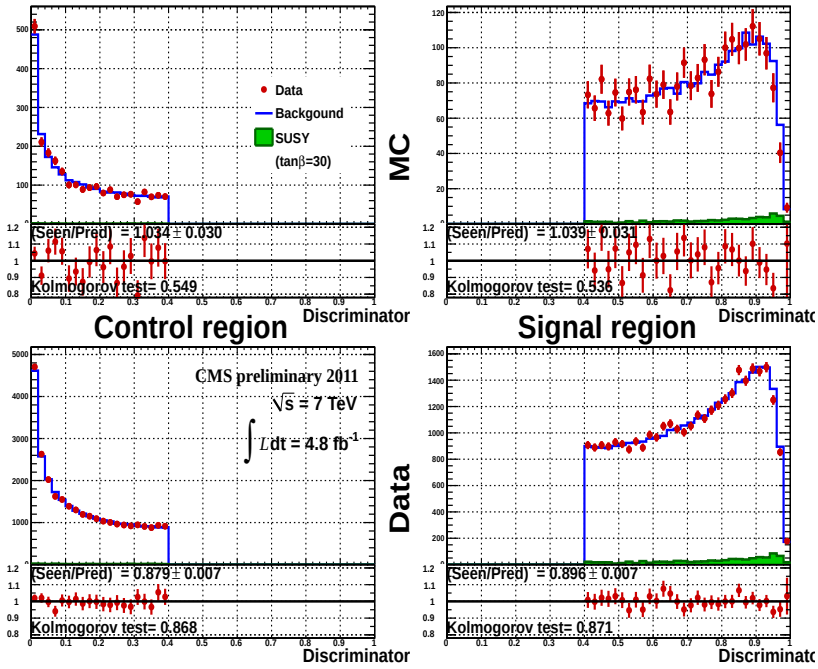


Figure 5.23: Discriminator variable, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the low mass region. The predicted distribution (blue histogram) from bbj sample is normalized to the observed distribution (red dots) in the bbb sample. The ratio (lower distribution in each plot) is between observed and predicted distributions. The Higgs signal (green histogram) corresponds to $M_\Phi = 120\ GeV$ and $tan\beta = 30$.

5.5.8 The event-by-event rate interpolation

The estimation of a b-tag probability from the ratio of the number of bbb and bjj events in the hyper-ellissoid could result in a bias when the tested event has some variable near to its threshold. In this situation the tested event is not centered in the hyperellissoid, and the fraction of bbb events in the hyperellissoid itself is not a good estimation of the probability for the test event to have all three jets b-tagged.
This bias was cured by looking, for each test event, which variables among the transverse energies and pseudo-rapidities are near to the corresponding thresholds and performing a linear interpolation with those variables:

$$z = f + \sum_i a_i(x_i + y_{\alpha i}) \tag{5.14}$$

In the equation z is 1 for bbb events and 0 for the others, while x_i and $y_{\alpha i}$ are the same as in eq. 5.12, with i running only over the chosen variables. The intercept f is taken as the 3 b-tags probability for the test event. The transverse energies of the jets are included in the linear fit when smaller than 40, 35 and 25 GeV respectively for the first, second and third leading jet respectively. The pseudo-rapidities were included in the fit when their absolute value was bigger than 1.5.
When all the energies and pseudorapidities in the tested event were not close to the above defined thresholds, the 3 b-tags probability for the test event was

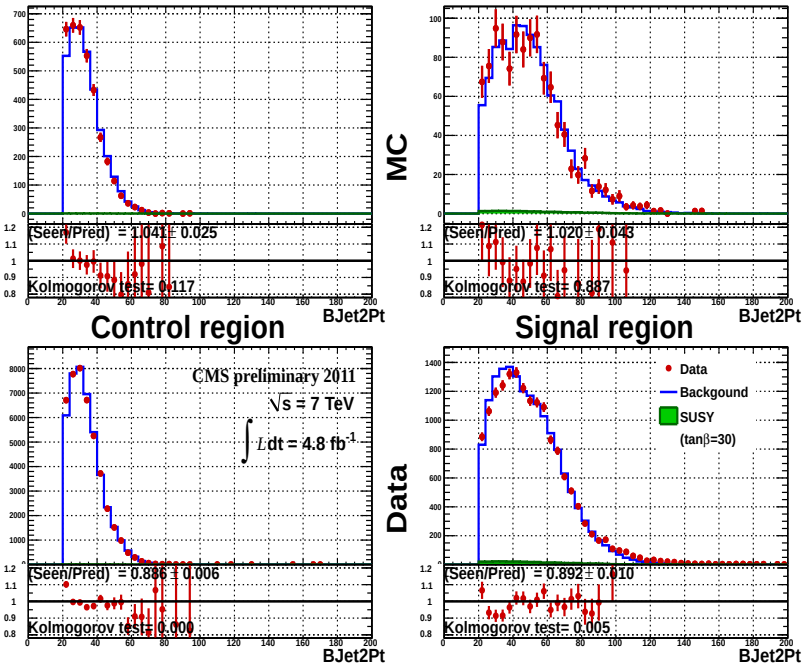


Figure 5.24: p_T of the third b-jet, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the high mass region. The predicted distribution (blue histogram) from bbj sample is normalized to the observed distribution (red dots) in the bbb sample. The ratio (lower distribution in each plot) is between observed and predicted distributions. The Higgs signal (green histogram) corresponds to $M_\Phi = 250\ GeV$ and $tan\beta = 30$.

computed by a weighted ratio:

$$P = \frac{\sum_\beta \frac{1}{D_\beta^2}}{\sum_\alpha \frac{1}{D_\alpha^2}} \tag{5.15}$$

where in the numerator the sum runs over bbb events and in the denominator it runs over bjj events; D is the distance between the test event and the training event as defined in eq. 5.12. In this way the most distant events have a smaller weight in the sum, so that any residual bias in the determination of P should be reduced.

5.5.9 Choosing and tuning of hyperball parameters

In order to apply the hyperball method to the real data, a tuning of the different parameters used in the method is needed. The different variables that can be set are the following:

- number of training events, N_T;
- number of events in the hyper-ellipsoid used for the tag-rate calculation, N_H;
- number of the observable quantities defining the multi-dimensional space, N_V;

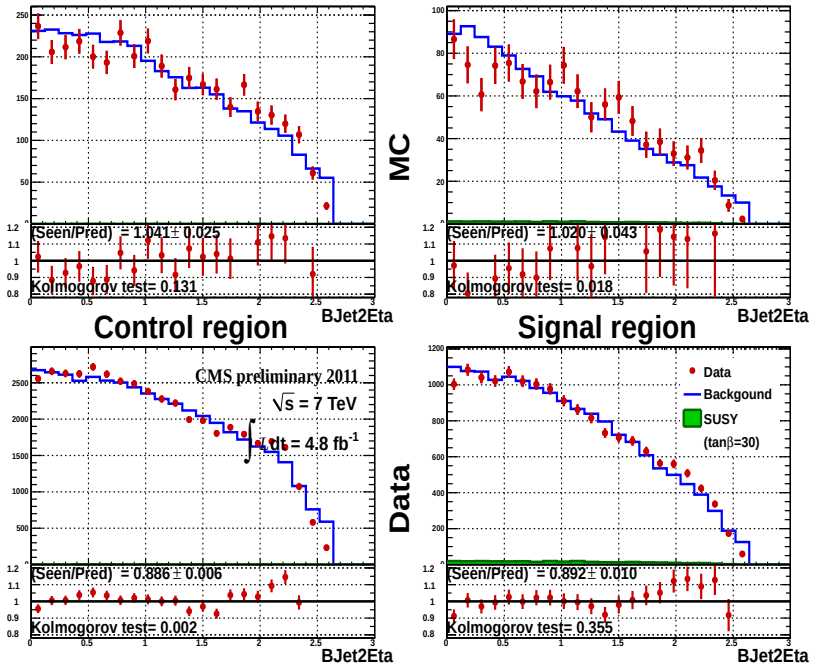


Figure 5.25: η of the third b-jet, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the high mass region. The predicted distribution (blue histogram) from bbj sample is normalized to the observed distribution (red dots) in the bbb sample. The ratio (lower distribution in each plot) is between observed and predicted distributions. The Higgs signal (green histogram) corresponds to $M_\Phi = 250\ GeV$ and $tan\beta = 30$.

- identity of the variables spanning the space;
- number of intervals of the invariant mass distributions between the two leading jets, each of one will contain the number of events predicted, N_{int};

The choise of the parameters listed above gives the operating point of the hyperball algorithm. As any other multivariate approach, the hyperball method is sensitive to the size of the training sample: for that reason, N_T should generally be chosen as large as possible. However, since the data sample has a finite size, we have to account for this while we optimize the other parameters of crucial importance: the hyperball size N_H, and the number of variables N_V. Tests with Monte Carlo simulated QCD events are performed, for which the number of possible training events is not larger than 400.000 after the cuts are applied, with the b-tagging requirement only in the leading jet. This is an adequate number since it roughly corresponds to the number of events in the data at the same level of selection, for each of the subsets of data collected by a single trigger version. N_H is quite important, and it is of course tightly connected to the total training sample N_T and the number of space dimensions N_V. For $N_T = 400.000$ and 10 space dimensions, already hyperballs of size $N_H > 20$ events will force the averaging of the b-tagging probability over a very wide subspace, as is clear by considering that a uniform weighting of the space dimensions would result in $(N_T/N_H)^{1/N_V} = 2.7$ "effective intervals" per variable. On the other hand, the b-tagging rate defined as a ratio between three- and one-b-tag events requires the

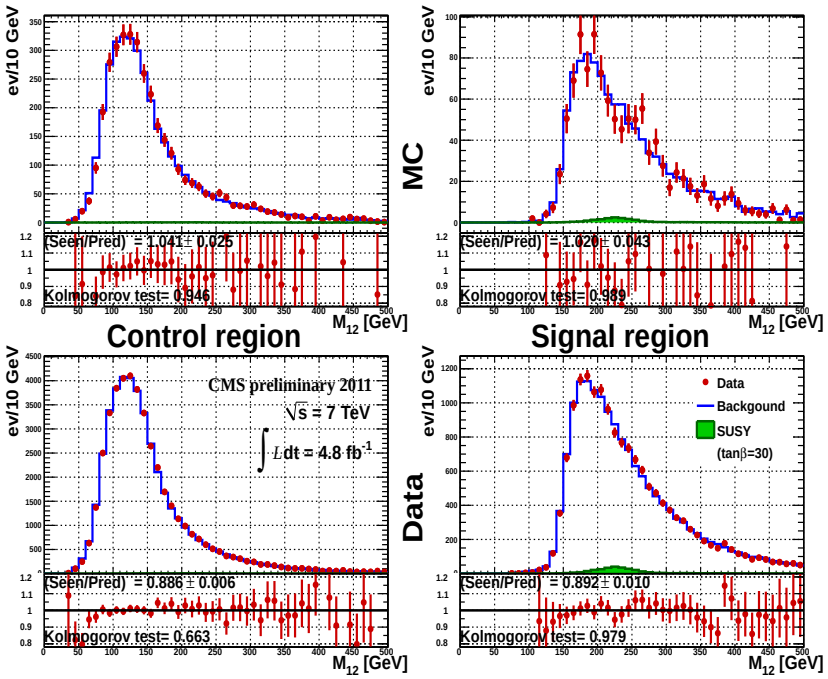


Figure 5.26: Invariant mass of the two leading b-jets, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the high mass region. The predicted distribution (blue histogram) from bbj sample is normalized to the observed distribution (red dots) in the bbb sample. The ratio (lower distribution in each plot) is between observed and predicted distributions. The Higgs signal (green histogram) corresponds to $M_\Phi = 250\ GeV$ and $tan\beta = 30$.

parametrization of, at the very least, the energy of jets and their pseudo-rapidity, plus a few angles between them (on which the physics of heavy flavor production is quite sensitive, as we have demonstrated in the previous sections). So with N_V needed of at least 10, N_H appears to be constrained to be small. On the other hand, a too small N_H can cause a very large statistical uncertainty in the b-tagging rate in each hyperball (the typical value of the bbb/bjj ratio is 0.05, meaning that for a 20 events hyperball on average only one event has three jets identified as b-jets). So the conflicting requirements of parametrizing accurately the variations of the b-tagging rate in the space -reducing possible systematic uncertainties- and reducing the statistical uncertainty of the estimate, point in the direction of $N_H \sim 100$ events.
The variables chosen to be used for the distance evaluation, 5.12, are the ones which exhibited the strongest variation of the bbb/bjj ratio. They are listed in Table 5.7. It has been verified that by increasing the number of variables the prediction rapidly worsens its power, especially for what concerns the normalization of the three b-tagged data, which becomes underpredicted: this is as expected, since by forcing a parametrization on too many variables we are effectively expanding the hyperellipsoids, which will end up containing events very far, in directions along which the b-tag probability changes rapidly, from the test point. Since we are parametrizing with a training set which has low values of the kinematic discriminant, and since overall the b-tagging rate in that region of

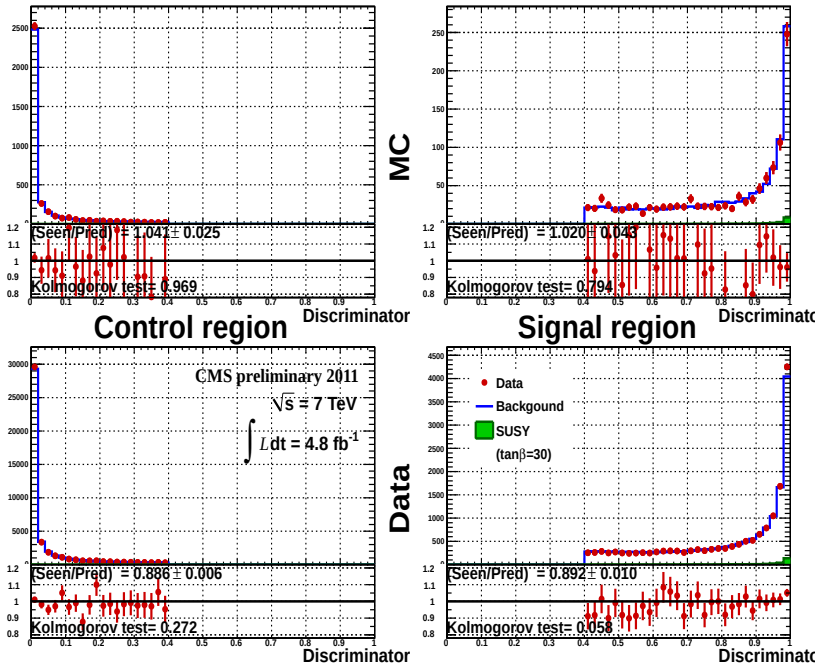


Figure 5.27: Discriminator variable, for Monte Carlo (up) and data (bottom) and for control (left) and signal (right) region, in the high mass region. The predicted distribution (blue histogram) from *bbj* sample is normalized to the observed distribution (red dots) in the *bbb* sample. The ratio (lower distribution in each plot) is between observed and predicted distributions. The Higgs signal (green histogram) corresponds to $M_\Phi = 250\ GeV$ and $tan\beta = 30$.

Table 5.7: List of variables used on the hyperball method to measure the *distance* between events, as defined in 5.12.

Name	Description
E_T^1	transverse energy of the leading jet
η^1	pseudo-rapidity of the leading jet
E_T^2	transverse energy of the sub-leading jet
η^2	pseudo-rapidity of the sub-leading jet
n_{trk}^2	track multiplicity of the sub-leading jet
E_T^3	transverse energy of the third sub-leading jet
η^3	pseudo-rapidity of the third sub-leading jet
n_{trk}^3	track multiplicity of the third sub-leading jet
$\Delta\phi_{12}$	angle between the leading and sub-leading jets in the transverse plane
ΔR_{23}	distance in the (η, ϕ) plane between the second and third leading jets
M_{12}	invariant mass of the two leading jets
p_T^{12}	transverse momentum of the two leading jets
p_T^{23}	transverse momentum of the second and third leading jets
p_T^{123}	transverse momentum of the three leading jets

the discriminant is smaller than in the signal region, the deficit in prediction is clearly expected.
In Figures 5.30 and 5.31, the comparison of the predicted and observed distri-

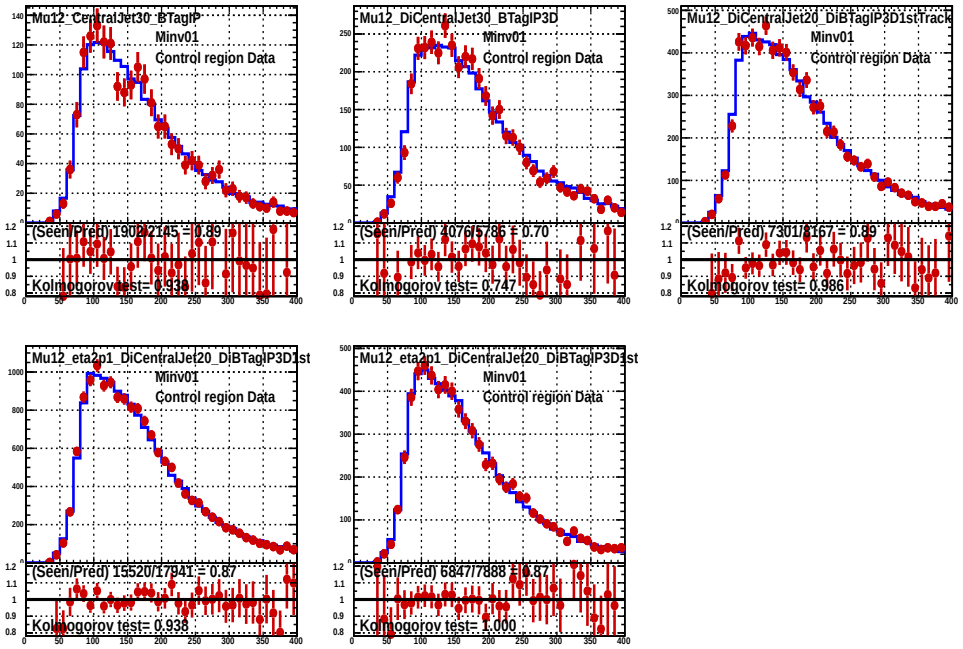


Figure 5.28: Invariant mass of the two leading b-tagged jets, for data control region, as predicted (blue histogram) and measured (red dots) for the five different trigger paths used in the analysis.

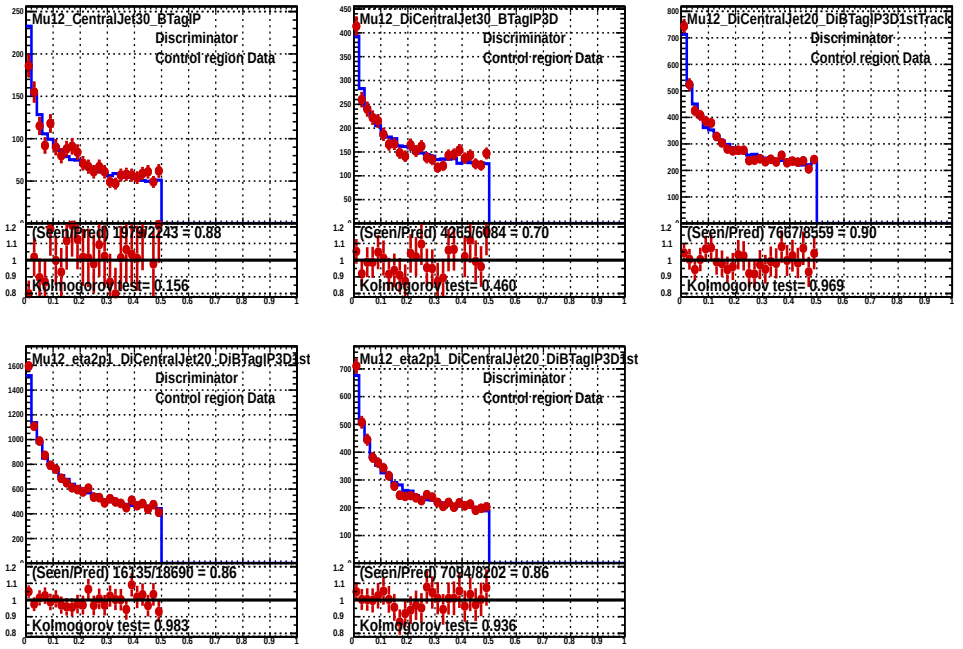


Figure 5.29: Discriminator variable, for data control region, as predicted (blue histogram) and measured (red dots) for the five different trigger paths used in the analysis.

bution for QCD simulated events are shown, for low and high mass regions, respectively, and for control and signal region. Similar results for data, in control

region only, are shown in Figure 5.32, for both low and high mass regions.

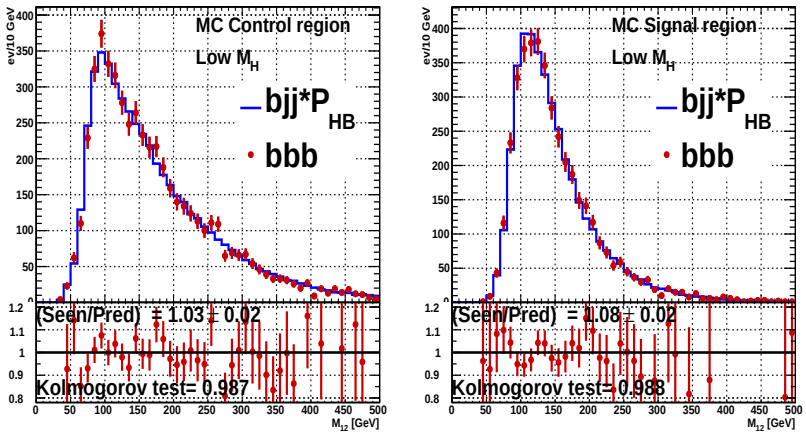


Figure 5.30: Invariant mass of the two leading jets as predicted by the Hyperball method (blue line) compared to the observed distribution (red dots) in QCD Monte Carlo simulated events. The results are shown for low Higgs mass region ($M_H < 200\ GeV/c^2$) in control (left) and signal (right) region. The predicted and observed distributions are normalized to the same area, and the ratio between the true events is shown on the bottom of the distributions together with the Kolmogorov test result.

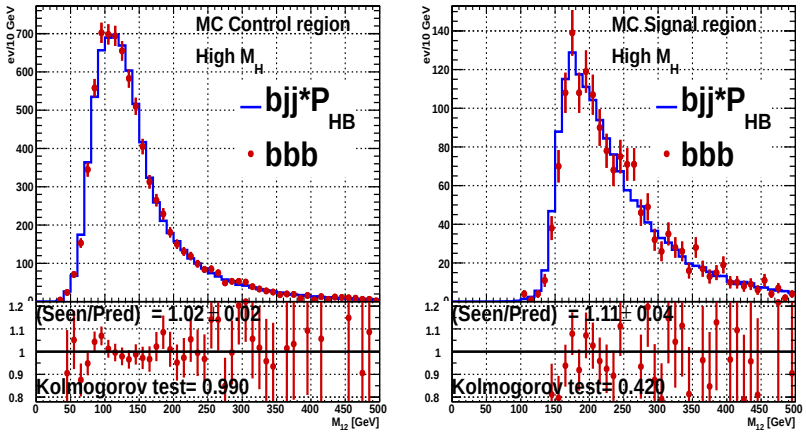


Figure 5.31: Invariant mass of the two leading jets as predicted by the Hyperball method (blue line) compared to the observed distribution (red dots) in QCD Monte Carlo simulated events. The results are shown for high Higgs mass region ($M_H \geq 200\ GeV/c^2$) in control (left) and signal (right) region. The predicted and observed distributions are normalized to the same area, and the ratio between the true events is shown on the bottom of the distributions together with the Kolmogorov test result.

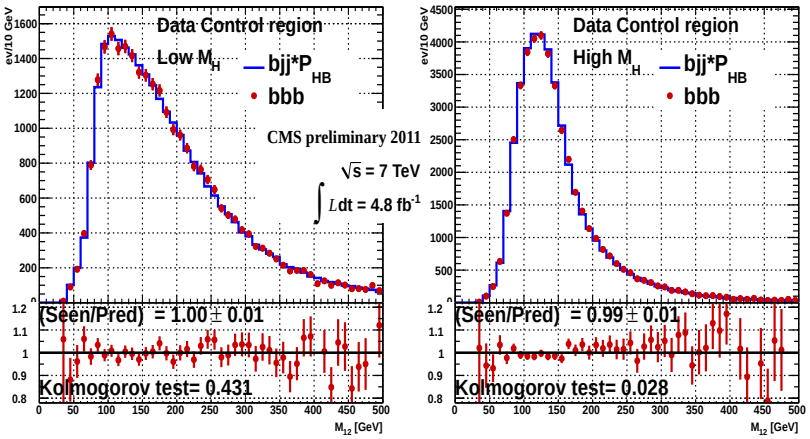


Figure 5.32: Invariant mass of the two leading jets as predicted by the Hyperball method (blue line) compared to the observed distribution (red dots) in data control region events. The results are shown for both low [left] ($M_H \leq 200\ GeV/c^2$) and high Higgs mass region [right] ($M_H > 200\ GeV/c^2$). The predicted and observed distributions are normalized to the same area, and the ratio between the true events is shown on the bottom of the distributions together with the Kolmogorov test result.

5.5.10 Combination of the background predictions using the two methods

Since the $B-Tagging$ Matrix and Hyperball methods are independent, the templates obtained for the background prediction exploiting the two data-driven methods can be combined. This leads to a unique predicted background distribution with reduced systematic on the shape.
The combination is done in the region where the signal search is performed, that is the signal region. The total normalization of the predicted background is chosen to be given by the $B-Tagging$ Matrix method, therefore the Hyperball background template is scaled in accordance. Then, the shapes are combined by performing a weighted bin-by-bin average of the two distributions:

$$x_W = \frac{x_{BM}\sigma_{BM}^2 + x_{HB}\sigma_{HB}^2}{\sigma_{BM}^2 + \sigma_{HB}^2} \tag{5.16}$$

where $x_{BM,HB}$ and $\sigma_{BM,HB}$ are the number of events that fall in a given bin of the background distribution and the associated statistical error, for $B-Tagging$ Matrix (BM) and Hyperball (HB) methods. The total statistical error is:

$$\sigma_{W,scal}^2 = \chi^2\sigma_W = \chi^2\Big(\frac{1}{\sigma_{BM}^2} + \frac{1}{\sigma_{HB}^2}\Big)^{-1} \tag{5.17}$$

which is the statistical error of two combined measures, σ_W, scaled by the χ^2 defined as:

$$\chi^2 = \frac{1}{N-1}\Big[\Big(\frac{x_{BM}-x_{HW}}{\sigma_{x_{BM}}}\Big)^2 + \Big(\frac{x_{HB}-x_{HW}}{\sigma_{x_{HB}}}\Big)^2\Big] \tag{5.18}$$

when the determinations with the two methods are incompatible in that bin, as it is described in the introduction of the Review of Particle Properties [3].

The difference between the unscaled and scaled statistical error ($(\chi^2 - 1)\sigma_W$) is interpreted as a systematic uncertainty on the shape of the background prediction. In Figure 5.33 are shown the predicted background distributions obtained with the two methods compared to the weighted average one, for the two leading jet invariant mass.

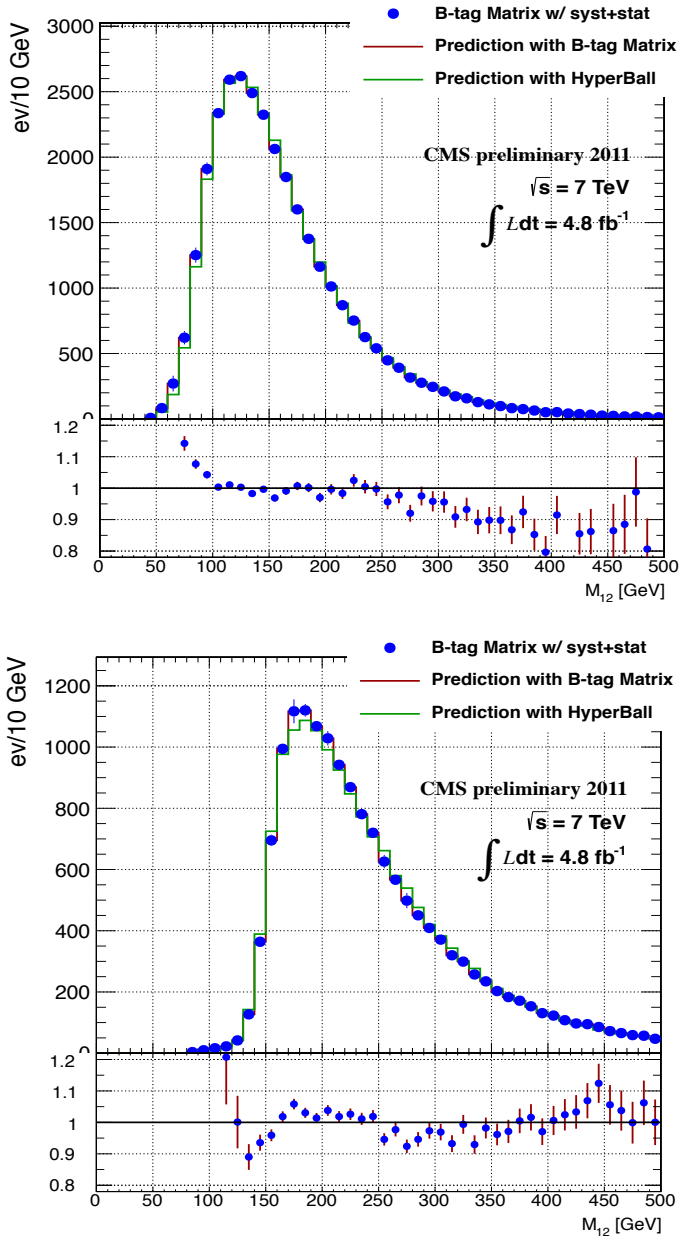


Figure 5.33: Invariant mass of the two leading jets: background predictions with $B - Matrix$ (red line) and Hyperball (green line) methods, compared to the weighted average (blue dots) in the data signal region, for data signal region in the low (left) and high (right) Higgs mass regions. The ratio of the distributions obtained with the two methods is shown in the bottom part of the graphs.

Chapter 6

Signal extraction and upper limits

6.1 Sources of uncertainties affecting the signal search

The sensitivity of the analysis is affected by uncertainties that limit the precision with which the signal and background are known. Two main categories of systematic uncertainties are identified: those affecting the signal yield and those related to the background prediction.

6.1.1 Systematics on signal yield

The signal yield is affected mainly by systematics uncertainty on the jet reconstruction and b-tagging, and by the uncertainty on the trigger efficiency turn-on curves. Additional uncertainty sources comes from how well the integrated luminosity is measured, the PDF modeling, and the precision of the lepton identification. These sources are listed in the following:

- *Trigger systematics*: in order to evaluate the systematics due to the measured trigger efficiency, the parameters parametrizing the turn-on curves, ϵ_μ and ϵ_{hadr}, are variated by plus and minus their error (see section 5.4). Once the scaled turn-on curves obtained by these variations are applied, the change of the number of events in the signal dataset is roughly $\approx 5\%$ and $\approx 3\%$ for the signal hypothesis with $m_\Phi = 120\ GeV/c^2$ and $m_\Phi = 250\ GeV/c^2$, respectively;

- *b-tagging efficiency*: a detailed study of this contribution has been performed using a b-enriched data sample from top decay [75]. The scale factor between MC and Data has been included in the efficiency estimated from the MC, and its error is used as a systematics: $\approx 4\%$ per jet, meaning $\approx 12\%$ for three jets events;

- *Jet Energy Scale (JEC)*: the effect of the uncertainty in the jet energy correction is estimated by scaling up and down the energy of all the jets in each event, according to the corresponding uncertainty for given p_T and η of the jets. The relative change in the amount of events passing the selection cuts is $^{+2.5\%}_{-3.1\%}$;

- *Jet Energy Resolution (JER)*: the contribution due to the resolution of the jet energy measurement is obtained by changing randomly the momenta of each jet, according to the corresponding uncertainty for given p_T and η of the jet. The procedure is repeated several times, obtaining different event yields. The corresponding uncertainty is then obtained from the observed standard deviation of the sample: $\pm 1.9\%$;

- *Muon momentum scale and resolution*: these effects are small given the precision of the muon measurement, estimated as $\approx 0.2\%$ and 0.6%, respectively [51];

- *Integrated luminosity*: the uncertainty on the measured integrated luminosity is 2.2% [76];

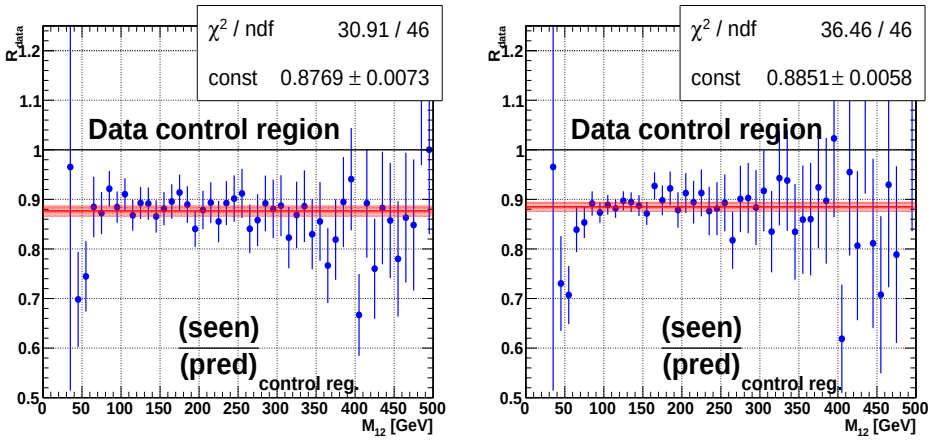
6.1.2 Systematics on the background

The background prediction is affected by the systematic uncertainties on its shape and normalization. The shape systematic uncertainties have been discussed in the section 5.5.10, where the background prediction obtained from the combination of the two data driven methods has been treated. Concerning the uncertainty on the background normalization, it can be obtained from the template obtained from the $B - Tagging$ Matrix method, since it gives the normalization of the predicted background. The possibility to compare the predicted background to the observed one in the control region, allows to obtain a scale factor for the background normalization. The comparison of the observed and predicted events in the data control region gives a scale factor with corresponding uncertainty, which is corrected by an extrapolation factor from control to signal region obtained from the Monte Carlo QCD sample. This leads to an additional contribution on the systematic uncertainty. The way this is done is described in the following:

- *Comparison of observed and predicted background in data control region*: the ratio of the predicted and observed distributions in the data control region is fitted with the lowest order polynomial compatible with the data, as it is shown in Figure 6.1. Given the good agreement between the *bbb* data and the prediction from *bbj*, a simple constant provide a good fit, as tested with a F-test with polynomials of increasing order. The normalization and the error are then constant across the distribution. The values are: 0.877 ± 0.007 and 0.885 ± 0.006 in the low and high Higgs mass regions, respectively. These values are used as normalization correction factors in the signal regions and as systematic uncertainties of the background normalization prediction;

- *Extrapolation from control to signal region*: the factor tacking into account the fact that we are obtaining a correction for the signal region normalization looking at the control region, is obtained from the QCD Monte Carlo simulated events. The correction factor is the ratio of ratios:

$$R_{extrap} = \frac{\left(\frac{N^{events}_{observed}}{N^{events}_{predicted}}\right)_{MC\ \ Signal\ \ Region}}{\left(\frac{N^{events}_{observed}}{N^{events}_{predicted}}\right)_{MC\ \ Control\ \ Region}} \tag{6.1}$$



(a) Low Higgs mass region ($M_\Phi < 200\ GeV/c^2$)

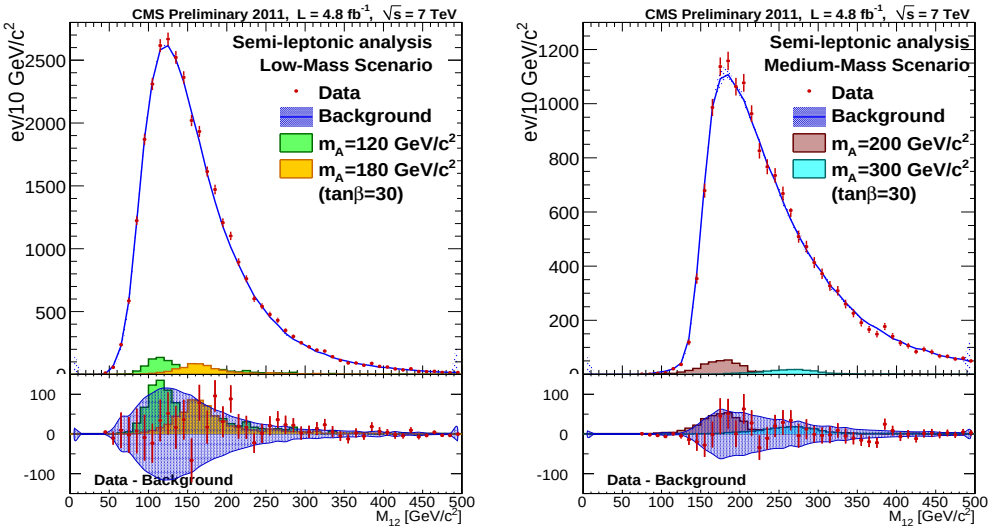
(b) High Higgs mass region ($M_\Phi \geq 200\ GeV/c^2$)

Figure 6.1: Ratio of the observed *bbb* and predicted from *bbj* events (blue dots) as a function of the invariant mass of the two leading jets in the data control region. The fit with a constant function is shown (red line).

obtained comparing the observed ($N^{events}_{observed}$) and predicted events ($N^{events}_{predicted}$) in the signal and control regions. These factors are: 1.01 ± 0.042 and 1.02 ± 0.05 in the low and high Higgs mass regions. They are used as correction factor for the background normalization and as systematic uncertainties in the background normalization prediction.

6.2 Signal extraction and upper limits

The presence of a possible MSSM Higgs boson signal in the triple b-tagged data sample selected through the selection described in section 5.3 is tested by a binned likelihood fit to the data using the background and signal templates. The background shape and normalization are obtained from the methods described in the section 5.5.10, whereas the signal templates are obtained from the Monte Carlo simulation, for different Higgs mass hypothesis. The variable that is exploited for this purpose is the invariant mass of the two leading jets (M_{12}), where a peaked structure is expected to pop up in case of a signal presence. In Figure 6.2 are shown the background predictions for the low and high mass range, the corresponding observed distributions in data, and the expected signal contribution for different Higgs mass hypothesis for $tan\beta = 30$. Since no evidence of signal has been found, upper limits at 95% of confidence level are set on the production cross section of the MSSM Higgs boson in association with b-quarks times the branching ratio of the Higgs decay in two b-quarks. The result is shown in Figure 6.3, where the values of cross section times branching ratio below the expected and observed limits curves are the expected and the actually excluded values, respectively, at 95% of confidence level. The statistical procedure followed to compute these upper limits is known as CL_s method, and is described in Appendix A. These limits can be interpreted as exclusion limits in the $(M_A, tan\beta)$ parameter space of the MSSM model. The values of cross section times branching ratio are converted in the corresponding $tan\beta$ value, for a given M_A, using the



(a) (b)

Figure 6.2: The distributions show the two leading jet invariant mass for the predicted backgrounds in the signal regions (blue line) compared to the observed data (red dots), for low (a) and high (b) Higgs mass regions. The expected signal contribution for different Higgs mass hypothesis is also shown. Below the distributions, the difference between the observed data and predicted background is shown, together with the systematics uncertainties on the background.

tools *bbh@nnlo* [67] and *FeynHiggs* [36, 68–70]. The results are shown in Figure 6.4, in the m_h^{max} benchmark scenario defined in [35].

6.2.1 Model dependent systematics

For the computation of the limits in the $(M_A, tan\beta)$ parameter space of the MSSM model, the following model dependent systematic uncertainties affecting the signal expectation have been taken into account, in addition to the ones listed in section 6.1:

- *Factorization and renormalization of QCD scale*: singularities that appear in perturbative QCD calculation are removed by factorization and renormalization techniques [77]. The theoretical systematic uncertainty on these calculation for the $pp \to b + \Phi$ process is in the range $6 - 28\%$, depending on the Higgs mass;

- *Parton density functions (PDF)*: the *PDF* describe the distributions of the parton momenta in the proton and typically depend on $N \simeq 20$ parameters. The systematic uncertainty related to the *PDF* is estimated by re-weighting event-by-event the Higgs signal Monte Carlo samples using the *PDF* suggested by the *PDF4LHC* working group [78] and varying their parameters by plus and minus the corresponding errors. The uncertainty is then represented by the number of events passing the offline selection in the two cases: $^{+2.5}_{-2.7}\%$ for $M_\Phi = 120\ GeV/c^2$ and $^{+4.7}_{-4.4}\%$ for $M_\Phi = 250\ GeV/c^2$;

- *Underlying event and parton showering*: the modeling of the underlying event, i.e. the particles in a *pp* collision not originated from the hard inter-

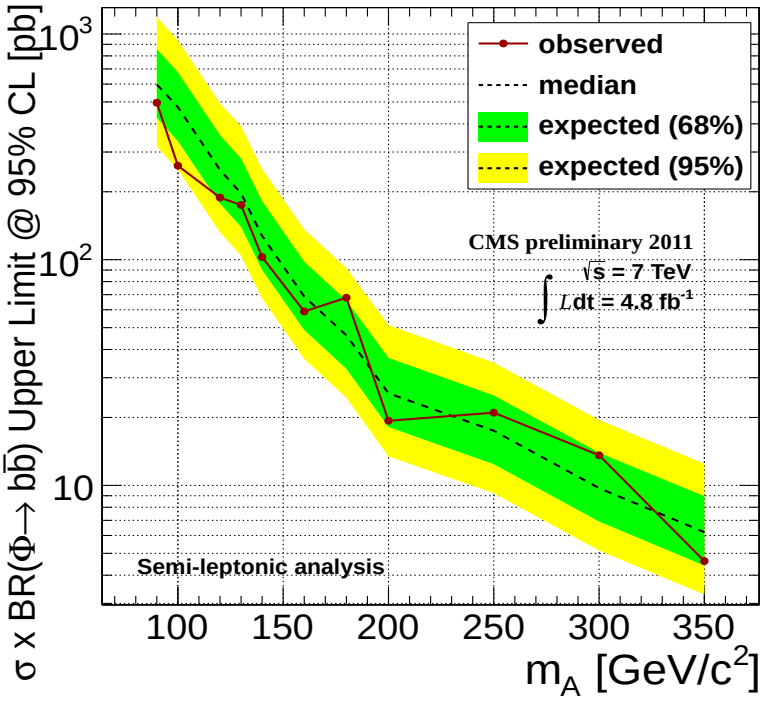


Figure 6.3: The 95% CL upper limits on the signal cross section times the branching ratio, as a function of the pseudo-scalar A mass, in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel, with $\Phi = h, H, A$. These limits correspond to a total amount of data of $\mathscr{L} = 4.8\ fb^{-1}$. The red line corresponds to the observed limit in the data. The dashed line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

action, and of the parton showering process, the radiation of partons from the primary parton produced in the pp hard interaction and underlying event, are affected by a systematic uncertainty estimated to be 4% [78].

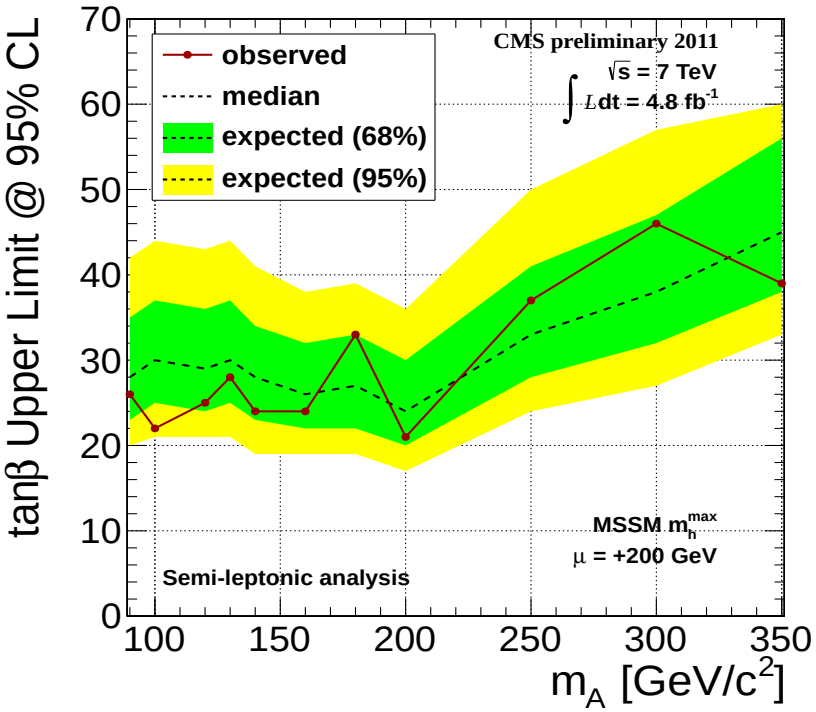


Figure 6.4: The 95% CL upper limits on the $tan\beta$, as a function of the pseudo-scalar A mass, in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel, with $\Phi = h, H, A$. The m_h^{max} is considered as benchmark scenario. These limits correspond to a total amount of data of $\mathscr{L} = 4.8\ fb^{-1}$. The red line corresponds to the observed limit in the data. The dashed line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

Chapter 7

Combination of the result with the all-hadronic analysis

7.1 Introduction

A search for the MSSM Higgs boson in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel has been also performed by the DESY CMS group in Hamburg, following a different analysis strategy with respect to the one presented in this thesis. The analysis is based on a data sample collected during 2011 LHC operations, almost completely statistically independent from the one used in this thesis. In the following section the analysis strategy and results will be briefly discussed, for details see [79]. The analysis, which shows no evidence for a Higgs signal, sets limits on the production cross section times the branching ratio of the considered process, with a sensitivity comparable to the one achieved in the analysis presented here. These limits are combined with the one presented in section 6.2, improving the overall sensitivity of the CMS experiment to the process under search, as described in section 7.3.

7.2 All-hadronic analysis strategy

The analysis presented in this thesis exploits the semi-leptonic decay of the b-quark in muon, as it has been discussed in Chapter 5. For this reason we call it "semi-leptonic analysis". The muon selection allows to reduces the QCD background and helps to keep the HLT trigger rate at reasonable levels, in the range $5 - 10\ Hz$. The strategy adopted by the DESY CMS group is to look for final states containing three high transverse momentum jets identified as originating from b-quark, without requiring the b-quark semi-leptonic decay. We will call this signature "all-hadronic". The first consequence of the all-hadronic strategy is on the design of the HLT trigger used for the online data selection. The muon selection is not applied, and in order to keep the trigger rate around $5 - 10\ Hz$ tighter requirements were applied on the jets transverse energy. Depending on the data-taking period two leading jets with $E_T^{1,2} \geq 46, 38\ GeV$ or $E_T^{1,2} \geq 60, 53\ GeV$ are selected. Part of the data are selected requiring a third jet with $E_T \geq 20\ GeV$. This has to be compared with the thresholds adopted in the trigger paths designed for the semi-leptonic analysis, for which at least a pair of jets with $E_T = 30\ GeV$ is required.

7.2.1 Data samples and offline selection

The data sample selected with the trigger paths with lower jet transverse momentum selection is suitable for searching a low mass Higgs boson, whereas the remaining data sample is suitable for searching at high masses. The analysis was performed separately for the two data samples: 2.7 fb^{-1} of data selected with the lower threshold trigger paths were used for searching in the Higgs boson mass region $m_\Phi < 180\ GeV$ (low-mass), and 4 fb^{-1} of data selected with the higher threshold trigger paths were used for searching in the Higgs boson mass region $m_\Phi \geq 180\ GeV$ (medium-mass). Offline selections are used to further filter the data, requiring three leading jets with $E_T^{1,2,3} = 46, 38, 20\ GeV$ and $E_T^{1,2,3} = 60, 53, 20\ GeV$ in the low and medium mass regions, respectively. These jets have to be identified as originating from b-quark exploiting the CSV algorithm (see section 4.2.4).

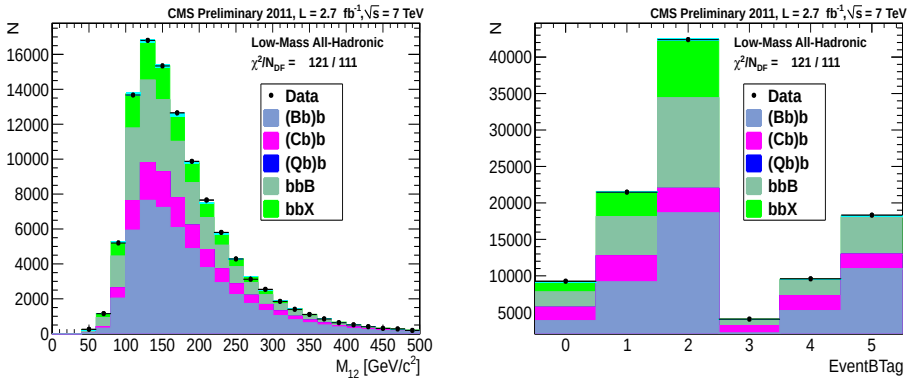
7.2.2 Background estimation

As for the semi-leptonic analysis, also for the all-hadronic the dominant background comes from QCD multijet production, where two or more jets are originated from b-quarks. A data-driven method is developed in order to estimate the background. The data sample used for the background modeling is selected using the offline selections, requiring only two jets to be b-tagged. This sample is divided in three categories: bbx, where the 3^{rd} leading jet is not b-tagged, bxb, where the 2^{nd} leading jet is not b-tagged and xbb, where the leading jet is not b-tagged. The jet which is not b-tagged can originate from light quarks or gluons, c-quark or b-quark. A total of nine background templates, three for each category, are constructed weighting the not b-tagged jet with the probability to be originated from light-quarks or gluon, c-quark or b-quark. These probabilities are determined using QCD Monte Carlo sample. The background templates are two dimensional distributions in the invariant mass of the two leading b-tagged jets and the so called event b-tag variable, built from the secondary vertex mass of the jets in each event. The total number of templates was reduced from nine to five, since some templates are similar each other in their shapes.
Possible distortions of the background model were evaluated and corrections to the background templates were applied. A distortion could arise from events in the three categories where one or both the two b-tagged jets are not originated from b-quark. The contamination, of the order of $3-4\%$, was estimated from the data. Additional correction is due to the b-tagging performed at trigger level with respect to the one performed with the offline selection: in the two b-tagged samples used for the background prediction the two b-jets coincide with the b-tagged jets at trigger level in $90-95\%$ of the events, whereas for the triple b-tagged sample used for the signal search the two b-tagged jets at trigger level could be any pair of the three jets. This produce a bias in the background determination which is estimated from the Monte Carlo simulation.

7.2.3 Results

The Higgs boson signal templates are obtained from Monte Carlo simulation, and the same 11 mass points in the range $m_\Phi = 90 - 350\ GeV$ used in the semi-leptonic analysis are considered. The presence of a possible MSSM Higgs boson is tested performing a fit to the data on the invariant mass distribution of the

two leading b-tagged jets and on the b-tag variable distribution, using a linear combination of the signal and background templates. The normalization of the background, not predicted with the data-driven method, is obtained from the fit. The background only hypothesis is tested performing a fit with only the background templates. The fit shows a compatibility of the background model to the data within the systematic uncertainties. The result of the fit using background templates is shown in Figure 7.1 for the low Higgs boson mass region. Similar distributions are obtained for the medium mass search. Afterwards, a signal template is included in the fit, for each mass point separately. Since no evidence



(a) Background only fit to the invariant mass of the two leading b-tagged jets in the low Higgs mass region ($M_\Phi < 180\ GeV/c^2$)

(b) Background only fit to the event b-tag variable (see text) in the low Higgs mass region ($M_\Phi \geq 180\ GeV/c^2$)

Figure 7.1: Results of the background only fit in data (black dots) using only the background templates (full distributions). The fit is performed in the two dimensional space defined by the invariant mass of the two leading b-tagged jets and the event b-tag variable (see text). The projection of the fit in each variable is shown in the plots. Each background template is indicated in the legend with the corresponding category name, $bbx/bxb/xbb$ (see text), with $x = Q, B, C$ depending on the true flavor assumed for the not b-tagged jet ($Q = light\ quark/gluon$, $B = b - quark$ and $C = c - quark$). The bbX template results from the combination of the bbQ and bbC templates.

for a Higgs boson signal has been found, upper limits at 95% of confidence level are set on the production cross section of the MSSM Higgs boson in association with b-quarks times the branching ratio of the Higgs decay in two b-quarks. The adopted method is the CL_s, as for the semi-leptonic analysis results discussed in section 6.2. These limits were interpreted in the $(M_A, tan\beta)$ parameter space of the MSSM, following the procedure described in section 6.2. The results are shown in Figures 7.2 and 7.3 for the limits on the cross section times the branching ratio and the limits on the $(M_A, tan\beta)$ parameter space, respectively. These limits are similar to those obtained by the semi-leptonic analysis and shown in section 6.2. As an example, for a Higgs boson mass of $m_\Phi = 120\ GeV$ the expected and observed upper limit on the $tan\beta$ parameter are $tan\beta_{expected} = 29$ and $tan\beta_{observed} = 25$ for the semi-leptonic analysis and $tan\beta_{expected} = 30$ and $tan\beta_{observed} = 25$ for the all-hadronic analysis.

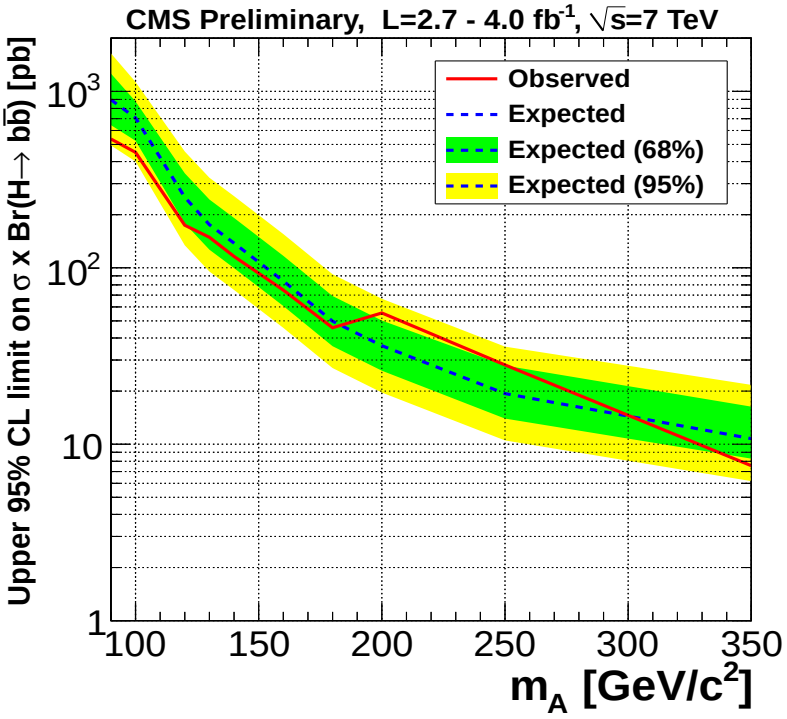


Figure 7.2: The 95% CL upper limits on the signal cross section times the branching ratio, as a function of the pseudo-scalar A mass, in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel ($\Phi = h, H, A$) in the all-hadronic search. These limits correspond to a total amount of data of $\mathscr{L} = 2.7\ fb^{-1}$ in the low Higgs mass region and $\mathscr{L} = 4\ fb^{-1}$ in the medium mass region. The red line corresponds to the observed limit on data. The dashed line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

7.3 Combined upper limits for the semi-leptonic and all-hadronic analysis

More stringent upper limits on the cross section times the branching ratio of the MSSM Higgs boson in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel can be obtained from the combination of the semi-leptonic and all-hadronic analysis. Again, the method used for the calculation is the CL_s. A unique Likelihood function according to eq. A.1 is constructed from the product of the function in eq. A.2 calculated for the semi-leptonic analysis times the one calculated for the all-hadronic analysis. The combination of the information from the two analysis leads to $pdfs$ for the test statistics (eq. A.3) which better discriminate the background only hypothesis from the signal plus background one. This leads to improved limits with respect to the single analysis. To avoid bias due to overcounting of events, the overlap of the data samples between the two analysis was quantified and removed in the all-hadronic analysis from the three b-tagged samples. This overlap resulted small: 2.3% and 2.7% of events in the low and medium Higgs mass regions, respectively. The reason of the smallness of this overlap is the large difference in the triggering and event selection criteria between the two analysis: in particular,

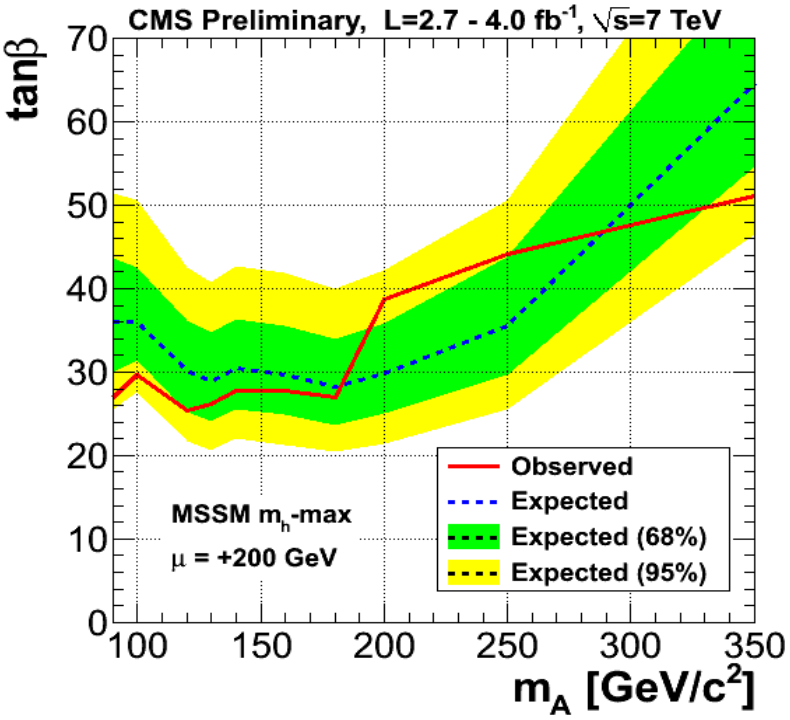


Figure 7.3: The 95% CL upper limits on the $tan\beta$, as a function of the pseudo-scalar A mass, in the $pp \to b+\Phi, \Phi \to b\bar{b}$ channel ($\Phi = h, H, A$) in the all-hadronic search in the m_h^{max} benchmark scenario. These limits correspond to a total amount of data of $\mathscr{L} = 2.7\ fb^{-1}$ in the low Higgs mass region and $\mathscr{L} = 4\ fb^{-1}$ in the medium mass region. The red line corresponds to the observed limit on data. The dashed line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

the requirement of a muon in the semi-leptonic analysis and the hard kinematic selections on jets in the all-hadronic analysis.

7.3.1 Systematic uncertainties

The systematic uncertainties of both the analyses contribute to worsen the sensitivity of the experiment to the searched channel. In Table 7.1 the systematic uncertainties affecting the signal yield for each analysis are summarized. As already was pointed out in section 6.2, some of these systematics are not dependent on the theoretical model describing the Higgs boson production, whereas other systematic uncertainties depends on the considered model. The latter have been taken into account only for the calculation of the upper limit on the $tan\beta$ parameter of the MSSM. The values for the systematic uncertainties listed in Table 7.1 for the semi-leptonic analysis are the same presented in sections 6.1.1 and 6.2.1. The same sources affects also the all-hadronic analysis, except for the muon momentum scale and resolution. For the all-hadronic analysis, two additional systematic sources are considered: the online b-tagging efficiency, which in the semi-leptonic case was determined from the data, and the b-tagging efficiency dependence on topology. The systematics on the online b-tagging efficiency ac-

Table 7.1: Systematic uncertainties affecting the signal yield for the semi-leptonic and all-hadronic analysis, divided in model independent and model dependent (see text) sources. The sources are further classified in systematics affecting only the signal rate and in systematics affecting both the signal shape and rate. Some of the systematics depends on the Higgs boson mass considered for the signal extraction. Their values are given in a range.

Model independent - signal shape and rate affected		
Systematic source	semi-leptonic	all-hadronic
Offline b-tagging efficiency	12%	$10 - 13\%$
Jet energy scale	$^{+2.5}_{-3.1}\%$	$1.4 - 6.8\%$
Jet energy resolution	1.9%	$0.6 - 1.3\%$
Model independent - signal rate affected		
Systematic source	semi-leptonic	all-hadronic
Trigger efficiency	$3 - 5\%$	10%
Online b-tagging efficiency	–	32%
b-tagging efficiency dependence on topology	–	6%
Muon momentum scale and resolution	0.2/0.6%	–
Signal Monte Carlo statistics	$1.1 - 2.6\%$	
Integrated luminosity	2.2%	
Model dependent - signal rate affected		
Systematic source	semi-leptonic	all-hadronic
PDF and α_s uncertainties	$^{+2.5}_{-2.7}$ $-^{+4.7}_{-4.4}$ %	$3 - 6\%$
Factorization and renormalization of QCD scale	$6 - 28\%$	
Underlying event and parton showering	4%	

counts for the uncertainty on the efficiency of b-tagging performed at the trigger level with respect to the data passing the offline b-tagging selection. This contribution is separated from the trigger efficiency, which evaluates only the efficiency to trigger jets with the kinematic characteristics required from the all-hadronic analysis selection. The systematics from the uncertainty on the observed dependence of the offline b-tagging efficiency on the event topologies defined in the all-hadronic analysis must also be considered, as described in detail in [79].
The systematic uncertainties related to the background prediction are included in the combined limit calculation. For the semi-leptonic analysis they are discussed in section 6.1.2. In the all-hadronic analysis the following sources of systematic uncertainties on the background prediction are considered:

- *Offline b-tagging efficiency*: the probabilities to b-tag a jet originating from b, c or light quark are estimated from Monte Carlo simulation. The uncertainties are 10% and 13% for the efficiencies to b-tag a jet originating from b and c quark, in the low and medium Higgs mass regions, respectively. The uncertainties are 10% and 20% for the efficiencies to b-tag jet originating from light quarks (also known as mistag rate), in the low and medium Higgs mass regions, respectively;

- *bb purity corrections*: the bbx, bxb and xbb categories used to estimate the

background contain contamination from fake events where one or both of the two b-tagged jets are originating from c or light quarks. This contamination is estimated and its uncertainties are quantified by their effect on the estimated signal fraction[1]: $0.1 - 0.3\%$ in the low Higgs mass region $m_\Phi = 90 - 130\ GeV$, and below 0.1% for the other Higgs mass points;

- *online b-tag template correction*: a background bias due to the differences in online b-tagging and offline b-tagging in the two b-tagged and the three b-tagged data sample is estimated. The systematic uncertainty on this bias is quantified by its effect on the estimated signal fraction: $0.1 - 0.4\%$ in the low Higgs mass region $m_\Phi = 90 - 130\ GeV$ and below 0.1% for the other Higgs mass points;

All the systematic uncertainties discussed here have been included in the combined limit calculation. All background systematics are specific for each analysis and have been treated as uncorrelated. The common signal systematics have been treated as 100% correlated, whereas specific signal systematics have been treated as uncorrelated.

7.3.2 Results of the combined limits

The result of the combined limits is shown in Figure 7.4, where the values of cross section times branching ratio below the expected and observed limits curves are the expected and the actually excluded values, respectively, at 95% of confidence level. These limits were interpreted as exclusion limits in the $(M_A, tan\beta)$ parameter space of the MSSM model. The values of cross section times branching ratio are converted in the corresponding $tan\beta$ value, for a given M_A, using the tools *bbh@nnlo* and FeynHiggs, as for the results presented in section 6.2. The results are shown in Figure 7.5 for the m_h^{max} benchmark scenario. In Figure 7.5 are also shown the expected limits from each of the two analysis. As can be seen, the combined limits significantly improve the limits from the individual analyses, leading to an exclusion region for $tan\beta$ values as low as ~ 20 for $M_A < 160\ GeV$.

[1]The signal fraction is the fitted signal fraction on the total number of three b-tagged events in data.

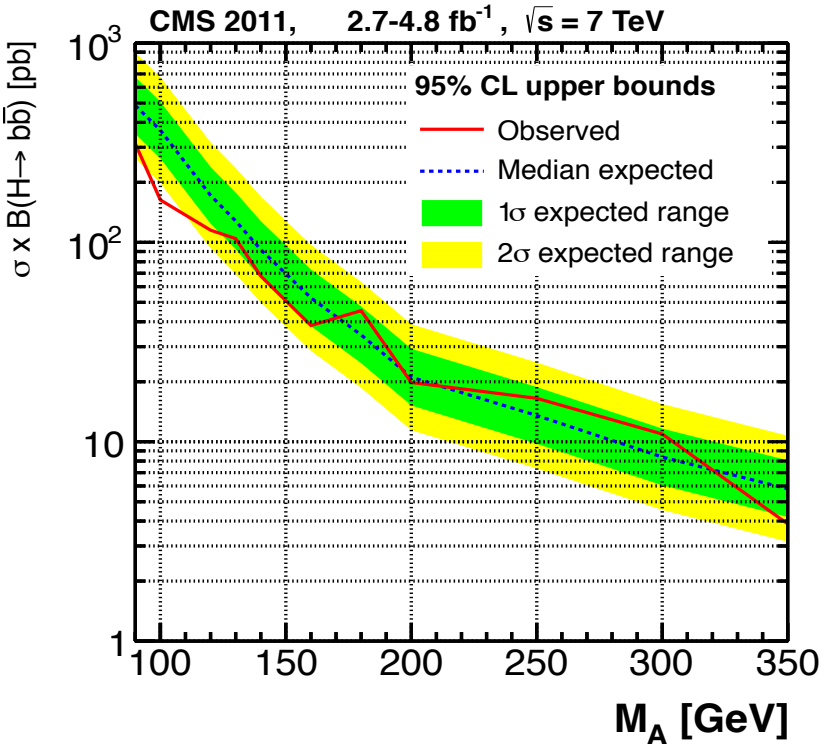


Figure 7.4: The 95% CL upper limits on the signal cross section times the branching ratio, as a function of the pseudo-scalar A mass, in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel ($\Phi = h, H, A$) for the combination of the semi-leptonic and the all-hadronic analyses. The red line corresponds to the observed limit on data. The dashed line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

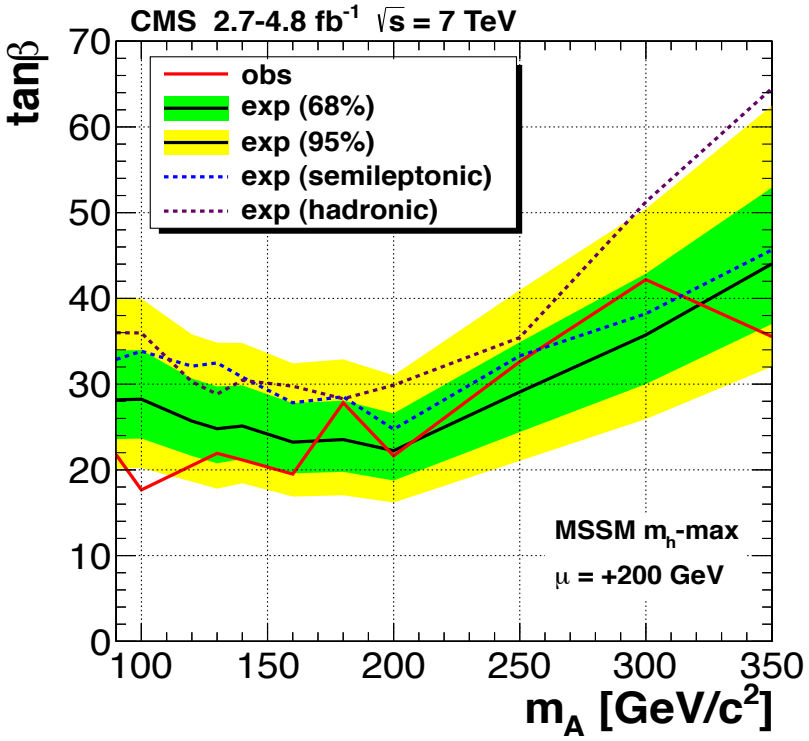


Figure 7.5: The 95% CL upper limits on the $tan\beta$, as a function of the pseudo-scalar A mass, in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel ($\Phi = h, H, A$) for the combination of the semi-leptonic and the all-hadronic analyses. The m_h^{max} is considered as benchmark scenario. The red line corresponds to the observed limit on data. The black line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median. The expected limits from the semi-leptonic analysis (blue dashed line) and from the all-hadronic analysis (magenta dashed line) are also shown.

Summary and conclusions

The first three years of LHC operations lead to remarkable results obtained by the four experiments installed along the pp collider. The discovery of a boson with a mass of $m_h = 125\ GeV$ was presented on July 2012 by the ATLAS and CMS collaborations with a significance of 5.9σ and 5σ, respectively. Although the production cross section of this particle is compatible, within uncertainties, with the Higgs boson predicted by the SM of the elementary particles, its nature has still to be tested in detail. The measurements of its characteristics (i.e. spin, couplings to other SM particles etc) could either confirm that the SM Higgs has been discovered or demonstrate that physics beyond the SM is needed to accommodate the new particle.
Meanwhile, the presence of a Higgs particle beyond the SM can be tested by searches involving production and decay channels predicted by these theories, and forbidden or suppressed in the SM. This thesis has reported the analysis performed for the search of Higgs boson in the $pp \to b + \Phi, \Phi \to b\bar{b}$ as predicted by the MSSM, a good candidate theory extending the SM. The production and decay of the Higgs boson through this channel is enhanced in MSSM for a large part of the parameter space of the model. This is in contrast to what happens for the SM Higgs, where the Yukawa coupling to the bottom quark is tiny. As an example, for the Higgs mass $m_\Phi = 120\ GeV$ the production cross section times the branching ratio in MSSM for the $pp \to b + \Phi, \Phi \to b\bar{b}$ process is 246 pb for $tan\beta = 30$, whereas in the SM is of the order of 1 pb.
The data used for the analysis, corresponding to a total integrated luminosity of 4.8 fb^{-1}, were taken during the 2011 LHC operations with HLT trigger designed for this specific signal search. The trigger path has been updated during data-taking with tighter requirements on some physics objects used for the data selection. This was done to cope with the increasing LHC luminosity during the 2011 run, from $\sim 2 \cdot 10^{32} cm^{-2} s^{-1}$ to $\sim 6 \cdot 10^{33} cm^{-2} s^{-1}$.
The signal search has been performed looking for final states with three energetic jets in the tracker region of the CMS detector, identified as originating from b-quarks. The major background contribution is given by multijets events generated by QCD processes. The semi-leptonic decay of the b-quark in muon has been exploited to reduce the QCD background and to keep the trigger rate at acceptable values of $\sim 5 - 10\ Hz$. Two background templates were obtained from data-driven methods. A Monte Carlo simulated QCD sample was used for studies and checks of the method robustness. Other Monte Carlo samples are used to estimate the contribution of electroweak background processes. The two background templates were combined to obtain a unique template with reduced systematic uncertainties on the background shape. This was possible since the two data-driven methods are independent. The background normalization is taken from one of the two methods, with a systematic uncertainty of the order

of 5% for the entire search mass region.
The presence of a Higgs boson signal in data was tested performing a binned likelihood fit to the invariant mass of the two leading b-jets. The data-driven background and signal from Monte Carlo simulation were used as templates for the fit. Since no evidence of Higgs boson signal was found, upper limits on the production cross section times the branching ratio of the considered process in the mass range $m_\Phi = [90, 350]\ GeV$ were computed. In particular, for a Higgs boson mass $m_\Phi = 120\ GeV$, the observed limit was $\sigma(pp \to b + \Phi) \times BR(\Phi \to b\bar{b})_{observed} = 188\ pb$, to be compared with the expected limit $\sigma(pp \to b+\Phi) \times BR(\Phi \to b\bar{b})_{expected} = 250\ pb$. These results were converted in exclusion plot in the $(tan\beta, M_A)$ parameter space of the MSSM, in the specific m_h^{max} benchmark scenario. For pseudo-scalar Higgs boson mass $M_A = 120\ GeV$ the observed limit on $tan\beta$ was $tan\beta_{observed} = 25$, to be compared with the expected limit $tan\beta_{expected} = 29$, considerably improving the previous limits obtained by the CDF [80] and D0 [81] collaborations.
The results obtained with the analysis presented in this thesis were combined with those obtained by the DESY CMS group. They searched for the Higgs boson in the $pp \to b + \Phi, \Phi \to b\bar{b}$ channel following a different strategy. Since no evidence for the presence of a signal was found in data, upper limits on the production cross section times the branching ratio of the considered process were calculated. From these limits an exclusion plot in the $(tan\beta, M_A)$ parameter space of the MSSM, in the specific m_h^{max} benchmark scenario, was obtained. The sensitivity of the two analysis is similar: as an example, for $m_\Phi = 120\ GeV$, the observed combined limit on $tan\beta$ was $tan\beta_{observed} = 25$, to be compared to the expected limit $tan\beta_{expected} = 30$. The combined upper limits significantly improve the sensitivity with respect to the single analyses: for the Higgs boson mass $m_\Phi = 120\ GeV$ the observed combined limit on $tan\beta$ was $tan\beta_{observed} = 20$, to be compared to the expected combined limit $tan\beta_{expected} = 26$.

Appendix A

The CL_s method

Two statistical approaches commonly used in high energy physics for characterizing the absence of a signal are the Bayesian and the classical frequentist approach. Both of them allow to quantify the level of incompatibility of data with a signal hypothesis, which is expressed as a confidence level (CL) on the strength of the signal or parameters of the model predicting the signal. A convention is to require 95% CL for excluding the signal presence. Here will focus on the frequentist approach, which is used by the Higgs groups at LHC [82].
In the following, the expected signal yield will be referred to as $s(\theta)$ and the background one as $b(\theta)$. They are function of a vector of *nuisance* parameters, $\theta = (\theta_1, ..., \theta_n)$, which take into account the systematic uncertainties affecting the signal and background. A *signal strength modifier* μ is defined in our case by $\sigma = \mu \cdot \sigma_{MSSM}$, where σ_{MSSM} is the signal cross section for given values of the MSSM parameters. It is a factor scaling the MSSM signal yield. The following likelihood function is constructed:

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

where *data* stands either for the actual experimental observed data or for pseudo-data obtained from toy Monte Carlo simulated experiments. The first factor on the right is a product of Poisson probabilities to observe n_i events in i-bin:

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

The second factor, the probability density function $p(\tilde{\theta}|\theta)$, is the probability of measuring a set of nuisance parameters $\tilde{\theta}$, given its true value θ.
The compatibility of the *data* with the background only and background plus signal hypothesis, are tested by constructing a test statistics $\tilde{q}_\mu$ from the profile likelihood ratio:

$$\tilde{q}_\mu = -2ln\frac{\mathscr{L}(data|\mu, \hat{\theta}_\mu)}{\mathscr{L}(data|\hat{\mu}, \hat{\theta})}, \qquad 0 \leq \hat{\mu} \leq \mu \tag{A.3}$$

where in the numerator $\hat{\theta}_\mu$ is the value of θ maximizing the likelihood for that given value of μ, and the denominator $\hat{\mu}$ and $\hat{\theta}$ are the values that maximize the likelihood. Since the signal rate is positive, this implies the lower constraint

$0 \leq \hat{\mu}$, while since upward fluctuation of the data such that $\hat{\mu} > \mu$ are not considered as evidence against the signal hypothesis, then the condition $\hat{\mu} \leq \mu$ is required.
The observed value of the test statistics on the data, $\tilde{q}_\mu^{obs}$, is determined for the given signal strength modifier μ under test. Also, the values of the nuisance parameters $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$ best describing the observed data, for the background only and background plus signal hypothesis, respectively, are found. Now, if the probability density functions for the test statistics are known for the background only, $f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})$, and for the background plus signal, $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$, hypothesis, the following two probabilities can be defined:

- probability to observe a value of the test statistics equal or greater than the observed one in the signal plus background hypothesis:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|signal + background) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \qquad (A.4)$$

- probability to observe a value of the test statistics equal or greater than the observed one in the background only hypothesis:

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|background) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu \qquad (A.5)$$

This probabilities are also known as p-values. The observed confidence level limit for that signal strength μ under test is then calculated as the ratio of these two probabilities:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} = \alpha \qquad (A.6)$$

that means that for the tested signal strength the Higgs boson is excluded at $(1 - \alpha)$ CL_s confidence level. The 95% confidence level upper limit on the signal strength is obtained by finding the value of μ for which $CL_s = 0.05$. This value of μ is then excluded at 95% confidence level.
The only missing pieces are at this point the *pdfs* $f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})$ and $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$. They can be generated by using toy Monte Carlo pseudo data, assuming a signal with strength μ in the signal plus background hypothesis and null signal ($\mu = 0$) in the case of the background only hypothesis. For each pseudo data sample a value of the test statistics is found in the two hypothesis. Generating a large amount of pseudo data samples, the distributions of the possible values of the test statistics, in the background only and background plus signal background hypothesis, are found. These distributions are the *pdfs* of the test statistic in the two hypothesis. An example of these *pdfs* for the background only and signal plus background hypothesis is shown in Figure A.1 for a Higgs mass of $m_\Phi = 120\ GeV$.
One would also know the upper limits in the case background only hypothesis, which is known as the expected median upper limit. The simplest way is to generate a large set of background only pseudo data samples and calculate the upper limits μ at 95% confidence level for each of them, as if they were real data. A distribution of possible upper limits on μ is obtained (Figure A.2 left). Then, the cumulative probability distribution of results can be built (Figure A.2 right),

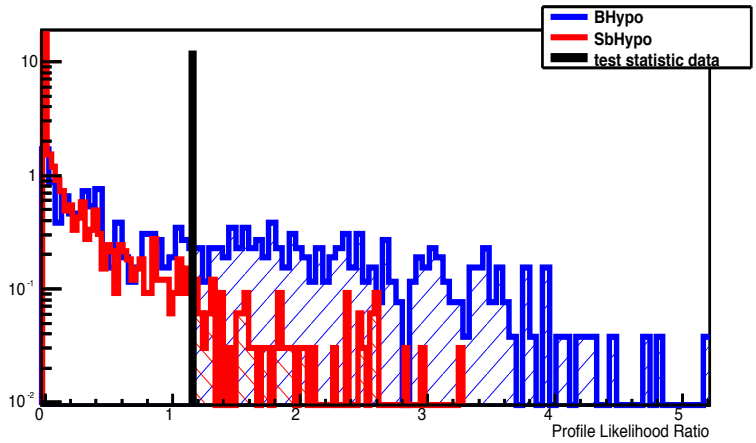


Figure A.1: Test statistic distributions for samples of pseudo data generated for signal plus background (red line) and background only (blue line) hypothesis for a Higgs signal of mass $m_\Phi = 120\ GeV$. The observed value is highlighted by the black vertical line.

and the point at which this distribution crosses the quantile of 50% is the median expected value. The $\pm 1\sigma$ (68%) band is defined by the crossings of the 16% and 84% quantiles, while the $\pm 2\sigma$ (95%) band is defined by the crossings of the 2.5% and 97.5% quantiles. These bands accounts for the possible statistical fluctuation in the data. The production of the Monte Carlo pseudo data samples used to

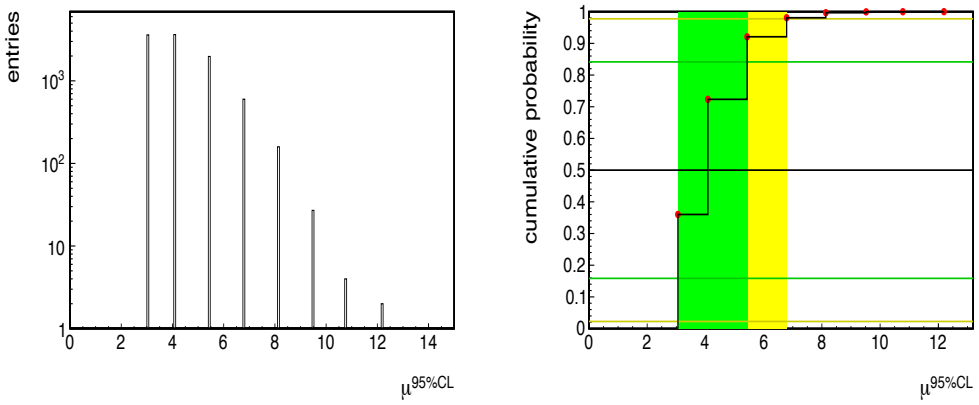


Figure A.2: On the left, an example of a possible distribution of limits on μ obtained in samples of background only pseudo data is shown. The right distribution, is the cumulative probability distribution of the left distribution. The 2.5%, 16%, 50%, 84%, and 97.5% quantiles (horizontal lines), defining the $\pm 1\sigma$ and $\pm 2\sigma$ bands are also shown.

build the test statistics *pdfs* and to compute the expected median upper limits and corresponding $\pm 1 - 2\sigma$ bands is demanding from the computational point of view. Though this approach is the only one possible when the number of events in the considered analysis is low, in the case of high statistics (asymptotic limit) mathematical expressions for the test statistics *pdfs* and for the computation of the expected median upper limits with corresponding error bars can be derived [83].

Bibliography

[1] Noether E. "Der Endlichkeitsatz der Invarianten endlicher linearer Gruppen der Charakteristik p". In: *Nachr. Ges. Wiss. Göttingen* (1926), pp. 28–35.

[2] D. V. Schroeder M. E. Peskin. *An Introduction To Quantum Field Theory*. Frontiers in Physics. Westview Press, 1995.

[3] Particle Data Group Collaboration. "The Review of Particle Physics". In: *Phys. Rev.* D.86 (2012), p. 1.

[4] R. Barate et al. "Search for the standard model Higgs boson at LEP". In: *Phys. Lett.* B565 (2003), pp. 61–75. DOI: 10.1016/S0370-2693(03)00614-2. eprint: hep-ex/0306033.

[5] T. Aaltonen et al. "Combination of Tevatron searches for the standard model Higgs boson in the W+W- decay mode". In: *Phys. Rev. Lett.* 104 (2010), p. 061802. DOI: 10.1103/PhysRevLett.104.061802. eprint: 1001.4162.

[6] D0 Collaboration DELPHI Collaboration L3 Collaboration OPAL Collaboration SLD Collaboration LEP Electroweak Working Group Tevatron Electroweak Working Group SLD Electroweak ALEPH Collaboration CDF Collaboration and Heavy Flavour Groups. "Precision Electroweak Measurements and Constraints on the Standard Model". In: *CERN-PH-EP-2010-095, FERMILAB-TM-2480-PPD, LEPEWWG-2010-01, TEVEWWG-2010-01, ALEPH-2010-001-PHYSICS-2010-001, CDF-NOTE-10338, D0-NOTE-6130, DELPHI-DELPHI-2010-001-PHYS-952, L3-NOTE-2837, OPAL-PR432, SLAC-PUB-14301* (2010). eprint: 1012.2367.

[7] S. Schael et al. "Precision electroweak measurements on the Z resonance". In: *Phys. Rept.* 427 (2006), pp. 257–454. DOI: 10.1016/j.physrep.2005.12.006. eprint: hep-ex/0509008.

[8] CMS Collaboration. "Combination of SM, SM4, FP Higgs boson searches". In: *CMS-PAS-HIG-12-008* (2012).

[9] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. eprint: 1207.7235.

[10] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. eprint: 1207.7214.

[11] Steven Weinberg. "Implications of dynamical symmetry breaking". In: *Phys. Rev. D* 13 (4 1976), pp. 974–996. DOI: 10.1103/PhysRevD.13.974. URL: http://link.aps.org/doi/10.1103/PhysRevD.13.974.

[12] H. Georgi, Helen R. Quinn, and Steven Weinberg. "Hierarchy of Interactions in Unified Gauge Theories". In: *Phys. Rev. Lett.* 33 (1974), pp. 451–454. DOI: 10.1103/PhysRevLett.33.451.

[13] Andrew G. Cohen, A. De Rujula, and S.L. Glashow. "A Matter - antimatter universe?" In: *Astrophys. J.* 495 (1998), pp. 539–549. DOI: 10.1086/305328. eprint: astro-ph/9707087.

[14] S. Sarkar B.D. Fields. "Big-Bang Nucleosynthesis". In: *Phys. Rev.* D.86 (2012), pp. 275–279.

[15] D. Dodelson. *Modern Cosmology*. Academic Press, 2003.

[16] A.D. Sakharov. "Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[17] M.E. Shaposhnikov. "Structure of the High Temperature Gauge Ground State and Electroweak Production of the Baryon Asymmetry". In: *Nucl. Phys.* B299 (1988), p. 797. DOI: 10.1016/0550-3213(88)90373-2.

[18] M.B. Gavela et al. "Standard model CP violation and baryon asymmetry. Part 1: Zero temperature". In: *Nucl. Phys.* B430 (1994), pp. 345–381. DOI: 10.1016/0550-3213(94)00409-9. eprint: hep-ph/9406288.

[19] M.B. Gavela et al. "Standard model CP violation and baryon asymmetry. Part 2: Finite temperature". In: *Nucl. Phys.* B430 (1994), pp. 382–426. DOI: 10.1016/0550-3213(94)00410-2. eprint: hep-ph/9406289.

[20] K. Kajantie et al. "Is there a hot electroweak phase transition at m(H) larger or equal to m(W)?" In: *Phys. Rev. Lett.* 77 (1996), pp. 2887–2890. DOI: 10.1103/PhysRevLett.77.2887. eprint: hep-ph/9605288.

[21] J. Bagger J. Wess. *Supersymmetry and Supergravity: Revised and Expanded Edition*. Princeton University Press, 1992.

[22] Rudolf Haag, Jan T. Lopuszanski, and Martin Sohnius. "All Possible Generators of Supersymmetries of the s Matrix". In: *Nucl. Phys.* B88 (1975), p. 257. DOI: 10.1016/0550-3213(75)90279-5.

[23] Sidney R. Coleman and J. Mandula. "ALL POSSIBLE SYMMETRIES OF THE S MATRIX". In: *Phys. Rev.* 159 (1967), pp. 1251–1256. DOI: 10.1103/PhysRev.159.1251.

[24] Patrick Huet and Ann E. Nelson. "Electroweak baryogenesis in supersymmetric models". In: *Phys. Rev.* D53 (1996), pp. 4578–4597. DOI: 10.1103/PhysRevD.53.4578. eprint: hep-ph/9506477.

[25] D. Comelli, M. Pietroni, and A. Riotto. "Spontaneous CP violation and baryogenesis in the minimal supersymmetric Standard Model". In: *Nucl. Phys.* B412 (1994), pp. 441–458. DOI: 10.1016/0550-3213(94)90511-8. eprint: hep-ph/9304267.

[26] Marcela S. Carena, M. Quiros, and C.E.M. Wagner. "Opening the window for electroweak baryogenesis". In: *Phys. Lett.* B380 (1996), pp. 81–91. DOI: 10.1016/0370-2693(96)00475-3. eprint: hep-ph/9603420.

[27] M. Carena et al. “The Baryogenesis Window in the MSSM”. In: *Nucl. Phys.* B812 (2009), pp. 243–263. DOI: 10.1016/j.nuclphysb.2008.12.014. eprint: 0809.3760.

[28] L. Girardello and Marcus T. Grisaru. “Soft Breaking of Supersymmetry”. In: *Nucl. Phys.* B194 (1982), p. 65. DOI: 10.1016/0550-3213(82)90512-0.

[29] Savas Dimopoulos and David W. Sutter. “The Supersymmetric flavor problem”. In: *Nucl. Phys.* B452 (1995), pp. 496–512. DOI: 10.1016/0550-3213(95)00421-N. eprint: hep-ph/9504415.

[30] G. L. Kane et al. “Study of constrained minimal supersymmetry”. In: *Phys. Rev. D* 49 (11 1994), pp. 6173–6210. DOI: 10.1103/PhysRevD.49.6173. URL: http://link.aps.org/doi/10.1103/PhysRevD.49.6173.

[31] Yasuhiro Okada, Masahiro Yamaguchi, and Tsutomu Yanagida. “Upper Bound of the Lightest Higgs Boson Mass in the Minimal Supersymmetric Standard Model”. In: *Progress of Theoretical Physics* 85.1 (1991), pp. 1–5. DOI: 10.1143/PTP.85.1. URL: http://ptp.ipap.jp/link?PTP/85/1/.

[32] R. Barbieri, M. Frigeni, and F. Caravaglios. “The supersymmetric Higgs for heavy superpartners”. In: *Physics Letters B* 258.12 (1991), pp. 167 – 170. ISSN: 0370-2693. DOI: 10.1016/0370-2693(91)91226-L. URL: http://www.sciencedirect.com/science/article/pii/037026939191226L.

[33] Howard E. Haber and Ralf Hempfling. “Can the mass of the lightest Higgs boson of the minimal supersymmetric model be larger than m_Z?” In: *Phys. Rev. Lett.* 66 (14 1991), pp. 1815–1818. DOI: 10.1103/PhysRevLett.66.1815. URL: http://link.aps.org/doi/10.1103/PhysRevLett.66.1815.

[34] John Ellis, Giovanni Ridolfi, and Fabio Zwirner. “Radiative corrections to the masses of supersymmetric Higgs bosons”. In: *Physics Letters B* 257.12 (1991), pp. 83 –91. ISSN: 0370-2693. DOI: 10.1016/0370-2693(91)90863-L. URL: http://www.sciencedirect.com/science/article/pii/037026939190863L.

[35] Marcela S. Carena et al. “Suggestions for benchmark scenarios for MSSM Higgs boson searches at hadron colliders”. In: *Eur. Phys. J.* C26 (2003), pp. 601–607. DOI: 10.1140/epjc/s2002-01084-3. eprint: hep-ph/0202167.

[36] G. Degrassi et al. “Towards high precision predictions for the MSSM Higgs sector”. In: *Eur. Phys. J.* C28 (2003), pp. 133–143. DOI: 10.1140/epjc/s2003-01152-2. eprint: hep-ph/0212020.

[37] (ed.) Evans Lyndon and (ed.) Bryant Philip. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[38] “Design study of the large hadron collider (LHC): A Multiparticle collider in the LEP tunnel”. In: *CERN-91-03, CERN-YELLOW-91-03* (1991).

[39] LHC Study Group. “The Large Hadron Collider: Conceptual design”. In: *CERN-AC-95-05-LHC, CM-P00047618* (1995).

[40] ALICE Collaboration. “ALICE: Technical proposal for a large ion collider experiment at the CERN LHC”. In: *CERN-LHCC-95-71, CERN-LHCC-P-3* (1995).

[41] ATLAS Collaboration. "ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN". In: *CERN-LHCC-94-43* (1994).

[42] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[43] LHCb Collaboration. "Technical Proposal". In: *CERN-LHCC-98-4* (1998).

[44] TOTEM Collaboration. "The TOTEM experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08007. DOI: 10.1088/1748-0221/3/08/S08007.

[45] LHCf Collaboration. "The LHCf detector at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08006. DOI: 10.1088/1748-0221/3/08/S08006.

[46] CMS Collaboration. "Measurement of Momentum Scale and Resolution of the CMS Detector using Low-mass Resonances and Cosmic Ray Muons". In: *CMS-PAS-TRK-10-004* (2010).

[47] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997.

[48] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. The following files are from ¡a href=. Geneva: CERN, 1997.

[49] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. Geneva: CERN, 2006.

[50] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997.

[51] CMS Collaboration. "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. eprint: 1206.4071.

[52] M. Aguilar-Benitez et al. "Construction and test of the final CMS Barrel Drift Tube Muon Chamber prototype". In: *Nucl. Instrum. Meth.* A480 (2002), pp. 658–669. DOI: 10.1016/S0168-9002(01)01227-X.

[53] M. Andlinger et al. "Bunch crossing identification at LHC using a mean timer technique". In: *Nucl. Instrum. Meth.* A336 (1993), pp. 91–97. DOI: 10.1016/0168-9002(93)91082-X.

[54] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: TheTrigger Systems*. Technical Design Report CMS.

[55] Sergio Cittolin, Attila Rcz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002.

[56] CMS Collaboration. "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". In: *CMS-PAS-PFT-09-001* (2009).

[57] CMS Collaboration. "Track reconstruction in the CMS Tracker". In: *CMS-PAS-TRK-09-001* (2010).

[58] CMS Collaboration. "Tracking and Vertexing Results from First Collisions". In: *CMS-PAS-TRK-10-001* (2010).

[59] CMS Collaboration. "Performance of Jet Algorithms in CMS". In: *CMS-PAS-JME-07-003* (2007).

[60] CMS Collaboration. "Jet Performance in pp Collisions at 7 TeV". In: *CMS-PAS-JME-10-003* (2010).

[61] CMS Collaboration. "b-Jet Identification in the CMS Experiment". In: *CMS-PAS-BTV-11-004* (2012).

[62] CMS Collaboration. "Identification of b-quark jets with the CMS experiment". In: *CMS-BTV-12-001, CERN-PH-EP-2012-262* (2012). eprint: 1211.4462.

[63] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual". In: *JHEP* 0605 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. eprint: hep-ph/0603175.

[64] Johan Alwall et al. "MadGraph 5 : Going Beyond". In: *JHEP* 1106 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. eprint: 1106.0522.

[65] G. Corcella et al. "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)". In: *JHEP* 0101 (2001), p. 010. eprint: hep-ph/0011363.

[66] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[67] Robert V. Harlander and William B. Kilgore. "Higgs boson production in bottom quark fusion at next-to-next-to-leading order". In: *Phys. Rev. D* 68 (1 2003), p. 013001. DOI: 10.1103/PhysRevD.68.013001. URL: http://link.aps.org/doi/10.1103/PhysRevD.68.013001.

[68] S. Heinemeyer, W. Hollik, and G. Weiglein. "FeynHiggs: a program for the calculation of the masses of the neutral CP-even Higgs bosons in the MSSM". In: *Computer Physics Communications* 124.1 (2000), pp. 76 – 89. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(99)00364-1. URL: http://www.sciencedirect.com/science/article/pii/S0010465599003641.

[69] S Heinemeyer, Wolfgang F L Hollik, and Georg Weiglein. "The Masses of the Neutral CP-even Higgs Bosons in the MSSM: Accurate Analysis at the Two-Loop Level". In: *Eur. Phys. J. C* 9.hep-ph/9812472. KA-TP-98-17. DESY-98-194. CERN-TH-98-405 (1998), 343–366. 45 p.

[70] M. Frank et al. "The Higgs Boson Masses and Mixings of the Complex MSSM in the Feynman-Diagrammatic Approach". In: *JHEP* 0702 (2007), p. 047. DOI: 10.1088/1126-6708/2007/02/047. eprint: hep-ph/0611326.

[71] Z. Was. "TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report". In: *Nucl. Phys. Proc. Suppl.* 98 (2001), pp. 96–102. DOI: 10.1016/S0920-5632(01)01200-2. eprint: hep-ph/0011305.

[72] CMS Collaboration. "Measurement of the $t\bar{t}$ Production Cross Section in pp Collisions at 7 TeV in Lepton + Jets Events Using b-quark Jet Identification". In: *Phys. Rev.* D84 (2011), p. 092004. DOI: 10.1103/PhysRevD.84.092004. eprint: 1108.3773.

[73] CMS Collaboration. "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 1110 (2011), p. 132. DOI: 10.1007/JHEP10(2011)132. eprint: 1107.4789.

[74] CMS Collaboration. "Measurement of the WW, WZ and ZZ cross sections at CMS". In: *CMS-PAS-EWK-11-010* (2011).

[75] CMS Collaboration. "Performance of the b-jet identification in CMS". In: *CMS-PAS-BTV-11-001* (2011).

[76] CMS Collaboration. "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update". In: *CMS-PAS-SMP-12-008* (2012).

[77] B. R. Webber R. K. Ellis W. J. Stirling. *QCD and Collider Physics*. Cambridge University Press, 2003.

[78] Michiel Botje et al. "The PDF4LHC Working Group Interim Recommendations". In: (2011). eprint: 1101.0538.

[79] CMS Collaboration. "MSSM Higgs production in association with b quarks - all hadronic". In: *CMS-PAS-HIG-12-026* (2012).

[80] T. Aaltonen et al. "Search for Higgs Bosons Produced in Association with b-quarks". In: *Phys. Rev.* D85 (2012), p. 032005. DOI: 10.1103/PhysRevD.85.032005. eprint: 1106.4782.

[81] Victor Mukhamedovich Abazov et al. "Search for neutral Higgs bosons in the multi-b-jet topology in 5.2fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Lett.* B698 (2011), pp. 97–104. DOI: 10.1016/j.physletb.2011.02.062. eprint: 1011.1931.

[82] A L Read. "Modified frequentist analysis of search results (the CL_s method)". In: CERN-OPEN-2000-205 (2000).

[83] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J.* C71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. eprint: 1007.1727.