Université catholique de Louvain

Faculté des sciences

École de physique

Improved analysis method for the search of $H \to Z(l\bar{l})A(b\bar{b})$ with the CMS detector

Auteur:
Auriane Canesse

Promoteur:
Prof. Christophe Delaere

Lecteurs:
Prof. Vincent Lemaitre
Prof. Giacomo Bruno

Mémoire présenté en vue de l'obtention du grade académique de Master en sciences physiques, finalité approfondie.

Année académique 2016-2017

Acknowledgements

This master thesis would not have been possible without the support of many people.

I would first like to thank my supervisor, Prof. Christophe Delaere, whose guidance helped me a lot in my research and in the writing of this thesis.

Many thanks to Alexandre Mertens whose help was crucial in the implementation and comprehension of this project.

My sincere thanks also goes to Jérôme de Favereau, Pavel Demin, Sébastien Brochet, Sébastien Wertz for their support as well as to the entire llbb group for their advice and comments on my results.

Also thanks to my readers, Prof. Vincent Lemaitre and Prof. Giacomo Bruno, who offered interesting advice at my pre-defence.

Last but not least I would like to thank all my friends for their emotional support and encouragement throughout my years of study.

Contents

Introduction

In 2012 the Higgs boson was discovered by the ATLAS and CMS experiments [1, 2]. This particle was the last one predicted by the Standard Model (SM) and its discovery made the theory self-consistent. Moreover, recent studies [3, 4] have shown that the new boson's spin, parity and branching ratios are consistent with the SM Higgs boson.

Unfortunately the Standard Model – although yielding accurate descriptions of collider experiments – doesn't account for observations such as dark matter, neutrino masses or baryonic asymmetry. It also suffers from theoretical hierarchy problems an doesn't include gravity. Therefore the SM is believed to be incomplete and numerous theoretical extensions have been suggested over the years to address its shortfalls. Many of these theories require an extended scalar sector.

The simplest extension of the SM scalar sector consists in adding a second Higgs doublet, thus obtaining a Two Higgs Doublet Model (2HDM). This model predicts five physical Higgs bosons: two neutral CP-even scalar (h, H), one neutral CP-odd pseudo-scalar (A) and two charged scalars ($H^{\pm}$). 2HDMs taken in the so-called alignment limit are especially relevant since the lighter scalar h may be identified to the boson observed at the LHC.

This thesis is the search for a signature of such a 2HDM. The studied decay is $H \to ZA \to l\bar{l}b\bar{b}$. A similar study has been carried out by CMS in 2014 [5] with run 1 data and this thesis consists in the study of the potential of a new analysis method.

This thesis is organised as follows: the first chapter will develop the theoretical aspects of 2HDMs, their phenomenology at the LHC and the motivation of the study. Then the second chapter is dedicated to the CMS detector and the particle reconstruction. Finally the third chapter presents the analysis method and the results of the search.

Chapter 1

Theory

This chapter describes the theory and phenomenology of a two Higgs Doublet Model (2HDM). The reader is assumed to be familiar with the Standard Model (SM) of particle physics and especially of it's scalar sector. A detailed description of the SM may be found in many books [6–8].

Section 1.1 begins by describing the theoretical aspects of 2HDMs – imposed symmetries, new particles, modification of the couplings to fermions & gauge bosons. Then section 1.2 discusses the Higgs and 2HDM's phenomenology at the LHC. It also explains the choice of decay channel used for this thesis. Finally 1.3 gives a reminder of the SM main shortfalls motivating 2HDMs studies .

1.1 Theory of the 2HDM

1.1.1 The new potential

Standard Model reminder

In the Standard Model (SM) the Higgs doublet ϕ must satisfy the Lagrangian:

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\mid \phi \mid^2) \tag{1.1}$$

The potential V must satisfy $SU(2)$ and $U(1)$ symmetries and therefore has the structure shown below in eq. 1.2. In order to have spontaneous symmetry breaking mechanism, the parameters μ^2 and λ are positive.

$$V(\mid \phi \mid^2) = -\frac{\mu^2}{2} \mid \phi \mid^2 + \frac{\lambda}{4} \mid \phi \mid^4 \tag{1.2}$$

The minimum of the potential is reached for $v^2 = \frac{\mu^2}{\lambda}$ where v is called the vacuum expectation value (vev) of the potential. In SM the vev has been measured and is $v = 246\text{GeV}$.

2HDM general potential

In the case of a 2 Higgs Doublet Model, the potential V has a more complex structure. There are now two complex scalar fields charged under $SU(2)_L \times U(1)_Y$: ϕ_1 and ϕ_2. The most general potential V satisfying $SU(2)$ and $U(1)$ symmetry is then [9]:

$$\begin{aligned} V(\phi_1, \phi_2) &= m_1^2 \phi_1^\dagger \phi_1 + m_2^2 \phi_2^\dagger \phi_2 - (m_{12}^2 \phi_1^\dagger \phi_2 + h.c.) \\ &+ \frac{1}{2}\lambda_1 (\phi_1^\dagger \phi_1)^2 + \frac{1}{2}\lambda_2 (\phi_2^\dagger \phi_2)^2 + \lambda_3 (\phi_1^\dagger \phi_1)(\phi_2^\dagger \phi_2) \\ &+ \lambda_4 (\phi_1^\dagger \phi_2)(\phi_2^\dagger \phi_1) + (\frac{1}{2}\lambda_5 (\phi_1^\dagger \phi_2)^2 + \lambda_6 (\phi_1^\dagger \phi_1)(\phi_1^\dagger \phi_2) \\ &+ \lambda_7 (\phi_1^\dagger \phi_2)(\phi_2^\dagger \phi_2) + h.c.) \end{aligned} \tag{1.3}$$

Where $m_{12}^2,\ \lambda_5,\ \lambda_6,\ \lambda_7$ might be complex. From 2 free parameters in the SM case ($\lambda,\ \mu$), we have increased to 14 free parameters. A first step in any 2HDM will be to reduce the number of these parameters by imposing additional symmetries.

1.1.2 Additional symmetries

CP conservation and $\mathbb{Z}_2$ symmetry

Since no important CP violation is observed, it is rational to want a CP conserving potential for the scalar sector. A first step for achieving CP conservation is to require all the potential's (1.3) parameters to be real. This decreases the number of free parameters from 14 to 10. The mass eigenstates of the potential are then the CP eigenstates.

Then, the general potential may yield strong Flavour Changing Neutral Currents (FCNC). However FCNC are not observed experimentally and the suppression of tree level FCNC in the new potential has to be imposed by hand. This is done by adding a softly broken $\mathbb{Z}_2$ symmetry, implying $\lambda_6 = \lambda_7 = 0$ (8 free parameters left). This symmetry transforms the fields as: $\phi_1 \to -\phi_1,\ \phi_2 \to \phi_2$ which makes the 2 doublets distinguishable. It may then be required that each fermion family (charged leptons, up quarks and down quarks respectively) couple to one of the two fields only, which suppresses tree level FCNC.

The new potential at this stage is then:

$$\begin{aligned} V(\phi_1,\phi_2) &= m_1^2\phi_1^\dagger\phi_1 + m_2^2\phi_2^\dagger\phi_2 - m_{12}^2(\phi_1^\dagger\phi_2 + \phi_2^\dagger\phi_1) \\ &+ \frac{1}{2}\lambda_1(\phi_1^\dagger\phi_1)^2 + \frac{1}{2}\lambda_2(\phi_2^\dagger\phi_2)^2 + \lambda_3(\phi_1^\dagger\phi_1)(\phi_2^\dagger\phi_2) \\ &+ \lambda_4(\phi_1^\dagger\phi_2)(\phi_2^\dagger\phi_1) + (\frac{1}{2}\lambda_5(\phi_1^\dagger\phi_2)^2 + \phi_2^\dagger\phi_1)^2 \end{aligned} \tag{1.4}$$

where all the parameters are real. As in the SM, the fields ϕ_i will acquire a vev v_i and the parameters m_1 and m_2 can be expressed in function of the vevs. Knowing that the 2HDM vevs are related to the SM vev v by $v^2 = v_1^2 + v_2^2$ and defining $\tan\beta = \frac{v_2}{V_1}$, a change of variable can be used to re-express the 7 parameters of potential (1.4) in the so-called *general basis*:

$$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad m_{12} \quad \tan\beta \tag{1.5}$$

Custodial symmetry

One more symmetry must be added in order to keep the parameter $\rho = \frac{M_W^2}{M_Z^2\cos\theta_W}$ close to one. It is indeed measured to be $\rho = 1.00037 \pm 0.00023$ [10].

In SM, the relation $\rho = 1$ is insured by an "accidental" $SO(4)$ symmetry of the Higgs potential. By writing the Higgs doublet as:

$$\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.6}$$

we may see that the the potential V (eq. 1.2) only depends on: $\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2$. The Higgs potential is therefore invariant under $SO(4)$ which itself is isomorphic to $SU(2) \times SU(2)$. Note that symmetry is approximate only and is not respected by the kinematic terms.

The general 2HDM doesn't have that same $SO(4)$ symmetry implying $\rho = 1$, and one should be imposed by hand. There are several ways to achieve that and two cases will be described here.

- A custodial $SU(2)$ symmetry may be imposed as in the SM [11]. It yields the relation: $\lambda_4 = \lambda_5$
- A twisted custodial symmetry may also be used [12]. It yields the relation $\lambda_4 = -\lambda_5$

After imposing that last symmetry, the Higgs potential is left with 6 free parameters. Note that in supersymmetric models the situation is different and MSSM models are left with only 2 free parameters.

1.1.3 Scalar particles

Now that the general 2HDM potential has been simplified and only depends on 6 free parameters, the mass terms of the physical scalar particles may be derived. The fields ϕ_i may be rewritten around their minimums as follows:

$$\phi_1 = \begin{pmatrix} \phi_1^+ \\ \frac{1}{\sqrt{2}}(v_1 + \rho_1 + i\eta_1) \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} \phi_2^+ \\ \frac{1}{\sqrt{2}}(v_2 + \rho_2 + i\eta_2) \end{pmatrix} \tag{1.7}$$

where the ϕ_i^+ fields are complex and the ρ_i and η_i fields are real. These fields have eight degrees of freedom, 3 of which mix with the weak bosons to give them a mass, the 5 others give rise to 5 physical Higgs particles: 2 charged particles ($H^{\pm}$,) 2 neutral scalars (H, h) and 1 pseudo-scalar (A). This may be shown by inserting the expressions (1.7) in equation (1.4). The mass terms obtained for the different particles are described bellow [9].

Charged Higgs: The term from eq. 1.4 contributing to the term are:

$$\mathcal{L}_{Mcharged} = (m_{12}^2 - (\lambda_4 + \lambda_5)v_1 v_2) \begin{pmatrix} \phi_1^- & \phi_2^- \end{pmatrix} \begin{pmatrix} \frac{v_2}{v_1} & -1 \\ -1 & \frac{v_1}{v_2} \end{pmatrix} \begin{pmatrix} \phi_1^+ \\ \phi_2^+ \end{pmatrix} \tag{1.8}$$

Note that the mass matrix has one zero eigenvalue. This means that out of the initial 4 degrees of freedom, two contribute to give the masses of the $W^{\pm}$ and the 2 remaining degrees of freedom give rise to the charged Higgs $H^{\pm}$. Their mass is:

$$M_{H^{\pm}}^2 = (\frac{m_{12}^2}{v_1 v_2} - \lambda_4 - \lambda_5)(v_1^2 + v_2^2) \tag{1.9}$$

Pseudo-scalar Higgs: The term from eq. 1.4 contributing to the mass are:

$$\mathcal{L}_{Mpseudo} = (m_{12}^2 - 2\lambda_5 v_1 v_2) \begin{pmatrix} \eta_1 & \eta_2 \end{pmatrix} \begin{pmatrix} \frac{v_2}{v_1} & -1 \\ -1 & \frac{v_1}{v_2} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \tag{1.10}$$

Again the mass matrix has one zero eigenvalue which gives its mass to the Z boson and the other is yields the pseudo-scalar particle A mass:

$$m_A = (\frac{m_{12}^2}{v_1 v_2} - 2\lambda_5)(v_1^2 + v_2^2) \tag{1.11}$$

Note that both mass matrices 1.8 and 1.10 are diagonalized by the same angle $\tan\beta = \frac{v_2}{v_1}$. What's more, in the case of a normal custodial symmetry ($\lambda_4 = \lambda_5$), the pseudo-scalar and charged Higgs masses are equal: $m_A = m_{H\pm}$.

Scalar Higgs: The term from eq. 1.4 contributing to the mass are:

$$\mathcal{L}_{Mscalar} = \begin{pmatrix}\rho_1 & \rho_2\end{pmatrix} \begin{pmatrix} m_{12}^2\frac{v_2}{v_1} + \lambda_1 v_1^2 & -m_{12} + (\lambda_{345})v_1 v_2 \\ -m_{12} + (\lambda_{345})v_1 v_2 & m_{12}^2\frac{v_1}{v_2} + \lambda_2 v_2^2 \end{pmatrix} \begin{pmatrix}\rho_1 \\ \rho_2\end{pmatrix} \quad (1.12)$$

where $\lambda_{345} = \lambda_3 + \lambda_4 + \lambda_5$ This mass matrix has 2 non zero eigenvalues and gives therefore rise to two neutral scalar Higgs particles H and h. The matrix may be diagonalized and the physical field may be seen as a rotation of angle α of the ρ_i fields. The physical fields are then defined as:

$$\begin{aligned} h &= \rho_1 \sin\alpha - \rho_2 \cos\alpha \\ H &= -\rho_1 \cos\alpha - \rho_2 \sin\alpha \end{aligned} \quad (1.13)$$

In the case of a twisted custodial symmetry, the mass degeneracy is between the charged Higgs and the heavier neutral scalar: $m_H = m_{H\pm}$.

1.1.4 Couplings to SM particles

Types of 2HDMs

There is now several possibilities for the Yukawa couplings giving their masses to the fermions:

$$\mathcal{L}_{mass} = -\lambda_l \overline{L_L} \cdot \phi_i l_R - \lambda_d \overline{Q_L} \cdot \phi_j d_R - \lambda_u \epsilon^{ab} \overline{Q_{La}} \phi_{kb}^\dagger u_R + hc \quad (1.14)$$

where $i, j, k = 1, 2$. Different types of models mare defined and their classification depends on which doublet couples to which fermion family. The main types and their couplings are listed in Table 1.1.

- In the type I model, the field ϕ_1 is totally decoupled from fermionic matter and may be a candidate for dark matter.
- The type II model is the one required by supersymmetry. Here the up quark couple differently from the down quark.
- A type III model also exists: in that case no symmetry is imposed to preserve fermions from coupling to a mixture of scalar doublets. As mentioned above, such a model gives rise to tree level FCNC and will not be discussed here.

- The model X is said to be lepton specific since the lepton and quark couple differently.
- Finally the model Y is also called "flipped" since it has the opposite coupling for the quarks compared to the type II

Model type	up quark	down quark	charged lepton
I	ϕ_2	ϕ_2	ϕ_2
II	ϕ_2	ϕ_1	ϕ_1
X	ϕ_2	ϕ_2	ϕ_1
Y	ϕ_2	ϕ_1	ϕ_2

Table 1.1: Couplings of fermions to the different Higgs doublets

Coupling to fermions

The fermions' couplings to the Higgs doublets may be re-expressed as a function of the physical Higgs A, H, h and $H^{\pm}$ instead of ϕ_1, ϕ_2 as follows [13]:

$$\begin{aligned}\mathcal{L}_{Yukawa}^{2HDH} &= -\sum_{f=u,d,l}\frac{mf}{2}(\xi_h^f\overline{f}fh+\xi_H^f\overline{f}fH-i\xi_A^f\gamma_5\overline{f}fA)\\ &- \Big(\frac{\sqrt{2}V_{ud}}{v}\overline{u}(m_u\xi_A^uP_L+m_d\xi_A^dP_R\Big)dH^+\\ &+ \frac{\sqrt{2}m_l\xi_A^l}{fv}\overline{\nu}_Ll_RH^++hc)\end{aligned} \tag{1.15}$$

The expressions of the parameters ξ are model dependent an they are listed for a type II 2HDM in Table 1.2

ξ_h^u	ξ_h^d	ξ_h^l	ξ_H^u	ξ_H^d	ξ_H^l	ξ_A^u	ξ_A^d	ξ_A^h
$\frac{\cos\alpha}{\sin\beta}$	$\frac{-\sin\alpha}{\cos\beta}$	$\frac{-\sin\alpha}{\cos\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\cot\beta$	$\tan\beta$	$\tan\beta$

Table 1.2: Yukawa couplings of u, d and l to the Higgs bosons h, H, A and $H_{\pm}$ for a type II 2HDM [9]

Coupling to gauge bosons

The couplings of the neutral scalars h, H to the vector bosons $V = W^{\pm}, Z$ are given by [14]:

$$g_{hVV} = \frac{m_V^2}{v} \sin(\beta - \alpha) \tag{1.16}$$

$$g_{HVV} = \frac{m_V^2}{v} \cos(\beta - \alpha) \tag{1.17}$$

And the coupling between h/H and Z, A are:

$$g_{ZAh} = -\frac{g \cos(\beta - \alpha)}{2 \cos\theta_W} (p_h - p_A)_\mu \tag{1.18}$$

$$g_{ZAH} = -\frac{g \sin(\beta - \alpha)}{2 \cos\theta_W} (p_H - p_A)_\mu \tag{1.19}$$

1.2 Higgs phenomenology at the LHC

1.2.1 Higgs measurements

In 2012, a new boson was observed by both ATLAS and CMS experiments at CERN [1,2]. The current measurement of it's mass is $m_h = 125.09 \pm 0.24$GeV [3,4]. This boson, noted h_{125} here, has been observed in the ZZ [15,16] and $\gamma\gamma$ [17,18] final states and evidences have been found in the WW [19,20] and $\tau\tau$ [21,22] final states during the LHC run 1.

Detailed analyses of h_{125} spin, CP and decay width have been carried out [23–26] and its coupling properties have been measured as well [3,4]. More specifically, h_{125} couplings to fermions are proportional to the fermion mass and the couplings to vector bosons are proportional to their squared mass. All of these results are consistent with the SM Higgs boson to the order of 10% [27].

Let's now elaborate briefly on the Higgs measured coupling properties. The SM predicts precise branching ratios (BR) for the Higgs boson decays to different particles (see figure 1.1). These predicted BR may be compared to the measured Higgs h_{125} decay rates in the different decay channels. In order to quantify eventual deviations from the SM predictions, coupling modifiers may be introduced:

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_i^{SM}} \tag{1.20}$$

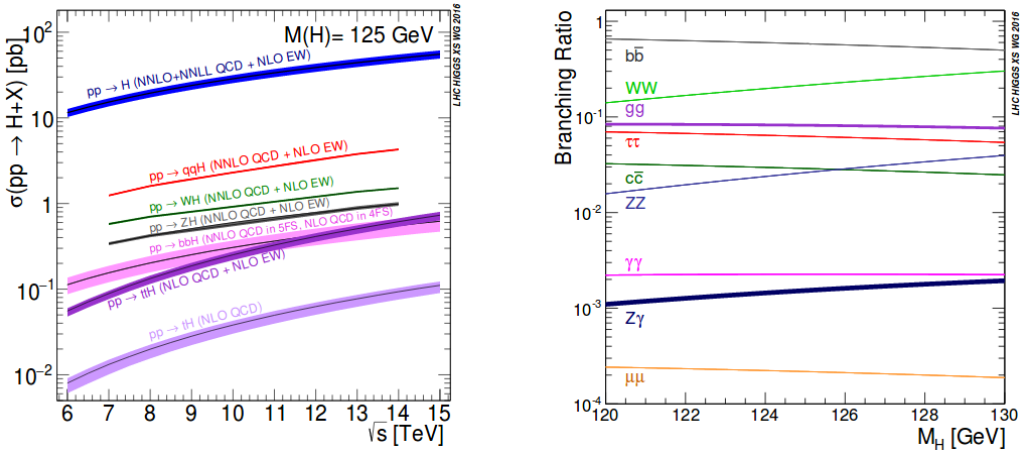


Figure 1.1: Predicted Higgs production cross-sections as a function of the center of mass energy, $\sqrt{s}$, for p collisions (left) and branching ratios for the main decays of the SM Higgs boson near $m_H = 125$ GeV (right) [28]

where σ_i and σ_i^{SM} are respectively the observed cross-section and the cross-section predicted by the SM. The values of these parameters for the Higgs coupling to vector bosons, heavy fermions (t, b, τ), gluons and photons are plotted on figure 1.2. Their closeness to 1 yields constrains on all Beyond Standard Model theories. The specific constrains on 2HDMs are described in the next subsection.

1.2.2 Constraints on 2HDM

The first obvious constrain on any 2HDM coming from measurements is the necessity to identify one of the scalars h or H with the observed boson h_{125}. In this thesis we will identify the lighter scalar particle to the observed boson: $h = h_{125}$.

The SM Higgs can be expressed in terms of h and H as:

$$h_{SM} = h \sin(\beta - \alpha) - H \cos(\beta - \alpha) \tag{1.21}$$

Since the observed Higgs h_{125} behaves as a SM Higgs ($h_{125} \sim h_{SM}$), the identification $h = h_{125}$ means that the angles α and β should be taken such as $\cos(\beta - \alpha) \to 0$ or, considering the other coefficient, $\sin(\beta - \alpha) \to 1$. This is known as the *alignment limit*. Note that in that case, the couplings of the heavier boson H to the SM particles will be strongly suppressed.

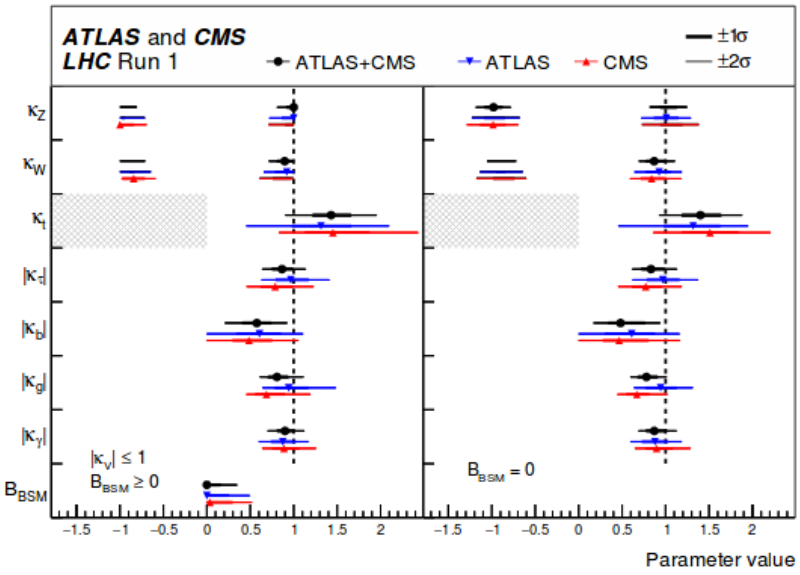


Figure 1.2: ATLAS-CMS combined measurements of coupling modifiers [28]

Limits on α and β can be derived from Higgs precision measurements. Since the 2HDM h decay rates are scaled by a ξ factor compared to the h_{SM} rates (recall table 1.2), itself depending on α and β, the coupling modifiers κ_i may be translated into limits on the 2HDM angles as shown on figure 1.3. For type II models, the constraints are important and $\cos(\alpha - \beta)$ has to be taken close to 0. The small lobe observed for $\tan\beta > 3$ and $\cos(\beta - \alpha) > 0$ is due to the opposite signs of κ_u and κ_d in that region.

1.2.3 Choice of the decay channel

Many direct searches for additional Higgs bosons have been carried out by the CMS experiment [27]. But the parameter space to probe is huge and even with these direct measurements and the constraints coming from precision measurements, an important part of the parameter space is still accessible.

When working with a type II 2HDM in the alignment limit, observing the heavier scalar H becomes difficult since it decouples from vector bosons (recall $g_{HVV} \sim \cos(\beta - \alpha)$, eq. 1.16). The decay into a pair of top quarks $H \to t\bar{t}$ is possible for $m_H > 2m_t$ but below that limit, the decay $H \to ZA$ is the most promising ($\sim \sin(\beta - \alpha)$). This decay is forbidden in MSSM scenarios as well as in generic type II 2HDM with a usual custodial symmetry because of the mass spectrum, but is natural in the case of the inverse mass

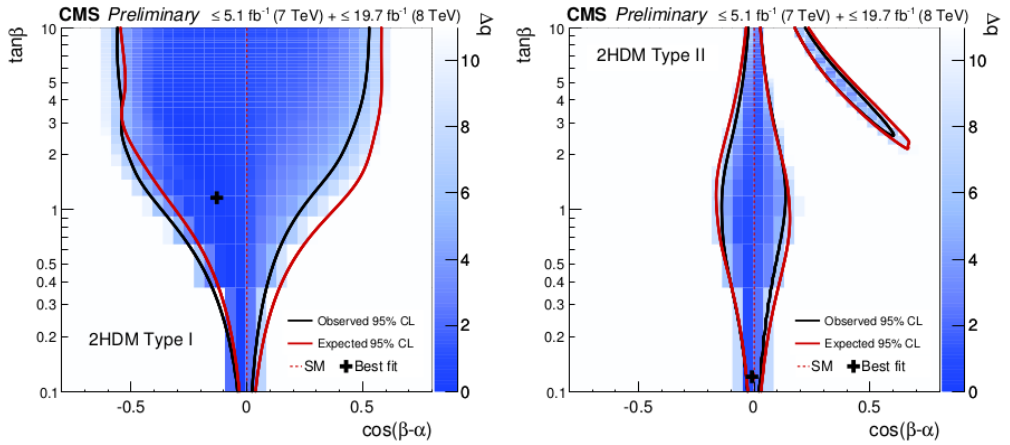


Figure 1.3: Allowed parameter space (in blue) for $\tan\beta$ and $\cos(\beta-\alpha)$ measured with the CMS detector for type-I 2HDM (left) and type-II 2HDM (right) [27]

spectrum characteristic of the "twisted" custodial symmetry [12].

In this thesis, the 2HDM studied is chosen to be of type II, CP-conserving, satisfying a $\mathbb{Z}_2$ symmetry and taken in the alignment limit. The parameters α and β are fixed such as: $\tan\beta = 1.5$, $\cos(\beta-\alpha) = 0.1$ and the masses m_H and m_A will be scanned.

The studied decay is $H \to ZA \to l\bar{l}b\bar{b}$. $H \to ZA$ has the highest branching fraction on most of the parameter space as illustrated on figure 1.4. The Z is chosen to decay into leptons for purity and for our choice of parameter, the pseudo-scalar decay into two b quarks has the highest BR just before τ leptons.

1.3 Motivation for a 2HDM study

1.3.1 Standard Model's shortfalls

The SM yields an excellent description of particle physics, especially in collider experiments. What's more the observed boson h_{125} is in good agreement with the SM expectations. But the Standard Model also suffers from several shortfalls and is therefore believed to be an incomplete theory. This subsection will briefly recall several of the SM problems.

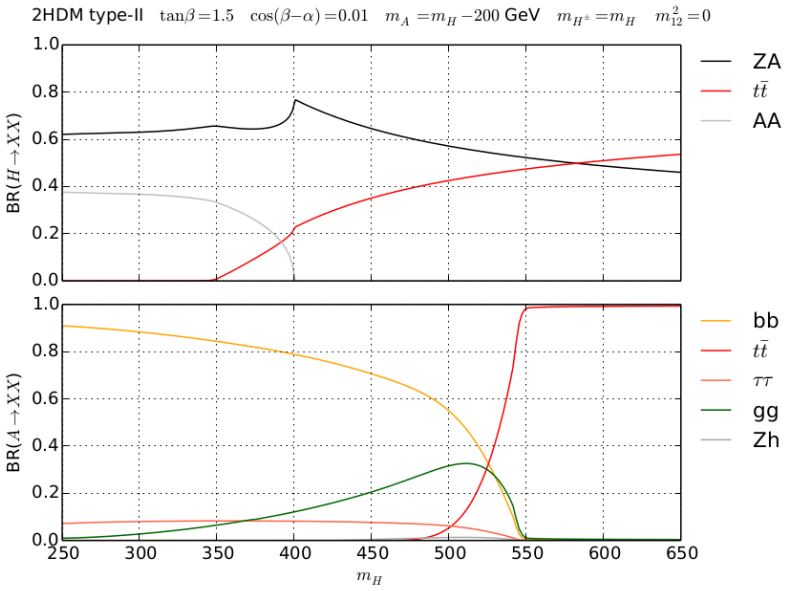


Figure 1.4: Evolution of the branching fraction as a function of m_H for $m_A = m_H - 200\text{GeV}$,$\tan\beta = 1.5$ and $\cos(\beta - \alpha)$. Top: $H \to XX$, bottom $A \to XX$ [29]

The most obvious problem is probably the absence of gravity from the SM. Gravity should become important at the Planck scale ($M_p = 1.2 \times 10^{19}$GeV). New physics are expected at the Planck scale (at least) and should affect the Higgs mass m_h through loop correction. The fact that the Higgs mass is observed at a scale far below the Planck scale requires fine tuning, cancellation between loop corrections. This fortuitous cancellation is known as the *hierarchy problem*.

Another important problem is the *baryonic asymmetry* in the universe. In order for the Big-Bang to create more matter than anti-matter, the Sakharov conditions must be met. These include a significant source of CP-violation, and no source is strong enough in the SM.

Then the *Dark Matter* (DM) problem must be mentioned: the presence of DM has been clearly shown by the study of the Cosmological Microwave Background (CMB). The SM offers no satisfying DM candidate and many theories and extension to the SM have been developed to explain the origin of DM.

Finally, an other problematic observation is the discovery of *neutrino oscillations*, that is, the observation of neutrino masses. In the minimal SM, neutrinos are massless and it is currently believed that neutrinos acquire a mass through the so called see-saw mechanism, which implies a more complex scalar structure than the SM's.

1.3.2 Advantages of 2HDMs

The Standard Model assumes the simplest scalar structure possible: a single Higgs doublet. But there is no reason to have such a simple structure (especially when knowing the complex structure of families in the fermionic sector). A second doublet may be introduced easily thus obtaining a 2HDM.

Two Higgs Doublet Models may be used as stand-alone models and it has been shown [30] that such models might explain baryonic asymmetry (by providing enough CP-violation). What's more a high mass gap between H and A has been shown to be in favour of strongly first order electro-weak phase transition. This high mass gap is rightly allowed by the choice of a twisted custodial symmetry made here.

Then 2HDMs can be required in broader theories. The most studied example is probably supersymmetry, in which the addition of super-partners particles naturally solves the hierarchy problem and also provides several Dark Matter candidates.
An extended Higgs sector is also required by axion theories which were introduced to solve the strong CP-problem and also provide dark matter candidates.

Chapter 2

Detector and particle reconstruction

This chapter gives a brief overview of the CMS detector operating at the LHC, of the processes of data taking and event reconstruction. Information about the LHC itself and the proton accelerating chain may be found in [31]

This chapter's first section describes CMS sub-detectors as well as the trigger system and the pile-up subtraction. Then the second sections explains the particle reconstruction and identification process. The focus is put on the objects required by the analysis: electrons, muons and b-jets. Finally the last section gives a brief overview of the different possibilities for data simulations.

2.1 CMS Detector

The Compact Muon Solenoid (CMS) is a multi-purpose detector operating at the Large Hadron Collider (LHC) at CERN. Its general structure is schematised on figure 2.1. CMS is composed of trackers, calorimeters and a superconducting solenoid magnet. The different sub-detectors are described in the coming sections from the inner layers to the outer ones. A more detailed description may be found in [32]

Coordinate system:
Let's first recall the coordinate system used by CMS: the x axis points radially inwards the LHC ring, the y axis points upwards and the z axis points along the beam-line such as $\vec{x} \times \vec{y} = \vec{z}$.
The azimuthal angle ϕ is measured from the x axis in the x-y plane, r is the

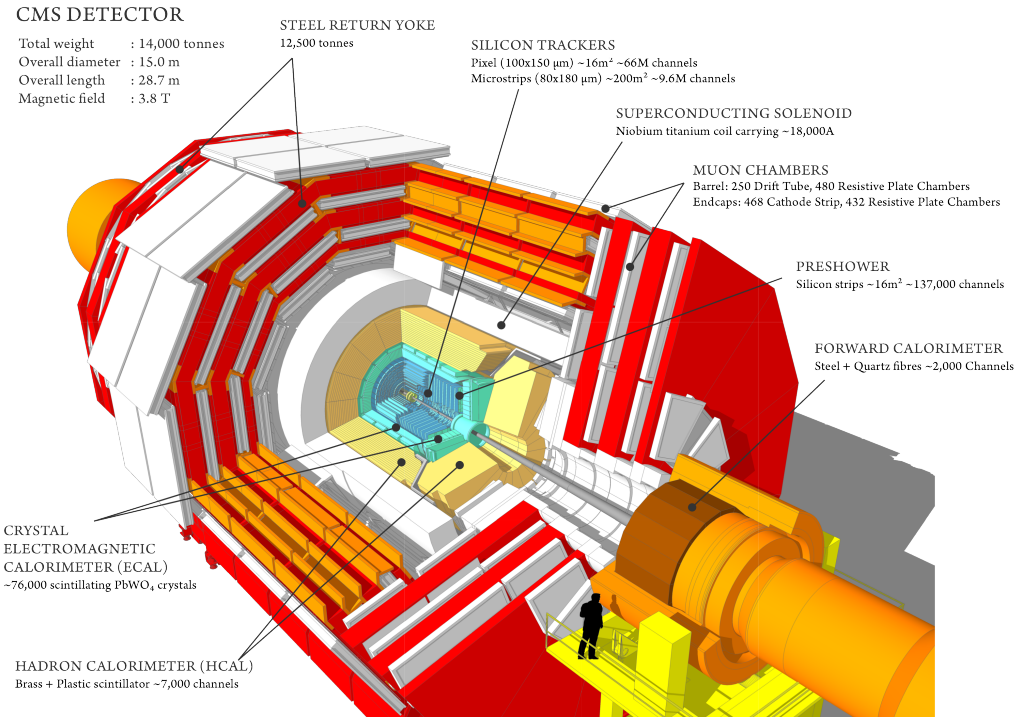


Figure 2.1: Diagram of the CMS detector and its different layers of sub-detectors

radial coordinate in that plane and the polar angle measured from the z axis is noted θ. The pseudo rapidity η defined as:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{2.1}$$

is often preferred to the polar angle θ. Indeed for $p_T >> m$ the difference of pseudo rapidity between two particles is invariant under a Lorentz boost along the z-axis in contrary to the angle θ. This is important since particles are produced in frames often boosted along the z axis.

Solenoid:

In order to measure precisely the momentum of highly energetic charged particles a magnetic field of high bending power is required. CMS therefore contains a 3.8-T superconducting solenoid of 13m long and 6m of diameter.

2.1.1 Sub-detectors description

Tracker

The inner tracker is meant to measure the trajectories of charged particles produced in the collisions as well as to reconstruct the primary and secondary vertices of interaction.
It is composed of a 3-layer pixel detector ($r = 4.4$ cm, 7.3 cm, 10.2 cm) and of a 10-layer silicon strip tracker (10.2 cm $< r <$ 110 cm) in the barrel region. Each system is completed by end-caps disks (2 pixel layers, 3+9 strip layers) extending the pseudo-rapidity coverage to $|\eta| < 2.5$

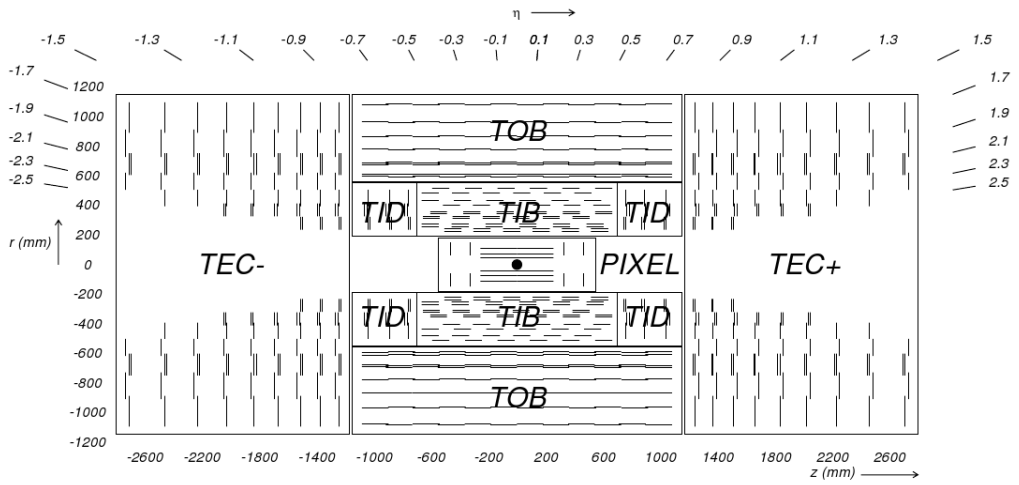


Figure 2.2: Detailed structure of CMS inner tracker [32]

Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is a hermetic and homogeneous lead tungstate ($PbWO_4$) crystal detector. It is designed to measure the energy of photons and electrons. The barrel part (EB) is made of 61 200 crystals and covers a pseudorapidity range of $|\eta| < 1.479$. The end-caps (EE) cover the pseudo rapidity range $1.479 < |\eta| < 3.0$ and is composed of 7 324 crystals. Avalanche photodiodes are used in the barrel region and vacuum photodiodes are used for the end-caps as photodetectors.

Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is designed to measure hadron jets. It is made of towers of sampling calorimeters: an alternation of scintillator tiles

and of absorber layers (brass). The HCAL is composed of 4 different regions:

- The barrel region (HB) goes from the ECAL to the solenoid magnet ($1.77m < r < 2.95m$) and covers $|\eta| < 1.4$
- The hadron end-caps (HE) cover $1.3 < |\eta| < 3.0$
- An outer HCAL (HO) is placed outside the solenoid to reduce the tail in the energy resolution distribution. The solenoid is then used as an absorber.
- The hadron forward (HF) is a different sampling calorimeter made of steel absorbers and Cerenkov quartz fibbers. It is radiation hard and covers $3.0 < |\eta| < 5.0$

Muon chambers

The muon detectors are located outside of the solenoid magnet between the flux-return yoke. The muon system has three functions: muon identification, momentum measurement and a triggering function. To do so, three types of gaseous detectors are used:

- In the barrel region where the muon rate is low and the magnetic field is uniform, regular drift tubes (DT) are used. The DT cover the region $|\eta| < 1.2$
- In the 2 end-cap regions, the muon and background rates are higher and cathode strip chambers (CSC) are used. The pseudorapidity range covered is $0.9 < |\eta| < 2.4$
- A complementary resistive plate chambers (RPC) is dedicated to the triggering system ($|\eta| < 1.6$). They have a faster response but coarser spatial resolution than DT and RPCs

2.1.2 Triggers

At the LHC, the proton bunches have a crossing frequency of 40MHz and for each crossing, 20 proton-proton interactions have been recorded on average at 8 TeV. The resulting quantity of collision data cannot be processed nor stored and a system of trigger has been implemented to reduce the data output rate.

The trigger system has two stages, the first is the *Level 1 (L1)* trigger. It reduces the output rate to a maximum of 100 kHz by using simple information from the calorimeters and muon system to select interesting events. The selection is done by an electronic system in 3 μs and during that time all the information from collisions is stored in a pipeline. If the event passes the L1, this information is transmitted to a computer farm for the second stage of selection.

The *High Level Trigger (HLT)* is a software based trigger that reduces the data output rate from 100 kHz to about 600 Hz. The process takes about 0.1 s per event and uses more complex objects. Several trigger paths are defined: they are meant to select different types of interesting events depending on criteria such as: number, type, p_T of particles. The two triggers used in this thesis are the di-electron and di-muon triggers where the events are kept if containing two electrons (muons) of traverse momentum higher than 20 GeV (leading) and 12 GeV (subleading).

2.1.3 Pile-up

When two proton bunches cross, several interactions happen and most of them are low energy uninteresting interactions. These interactions constitute a noise called pile up and the emerging particles should be identified and subtracted from the main interesting event.

This can be done in several ways: once the particles tracks and vertices have been reconstructed (see section 2.2), pile-up interactions are suppressed by removing the tracks and deposits coming from vertices situated far from the main hard interaction vertex. This procedure is called *charged hadron subtraction (CHS)*.

The deposit coming from neutral particles are more difficult to identify (since they don't interact with the tracker). The average contamination in jets can be estimated using the *Hybrid Jet Area Method* detailed in [33] and removed from the jet.

2.2 Event reconstruction and particle identification

2.2.1 Tracks reconstruction

Track reconstruction is a complex process and only its general structure is given here. More details may be found in [34].
The tracker is sensitive to charged particles only: they interact creating electron-holes pairs in the tracker's sensitive region. This creates a current then detected by the front-end electronics. The first step in track reconstruction is to suppress noise and to proceed to a local reconstruction.

In the pixel detector, clusters are formed from adjacent pixels (side-by-side, corner-by-corner adjacent cells). Clusters are kept only if they have a charge equivalent to 4000 electrons or more.
In the strip detector, a cluster is started by a strip having registered 3 times its expected noise. The adjacent strips with a signal of twice the noise are added to the cluster. The position of the cluster is given by the charge weighted average of the strips positions.

The hits obtained by local reconstruction are then used to reconstruct particles tracks. This is done by iterating six passes of the Combinatorial Track Finder (CTF), the CMS track reconstruction algorithm based on the Kalman filter [35]. The steps for each iteration are the following:

1. *Seed generation:* An initial track candidate is defined from two or three hits.

2. *Track finding:* The track is extrapolated using Kalman filter.

3. *Track fitting:* This step yields the track parameters

4. *Track selection:* The track is kept if it matches quality requirements.

This process starts with tracks of relatively large p_T produced near the interaction region. After one iteration the reconstructed tracks hits are hidden and the CTF is run again on lower p_T or displaced tracks.

The typical track reconstruction efficiency for charged particles of $p_T > 0.9$ GeV and $\mid \eta \mid < 0.9$ is of 94% and of 85% for $0.9 < \mid \eta \mid < 2.5$. For isolated muons the efficiency rises to 100% and at 100 GeV the resolution on the mouns p_T is of 1.8%.

2.2.2 Vertices reconstruction

Vertices are reconstructed from the tracks: primary vertices (PV) correspond to $p - p$ interactions and secondary vertices (SV) to the decay of particles produced in the initial collision. Vertices from uninteresting events corresponding to pile up are removed from the event as described in 2.1.3.

The tracks used to reconstruct vertices must have a minimum of 2 hits in the pixel detector, hits in total and $\chi^2 < 20$ from the track fit. They must also come from the luminous region where the collision come from. The accuracy on the reconstructed vertices positions is of 10-12μm

2.2.3 Calorimeter clustering

Calorimeter clustering has the purpose to detect the energy deposit coming from charged and neutral particles, to separate these deposits form different particles, to reconstruct electrons (including bremsstrahlung photons) and to improve the energy resolution associated to poor quality tracks [36]. The clustering is done in each sub-detector separately.

The general process is the following: first the clustering algorithm finds a *cluster seed* which is a local energy maximum. Then topological clusters are grown by adding adjacent cells with an energy above threshold (80MeV in the EB, up to 300 MeV in the EE and up to 800 MeV in the HCAL).

2.2.4 Particle reconstruction and identification

To reconstruct particles, the right tracks and/or calorimeter clusters must be linked together. The linking algorithm iterates on all elements (tracks and energy clusters) and evaluates the likelihood of the link between elements. Once the elements are linked, they may be identified as different types of particles, thus creating *particle flow* (PF) objects [36]. The identification process varies with the particle type and three cases will be detailed here: electron, muon and b-jet reconstruction.

In general, charged particles tracks are reconstructed from the tracker then extrapolated to the ECAL and eventually to the HCAL. The track is linked to a calorimeter cluster if the extrapolated position falls into a cluster boundaries. The cluster boundaries might be extended of one cell in order to account for errors coming from hermicity imperfections, uncertainties on shower maximum and for multiple scattering. For neutral particles that

didn't leave hits in the tracker, the clusters from the ECAL can also be extrapolated to the HCAL and the objects are linked with a similar method.

Muons

Muons are reconstructed from the muon chambers tracks combined with the inner tracker tracks. The hits from the muon chambers are used to create *stand alone muon tracks*. These tracks are extrapolated and fitted to inner tracker tracks. If the fit is successful, the reconstruction algorithm yields a *global muon*.

Lower energy muons might not create enough hits in the muon chambers to create a stand alone muon. Therefore another algorithm may be used: all the inner tracker tracks are extrapolated to the muon chambers. If they match a hit, a *tracker muons* is created.

An isolation criteria is usually defined for muons (and electrons) in order to differentiate prompt muons from background (for example decay products of b or c-jets). The particles contained in a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around the muon track should be sparse and of low energy. Limits are imposed on parameters of the particles contained in the cone such as the sum of p_T, the number of tracks or calorimetric variables.

Electrons

The electron tracks are more difficult to reconstruct than muon tracks because of the important energy loss undergone by electrons before reaching the ECAL. Energy losses come from bremsstrahlung and showering due to photon conversion. This leads to non-Gaussian event by event fluctuations and reduces the hit efficiency and the accuracy of the tracks parameters given by the Kalman Filter. Another algorithm is then used to re-fit the tracks: the Gaussian-sum Filter (GSF) [34].

Clustering in the ECAL is also complicated by bremstrahlung photons. Electron clusters need to be expanded into super clusters (SC) containing the initial electron and its bremsstrahlung photons. Two algorithms are used in the barrel and end-caps respectively, they both add neighbouring cells with an energy above a given threshold to an initial seed (local energy maximum) [37]. In the EB, the *hybrid algorithm* is used. It is meant to include the bremstrahlung photons spread in ϕ to the electron cluster by adding arrays of 5×1 arrays of crystals (in $\eta \times \phi$ to the original seed. In

the EE the *multi-*5×5 *algorithm* is used. In adds arrays of 5×5 cells to the original seed.

The SC and tracks are then combined together to derive the final track parameters. And as for muons, isolation parameters are defined to differentiate prompt electrons from background.

b-jets

High energy quarks created by the p-p collisions hadronise into showers of particles, thus a specific observable must be defined to include all secondary particles: a jet. Jets are cone-like structures containing different particles and energy deposits assumed to come from the hadronisation of a high energy quark or gluon. Associating the right particles together is not an easy tasks and several algorithms have been developed to define the jets limits.

Energy deposits from ECAL and HCAL can directly be clustered into jets (CALO jets) or particle flow candidates may be used (PF jets) [38]. This latter option is currently used by CMS and the sub particles are combined using the *anti-*k_T algorithm [39]. The distance between two PF i and j is defined as:

$$d_{ij} = min(E_{Ti}^{2p}, E_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{2.2}$$

where $p = -1$ for antik_T. Jets are built by an iterative process, the hardest object of an event is used as a first seed and surrounding particles are associated to it if they are close enough (typically $d_{ij} < 0.5$). Then the jet center is recomputed – and will change significantly if other hard particles have been added to the jet. The new center is used as a seed and the process is re-iterated.

Once all the event's jets have been reconstructed, the jets coming from b quarks may be identified. B hadrons typically travel a few mm in the detector before decaying. The decays usually produce several charged particles. This particles reconstructed by the particle flow will lead to the reconstruction of a secondary vertex a few mm away from the primary vertex. The position of the vertex is a crucial parameter to determine if a jet comes from a heavy quark. To increase the b-tagging efficiency, only good quality jets are considered and several cut are imposed on the jets candidates such as $p_T > 1$ GeV, a minimum of 8 hits in the tracker (including 2 hits in the pixel detector) and a track fit that satisfies $\chi^2/n.d.o.f < 5$ [40].

To determine if a jet comes from a b quark, several b-tagging algorithms have been developed. They are based on either tracks impact parameters or vertex position or both. Note that all b-tagging algorithms will eventually b-tag jets coming from c or light quarks. This is called mistagging and the percentage tolerated depends on every analysis requirements. Three working points are usually defined: loose, medium and tight where 10%, 1% and 0.1% of the jets are mistagged respectively.
In this thesis, the *Combined Secondary Vertex* (CSV) algorithm has been used. As shown of figure 2.3 it is the most efficient algorithm at the medium working point.

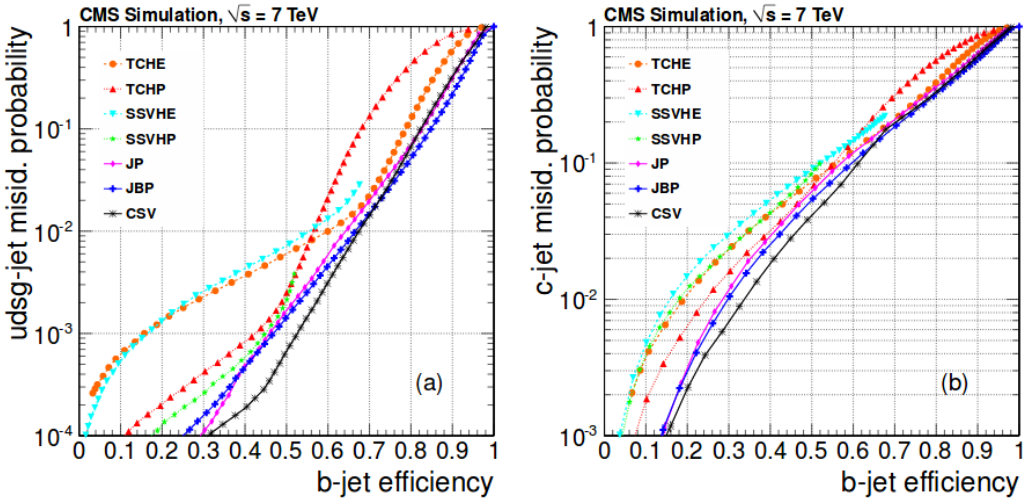


Figure 2.3: Comparison of tagging and mistagging efficiency for different b-tagging algorithms for jets of $p_T > 60$ GeV (from [40])

2.3 Event simulation

The data obtained by CMS needs to be compared to predictions by using a precise modelling of the p-p interactions and of the detector response. This is done by using Monte Carlo (MC) simulations to compare to the data or to estimate the background for a given analysis.

Full simulations are the most precise, they are computed with the following steps:

1. An event generator such as aMC@NLO [41] is used to generate the p-p collision. It uses the observed PDFs, and uses Matrix element method

to generate interactions at Leading Order (LO) or Next to Leading Order (NLO). Hadronisation and showering are either computed by the generator or by a dedicated program.

2. Once the stable particles (stable enough to propagate through the detector) have been computed they are propagated through the detector. This is done by using a simulation of the detector using the GEANT4 software [42]. The program contains an accurate description of CMS structure and it simulates the matter-particles interactions in each detector layer.

3. Then the different simulated energy deposits are processed and digitised into the expected electronic signals.

4. The result of the last step yields simulated data in the same format as real collision data. The physical objects are therefore reconstructed using the same reconstruction processes as for data – as explained in section 2.2.

This kind of simulation is especially time consuming (around 1 min per event) and simplifications may be used to speed up the simulation process.

Fast simulations use simplifications such as approximate detector geometry or parametric responses of certain components to speed up the processing time. Such software are especially useful to scan wide parameter spaces like the one available in 2HDM models.
The DELPHES [43] software has been used in this thesis. Its uses a very simple detector response simulation (tracks and calorimeter deposits) and the particle flow is simplified as well. The reconstructed particles are smeared with a parametric approach to reflect the full detector response. The time per simulated event is of around 10ms.

Chapter 3

Analysis

As explained in the first chapter, 2HDMs are a compelling extensions of the Standard Model and the decay $H \to Z(l\bar{l})\ A(b\bar{b})$ is an interesting signature to study. The core of this thesis is the search for this double resonance in the (m_{bb}, m_{llbb}) mass plane with CMS 8 TeV data. The analysis is kept as model independent as possible in order to eventually be re-interpret it in an other framework than the specific type II 2HDM used here.
A similar search has already been carried out by CMS [5] with 8 TeV data and the goal here is to test the potential of a new analysis method.

3.1 Analysis strategy

The 2HDM model used for this analysis has two free parameters impacting the kinematics: m_A and m_H, the masses of the pseudo-scalar and of the heavier neutral scalar respectively. Therefore in order to find the double resonance $H \to Z(l\bar{l})\ A(b\bar{b})$, the (m_{bb}, m_{llbb}) mass plane is scanned with $m_{bb} \in [30, 1000]$ GeV and $m_{llbb} \in [100, 1000]$ GeV. Note that the only observables used here are invariant masses that do not depend on the scalar or pseudo-scalar nature of H and A, therefore this analysis can also be reinterpreted as the search for $A \to Z(l\bar{l})\ H(b\bar{b})$. The parameters α and β influence the cross section and will be kept fixed.

The search consists in a cut and count analysis and the mass plane has been paved with overlapping bins. The novel implementation of this thesis is the choice of elliptical bins to scan the parameter space. The definition of the binning is described in detail in section 3.3. The background contained in each bin is estimated thanks to a data driven method described in section 3.4.

The constraints coming from the model are kept minimal and the only two constraints on the masses m_A and m_H are:

- $m_b \approx 5$ GeV $\rightarrow$ $m_A > 10$ GeV
- $m_H > m_A + m_Z$

The width of the resonances in the (m_{bb}, m_{llbb}) mass plane are dominated by detector resolution. Note that for important mass differences between H and A, the daughter particle is produced with an important boost resulting in close b-jets. For $m_H > 5 \times m_A$ the jets begin to overlap and for $m_H > 10 \times m_A$ the analysis efficiency is zero. This region of the mass plane is referred as the *forward region*.

3.1.1 Statistical method

To determine if the observed number of events is compatible with the SM background or with a 2HDM, statistical tests are used. In order to explain the different tests, let's define the null hypothesis H_0 as corresponding to the background only hypothesis and the alternative hypothesis H_1 as corresponding to the background+signal hypothesis.

The **p-value** quantifies the level of agreement between the number of observed events n and the number of events expected by the H_0 hypothesis. Assuming H_0 true, the p-value is the probability of observing n or more events in data. The H_0 hypothesis is excluded if the p-value is below a given threshold. This threshold is often taken at 0.05 and corresponds to the 95% confidence limit.
The p-value can also be converted into the significance Z. Supposing H_0 distribution Gaussian, Z is the number of standard deviation from the mean that gives a one sided probability p.

The limits on the 2HDM decay rate are set using a frequentist significance test on a **likelihood ratio**. The likelihood function is defined the product of the Poisson probability for each signal bins times constraint terms [44]:

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} \exp\left(-\mu s_j + b_j\right) \prod_{k=1}^{M} f_p(a_p \mid \theta_p) \tag{3.1}$$

where μ is the signal strength, $s_j(b_j)$ the expected signal(background) rate and n_j the observed rate in the signal bin j. Note that here only one signal bin

will be considered at a time (N=1). $\boldsymbol{\theta}$ represents all the nuisance parameters and the second product term corresponds to the systematic uncertainties that should be taken into account. Here distribution f_p will be taken to be a log-normal.
The likelihood ratio used as a test statistics is then:

$$q(\mu) = -2 \ln \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{3.2}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the values of the parameters that maximise L and $\hat{\hat{\boldsymbol{\theta}}}$ are the parameter that maximises L with μ fixed.

To set an upper limit on the signal strength, the CL_S method is used [45]. The upper limit on the signal strength μ is computed by solving $p'_\mu = 0.05$ where p' is a modified p-value defined as:

$$p'_\mu = \frac{p_\mu}{1 - p_b} \tag{3.3}$$

$$p_\mu = \int_{q_{\mu obs}}^{\inf} f(q|\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu, obs)) dq_\mu \tag{3.4}$$

$$p_b = 1 - \int_{q_{\mu obs}}^{\inf} f(q|0, \hat{\hat{\boldsymbol{\theta}}}(0, obs)) dq_\mu \tag{3.5}$$

Here the p-value p_μ is defined to quantify the agreement between data and H_1.

These tests are done using the COMBINE tool from CMS [46].

3.2 Setup

3.2.1 Event selection

This analysis uses CMS run I data collected at $\sqrt{s} = 8$ TeV and corresponding to an integrated luminosity of 19.8 fb^{-1}. The datasets used contain the events having passed the di-muon or di-electron HLT described in section 2.1.2. Signal events contain 2 b-jets and 2 opposite sign leptons (e^-e^+ or $\mu^-\mu^+$) hence the SM background comes mostly from $t\bar{t}$ (decaying into 2 b-jets and 2 W), $Z + jets$ and ZZ production.

To discriminate events coming from background processes, cuts are applied on the data events. They are listed in Table 3.1.
First the two leptons must come from a Z, they are therefore required to be of opposite sign and their invariant mass must be compatible with the Z mass. Cuts on the leptons rapidity and transverse momentum are applied to improve purity. The isolation variable for electrons is set at the medium working point and at the tight working point for muons.
Then 2 b-jets are obviously required. They are tagged using the CSV algorithm and the middle working point is used (the tight working point would lead to too little statistics). A cut on the jets transverse momentum p_T^{jet} and rapidity $\mid \eta_{jet} \mid$ is set in order to reduce the contamination from pile-up.
Since the cone size of jets is defined to be $R = 0.5$, the angle between jets and leptons is required to be superior to that to avoid overlap.
Finally the expected signal does not produce any invisible particle and no missing transverse energy (MET) is expected. Hence a cut on the MET significance E_T^{miss} is set to discriminate events with neutrinos coming from $t\bar{t}$.

76 GeV $< m_{ll} <$106 GeV
$n_b \geq 2$
CSV$_b > 0.679$
$p_T^{e,\mu} > 20$ GeV
$\mid \eta_\mu \mid < 2.4, \mid \eta_e \mid < 2.5$
$p_T^{jet} > 30$ GeV
$\mid \eta_{jet} \mid < 2.4$
$\Delta R(l,j) > 0.5$
$p_T^b > 30$ GeV
E_T^{miss} significance < 10

Table 3.1: Event selection for the objects used in the analysis.

3.2.2 Background simulations

Although the background subtraction is data driven, simulated samples are required to implement the subtraction as explained in section 3.4. They were also used to derive the analysis expected limits before looking at the data. Four sets of samples are used (see Table 3.2): one sample of $Z + jets$, one sample of $t\bar{t}$ decaying in the dileptonic channel ($bbll\nu\nu$), one sample of $t\bar{t}$ decaying in the lepton+jet channel ($bbl\nu jj$) and a last sample of ZZ.

The samples are reused from the previous analysis [5]: the ZZ sample was initially generated with PYTHIA 6 [47] and interfaced with TAULA [48]. The Z+jets samples were generated with MADGRAPH 5 [41] and interfaced with TAULA.

Samples	Cross section (pb)	Number of events
Z+jets (inclusive)	3503.7 (NNLO)	30 459 503
$t\bar{t}$ (dileptonic)	27.3 (NNLO)	12 119 013
$t\bar{t}$ (lepton + jet)	109.2 (NNLO)	25 414 818
ZZ	8.2 (CMS)	9 799 908

Table 3.2: List of samples used and the corresponding cross section computed at a given order. "CMS" means that the theoretical cross section is rescaled to the best CMS measurement.

3.2.3 Signal generation

Around 300 signal samples generated for the previous analysis [5] have been reused. Each sample contains 100 000 events and was generated for the decay $H \to Z(l\bar{l})\ A(b\bar{b})$ of a type II 2HDM of inverted mass spectrum. The samples were generated for different values of m_A and m_H and the model parameters used are listed in Table 3.3. These samples are used to derive the optimal binning (see section 3.3) and to compute the limits on cross sections by giving the number of signal events expected in each bin.

The 2HDM description and branching ratios were obtained with the 2HDMC calculator [49] and the events were generated with MadGraph 5. The corresponding cross sections were computed at NNLO thanks to SUSHI 1.4.1 [50]. Then detector response was simulated with DELPHES 3 [43]. The consistency between DELPHES simulations and CMS official simulations was tested [29]. The kinematics of both simulations are in good agreement but there is a slight discrepancy in the overall efficiency. This is taken into account when deriving the expected cross section by adding an correction coefficient.

3.3 Binning

When looking at the shape of the simulated signal in the (m_{bb}, m_{llbb}) mass plane (Fig.3.1) one clearly sees that the signal shows a roughly elliptical shape because of the correlation between the two mass observables. Therefore, if usual rectangular bins are defined to scan the phase space, part of the bins

$m_A \in [10, 1000]$ GeV
$m_H \in [100, 1000]$ GeV
$m_H > m_A + m_Z$
$m_{H\pm} = m_H$
$m_h = 125$ GeV
$\tan\beta = 1.5$
$\cos(\beta - \alpha) = 0.01$
$m_{12}^2 = m_{H\pm}^2 \cos\beta \sin\beta$

Table 3.3: List of parameters used to generate the signal samples. The values of α and β come from a benchmark point chosen in order to be consistent with other 2HDM searches.

contains only background. To increase the signal over background ratio, one can take advantage of the spacial distribution of the signal by defining elliptical bins.

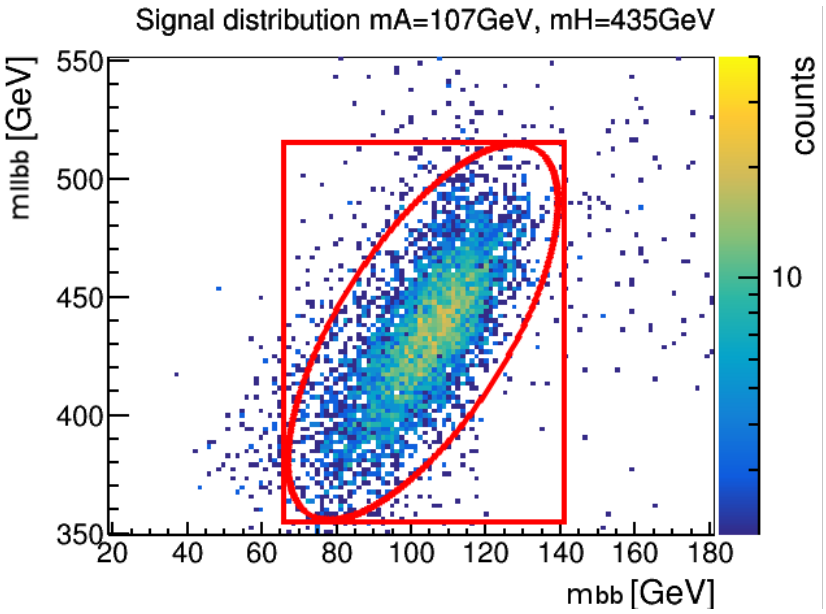


Figure 3.1: Signal distribution for $m_A = 107$ GeV, $m_H = 435$ GeV. Note the the distribution is roughly elliptical. A well fitted elliptical bin contains as much signal as a rectangular bin while being smaller, thus containing less background.

This choice of elliptical bins leads to several complications: first, the signal distribution varies with m_A and m_H and so does the shape of the

optimal ellipse containing it. Secondly paving the entire mass plane with (varying) ellipses is a complex task. Both of these issues have been addressed and the approach taken is detailed in the two following subsections.

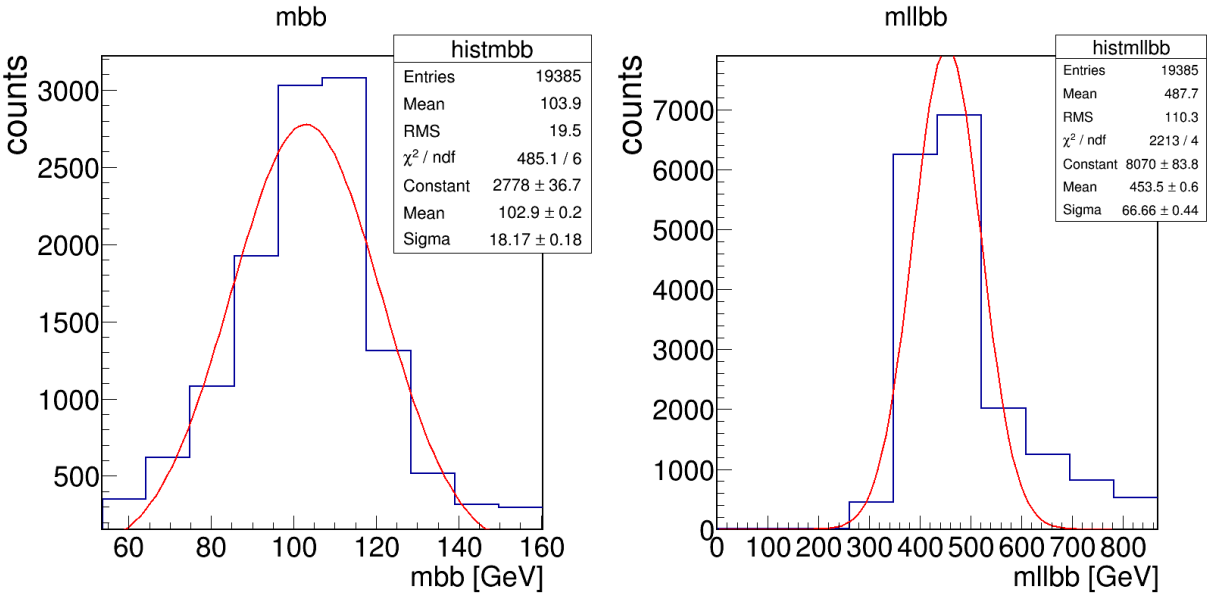


Figure 3.2: Mass fits for the sample generated at $m_A = 107$ GeV, $m_H = 435$ GeV. The reconstructed masses from the fit are $m_{bb} = 103$ GeV and $m_{llbb} = 453$ GeV. The fitted relative widths are $\sigma_{bb} = 17\%$ and $\sigma_{llbb} = 14\%$.

3.3.1 Deriving the ellipses parameters

The signal simulations have been used to derive a map of the ellipses shapes as a function of the mass. For each of the 300 samples, the events passing the selection are plotted on the m_{bb}, m_{llbb} mass plane where the x-axis is m_{bb} and the y-axis m_{llbb}. The procedure to define the optimal ellipse shape for each sample is the following:

1. For each signal hypothesis, a (1D) fit of the each invariant mass is done to determine the center of the signal distribution in the mass plane (see Fig. 3.2). The reconstructed masses m_{bb} and m_{llbb} can be shifted compared to the initial masses of the resonance m_A and m_H. This effect is especially pronounced at high mass and in the forward region where the efficiency is low.

2. The 1D fits are used to determine the center of the 2D signal distribution. A two-dimensional Gaussian function is then fitted around that

center to the signal distribution in the (m_{bb},m_{llbb}) mass plane. The variance and covariance of the fit are extracted yielding a covariance matrix of the form:

$$\begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{yx}^2 & \sigma_y^2 \end{pmatrix} \tag{3.6}$$

3. The matrix is then diagonalised thanks to a rotation of angle θ:

$$\begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix} \tag{3.7}$$

The three parameters a, b and θ will be taken respectively to be the semi-major axis, the semi-minor axis and the angle of inclination of the ellipse centred in m_{bb}, m_{llbb}. This procedure is iterated on all mass samples and the values of a, b and θ are saved to form a primary map as a function of m_{bb} and m_{llbb}.

That parameter map is used to derive the shape of an ellipse located anywhere in the mass plane by interpolation. Because of ROOT interpolation algorithm (based on Delaunay triangulation) the map has to be extended along the edges to avoid interpolation errors. This procedure is schematised on Fig. 3.3: the primary map is symmetrised along the y=x axis and the points close to the edges copied to $m_{bb} = 0$ and $m_{llbb} = 1000$. The final parameter maps of a, b and θ are shown on Fig. 3.4.

When fitting the reconstructed masses for each sample, the observed relative widths σ_{bb} and σ_{llbb} are also extracted and saved as a map of the resolution as a function of m_A, m_H. They will be used to derive the pavement as explained in the next subsection.

3.3.2 Paving definition

Once the ellipses shapes can be computed anywhere in the parameter space, their positions and an eventual dilatation coefficient must be defined.

Dilation coefficient

The parameters a and b computed previously define ellipses containing one standard deviation of signal (if it were Gaussian). But signals are not Gaussian and less than 60% of signal events are generally contained in an ellipse. A dilatation coefficient can be applied to the ellipses (axis are scaled uniformly) in order for it to contain more signal. The evolution of the signal

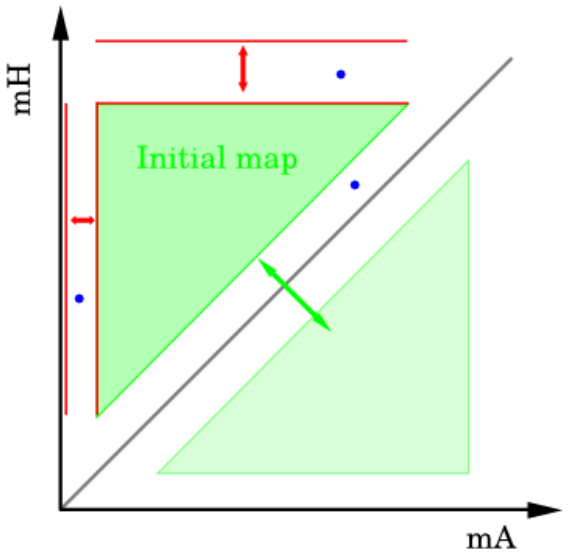


Figure 3.3: The signal samples are used to compute an initial map (upper green triangle) but the value of the ellipse parameters cannot be computed correctly for the points in blue, therefore the map is extended at low m_A and high m_H by coping the edges (in red). The entire map is copied to the the lower part of the mass plane in order to extend the coverage to the diagonal region.

percentage contained in an ellipse versus dilatation coefficient ρ if shown on Fig. 3.5 for a few signal examples.
The expansion coefficient has been chosen to be $\rho = 2$. It leads to the acceptance shown on Fig. 3.7 which ranges from 50% for high masses to around 30% for $m_{bb} \in [50, 400]$ GeV approximately. In the forward region the acceptance drops to 10% which is expected. While this choice for ρ seems reasonable, there was no attempt to optimise this parameter in order to gain in significance.

Ellipse center

Now that the shape and size of the ellipses are fixed, their position must be defined. Many different paving options have been tried and the one kept is shown on Fig. 3.6. The iteration used to derive it is the following:

- $m_{bb} \in [30, m_{llbb} - 50]$ GeV
- $m_{llbb} \in [120, 1000]$ GeV
- $m_{bb}^{i+1} = m_{bb}^{i}(1 + \sigma_{bb}\rho)$
- $m_{llbb}^{i+1} = m_{llbb}^{i}(1 + \sigma_{llbb}\rho)$

where $\sigma_{bb/llbb}$ is the real width of the resonances derived from the earlier fits.

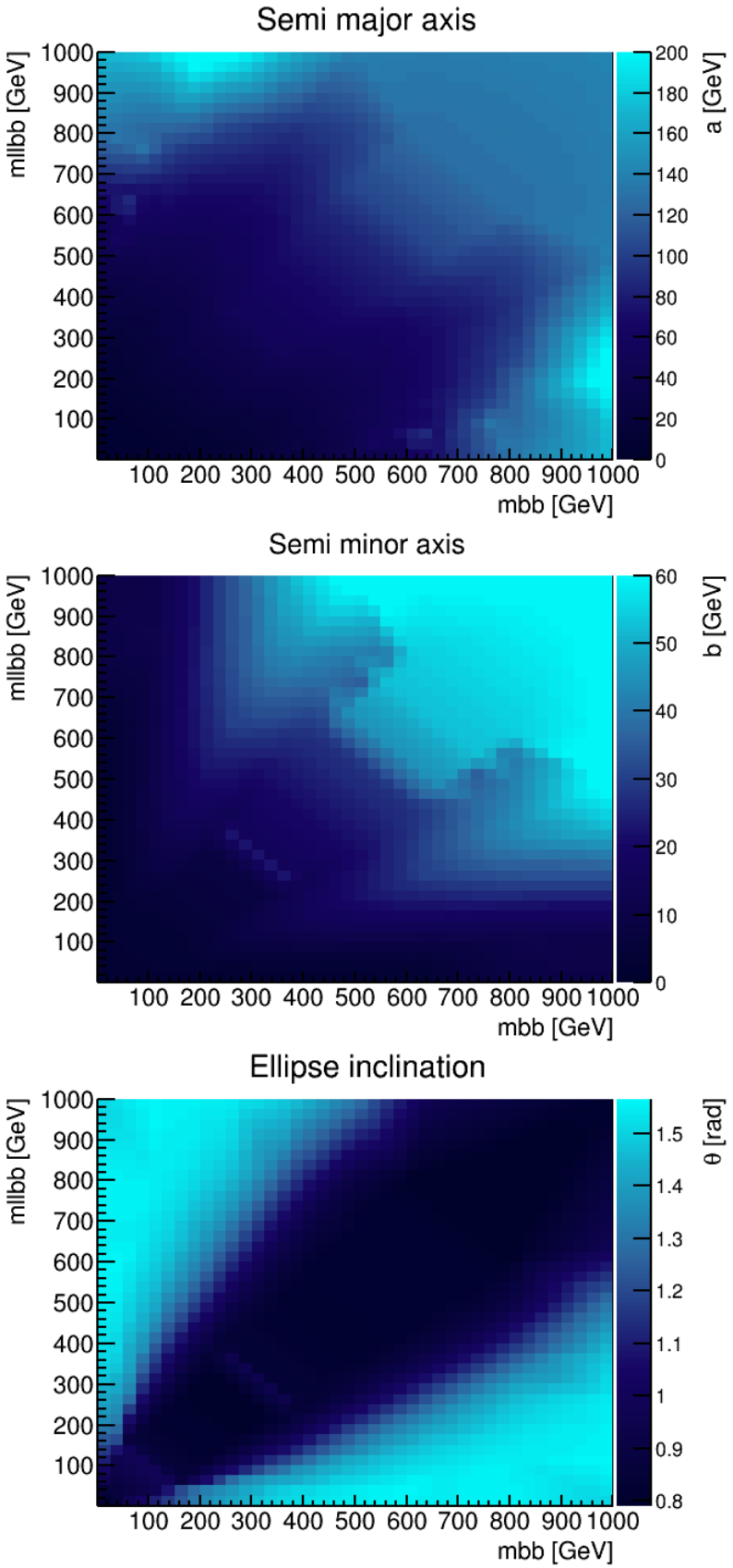


Figure 3.4: Final parameter maps of a, b and θ defining the shape of an ellipse anywhere in the m_{bb}, m_{llbb} mass plane.

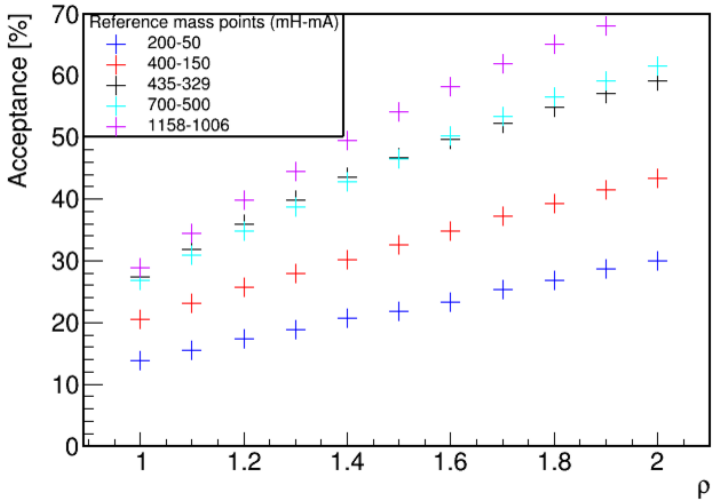


Figure 3.5: Evolution of the percentage of signal contained in the optimal ellipse versus dilatation coefficient ρ for different signal hypothesis.

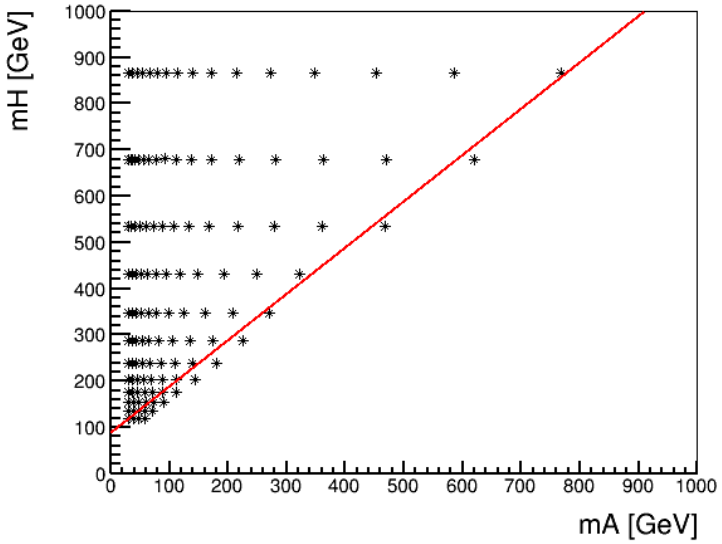


Figure 3.6: Position of the elliptical bins centres in the m_A m_H mass plane. For a better visualisation, the shapes of the ellipses themselves are represented on Fig. 3.19 in annex.

The parameter used to chose the paving is the coverage of the phase space: all points must fall in at least one ellipse. But the pavement shouldn't be too dense in order to keep the process time low when evaluating the limits as well as to avoid the look-elsewhere effect. The results of the coverage tests for the chosen paving are shown on figure 3.8: the region of interest of the m_{bb}-m_{llbb} mass plane is fully covered by the elliptical bins and the overlap between bins is reasonable.

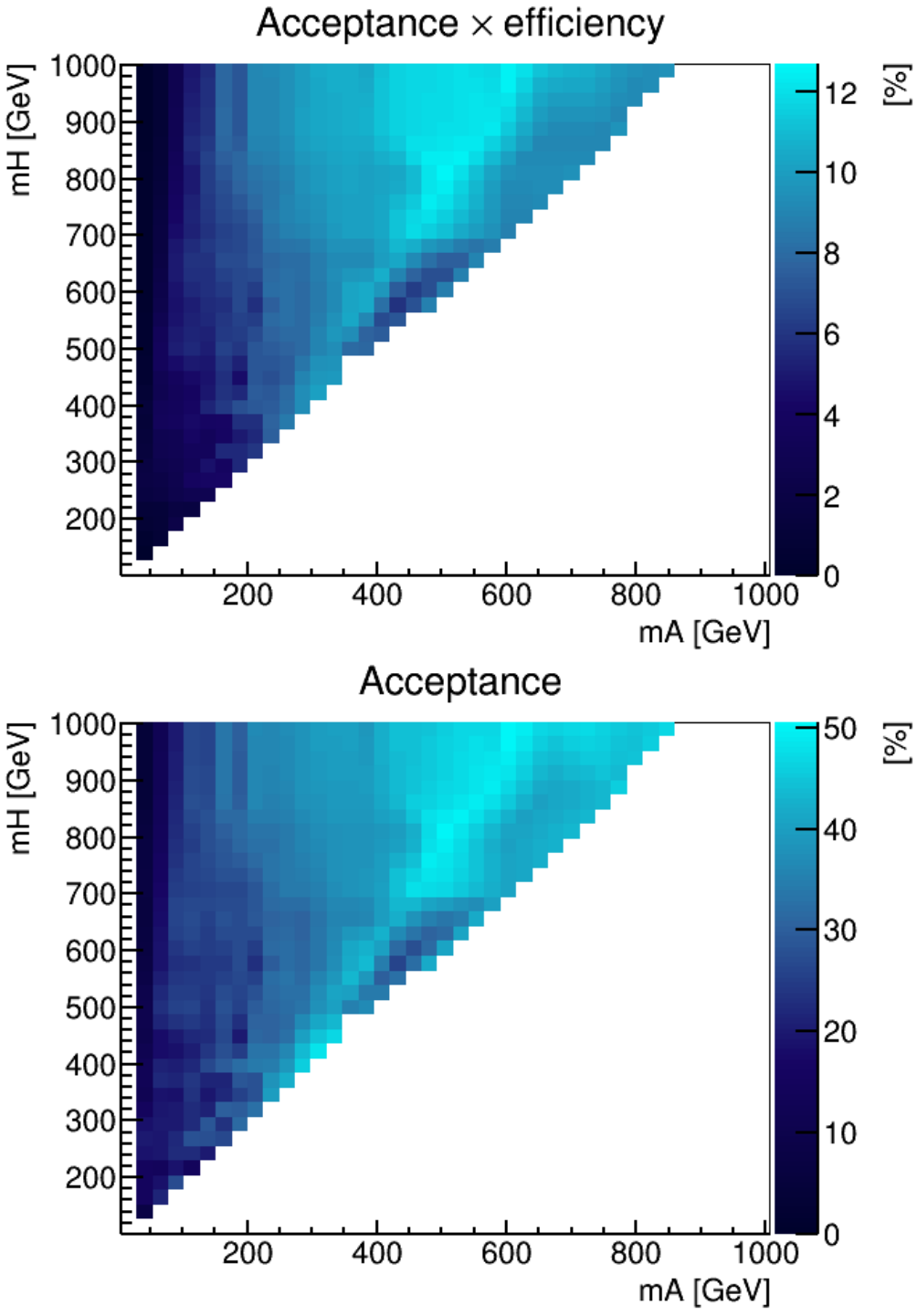


Figure 3.7: Top: signal acceptance × efficiency as a function of m_A, m_H. Bottom signal acceptance alone as a function of m_A, m_H. The lower acceptance in the low mass region is due to the important distribution tails.

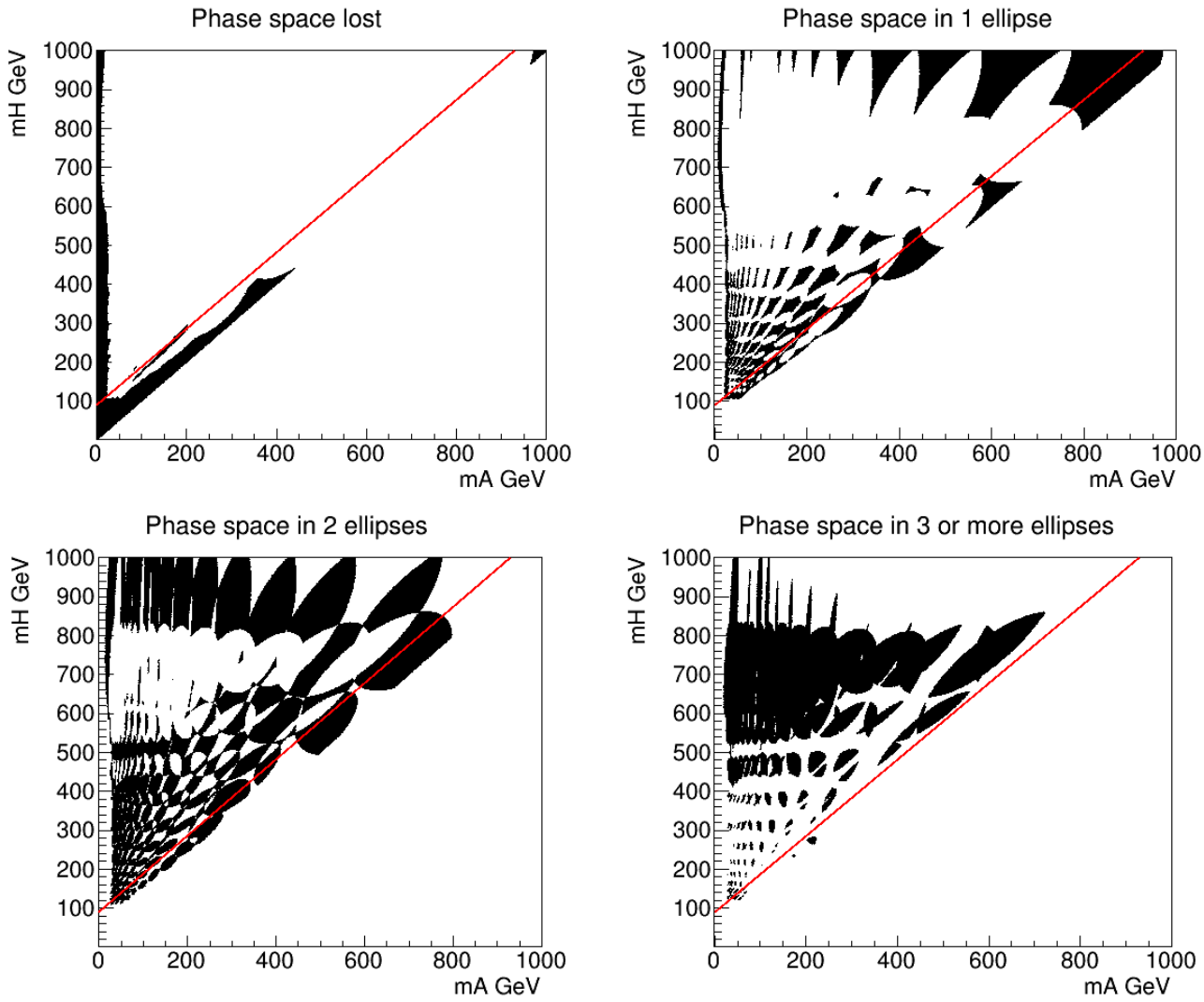


Figure 3.8: Top left plot: points of the phase space falling into no elliptical bin (in black), note that they are situated in the forward region where the analysis efficiency is zero and in the kinematically forbidden region. Top right: points of the phase space falling into exactly one elliptical bin. Bottom left: points falling into exactly two elliptical bins. Bottom right: points falling into at least three elliptical bins.

3.4 Background Estimation

3.4.1 General method

In this analysis, the background is evaluated directly from data. The idea is to count the number of events in a control region in order to estimate the background contribution in a given signal region. Here the amount of background contained in each elliptical bin is estimated by defining a ring-shaped control region as shown on Fig 3.9.

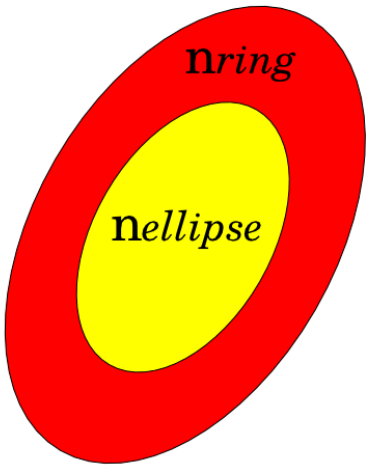


Figure 3.9: For a given bin, the signal region is defined as the elliptical bin (in yellow) while the control region is the elliptical ring surrounding it (in red). The control region is defined to have the same area as the signal region. The number of events observed in the signal region is noted $n_{ellipse}$ and the number of events observed in the control region is noted n_{ring}.

The true number of signal events contained in the ellipse is noted n_S and the true number of background events contained in the ellipse is n_{B1}. They are related as: $n_{ellipse} = n_S + n_{B1}$. The true number of background events contained in the control region is noted n_{B2}.

The estimation of the background in the signal region is based on two hypothesis:

- The signal is entirely contained in the signal region, i.e:

$$n_{B2} = n_{ring} \tag{3.8}$$

- The background is linear over the signal and control region and n_{B1} and n_{B2} are equal up to a proportional coefficient τ:

$$n_{B1} = \tau n_{B2} \tag{3.9}$$

Combing the two hypothesis yields $n_{B1} = \tau n_{ring}$. Knowing τ, the number of signal events can then be estimated from the two observables $n_{ellipse}$ and n_{ring} by:

$$n_S = n_{ellipse} - \tau n_{ring} \tag{3.10}$$

3.4.2 Application

Since the control regions were taken to be of equal area, one could expect to have $\tau = 1$. But this is not the case for two reasons:

- The background is not linear and taking signal and control regions of the same areas do not ensure $n_{B1} = n_{B2}$.
- Important border effects have to be taken into account: the ellipses situated near the mass plane edges can span partially over negative mass regions (non-physical and empty) so do the control regions – in bigger proportions.

Hopefully these two points can be taken into account: the background distribution is known and the border effect is of geometrical nature and both effects can be estimated. For each bin, the coefficient τ has been computed. The procedure is the following: Monte Carlo samples of background processes defined in section 3.2.2 have been used to measure $n_{ellipse}$ and n_{ring} to compute τ as:

$$\tau = \frac{n_{ellipse}}{n_{ring}} \tag{3.11}$$

The coefficient τ was computed as a function of m_A, m_H for different background processes separately. The measured values didn't depend much on the nature of the background process (see Figures 3.16-3.18 in annex). The ellipse position in the m_A-m_H mass plane in contrary do influences the value of τ . The final map of τ was derived using both the $t\bar{t}$ and $Z + jets$ samples combined to increases statistics. The results is shown of Fig. 3.10: τ varies from 0.4-0.5 in the high m_H low m_A region (due to the shapes of the bins in that area) to 1.2-1.1 for a few specific bins at low mass close to the diagonal. Appart from these bins, the coefficient stays close to 1 for $m_H < 500$GeV.

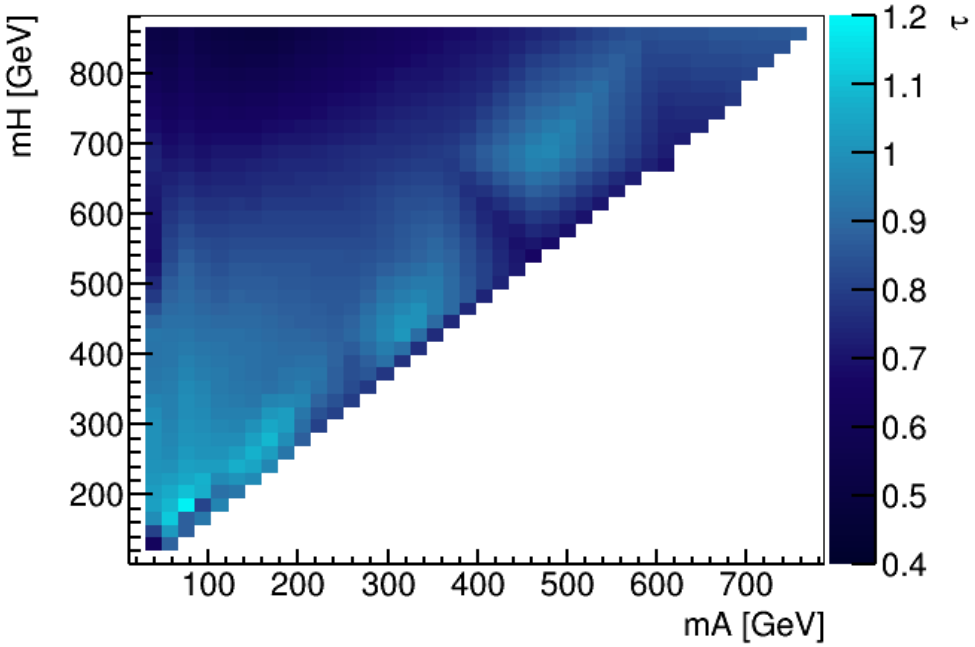


Figure 3.10: Map of the τ coefficient. Its value is close to one in the high statistics regions where the background is uniform and even above one where the background is peaked. The especially low values at low m_A and high m_H are explained by the border effect.

Thanks to that coefficient map derived from simulations, the signal contained in a signal region in data can be computed by:

$$n_S^i = n_{ellipse}^i - \tau^i n_{ring}^i \tag{3.12}$$

where i stands for the ellipse number and $n_{ellipse}$ and n_{ring} are the number of data events observed in the signal and control regions respectively.

3.5 Systematics

Systematic uncertainties affecting the computed number of signal events n_s and the estimated background n_B have been derived and are listed in Table 3.4. The first six uncertainties have been reused from the previous analysis [5]. The derivation of the uncertainties on τ is exposed in the section 3.5.1

The *luminosity* uncertainty affects the overall normalisation of the signal and of the background. The *lepton reconstruction and trigger efficiency*, accounts for the uncertainties on the lepton reconstruction, isolation and trigger efficiency. The *signal PDF* was derived by varying the PDF and the *signal scale* was derived by varying the renormalisation scale μ_R and the factorisation scale μ_F. The *signal efficiency* accounts for the uncertainty coming from the use of DELPHES instead of a CMS full simulation.
These uncertainties are reused since they shouldn't be too much affected by the new bin shape used.

Systematic	value
Luminosity (from [5])	2.6%
Lepton reconstruction and trigger efficiency (from [5])	3%
Signal PDF (from [5])	5%
Signal scale (from [5])	6%
Signal efficiency (from [5])	varies
Statistical error on τ	varies
Systematic error on τ	varies

Table 3.4: List of the uncertainties on the background and signal estimations

Several uncertainties haven't been taken into account in this analysis. The uncertainties on b-tagging, jet energy resolution and scale for the background were not considered. Since the background is estimated from data the influence of these uncertainties should affect the estimation as much as the data, thus having a small overall impact.
The jet energy scale should also affect the signal estimation by shifting the reconstructed masses m_{bb}, m_{llbb} compared to the generated masses m_A, m_H. But that shift shouldn't affect to much the results since the signals would simply be shifted to a adjacent bin. The jet energy resolution would affect the spread of the signal and affect the acceptance. That systematic wasn't computed because of the use of DELPHES but should be considered in future analyses.

3.5.1 Uncertainties on τ

The uncertainty on the background estimation derived from data is given by the uncertainty on τ. Two different sources of systematic uncertainties have been computed.

The first source of uncertainty on τ comes from the statistical error when computing τ. It will be noted σ_τ^{stat} and accounts for the statistical fluctuations on the number of events $n_{ellipse}$ falling into the signal region and n_{ring} the number of events falling in the control regions. It has been computed using error usual propagation and yields:

$$\sigma_\tau^{stat} = \frac{n_{ellipse}}{n_{ring}} \sqrt{\frac{1}{n_{ellipse}} + \frac{1}{n_{ring}}} \tag{3.13}$$

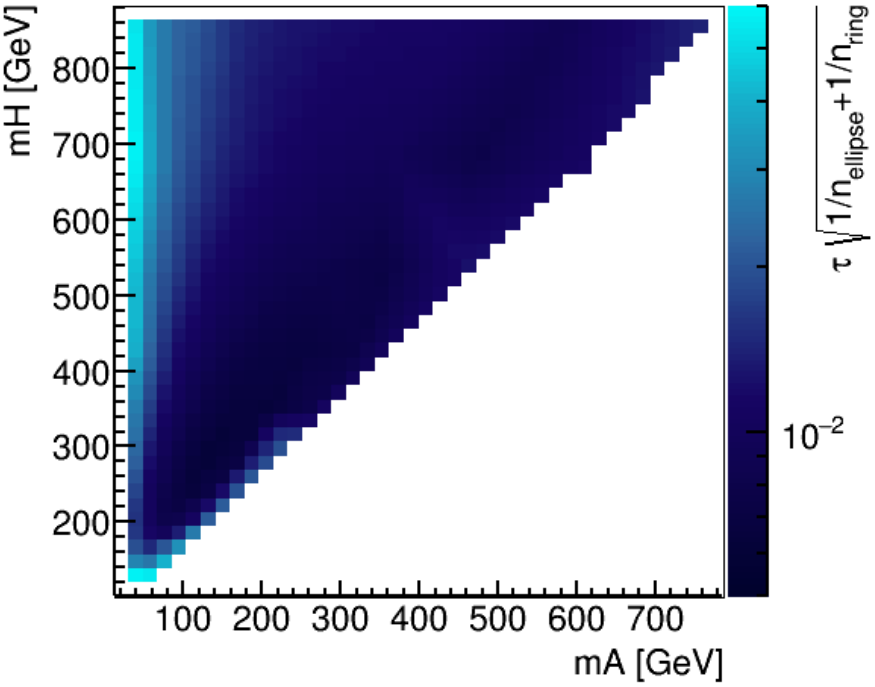
where $n_{ellipse}$ and n_{ring} are from the combination of $t\bar{t}$ and $Z + jets$ MC samples used to derive the τ map (Fig. 3.10). The map of the statistical error is shown of Fig. 3.11a.

The second source of uncertainty on τ comes from the systematic error of the method itself. It has been estimated by comparing the results obtained for τ when using different MC samples. The $t\bar{t}$ sample was used to derive the ratio τ and the DY+jets sample was used as "data" to measure $n_{ellipse}$ and n_{ring} . The error was then defined as:

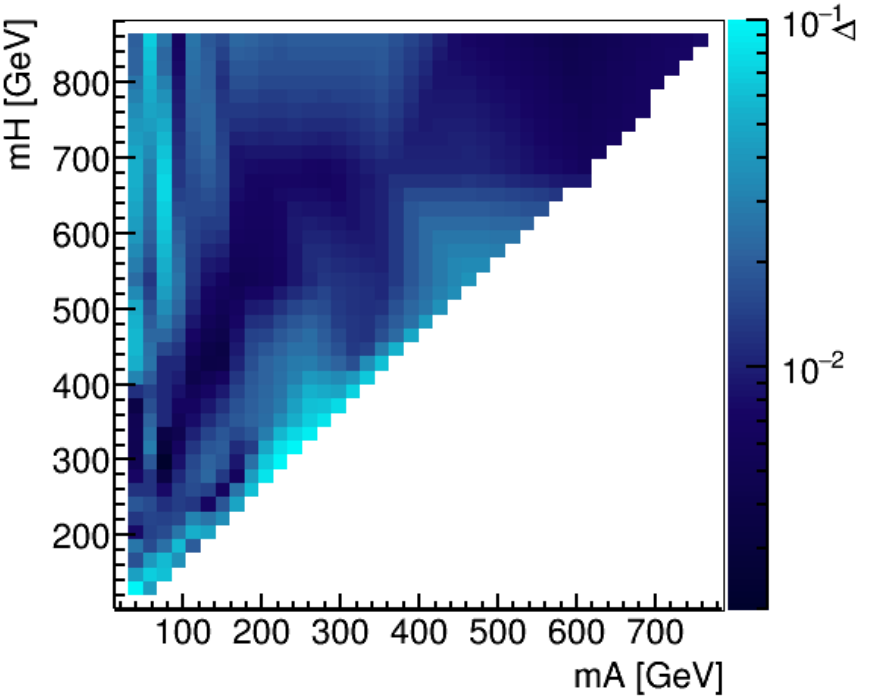
$$\Delta = \mid \frac{n_{ellipse} - \tau \; n_{ring}}{\frac{1}{2}(n_{ellipse} + n_{ring})} \mid \tag{3.14}$$

The results are shown on Fig. 3.11b.

For most of the mass plane, the dominating uncertainty is Δ. Its value stays below 5% in most of the mass plane and reaches a bit above 10% in the region m_A-m_H 250-350 GeV. The uncertainty is also higher in the forward region where it reaches 8%, this is due to the poor statistics in that area. The uncertainty σ_τ^{stat} stays below 2% in the majority of the mass plane and reaches 6% in the forward region.



(a) Map of the statistical error on τ. Note that the error is important in the forward region since the efficiency is low.



(b) Map of the systematic error on τ. The uncertainty is rather uniform excepted in the forward region where it is dominated by statistics.

3.6 Results

3.6.1 Limits

The number of observed events in each bin covering the mass plane has been computed and the compatibility of the observed data with the expectation has been computed with the COMBINE tool from CMS [46].

The significance of a potential excesses observed is given by the p-value. The p-value was computed for each bin and the results are shown on Fig. 3.12: no significant excess is observed.
The minimum p-value is observed for the bin 112-203 with a value of $p = 0.038$ corresponding to a local significance of 1.8σ. A few other excesses are observed in the forward region (where statistics are poor) and the one situated at 678-171 has a p-value of $p = 0.044$. None of these reaches a local significance of 3σ.

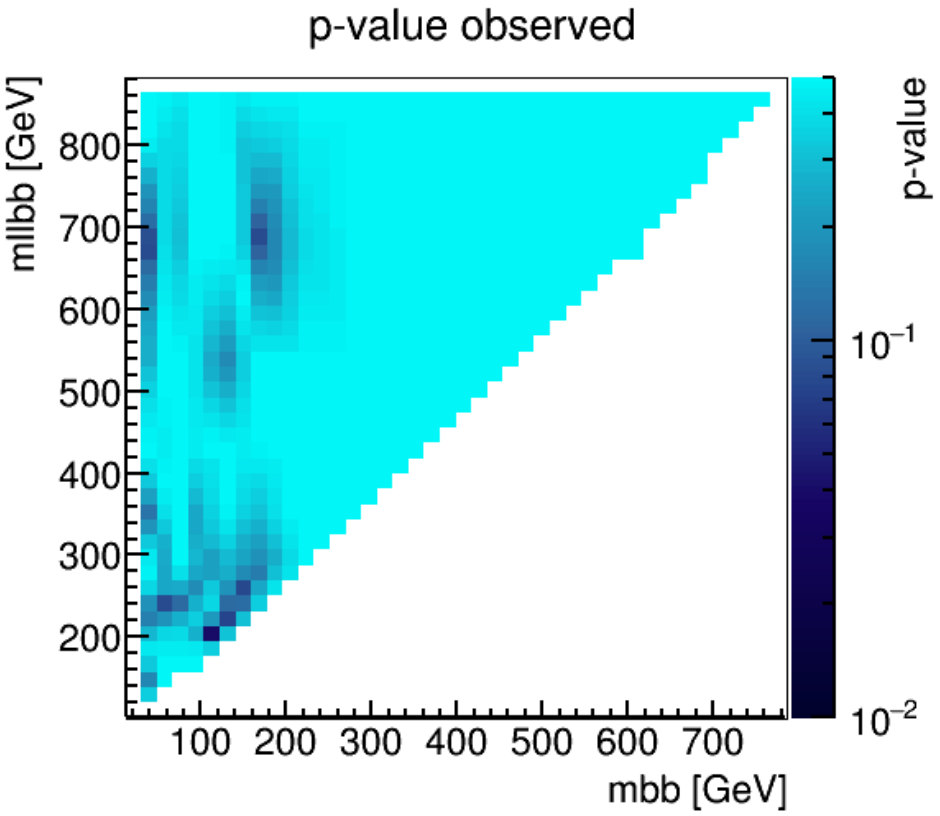
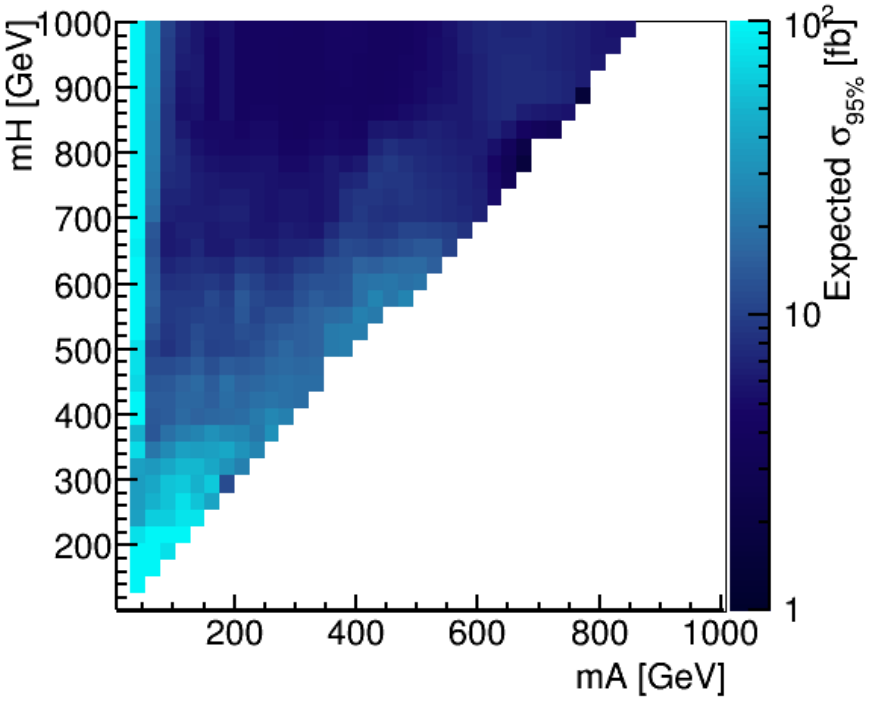
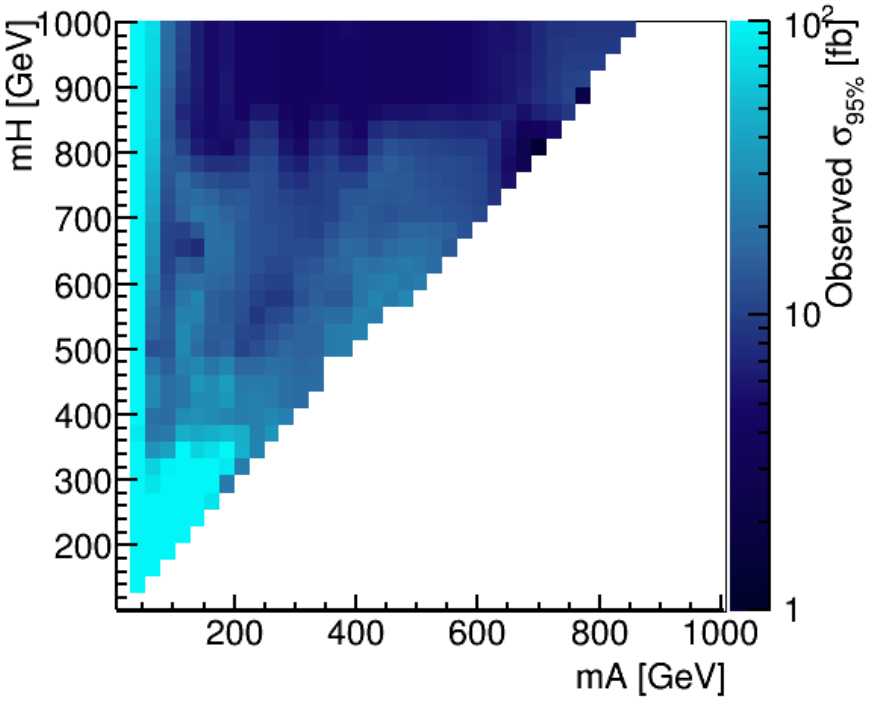


Figure 3.12: Observed p-value as a function of m_{bb} and m_{llbb}. The highest excess (bin 112-203) has a p-value of 0.038

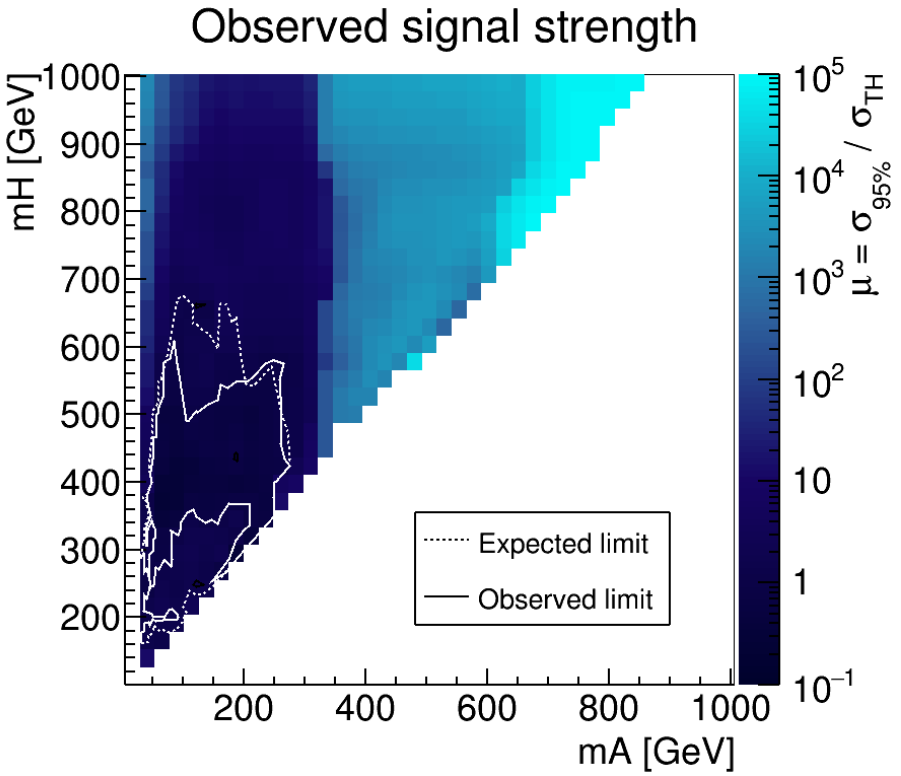


(a) Expected exclusion limits on the signal cross section as a function of m_A and m_H

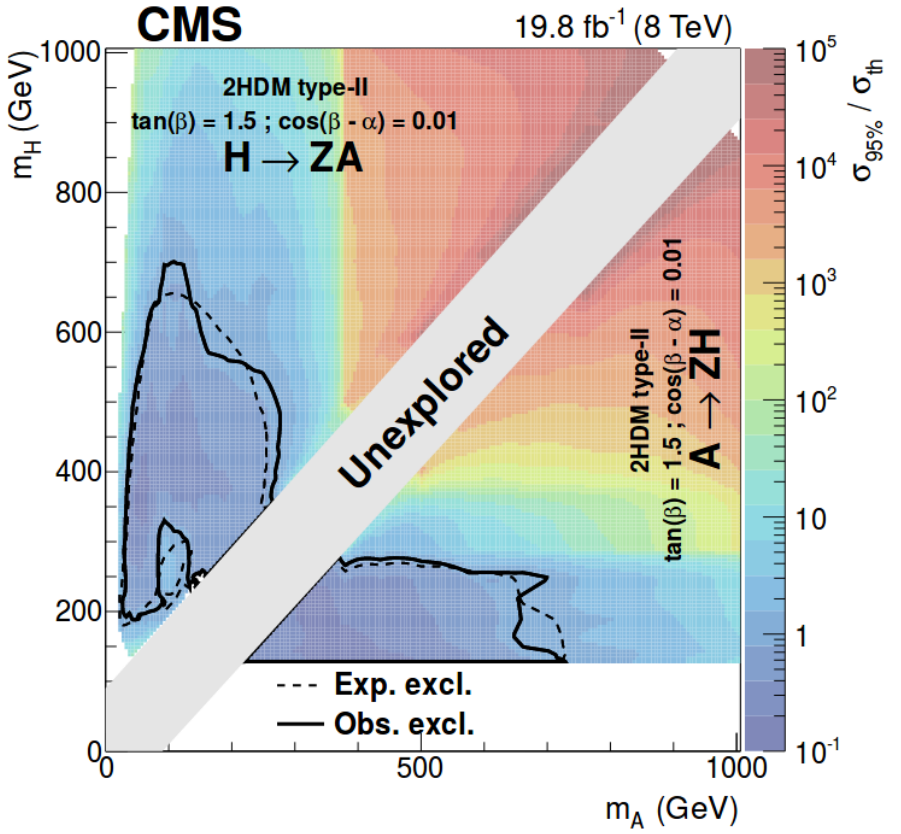


(b) Observed exclusion limits on the signal cross section as a function of m_A and m_H

Figure 3.13



(a) Observed signal strength as a function of m_A and m_H The contours show the expected limit (dashed) and observed limit (continuous) on μ.



(b) Past analysis result: expected and observed limit on μ as a function of m_A and m_H from simulation

The number of expected/observed signal events may be used to set an upper exclusion limit on $\sigma \times BR$ for a double resonance such as $H \to Z(l\bar{l})A(b\bar{b})$. The maps of the upper limit at the 95% level on the cross sections as a function of m_A and m_H are shown on Fig 3.13.
The cross section excluded varies from a 100 fb at low mass to a few fb at higher mass. The limits on the forward region where the efficiency is low is also of around 100 fb.

Finally, limits on the signal strength μ can be derived. The signal strength is defined as:

$$\mu = \frac{\sigma_{95\%}}{\sigma_{TH}} \tag{3.15}$$

with $\sigma_{95\%}$ the observed upper exclusion limit on the cross section and σ_{TH} the theoretical cross section of our specific 2HDM. The part of the mass plane where $\mu < 1$ is excluded and is showed on figure 3.14a. The expected exclusion region covers the entire low mass region ranging approximately from $m_H = 120$ GeV to $m_H = 600$ GeV and $m_A = 30$ GeV to $m_A = 300$ GeV.
The observed exclusion region is smaller and not as continuous. The region centred in 150-300 GeV was not excluded.

3.6.2 Comparison to previous analysis

The previous analysis had found two excesses for ($m_A = 575$, $m_H = 662$) GeV and ($m_A = 93$, $m_H = 256$) GeV with a local (global) significance of 2.9 (1.9) and 2.6 (1.5) respectively. The second excess was compatible with the signal hypothesis but not significant enough to separate the background-only and signal-plus-background hypotheses. This excess wasn't observed again as shown by the p-value map: the closest bin centred in 84-239 has a p-value of $p = 0.17$.

Fig.3.14b shows the exclusion limits on the signal strength from the previous analysis. The new analysis yields a better expected exclusion region covers a bigger area and including the position of the past excess. However the observed limit on signal strength covers a smaller region of the mass plane.

The shapes and positions of the bins centred on the excess are shown on Fig. 3.15. The rectangular bin spans from 72 to 114GeV in m_A and from 222 to 350GeV in m_H. The six nearby elliptical bins are drawn, the two

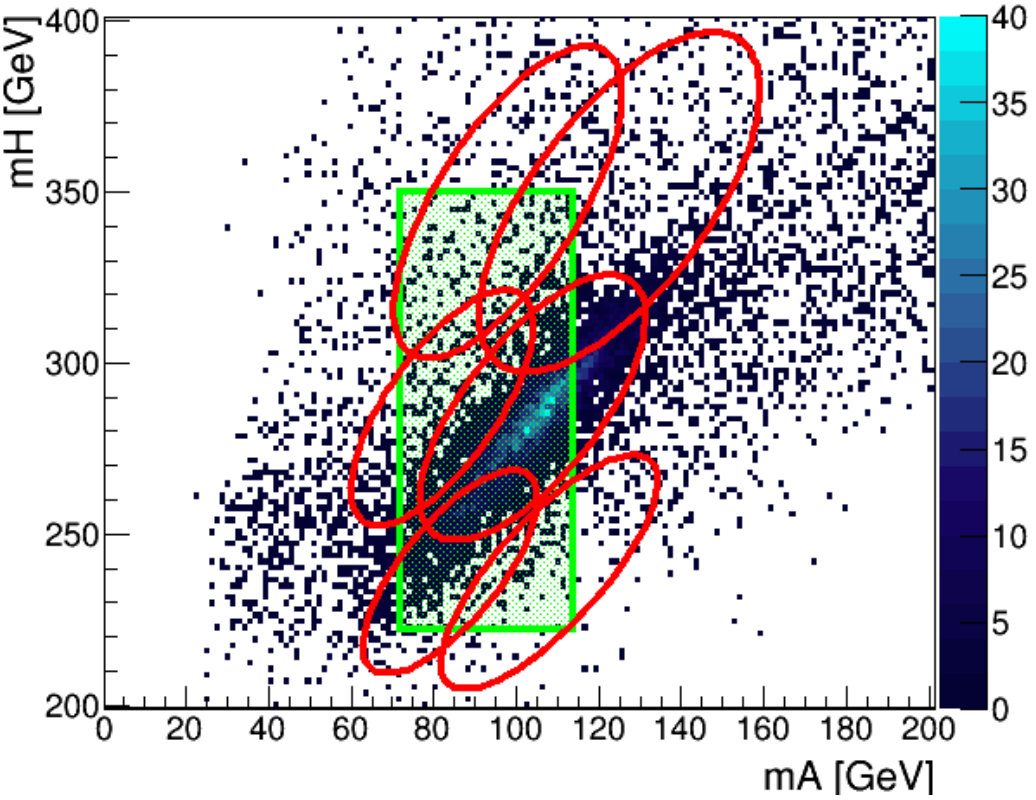


Figure 3.15: Layout of the bins of both analysis close to the maximum excess located in $m_A = 93$GeV, $m_H = 256$GeV. The signal shown was generated for $m_A = 104$GeV and $m_H = 270$GeV.

closest bins are centred in 84-287 and 104-287 GeV respectively. Their semi major(minor) axis are of 40(15.3) and 44(17.4) GeV. The new bins are clearly smaller than the old rectangular one which should increase the $\frac{s}{\sqrt{b}}$ ratio if an excess is well centred in an ellipse. But if an excess is centred between elliptical bins, it is possible that its significance would decrease. A denser paving, especially in m_H could be considered.

Conclusion

Two Higgs Doublets Models are compelling extensions of the Standard Model and in the alignment limit, the exotic decay $H \to ZA \to l\bar{l}b\bar{b}$ is enhanced. The search for this decay was carried out using CMS run I data at 8TeV corresponding to an integrated luminosity of 19.8fb^{-1}.

The m_{bb}-m_{llbb} mass plane was scanned for a double resonance using a new analysis strategy. It involved elliptical bins to take advantage of the signal correlation in order to increase the signal over background ratio. The varying ellipses shapes were derived from signal simulations to cover the mass plane optimally. The background was estimated using a new data driven method. No excess was found and a generic limit on the signal cross section was set: it varies from a few fb to 100fb depending on the region of the mass plane. This limit may be interpreted in other models. Exclusion limits were set on the signal strength for a type II 2HDM of inverted mass spectrum and compared to the results of a previous analysis. The analysis sensitivity has been improved and the expected exclusion region was extended. The excess of local significance of 2.6 previously of found in $m_A = 93$GeV, $m_H = 256$GeV was not observed again.

To improve the sensitivity of the analysis further, several points should be considered. First a better estimation of the systematic uncertainties would be necessary since some of the uncertainties have been neglected or derived for the previous analysis. The use of full simulations instead of DELPHES to generate signal samples would also suppress some uncertainties. Then the dilation coefficient ρ defining the ellipse size as well as the size of the control region could be optimised in order to obtain the best significance. Next, new developments in the simulation of the Drell-Yann+jets process should be taken into account. The present data-driven background estimation could eventually be improved by the implementation of a shape analysis. Finally, the analysis of the CMS run II data-set with much higher statistics would allow to set even more stringent limits on the signal cross section.

Bibliography

[1] G. Aad et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012.

[2] S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.

[3] Georges Aad et al. Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment. *Eur. Phys. J.*, C76(1):6, 2016.

[4] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J.*, C75(5):212, 2015.

[5] Vardan Khachatryan et al. Search for neutral resonances decaying into a Z boson and a pair of b jets or tau leptons. *Phys. Lett.*, B759:369–394, 2016.

[6] Guy Moore Cliff Burgess. *the Standard Model, a primer*. Cambridge university press, 2007.

[7] Michael E. Peskin and Daniel V. Schroeder. *An Introduction To Quantum Field Theory*. Westview Press Incorporated, 1995.

[8] Donald H. Perkins, Donald H. Perkins, and Donald H. Perkins. *Introduction to high energy physics*. Cambridge University Press, 2012.

[9] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-higgs-doublet models. *Physics Reports*, 516(1–2):1 – 102, 2012. Theory and phenomenology of two-Higgs-doublet models.

[10] C. Patrignani et al. Review of particle physics. *Chin. Phys.*, C40(10):100001, 2016.

[11] Alex Pomarol and Roberto Vega. Constraints on cp violation in the higgs sector from the parameter. *Nuclear Physics B*, 413(1):3 – 15, 1994.

[12] J. M. Gerard and M. Herquet. A twisted custodial symmetry in the two-higgs-doublet model. *Phys. Rev. Lett.*, 98:251802, 2007.

[13] Mayumi Aoki, Shinya Kanemura, Koji Tsumura, and Kei Yagyu. Models of yukawa interaction in the two higgs doublet model, and their collider phenomenology. *Phys. Rev. D*, 80:015017, Jul 2009.

[14] Alexandre Mertens. *Search for 2HDM extensions of the scalar sector close to the alignment limit with the CMS detector*. PhD thesis, Université catholique de Louvain, 2017.

[15] G. Aad et al. Measurements of higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 tev with the atlas detector. *Phys. Rev. D*, 91:012006, January 2015.

[16] S. et all Chatrchyan. Measurement of the properties of a higgs boson in the four-lepton final state. *Phys. Rev. D*, 89:092007, May 2014.

[17] G. Aad et al. Measurement of higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 tev with the atlas detecto. *Phys. Rev. D*, 90:112015, 2014.

[18] V. et al Khachatryan. Observation of the diphoton decay of the higgs boson and measurement of its properties. *The European Physical Journal C*, 74(10):3076, 2014.

[19] G. Aad et al. Observation and measurement of higgs boson decays to ww with the atlas detector. *Phys. Rev. D*, 92:012006, 2015.

[20] S. Chatrchyan and etal. Measurement of higgs boson production and properties in the ww decay channel with leptonic final states. *Journal of High Energy Physics*, 2014(1):96, 2014.

[21] G.and etal Aad. Evidence for the higgs-boson yukawa coupling to tau leptons with the atlas detector. *Journal of High Energy Physics*, 2015(4):117, 2015.

[22] S. Chatrchyan and etal. Evidence for the 125 gev higgs boson decaying to a pair of τ leptons. *Journal of High Energy Physics*, 2014(5):104, 2014.

[23] ATLAS Collaboration. Evidence for the spin-0 nature of the higgs boson using atlas data. *Phys. Lett. B*, 726 120–144, 2013.

[24] CMS Collaboration. Constraints on the spin-parity and anomalous hvv couplings of the higgs boson in proton collisions at 7 and 8 tev. *Phys. Rev. D*, 92 no. 1 ,012004, 2015.

[25] ATLAS Collaboration. Constraints on the off-shell higgs boson signal strength in the high-mass zz and ww final states with the atlas detector. *Eur. Phys. J. C*, 75,no. 7, 335, 2015.

[26] CMS Collaboration. Constraints on the higgs boson width from off-shell production and decay to z-boson pairs. *Phys. Lett. B*, 736 64–85, 2014.

[27] CMS Collaboration. Summary results of high mass BSM Higgs searches using CMS run-I data. Technical Report CMS-PAS-HIG-16-007, CERN, Geneva, 2016.

[28] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

[29] Adrien Caudron. *The final state with two b jets and two leptons at the LHC as a probe of the scalar sector*. PhD thesis, Université catholique de Louvain, 2016.

[30] G. C. Dorsch, S. J. Huber, K. Mimasu, and J. M. No. Echoes of the Electroweak Phase Transition: Discovering a Second Higgs Doublet through $A_0 \to ZH_0$. *Physical Review Letters*, 113(21):211802, November 2014.

[31] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[32] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[33] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011.

[34] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, 2014.

[35] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987.

[36] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.

[37] CMS Collaboration. CMS physics: Technical design report. 2006.

[38] CMS collaboration. Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev. *Journal of Instrumentation*, 12(02):P02014, 2017.

[39] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.

[40] The CMS collaboration. Identification of b-quark jets with the cms experiment. *Journal of Instrumentation*, 8(04):P04013, 2013.

[41] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, 2014.

[42] J. Allison and etal. Amako. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278, 2006.

[43] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. Delphes 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2):57, 2014.

[44] Kyle Cranmer. Practical Statistics for the LHC. pages 267–308, 2015.

[45] A L Read. Modified frequentist analysis of search results (the CL_s method). (CERN-OPEN-2000-205), 2000.

[46] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011.

[47] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[48] S. Jadach, Z. Was, R. Decker, and J.H. Kühn. The τ decay library tauola, version 2.4. *Computer Physics Communications*, 76(3):361 – 380, 1993.

[49] David Eriksson, Johan Rathsman, and Oscar Stal. 2hdmc – two-higgs-doublet model calculator. *Computer Physics Communications*, 181(1):189 – 205, 2010.

[50] Robert V. Harlander, Stefan Liebler, and Hendrik Mantler. Sushi: A program for the calculation of higgs production in gluon fusion and bottom-quark annihilation in the standard model and the {MSSM}. *Computer Physics Communications*, 184(6):1605 – 1617, 2013.

Annex

τ maps for the separate backgrounds

The maps of the extrapolation coefficient τ have been computed for three different background samples: $t\bar{t}$, DY+jets and DY+b.

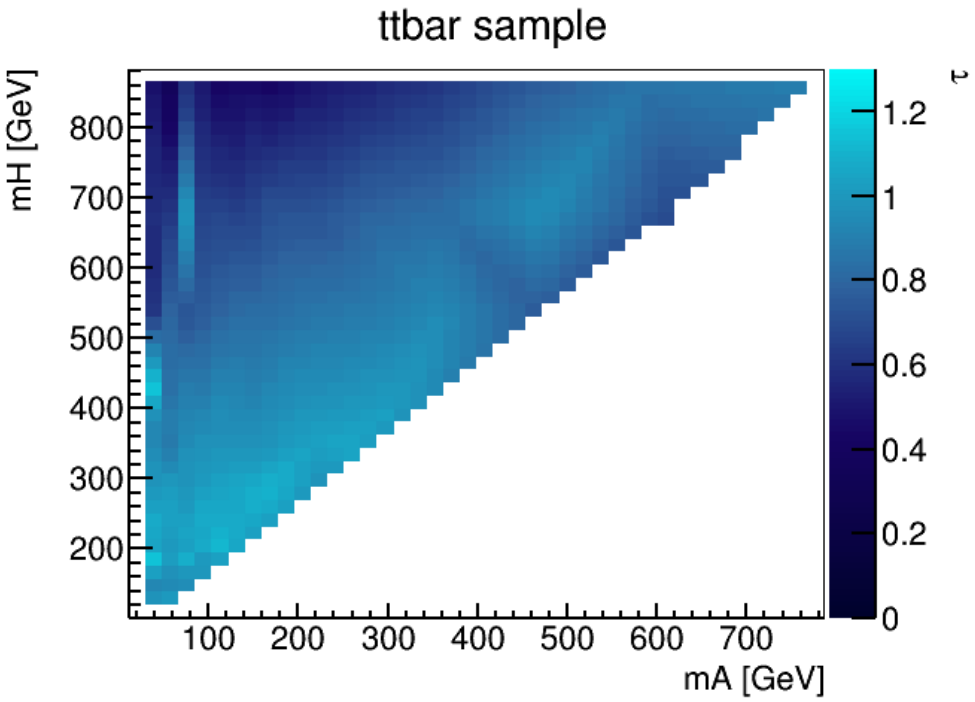


Figure 3.16: Map of the τ coefficient for $t\bar{t}$ events only

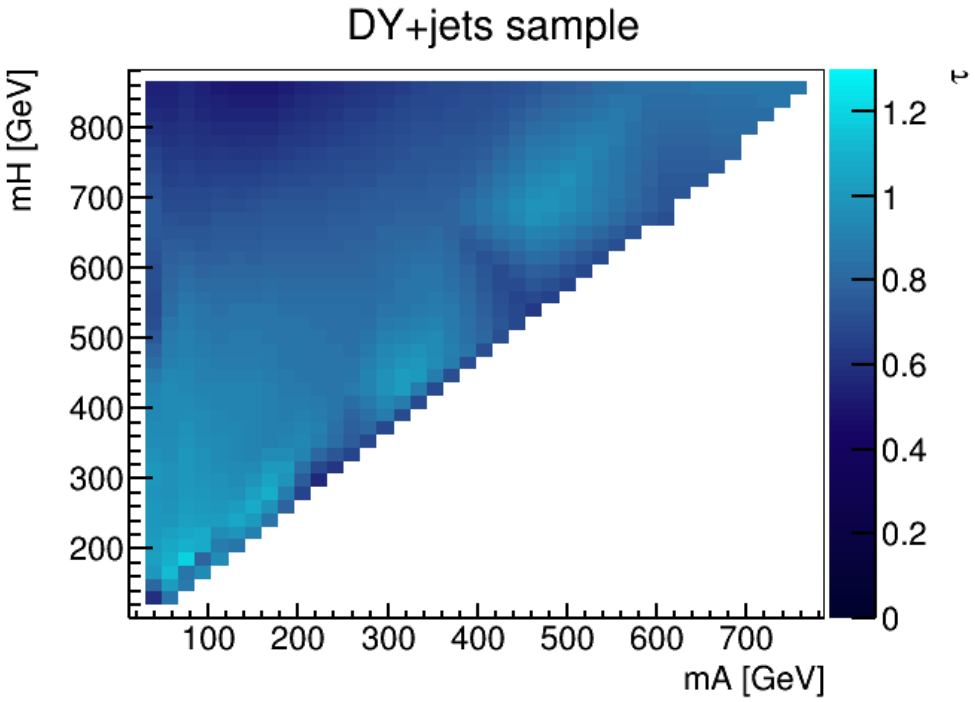


Figure 3.17: Map of the τ coefficient for DY+jets events only

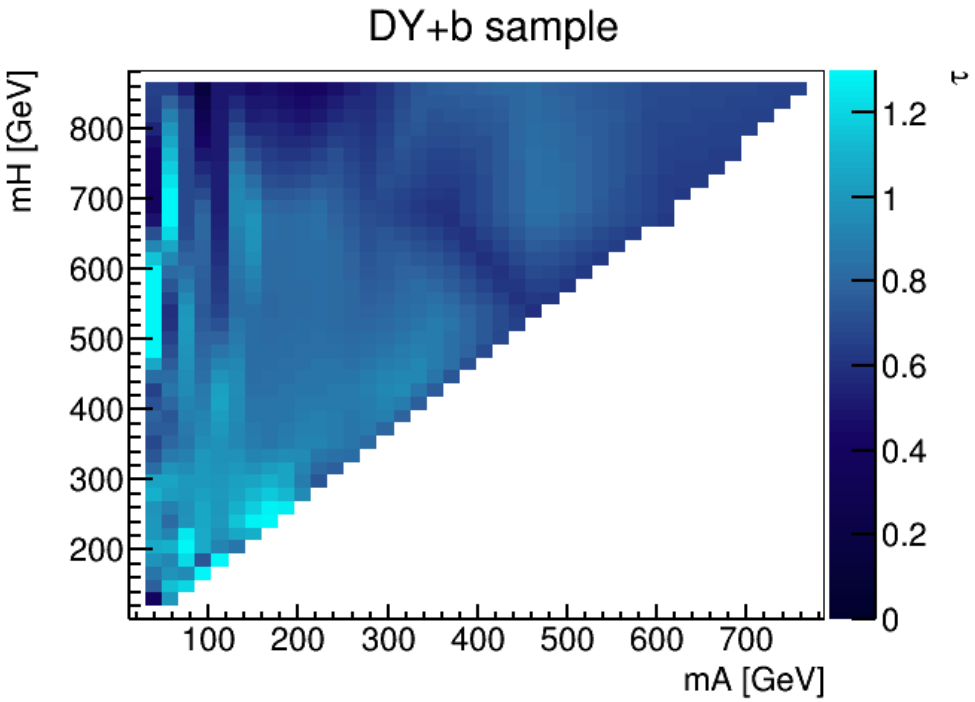


Figure 3.18: Map of the τ coefficient for DY+b events only. Note that the statistic of this sample is much lower than the two previous ones which explains the higher fluctuations in the forward region

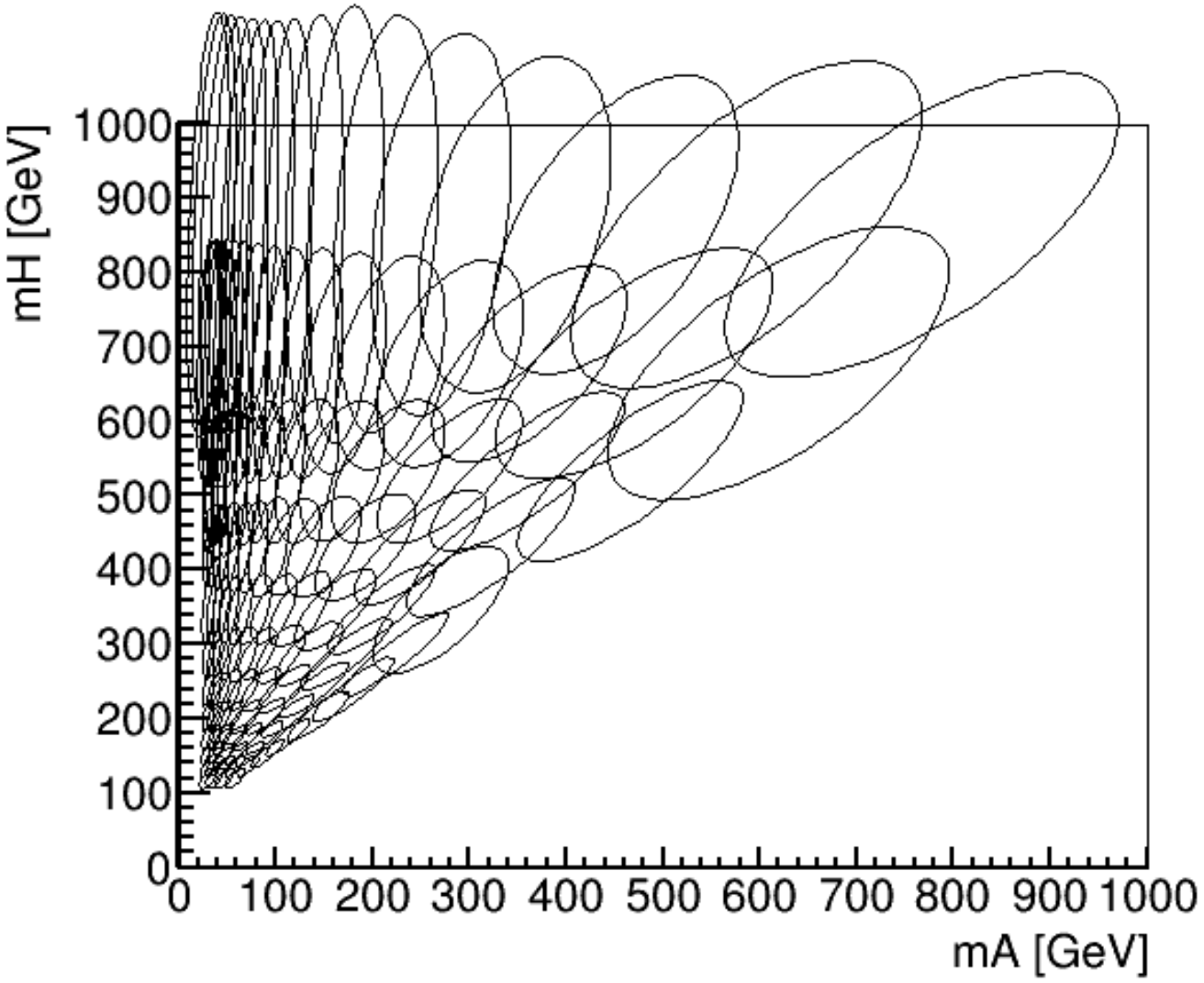


Figure 3.19: Ellipse pavement covering the mass plane