Université catholique de Louvain
Secteur des Sciences et Technologies
Institut de Recherche en Mathtématique et Physique
Centre for Cosmology, Particle Physics and Phenomenology

Search for 2HDM extensions of the scalar sector close to the alignment limit with the CMS detector.

Doctoral dissertation presented by

Alexandre MERTENS

in fulfillment of the requirement for the degree of Doctor in Sciences

Jury de thèse

Pr. Christophe DELAERE (*Advisor*)	UCL, Belgium
Pr. Vincent LEMAÎTRE (*Chairman*)	UCL, Belgium
Pr. Fabio MALTONI	UCL, Belgium
Dr. Muriel VANDER DONCKT	IPNL, France
Dr. Maria CEPEDA	CERN, Switzerland

Remerciements

Pour commencer, je tiens à remercier chaleureusement Christophe Delaere, mon promoteur, pour m'avoir offert l'opportunité de réaliser cette thèse. Sa disponibilité et son expérience m'ont permis de progresser et de résoudre de nombreux problèmes. Son engagement et sa motivation pour la recherche expérimentale ont été, pour moi, source d'inspiration.

Ensuite, je tiens à remercier les membres du jury. En particulier, Vincent Lemaitre pour son esprit critique. Les nombreuses discussions qui ont eu lieu lors de nos meetings hebdomadaires ont été fortement enrichissantes. Merci également à Muriel Vander Donckt, qui après avoir partagé avec moi sa passion pour la physique expérimentale lors de stages à Lyon, a accepté de faire partie de mon jury. Enfin, merci à Fabio Maltoni et Maria Cepeda pour leurs commentaires sur le texte qui ont été très constructifs.

La vie de tous les jours au CP3 a été particulièrement agréable grâce à de nombreux collègues que je me dois de remercier. J'aimerais commencer par Michele S. avec qui j'ai partagé le bureau durant quatre années. Ses qualités de scientifique, sa patience et sa bonne humeur ont rendu notre collaboration particulièrement motivante et productive. D'ailleurs, le projet DELPHES sur lequel nous avons travaillé ensemble est pour moi la plus grande réussite de cette thèse. J'en profite pour remercier également Pavel pour sa contribution importante et ses nombreux conseils.

Analyser des données de CMS se fait difficilement seul, j'ai eu la chance de travailler avec de nombreux collègues de qualité. Adrien C. a joué un rôle majeur

dans l'analyse des données de 2012. Il m'a appris énormément sur CMSSW et sur les méthodes de b-tagging, j'ai particulièrement apprécié notre collaboration. Thanks to Roberto and Tristan for their help and the numerous discussions about 2HDM theory and results. C'était un plaisir de travailler avec Loic et Lucia pour combiner nos analyses. Avec Simon, nous avons analysé les données de 2015. Un énorme merci pour le partage de son expérience et pour toutes les longues soirées et nuits de travail, notamment celle pour rattraper mon oubli de changer les valeurs de b-tagging, la veille de notre approval... Nous avons fortement bénéficié du travail de l'équipe HH. D'ailleurs, merci à Alessia, Arnaud, Brieuc, Ludivine, Martin, MigWel, Olivier, Seb et Seb pour les conseils en CMSSW et analyses de données, pauses cafés, discussions de stats, longues discussions sur la méthode des éléments de matrice, afterwork, etc. (Choisissez ce qui vous convient)

Merci aussi à vous pour ces temps de midi, accompagnés de Jerôme, Chiara, Christophe R. *en* Pieter. C'était, chaque jour, un moment agréable et stimulant.

Merci aussi Carine, Carinne et Ginette pour le soutien administratif.

Ensuite, ce travail n'aurait pas été possible sans les amis avec qui j'ai partagé une maison, des vacances, des week-ends, des soirées, des balades ornithos, des matchs de tennis ou des entrainements de kung-fu. Un énorme merci à Adri T., Alban J., Antoinette J., Bob V., David F., Gauthier D., Jeremy R., Jerôme M., Julie C., Laurent F., Olivier d..., Pierre-Yves A., Thérèse T., Sylvain T., Thom F., Violaine S., Virginie D. Nos échanges m'enrichissent chaque jour.

À ma filleule, Camille, qui, je l'espère, saura très vite lire.

Je voudrais également remercier tous les membres de ma famille, qui m'ont toujours soutenu dans mes choix. Merci à mon frère, mes grand-parents, tontons, taties et cousins. Merci aussi à la famille Lecomte de Wideumont.

Je tiens à remercier mes parents qui ont toujours cru en moi et qui m'ont soutenu durant toutes mes années d'étude.

Enfin, il y a une personne qui a particulièrement amélioré mon quotidien, c'est Adeline. Ta gentillesse et ton soutien ont facilité chacune des étapes de l'élaboration de cette thèse.

Contents

Introduction

Almost half a century after its prediction, the Higgs boson has been discovered in 2012 by the ATLAS and the CMS collaborations with the data accumulated during the first run of the LHC. This was a major event in the particle physics community since now all the predicted particles from the standard model have been observed experimentally, making the theory self-consistent.

The standard model of particle physics is a theory developed in the 70s that aims at describing the fundamental particles and their interactions. The model contains three of the four known interactions: the electromagnetic, the weak and the strong interaction. It has accumulated successes to describe experimental results including precise measurements of theoretical predictions. However, physicists have evidence that the standard model is not the end of the story. For instance, it neither includes gravitation nor a dark matter candidate and contains some theoretical features.

The theoretical difficulties have been studied for decades and many ways of extending the standard model to accommodate for them have been explored and experimentally tested. Many of these predict a more complex scalar sector, as supersymmetry for instance, but the searches for such new physics have shown no success so far. The current situation is therefore completely particular to our field, we expect the theory to be incomplete, but no evidence of new physics are yet been found among the huge number of analyses at the LHC. The recent observation of the only scalar particle predicted in the stan-

dard model has quickly turned into precision measurements, which currently shows no deviation from the standard model either.

A simple extension of the scalar sector, the 2HDM, predicts the existence of five scalar bosons. While it is not considered as a complete theory, it provides a very rich phenomenology and therefore new channels to test the standard model. Special cases of the 2HDM, in the so-called *alignment limit*, are extremely relevant, since the lighter scalar behaves in such theories as the standard model Higgs boson.

It is in this context that this thesis has been written. The goal is to search for a typical signature of the *aligned* 2HDM: the production of a heavy scalar H decaying into a Z boson and a pseudoscalar A with the CMS detector. In addition to being well motivated theoretically, searching simultaneously for two new resonances brings a very clear signature in the detector that would have escaped other existing searches.

Searching for two unknown particles also brings practical difficulties: the two-dimensional masses spectrum of the heavy scalar and the pseudoscalar bosons has to be probed. It therefore requires many signal models and then many simulation which are time consuming.

More generally, the huge quantity of data taking at the LHC will bring fundamental changes in the usage of the detector simulation in particle physics. An alternative tool called DELPHES provides parametric simulation of generic detectors at the LHC and for future colliders. While less precise than the complete simulations, parametric simulations allow to gain a factor 10^4 in time consumption and has shown to be ideal for phenomenological studies. To face the variety of applications, significant improvement of the software were needed. Once performed, DELPHES appeared as a useful solution to perform the many detector simulations of 2HDM processes events needed to cover the full 2HDM A and H masses spectrum.

This thesis is divided into three chapters. The first chapter describes the theoretical context and motivations of this work. Among others, the standard model, the 2HDM and the scalar sector phenomenology at the LHC are reviewed. The experimental setup, including the description of the CMS detector at the LHC, is discussed in the second chapter. A discussion of the event recon-

struction is presented through a comparison between the parametric approach and the complete detector simulation. Finally, the first searches for new scalar bosons in the $llbb$ final state at the LHC are explained in the third chapter.

Chapter 1

Study of the scalar sector at proton-proton colliders

1.1 The standard model

The *standard model* is a theory describing the elementary particles and their interactions. It has provided so far a good modeling of the observed phenomena at the accessible scales. Therein, the elementary particles are divided in two distinct groups. The first gathers all the particles that obey the Bose-Einstein statistics. These particles are called *Bosons* and are the carriers of the different interactions. While gravitation is not included in the standard model, it includes the three other fundamental interactions:

- The *photon* (γ) mediates the *electromagnetic interaction*.
- The W^+, W^- and Z bosons are the carriers of the *weak interaction*. It is responsible, for instance, for the nuclear fusion inside the Sun.
- The eight *gluons* are the force carriers of the *strong interaction*. It ensures the stability of most ordinary matter.

The second group of particles, called *fermions*, are all the particles that obey the Fermi-Dirac statistics. The fermions are the matter component of the uni-

verse (e.g. the atoms) and are further divided into two sub-groups: the *leptons* that are interacting through the Electromagnetic or Weak interactions and the *quarks* which are also sensitive to the strong interaction. There exists three generations of quarks and leptons and for each of them there exists a partner with the same mass but opposite charge: its *anti-particle*.

The standard model is the result of decades of theoretical and experimental research. It started with the elaboration of the *Quantum Electro-Dynamics* (QED) at the end of the 40's by, among others, S. Tomonaga [1], J. Schwinger [2, 3], F. Dyson [4] and R.P. Feynman [5, 6]. At the beginning of the 50's, C.N. Yang and R. Mills generalized QED by developing the theories of the gauge fields [7]. This generalization was the first ingredient that led to the unification of the electromagnetic and weak interactions in 1961 by Glashow [8]. Unfortunately, at that time, the theory still had a huge issue. Indeed, the particles (bosons and fermions) were modeled without mass. A solution to give a mass to the particles has been developed by three independent groups (Robert Brout and François Englert [9], Peter Higgs [10] and Gerald Guralnik, C.R. Hagen and Tom Kibblel [11]) and introduced into the standard model still under construction by Weinberg [12] and Salam [13]. This mechanism, called *the Brout-Englert-Higgs mechanism* allows to provide a mass to the particles by adding an extra scalar field to the model. In 1970, the quarks were added to the electro-weak theory by Glashow Illiopoulos and Maiani [14]. Finally, *Quantum Chromo-Dynamic* (QCD) has been added to the model to describe the strong interaction. This work completed the standard model in the middle of the 70's after the work of D. Gross, F. Wilczek [15, 16] and H. David Politzer [17]. Since then, all the predicted particles have been observed and the observations are in agreement with the standard model expectation, as illustared in Fig 1.1.

Modern particle physics is based on the Lagrangian formalism. Any term in the lagrangian must be reference frame invariant, therefore it has to be invariant under the Poincaré group. In a local field theory, it is a function of one or more fields $\phi(x)$ and their derivatives $\partial_\mu \phi$

$$\mathcal{L} = \mathcal{L}(\phi(x), \partial_\mu \phi) \tag{1.1}$$

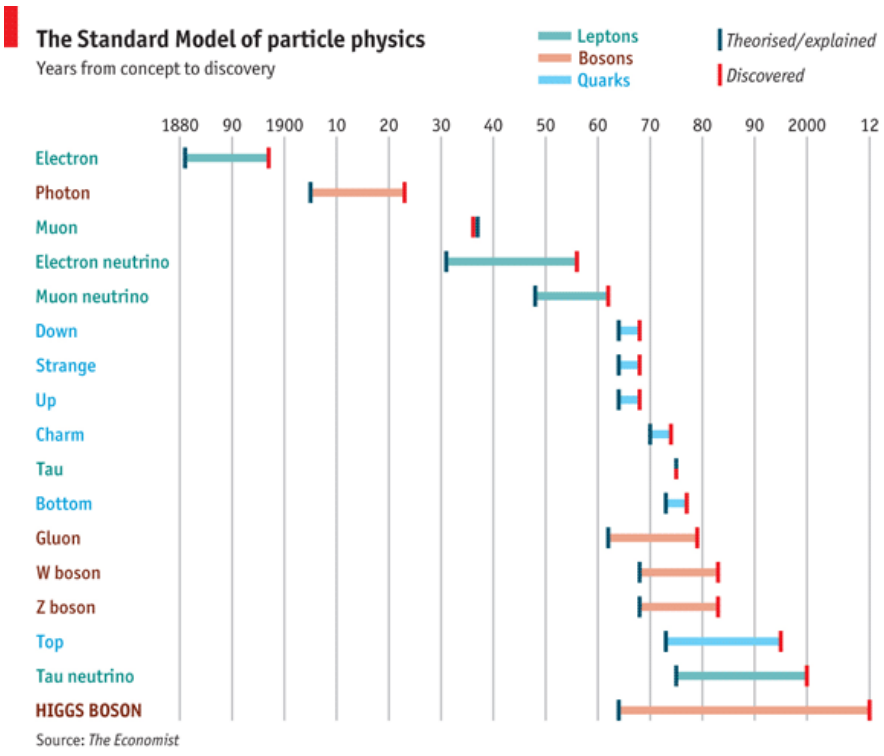

Figure 1.1: The standard model of particle physics, years from concept to discovery. The economist [18]

The gauge part of the standard model is based on the group of symmetry $\mathrm{SU(3)_C} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$. Each term of the Lagrangian is also invariant under this group of symmetry.

In this section, all the evolutions and the different steps of the standard model will be omitted to focus on the components and the standard model itself. To do so, the notations and derivations are taken from Peskin and Schroeder [19]. First, the fermions will be introduced in Sec 1.1.1. The gauge bosons and the description of the electroweak symmetry breaking mechanism are discussed in Sec 1.1.2. Sec 1.1.3 provides a second look at the leptons to introduce their mass and the different quantum numbers. More details about QCD, and in particular the notion of asymptotic freedom and confinement are introduced in Sec 1.1.4. To conclude this section, the standard model Lagrangian and its weaknesses are discussed in Sec 1.1.5.

1.1.1 Fermions, Part I: Unbroken symmetries

The fermions are defined as the particles with a half-integer spin. They obey to the Fermi-Dirac statistics. Two kinds of elementary fermions have been observed: the quarks, which are sensitive to the strong interaction, and the leptons, which are not.

- There are three lepton families, summarized in the first part of the Table 1.1. The first family is composed of the electron (e^-), the positron (e^+) and the electronic (anti-)neutrino(ν_e and $\bar{\nu_e}$). The muons ($\mu^\pm$) and the muonic-neutrinos (ν_μ and $\bar{\nu_\mu}$) belong to the second family. The third family contains the taus($\tau^\pm$) and the tauonic neutrinos (ν_τ and $\bar{\nu_\tau}$).
- There are also three families of quarks, summarized in the second part of Table 1.1. The up and down quarks, which are the main components of the proton and the neutron, are in the first family. The second family of quarks contains the strange (s and $\bar{s}$) and the charm (c and $\bar{c}$) quarks. The bottom (b and $\bar{b}$) and the top (t and $\bar{t}$) quarks are members of the third family.

A fermion of mass m is represented by a four-component field ψ called a *Dirac spinor*. A free lepton is described by the Dirac equation, derived from the Dirac Lagrangian

$$\mathcal{L}^{\text{free}} = \bar{\psi} i \gamma_\mu \partial^\mu \psi - m \bar{\psi} \psi \tag{1.2}$$

where $\bar{\psi} i \gamma_\mu \partial^\mu \psi$ is a *kinetic term* related to the motion of the field ψ and $m\bar{\psi}\psi$ is a *mass term*. The nature of the neutrinos has still not been revealed experimentally. They are often believed to be Majorana particles which could imply that a neutrino is its own anti-particle.

	Leptons			Charge
l	e 511 keV	μ 105 MeV	τ 1.77 GeV	-1
ν_l	ν_e < 2.5 eV	ν_μ < 170 keV	ν_τ < 18 MeV	0
	Quarks			
q_u	u 2.3 MeV	c 1.27 GeV	t 173 GeV	$\frac{2}{3}$
q_d	d 4.8 MeV	s 95 MeV	b 4.66 GeV	$-\frac{1}{3}$

Table 1.1: Lepton and quark list of physical states.

Leptons Quantum numbers

The main differences between quarks and leptons and between the up and down part in each family is driven by their behavior under the different interactions and therefore their quantum numbers.

While the leptons are singlet under $\mathrm{SU}(3)_\mathrm{C}$, the quarks are assigned with an extra quantum number: the color. The resulting consequences will be explained in Sec 1.1.4.

A fermion can be *right-handed* when its spin is parallel to its direction of motion or *left-handed* when it is opposite. From the point of view of the weak interaction, left- and right-handed fermions are different particles. It is therefore convenient to decompose a fermion into its left and right part.

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \tag{1.3}$$

In fact, only the left-handed particles are sensitive to the weak interaction, this is translated in terms of the associated quantum numbers, the weak-isospin, by assigning 0 to the right-handed fermions, and $\pm 1/2$ to the left-handed fermions.

The *hypercharge* is the quantum number associated with the $U(1)_Y$ gauge symmetry. It is intimately related to the charge of a particle, this will be demonstrated in the Sec 1.1.3.

	Left-Handed				Right-Handed			
	q_u	q_d	ν_l	l	q_u	q_d	ν_l	l
Hypercharge	$1/6$	$1/6$	$-1/2$	$-1/2$	$2/3$	$-1/3$	0	-1
Weak Isospin	$1/2$	$-1/2$	$1/2$	$-1/2$	0	0	0	0
Color	r-b-g	r-b-g			r-b-g	r-b-g		

Table 1.2: Lepton and quark quantum numbers.

1.1.2 Bosons and the electroweak symmetry breaking

We have seen in Eq. 1.1 that the Lagrangian can be written in term of the fields and their derivatives. The construction of a locally gauge invariant Lagrangian requires the definition of the *covariant derivative* $\mathrm{D}_\mu = \partial_\mu + \mathrm{igA}^a_\mu \mathrm{T}^a$, where g is the gauge coupling, A^a_μ is the connection and T^a is the generator of the group. Using the previous definition, one can construct the gauge field tensor $\mathrm{F}^{\mathrm{a}}_{\mu\nu}$ as

$$\mathrm{igF}^{\mathrm{a}}{}_{\mu\nu}\mathrm{T}^{\mathrm{a}} = [\mathrm{D}_\mu, \mathrm{D}_\nu] \quad . \tag{1.4}$$

In establishing the standard model, three particular cases are interesting. First, the $\mathrm{SU(3)_C}$ gauge group which is composed of eight bosons $\mathrm{G^a_\mu}$, with $\mathrm{a} = 1, ..., 8$. Using Eq. 1.4, and the $\mathrm{SU(3)}$ group generators $\mathrm{t^a}$, we find:

$$\mathrm{G^a_{\mu\nu}} = \partial_\mu \mathrm{G^a_\nu} - \partial_\nu \mathrm{G^a_\mu} + \mathrm{g_S} f^{abc} \mathrm{G^b_\mu G^c_\nu} \tag{1.5}$$

where f^{abc} is called the *structure constant* and $\mathrm{g_S}$ is the coupling constant of the strong interaction. Second, the $\mathrm{SU(2)_L}$ group induces three gauge bosons $\mathrm{A^i_\mu}$ with $i = 1, 2, 3$ and a structure constant being ϵ^{ijk}.

$$\mathrm{A^i_{\mu\nu}} = \partial_\mu \mathrm{A^a_\nu} - \partial_\nu \mathrm{A^a_\mu} + \mathrm{g} \epsilon^{ijk} \mathrm{A^j_\mu A^k_\nu} \tag{1.6}$$

where g is the associated coupling. The last and simplest group considered here is $\mathrm{U(1)}$. The coupling constant associated to this group is denoted $\mathrm{g'}$ and the gauge tensor is defined as

$$\mathrm{B_{\mu\nu}} = \partial_\mu \mathrm{B_\nu} - \partial_\nu \mathrm{B_\mu} \tag{1.7}$$

The most general locally invariant Lagrangian for the fermion field ψ contains all the kinetic terms of the gauge bosons

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}\mathrm{B_{\mu\nu}B^{\mu\nu}} - \frac{1}{4}\mathrm{A_{\mu\nu}A^{\mu\nu}} - \frac{1}{2}\mathrm{trG_{\mu\nu}G^{\mu\nu}} \tag{1.8}$$

the interaction terms between the fermions and the vector bosons

$$\mathcal{L}_{\text{fermion}} = \bar{\psi}(i\not{D})\psi \tag{1.9}$$

and the mass term for the leptons,

$$\mathcal{L}_{\text{fermion-mass}} = -\bar{\psi}\mathrm{m}\psi \tag{1.10}$$

where only the dimension four operators have been considered. It can also be concluded that no mass term for the gauge bosons is present (e.g. $\frac{1}{2}\mathrm{mA_\mu A^\mu}$) since it would violate gauge invariance.

Electro-Weak symmetry breaking Mechanism

From the previous section, we have seen that the assumption that the interactions are gauge invariant requires the existence of massless vector fields for

each symmetry generator, while it is known from the experiments that the weak interaction carriers the $W^{\pm}$ and the Z bosons are massive. Also, the construction of a fermion mass term such as

$$\Delta\mathcal{L} = -\mathrm{m_e}(\bar{\mathrm{e}}_\mathrm{L}\mathrm{e}_\mathrm{R} + \bar{\mathrm{e}}_\mathrm{R}\mathrm{e}_\mathrm{L}) \tag{1.11}$$

is forbidden because ψ_L and ψ_R have different gauge quantum numbers. Simple terms are therefore violating gauge invariance.

The solution proposed in the standard model is to add a complex scalar field ϕ doublet under $\mathrm{SU}(2)$, which has therefore four degrees of freedom. This new field interacts with itself and with the electro-weak gauge fields, where it is assumed that the scalar field has a charge $\frac{1}{2}$ under the $\mathrm{U}(1)$ symmetry. One can therefore consider the following Lagrangian:

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + |\mathrm{D}_\mu\phi|^2 - \mathrm{V}(\phi) \tag{1.12}$$

where one assumes the potential to have a *Mexican hat* form.

$$\mathrm{V}(\phi) = -\mu^2\phi^*\phi + \frac{\lambda}{2}(\phi^*\phi)^2 \tag{1.13}$$

with $\mu^2 > 0$. When minimizing the potential, the field acquires a non-zero vacuum expectation value and the electro-weak symmetry $\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y}$ is broken. The field ϕ acquires a vacuum expectation value, that can be chosen to be

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \quad . \tag{1.14}$$

where $v = \left(\frac{\mu^2}{\lambda}\right)^{1/2}$ is the minimum of the potential energy. The covariant derivative of ϕ is given by:

$$\mathrm{D}_\mu\phi = \left(\partial_\mu - \mathrm{igA}^\mathrm{a}_\mu\tau^\mathrm{a} - \mathrm{i}\frac{1}{2}\mathrm{g}'\mathrm{B}_\mu\right)\phi \tag{1.15}$$

and therefore the kinetic term $|\mathrm{D}_\mu\phi|^2$ of the Lagrangian can be re-written as following:

$$\Delta\mathcal{L} = \frac{1}{2}\frac{\mathrm{v}^2}{4}\left[\mathrm{g}^2(\mathrm{A}^1_\mu)^2 + \mathrm{g}^2(\mathrm{A}^2_\mu)^2 + (-\mathrm{gA}^3_\mu + \mathrm{g}'\mathrm{B}_\mu)^2\right] \quad . \tag{1.16}$$

Three massive vector bosons arise from this equation:

$$\mathrm{W}_\mu^\pm = \frac{1}{\sqrt{2}}(\mathrm{A}_\mu^1 \mp \mathrm{iA}_\mu^2) \quad \text{with mass} \quad \mathrm{m_W} = \mathrm{g}\frac{\mathrm{v}}{2} \tag{1.17}$$

$$\mathrm{Z}_\mu = \cos\theta_\mathrm{w}\mathrm{A}_\mu^3 - \sin\theta_\mathrm{w}\mathrm{B}_\mu \quad \text{with mass} \quad \mathrm{m_Z} = \sqrt{\mathrm{g}^2 + \mathrm{g}'^2}\frac{\mathrm{v}}{2} \tag{1.18}$$

and a massless boson

$$\mathrm{A}_\mu = \sin\theta_\mathrm{w}\mathrm{A}_\mu^3 + \cos\theta_\mathrm{w}\mathrm{B}_\mu \tag{1.19}$$

where we have defined the *weak mixing angle* θ_w as $\sin\theta_\mathrm{w} = \frac{g}{\sqrt{g^2+g'^2}}$ and $\cos\theta_\mathrm{w} = \frac{g'}{\sqrt{g^2+g'^2}}$. From the Eqs. 1.17 and 1.18, we see that:

$$\mathrm{m_W} = \mathrm{m_Z}\cos\theta_\mathrm{w} \tag{1.20}$$

From the four degrees of freedom introduced in the model, we see that three are "eaten" by the W and Z bosons to acquire a mass and the massless A is the photon field.

Let us investigate further the behavior of this scalar field ϕ by parametrizing it in the *unitarity gauge* as a doublet with a vacuum expectation value plus a real field h.

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.21}$$

The potential (Eq 1.13) becomes

$$\mathcal{L}_\mathrm{V} = -\frac{1}{2}\mathrm{m_h}^2 h^2 - \sqrt{\frac{\lambda}{2}}\mathrm{m_h}h^3 - \frac{1}{4}\lambda h^4 \tag{1.22}$$

where $\mathrm{m_h} = \sqrt{2\lambda}v$. The remaining degree of freedom is therefore an extra massive scalar particle. The same exercise applied on the kinetic term of Eq 1.12 shows that this new scalar field has a coupling with the gauge bosons which is proportional to their masses:

$$\mathcal{L}_\mathrm{kin} = \frac{1}{2}(\partial_\mu h)^2 + \left[\mathrm{m}_W^2\mathrm{W}^{\mu+}\mathrm{W}_{\mu-} + \frac{1}{2}\mathrm{m}_Z^2\mathrm{Z}^\mu\mathrm{Z}_\mu\right]\cdot\left(1+\frac{h}{v}\right)^2 \tag{1.23}$$

Therefore, this mechanism appeared so far with two important predictions:

- The relation between the Z and the W-bosons masses.
- The existence of an extra scalar field, the *Higgs Boson*, that couples to the gauge bosons proportionally to their masses.

Both have been confirmed and the status of the Higgs Boson observation is summarized in Sec. 1.3.3.

1.1.3 Fermions, Part II: Charges and Masses

For a fermion field in the $SU(2)$ representation with a $\mathrm{U}(1)$ charge Y, the covariant derivative can be re-written in terms of the physical fields:

$$\begin{aligned} \mathrm{D}_\mu &= \partial_\mu - \mathrm{i g A}^{\mathrm{a}}_\mu \mathrm{T}^{\mathrm{a}} - \mathrm{i g}' \mathrm{Y B}_\mu \\ &= \partial_\mu - \mathrm{i}\frac{\mathrm{g}}{\sqrt{2}}\left(\mathrm{W}^+_\mu \mathrm{T}^+ \mathrm{W}^-_\mu \mathrm{T}^-\right) - \mathrm{i}\frac{\mathrm{g}}{\cos\theta_{\mathrm{w}}}\mathrm{Z}_\mu(\mathrm{T}^3 - \sin^2\theta_{\mathrm{w}}\mathrm{Q}) \\ &\quad -\mathrm{ieA}_\mu(\mathrm{Q}) \end{aligned}$$

where $\mathrm{T}^\pm = (\mathrm{T}^1 \pm \mathrm{iT}^2)$ and where the coefficient of the electromagnetic interaction e and the electric charge quantum number Q have been identified.

$$\mathrm{e} = \frac{\mathrm{g g}'}{\sqrt{\mathrm{g}^2 + \mathrm{g}'^2}} \tag{1.24}$$

$$\mathrm{Q} = \mathrm{T}^3 + \mathrm{Y} \tag{1.25}$$

We have seen in Sec 1.1.2 that it was not possible to put an ordinary mass term without violating gauge invariance. Here also, the mechanism of spontaneous symmetry breaking brings a solution. Keeping in mind that ϕ is a spinor under $\mathrm{SU}(2)$ and has a hypercharge of $\frac{1}{2}$, we can write:

$$\Delta\mathcal{L}_{\mathrm{e}} = -\lambda_{\mathrm{e}}\bar{\mathrm{E}}_{\mathrm{L}} \cdot \phi \mathrm{e}_{\mathrm{R}} + \mathrm{h.c.} \tag{1.26}$$

$$= -\frac{1}{\sqrt{2}}\lambda_{\mathrm{e}}\mathrm{v}\bar{\mathrm{e}}_{\mathrm{L}}\mathrm{e}_{\mathrm{R}} + \mathrm{h.c.} + ... \tag{1.27}$$

where the λ_{e} is a dimensionless coupling constant that determines the mass of the fermion with the vacuum expectation value v.

$$\mathrm{m_e} = \frac{1}{\sqrt{2}}\lambda_{\mathrm{e}}\mathrm{v} \tag{1.28}$$

We find that for each fermions f, the Higgs boson couples as

$$\mathcal{L}_f = -\mathrm{m}_f \bar{f} f \left(1 + \frac{h}{v}\right) \qquad . \tag{1.29}$$

This last equation provides an extra prediction:

- The *Higgs Boson* couples to the fermions proportionally to their masses.

This prediction is also verified so far and will be discussed in Sec. 1.3.3.

1.1.4 Asymptotic freedom and confinement

The strong interaction also has some particular features related to the coupling. Unlike electromagnetism and weak interactions, the effective coupling α_s decreases as a function of the energy.

On the one hand, at high energy, $\alpha_s(q^2) \to 0$ and the quarks and gluons behave as free particles. This phenomenon is called *asymptotic freedom*.

On the other hand, at low energy, the coupling constant diverges. It is therefore not possible to develop the Feynman diagrams in series for QCD. This is most probably related to another feature of QCD called *confinement*. In fact, the colored particles have never been observed isolated. Only *hadrons*, which are *"white"* bound state of several (anti-)quarks, are observed. Two types of hadrons are commonly observed: the *mesons* are made of $q\bar{q}$ pair and the *baryons* are combinations of three quarks.

To better understand the behavior, let us use a classical picture with a potential of the form

$$V_s(r) = -\frac{4}{3}\frac{\alpha_s}{r} + kr \tag{1.30}$$

When a pair of quarks is produced at high energy and drifting away, the potential will start rising with the distance. Once the energy accumulated is sufficient, an extra pair of quarks will be created from the vacuum. After loosing enough energy by this procedure, quarks and anti-quarks will combine with each other and produce a cascade of hadrons dubbed *jets*. This process is called *hadronization* or *fragmentation*.

1.1.5 The standard model and some difficulties

Through the previous section, we have built all the lagrangian terms that satisfy the required symmetries at order four. The sum of these pieces defines the standard model Lagrangian:

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{Yukawa}} \tag{1.31}$$

Despite the experimental confirmations of the electro-weak symmetry breaking, some theoretical difficulties of the standard model could be related to the scalar sector.

The theory is *renormalizable* [20], and therefore well-defined and calculable up to infinite energies. Nevertheless, it is known, since it doesn't include gravitation, that the standard model is not the complete theory. In particular, gravity is supposed to become important around the Planck mass

$$M_{\text{P}} \approx 1.2 \times 10^{19}\text{GeV}. \tag{1.32}$$

Hierarchy problem

If the Planck mass is the scale of the new physics, or if there is another new physics scale $\Lambda \gg v$, new loop-diagrams will contribute to the self-energy of the of the Higgs bosons. Those one-loop corrections to $\mathrm{m_h}$ would be much greater than $\sqrt{2\lambda}v$ and would therefore require a huge cancellation mechanism. This difficulty of the standard model is called the *hierarchy problem*.

Neutrino masses

The neutrinos are considered in the standard model as massless particles. The mass measurement is currently still an upper limit but the observation of neutrino oscillations implies that neutrinos are massive particles.

Dark Matter

Even though the nature of the *dark matter* is still not known, much evidences of its existence has been observed. The rotation curves of the galaxies and the gravitational lensing are convincing observations, among others. In the relevant hypothesis that dark matter is a new particle, several conditions must be fulfilled: dark matter has to be electrically neutral, colorless and very long-lived or stable.

Since it has a mass, dark matter is expected to interact with the Higgs boson. In fact, a simple dark matter candidate arises already when adding a scalar singlet to the standard model [21].

Fermion masses hierarchy

While very elegant in the gauge sector, the Yukawa part of the standard model Lagrangian implies one new parameter per fermion related to its mass. This leads to a striking feature: the charged leptons/quarks masses range from 173 GeV for the top quark down to 0.5 MeV for the electron. This feature is far from being explained in the standard model.

Vacuum stability

When considering loop diagrams, corrections to the Higgs potential arise with two different contributions: the Higgs self-interaction and the top quark loop. Those effects modify the 'Mexican hat' shape of the potential and three different scenarios are possible:

- The vacuum is *stable* if there is only one minimum, or if there is another minimum higher than the electroweak minimum.
- If the additional minimum is lower or at the same level, quantum tunneling could make the vacuum decay. If the lifetime of the vacuum is larger than the age of our universe, the vacuum is *metastable*.
- If the lifetime of the vacuum is smaller than the age of the universe or if the vacuum is not bounded from below, it is *unstable*.

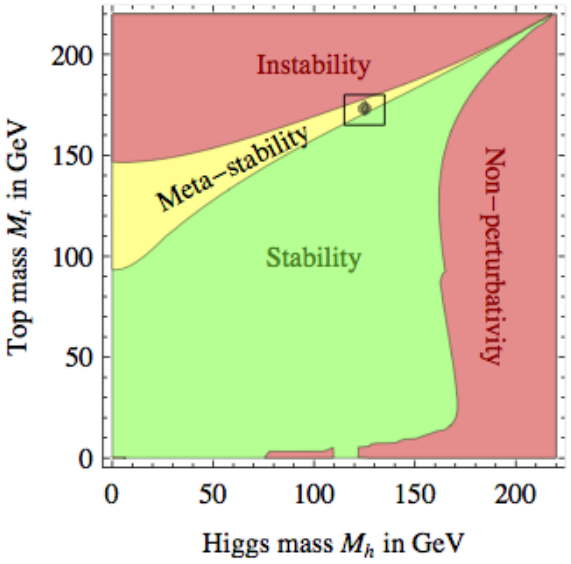

Figure 1.2: Regions of absolute stability, meta-stability and instability of the SM vacuum in the $\mathrm{m_t}$-$\mathrm{m_h}$ plane [22].

The measurement of the mass of both the Higgs and the top particles indicates that the universe is in the particular metastability region, as can be seen in Fig 1.2.

Matter-antimatter asymmetry

In the universe, almost only matter is observed rather than an equal amount of matter and antimatter. One of the challenges in particle physics is to understand the source of this asymmetry. Indeed, significant CP violation is necessary to produce matter and antimatter at a different rate. The standard model contains several sources of CP violation that are too small to explain this observation. Sources of CP violation could therefore arise from extensions of the standard model, in particular in the scalar sector.

1.2 Extending the scalar sector

The recent discovery of a Higgs look-alike particle appears as an evidence of the mechanism of spontaneous symmetry breaking and proved the existence of elementary scalar particles in the universe. It is also clear from Sec 1.1.5 that extensions of the standard model are expected and motivate, eventually, a more complex Brout-Englert-Higgs sector. In particular, one can add extra singlets, doublets, triplets, etc... In 1973, the two-Higgs-doublet model (2HDM) has been first proposed as an attempt to find new sources of CP violation [23].

In this context, the 2HDM is not taken as being the ultimate theory but rather as a part of the solution. Another example of such theories is the *Minimal Supersymmetric Standard Model(MSSM)* which contains, among other extra particles, a constrained 2HDM.

1.2.1 Electro weak symmetry breaking with two Higgs Doublets

In this work, we will consider that CP is conserved in the Higgs sector and not spontaneously broken. Following the reference [24], if we consider two doublets ϕ_1 and ϕ_2 with $Y = 1$, the most general scalar potential is

$$\begin{aligned} V(\phi_1,\phi_2) &= \mathrm{m}_{11}^2\phi_1^\dagger\phi_1 + \mathrm{m}_{22}^2\phi_2^\dagger\phi_2 - \mathrm{m}_{12}^2\left(\phi_1^\dagger\phi_2 + \phi_2^\dagger\phi_1\right) \\ &\quad + \frac{\lambda_1}{2}\left(\phi_1^\dagger\phi_1\right) + \frac{\lambda_2}{2}\left(\phi_2^\dagger\phi_2\right) \\ &\quad + \lambda_3\phi_1^\dagger\phi_1\phi_2^\dagger\phi_2 + \lambda_4\phi_1^\dagger\phi_2\phi_2^\dagger\phi_1 \\ &\quad + \frac{\lambda_S}{2}\left[\left(\phi_1^\dagger\phi_2\right)^2 + \left(\phi_2^\dagger\phi_1\right)^2\right] \end{aligned} \tag{1.33}$$

where the parameters are real. By analogy with the SM, we can assume that the minimization of this potential gives

$$\langle\phi_1\rangle_0 = \begin{pmatrix} 0 \\ \frac{v_1}{\sqrt{2}} \end{pmatrix} \qquad \langle\phi_2\rangle_0 = \begin{pmatrix} 0 \\ \frac{v_2}{\sqrt{2}} \end{pmatrix} \tag{1.34}$$

Since these are complex scalar SU(2) doublets, there are four fields for each doublet and they can be parametrized as:

$$\phi_\alpha = \begin{pmatrix} \phi_a^+ \\ (v_a + \rho_a + i\eta_a)/\sqrt{2} \end{pmatrix} \quad , \quad a = 1, 2 \tag{1.35}$$

By inserting Eq 1.35 in the Lagrangian 1.33, we can write the mass term for the charged scalars:

$$\mathcal{L}_{\phi^\pm \text{ mass}} = \frac{\mathrm{m}_\pm^2}{v_1^2 + v_2^2}(\phi_1^-, \phi_2^-) \begin{pmatrix} v_2^2 & -v_1 v_2 \\ -v_1 v_2 & v_1^2 \end{pmatrix} \begin{pmatrix} \phi_1^+ \\ \phi_2^+ \end{pmatrix} \tag{1.36}$$

where $\mathrm{m}_\pm^2 = \left[\frac{\mathrm{m}_{12}^2}{v_1 v_2} - \lambda_4 - \lambda_5\right](v_1^2 + v_2^2)$. The diagonalization of this mass matrix provides one zero eigenvalue and the mass to the $\mathrm{W}^\pm$ bosons. The mass term for the pseudoscalar has the same form:

$$\mathcal{L}_{\eta \text{ mass}} = \frac{{\mathrm{m}_\mathrm{A}}^2}{v_1^2 + v_2^2}(\eta_1, \eta_2) \begin{pmatrix} v_2^2 & -v_1 v_2 \\ -v_1 v_2 & v_1^2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \tag{1.37}$$

whith ${\mathrm{m}_\mathrm{A}}^2 = \left[\frac{\mathrm{m}_{12}^2}{v_1 v_2} - 2\lambda_5\right](v_1^2 + v_2^2)$. Interestingly, the diagonalization is done by applying the same angle β such that $\tan\beta \equiv \frac{v_1}{v_2}$. The zero eigenvalue provides a mass to the Z boson. Finally, the mass terms for the scalars are:

$$\mathcal{L}_{\rho \text{ mass}} = (\rho_1, \rho_2) \begin{pmatrix} \mathrm{m}_{12}^2 \frac{v_2}{v_1} + \lambda_1 v_1^2 & -\mathrm{m}_{12}^2 + \lambda_{345} v_1 v_2 \\ -\mathrm{m}_{12}^2 + \lambda_{345} v_1 v_2 & \mathrm{m}_{12}^2 \frac{v_1}{v_2} + \lambda_1 v_1^2 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \tag{1.38}$$

The rotation angle α is defined as the angle that performs the diagonalization of the mass-squared matrix of the scalars.

The physical scalars h and H are therefore defined as

$$\mathrm{h} = \rho_1 \sin\alpha - \rho_2 \cos\alpha \tag{1.39}$$

$$\mathrm{H} = -\rho_1 \cos\alpha - \rho_2 \sin\alpha \tag{1.40}$$

As a remark, the parameters m_{11} and m_{22} can be written in terms of v_1 and v_2, and the other parameters of the models. In the 2HDM model, v_1 and v_2

are constrained by the relation $v_1^2 + v_2^2 = v^2$ where $v \approx 246$ GeV. Therefore a change of variable can be performed from v_1 and v_2 to v at its SM value and $\tan\beta$ as a free parameter. This change of variable defines the *general basis* of a CP-conserved 2HDM:

$$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \mathrm{m}_{12} \quad \tan\beta$$

We also notice that the standard model Higgs boson can be written as a combination of both h and H. This is straightforward when using the *Higgs basis*, in which only one doublet acquires the vacuum expectation value v.

$$\mathrm{h}^{\mathrm{SM}} = \mathrm{h}\sin(\beta-\alpha) - \mathrm{H}\cos(\beta-\alpha) \tag{1.41}$$

The limit $\beta-\alpha \approx \frac{\pi}{2}$ in which the light h tends to the SM-Higgs boson is called *alignment limit*.

Couplings to vector bosons

The couplings of h and H to the vector boson $\mathrm{W}^{\pm}$ and Z are given by

$$g_{\mathrm{hVV}} = \frac{\mathrm{m}_{\mathrm{V}}^2}{v}\sin(\beta-\alpha) \quad \text{and} \quad g_{\mathrm{HVV}} = \frac{\mathrm{m}_{\mathrm{V}}^2}{v}\cos(\beta-\alpha) \quad . \tag{1.42}$$

In the alignment limit, we see that the heavy H is decoupled from the vector bosons. The couplings between a scalar, a pseudoscalar and the Z boson are also determined by the gauge coupling structure:

$$g_{\mathrm{ZAh}} = -\frac{g\cos(\beta-\alpha)}{2\cos\theta_w}(p_{\mathrm{h}}-p_{\mathrm{A}})_\mu \quad \text{and} \quad g_{\mathrm{ZAH}} = -\frac{g\sin(\beta-\alpha)}{2\cos\theta_w}(p_{\mathrm{H}}-p_{\mathrm{A}})_\mu \quad . \tag{1.43}$$

1.2.2 Couplings to fermions and flavor conservation

The presence of several Higgs doublets in a theory generally induces Flavor Changing neutral currents (FCNC) via the Yukawa couplings. This difficulty

can be solved by considering that all the fermions with the same quantum numbers couple to the same Higgs doublet. Four different possibilities have been proposed:

- All quarks and leptons couple to one Higgs doublet (ϕ_2 by convention) in 2HDM *type-I*.
- The $Q = \frac{2}{3}$ quarks couple to one Higgs doublet (ϕ_2 by convention) and the $Q = \frac{-1}{3}$ quarks and the charged leptons couple to the other. This possibility is called *type-II*.
- The *lepton specific* model assume that the quarks couple to ϕ_2 and the leptons to ϕ_1.
- The $Q = \frac{2}{3}$ quarks and the charged leptons couple to one Higgs doublet (ϕ_2 by convention) and the $Q = \frac{-1}{3}$ quarks couple to the other. This possibility is called *flipped*.

Those four naturally flavor-conserving models are summarized in Tab 1.2.2

Model	u	d	e
Type-I	ϕ_2	ϕ_2	ϕ_2
Type-II	ϕ_2	ϕ_1	ϕ_1
Lepton-specific	ϕ_2	ϕ_2	ϕ_1
Flipped	ϕ_2	ϕ_1	ϕ_2

Table 1.3: Two-Higgs-doublet models leading to natural flavor conservation.

1.2.3 Physical basis

So from the general basis with the seven parameters λ_{1-5}, m_{12} and $\tan\beta$, we can move on to the so-called physical basis with as parameters the five scalar bosons masses $\mathrm{m_h}$, $\mathrm{m_H}$, $\mathrm{m_A}$ and $\mathrm{m_H}^{\pm}$, the soft $\mathcal{Z}_2$-symmetry breaking parameter m_{12} and the angles α and $\tan\beta$.

The couplings to fermions are summarized in Tab 1.4 for the Type I and Type II.

	Type I	Type II
ϵ_h^u	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\sin\beta}$
ϵ_h^d	$\frac{\cos\alpha}{\sin\beta}$	$\frac{-\sin\alpha}{\cos\beta}$
ϵ_h^l	$\frac{\cos\alpha}{\sin\beta}$	$\frac{-\sin\alpha}{\cos\beta}$
ϵ_H^u	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$
ϵ_H^d	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\cos\beta}$
ϵ_H^l	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\cos\beta}$
ϵ_A^u	$\cot\beta$	$\cot\beta$
ϵ_A^d	$-\cot\beta$	$\tan\beta$
ϵ_A^l	$-\cot\beta$	$\tan\beta$

Table 1.4: List of couplings

1.2.4 Mass spectrum: Hierarchical vs Degenerate 2HDM

In the standard model, the ρ-parameter is a relation between $\mathrm{m_Z}$, $\mathrm{m_W}$ and θ_w and is equal to 1 at tree-level. It has been measured with a very good precision: $\rho = 1.00037 \pm 0.00023$ [25]. In the 2HDM, as for many models beyond the standard model, this relation is in general not satisfied. A *custodial symmetry* has therefore to be enforced by hand to protect the ρ parameter from significant corrections. Applied to the 2HDM, that extra symmetry entails two possible cases:

- The *usual* hierarchy where the $\mathrm{H}^\pm$ have to be degenerate in mass with A, therefore forming a mass triplet.
- The *inverted* mass spectrum, as proposed in [26, 27] with a triplet formed of H,$\mathrm{H}^\pm$. Such models are naturally arise from 2HDM invariant under twisted custodial symmetry and motivate the search for a light pseudoscalar state.

As proposed in [28], we will make the distinction between *degenerate* scenarios with A, H and $\mathrm{H}^\pm$ close in mass and *hierarchical* scenarios with a mass splitting $\Delta\mathrm{m}_{A-H} = |\mathrm{m_A} - \mathrm{m_H}|$ big compared to the vector boson masses.

It has also been shown that a significant $\Delta\mathrm{m}_{A-H}$ could contribute to the electroweak baryogenesis and therefore brings a key ingredient to explain the

cosmic baryon asymmetry. In fact, Δm_{A-H} favors strong EW phase transition [29], as shown in Fig 1.3. The physical region is defined as models that satisfy unitarity, perturbativity, electroweak precision constraints and collider bounds. Among the physical points, those with a strongly first order electroweak phase transition are added in the heat-map of the right part of Figure 1.3. In particular, models with $m_A \gtrsim 300$ GeV and lighter H close to the alignment limit are motivated.

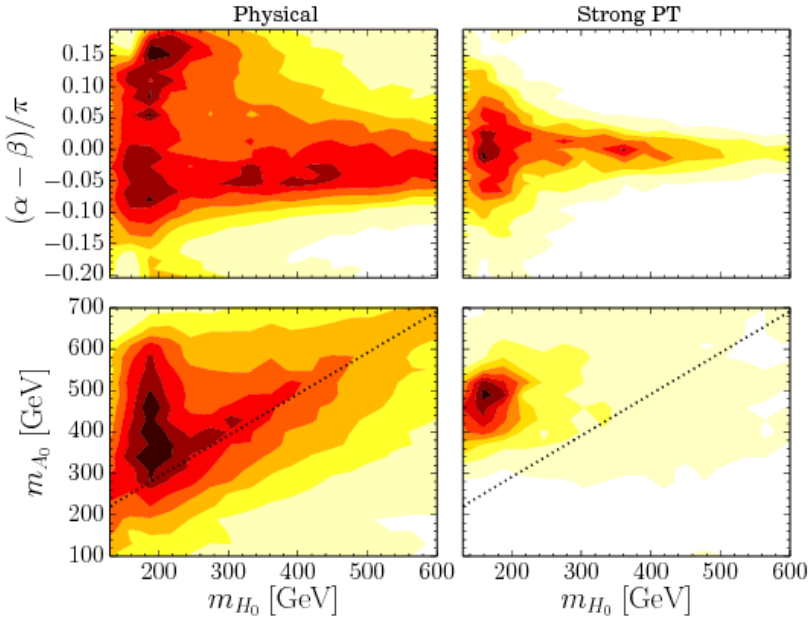

Figure 1.3: Heat-maps for the physical region (left) and region with a strongly first order EWPT (right). Top: (m_H,$\alpha-\beta$)-plane. Bottom: (m_H,m_A)-plane. The dotted-black line corresponds to $m_A = m_H + m_Z$.

These studies motivate, therefore, two different scenarios. The case where a heavy A and a lighter H are present and the twisted case with a light pseudoscalar.

1.3 Higgs bosons phenomenology at the LHC

The main result of the last years in particle physics is the discovery of the Higgs boson at the LHC. After decades of searches at LEP and at Tevatron, the choice of a proton-proton collider has been driven by the fact that the proton radiates significantly less than the electron when accelerated. Therefore it is much easier to have high energy protons with circular collider. Also, the production of protons is easier than anti-protons, allowing a greater parton luminosity at high energy. For this reason, proton-proton colliders are frequently called *discovery machines*, since they allow to study physics with a high energy in the centrer of mass and a high luminosity. Besides this argument, proton collisions and more generally hadrons collisions are subject to the complexity of the strong interaction already mentioned in Sec 1.1.4. Sec 1.3.1 is dedicated to this context and to the modelling of proton-proton collisions.

Currently, Higgs-related research is divided into two topics, the study of the H and the searches for extensions to the scalar sector. Sec 1.3.2 discusses the different production mechanisms of the standard model Higgs Boson and its decay modes. The current experimental status is summarized in Sec 1.3.3. Those results, and in particular the measurement of the couplings of the newly discovered boson have a significant impact on the 2HDM phenomenology. In Sec 1.3.4, these results are examined and some interesting features are reviewed.

1.3.1 Proton-proton collisions and their simulation

The proton and the partons

At high energy, a collision between protons is modeled as an interaction between their constituents called *partons*. The probability density to find a parton with an energy fraction x at a squared energy scale Q^2 is called "parton density function".

When computing hadronic cross-section, one needs to take into account all radiations or splittings of the initial-state partons, which induce collinear divergences. These IR divergences can be absorbed into a redefinition of the

parton densities at a scale μ_{F}. The formulation of the hadronic cross section is therefore given by

$$\sigma_{\mathrm{tot}} = \int_0^1 dx_1 \int_0^1 dx_2 f_i(x_i, \mu_{\mathrm{F}}) f_j(x_j, \mu_{\mathrm{F}}) \hat{\sigma}_{ij}(x_1 x_2 S, \mu_{\mathrm{R}}) \quad . \tag{1.44}$$

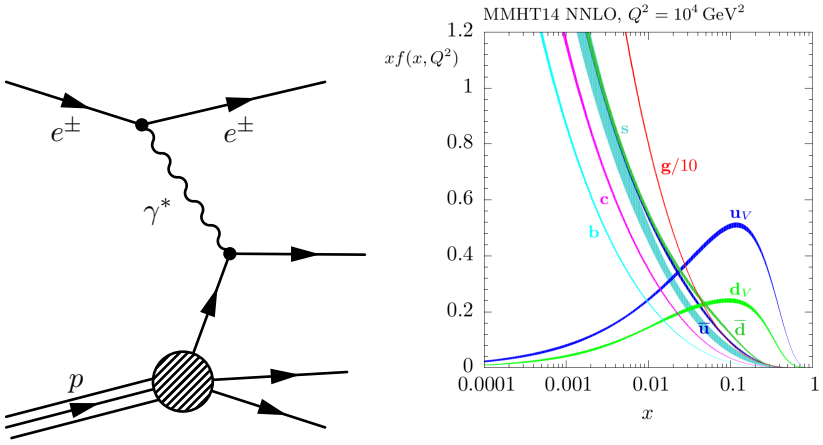

Figure 1.4: Left: Deep inelastic process Feynman diagram. Right: Parton density function obtained by the MMHT 2014 group.

Since the pdfs describe properties of hadrons in the non-perturbative regime, they cannot be calculated with perturbative QCD an need to be measured by experiment. The partonic structure of the proton can be studied with inelastic electron scattering off proton, where the lepton acts as a probe which transfers a momentum q to the nucleon in the collision. The deep inelastic process diagram is drawn in the the left part of Fig 1.4. The H1 and ZEUS collaborations have measured deep inelastic cross sections in neutral and charged current unpolarised $\mathrm{e}^{\pm}\mathrm{p}$ scattering at HERA electron-proton collider at DESY. Recently, the precise measurements from D0 and CDF at the Tevatron and from ATLAS, CMS and LHCb at the LHC on jet, vector boson and top quark pair production bring extra information to the estimations of the pdf.

Several collaborations provide pdf fits, as drawn in right part of Fig 1.4. Some of them and their inputs are summarized here.

- The NNPDF 3.0 [30] uses a dataset including HERA-II deep-inelastic inclusive cross-sections, jet production from ATLAS and CMS, vector boson measurements from ATLAS, CMS and LHCb and top quark pair production total cross sections from ATLAS and CMS. Results are based on LO, NLO and NNLO QCD theory and also include electroweak corrections.
- The CT10 [31] fit is obtained using a dataset from HERA and the asymmetry in the rapidity distribution of the charged lepton from W boson decay from the CDF and D0 collaborations.
- The MMHT 2014 pdf set [32] is based on HERA, Tevatron and ATLAS and CMS results at LHC. In particular, it is based on jet production, vector boson measurements and top quark pair production total cross sections from ATLAS and CMS.

Event generators and hard scattering

The high energy part of the interaction is called *hard scattering* and has the advantage to be in the perturbative regime. Cross-section can therefore be precisely computed with lowest orders of the series expansion only. Programs called *event generators* can generate events using the Matrix element information with up to the next-to-leading order precision. This therefore provides a fully differential cross-section.

- MADGRAPH_aMC@NLO [33] is the latest version of the MADGRAPH suite. Those are generic and automated generators since it is possible to generate any SM or BSM processes automatically. In the latest version, one can choose between LO and NLO precision in QCD.
- POWHEG [34] is another suite of event generator with NLO precision. It has the advantage to provide unweighted events.

Parton shower, hadronization and underlying events

Once the hard scattering event is generated, several contributions involving low energy QCD are added using tools such as PYTHIA or HERWIG.

- the emission of initial and final states radiation (ISR and FSR)
- the hadronization

Aside from the hard scattering process, other interactions may occur between the two protons. Those remnant are called *underlying events* and are also simulated by `PYTHIA` or `HERWIG`.

1.3.2 Higgs boson production mechanisms and decay modes

At p-p colliders, there are several Higgs boson production mechanisms. The mechanism with the highest cross-section is called *gluon fusion*. It consists of two gluons coming from each proton, and fusing into one Higgs boson through a top quark loop. The leading order diagram is drawn on the top left part of Fig 1.5. For $\mathrm{m_h} = 125.5$ GeV, the production cross-section computed at (NNLO+NNLL) QCD and NLO EW is $43.83^{+7.6}_{-8.1}$(scales) $\pm$ 3.1(PDF+α_S) pb [35]. This process has been studied a lot in the last years and is the first process calculated at next-to-next-to-next-to-leading order ($\mathrm{N^3LO}$) in perturbative QCD [36]. The $\mathrm{N^3LO}$ corrections are of the order of $+2.2\%$ and the scale variation of 3%.

The second highest cross-section mechanism is called *vector boson fusion*. In this case, two vector bosons, emitted by (anti-)quarks from the protons, fuse to produce a Higgs boson. The leading order diagrams are drawn on the top right part of Fig 1.5. For $\mathrm{m_h} = 125.5$ GeV, the production cross-section computed at approximated NNLO QCD and NLO EW is $3.767^{+0.4}_{-0.3}$(scales)± 2.1(PDF+α_S) pb [35].

The third mechanism is the associate production of a Higgs boson and a vector boson. The production is made of the annihilation of (anti-)quarks as incoming particles into an off-shell vector boson where the vector boson can be a $\mathrm{W^{\pm}}$ or a Z. The leading order diagram is drawn in the bottom left part of Fig 1.5. For $\mathrm{m_h} = 125.5$ GeV, the cross-section in association with a W boson calculated at NNLO QCD and NLO EW is $1.355^{+0.5}_{-0.7}$(scales) $\pm$ 1.9(PDF+α_S) pb [35]. The cross-section in association with a Z boson calculated at NNLO QCD and NLO EW is $0.8744^{+3.7}_{-3.1}$(scales) $\pm$ 1.6(PDF+α_S) pb [35].

The fourth mechanism is the associate production of Higgs boson and a pair of top quarks. The initial state particles are two gluons as drawn in the bottom right part of the Fig 1.5. For $\mathrm{m_h} = 125.5$ GeV, the cross-section calculated at NLO QCD and NLO EW is $0.5023^{+5.9}_{-9.3}$(scales) $\pm$ 3.6(PDF+α_S) pb [35].

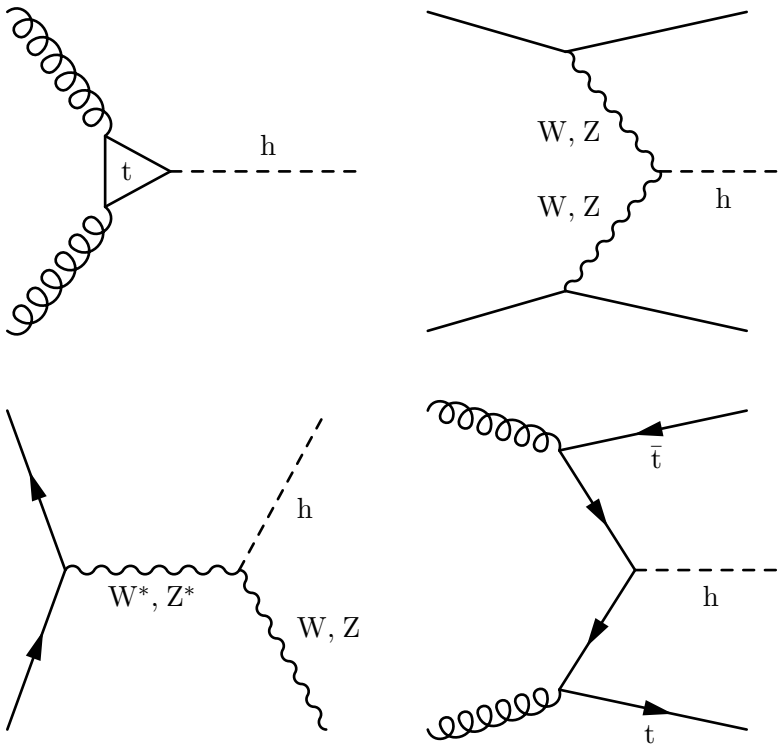

Figure 1.5: Top Left: Higgs boson production by gluon fusion. Top Right: Higgs boson production by vector boson production. Bottom Left: Associate production of a Higgs boson and a vector boson. Bottom Right: Associate production of a Higgs boson and a $t\bar{t}$ pair.

For masses close to 125 GeV, the Higgs boson can decay to a lot of different final states, including fermions and gauge bosons. The main fermionic decay mode is a pair of bottom quarks with a branching fraction of 0.575, followed by a pair of tau leptons with 0.062 [35]. The bosonic decays to $\mathrm{W^+W^-}$ leads with a branching fraction of 0.22 while the ZZ-decay reaches 0.027 and the

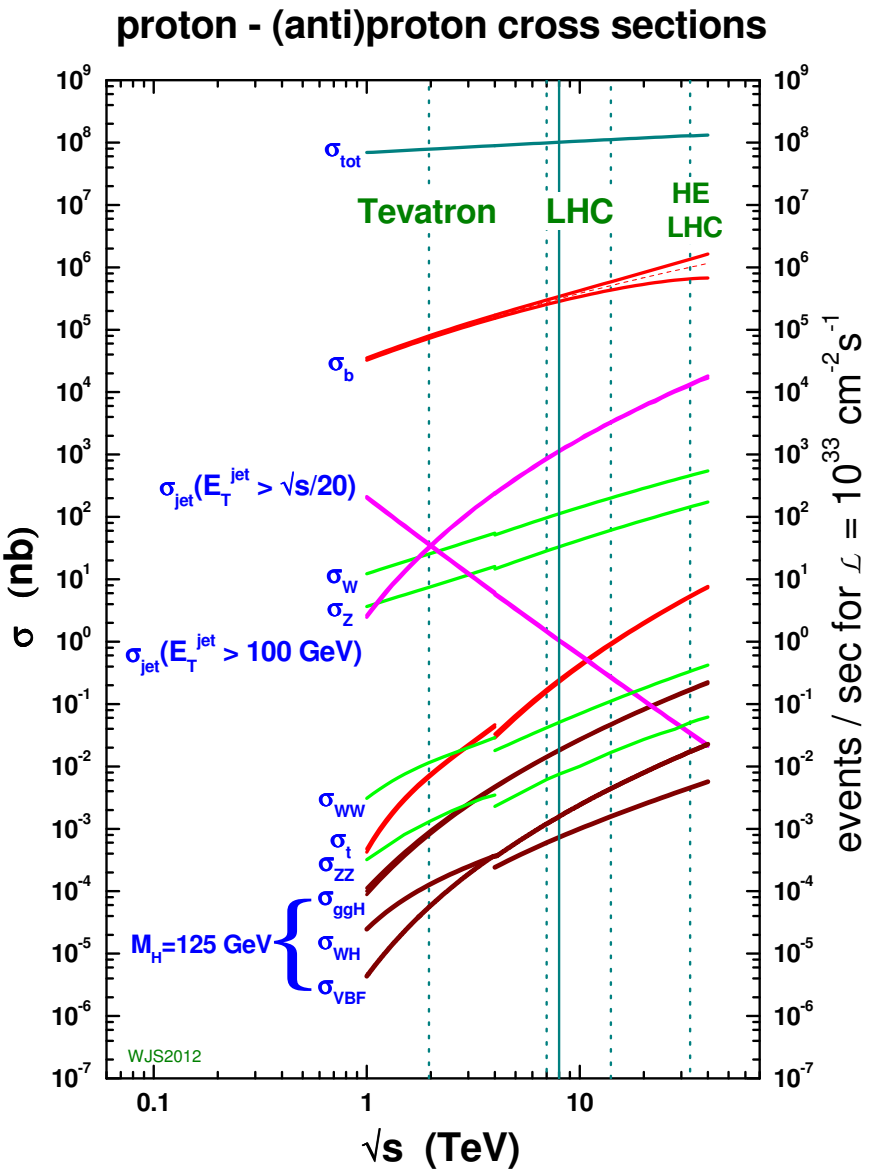

Figure 1.6: Standard model proton-(anti-)proton cross sections as a function of collider energy, including 33 TeV HE LHC [37].

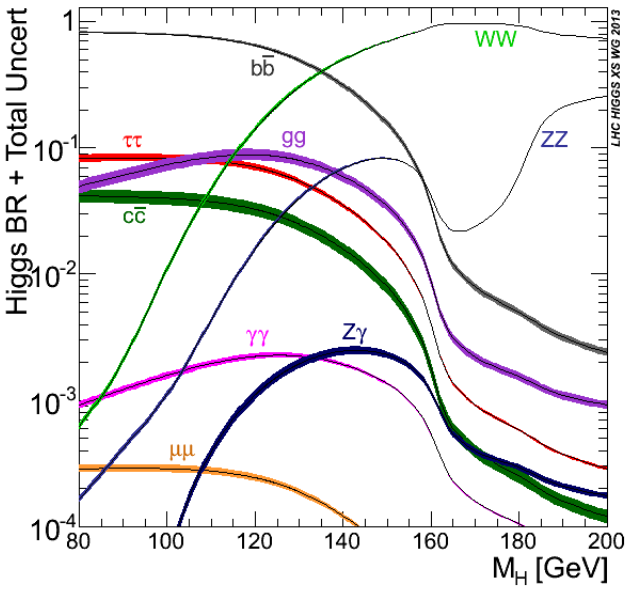

Figure 1.7: Standard model Higgs boson decay branching fractions for 80 GeV $\leq m_h \leq$ 200 GeV [35].

$h \rightarrow \gamma\gamma$ is 0.0023. Other decay modes and their variations with the Higgs boson mass are drawn in Fig 1.7.

1.3.3 Standard model Higgs boson: status

During the first run of data taking at the LHC, both CMS and ATLAS have observed a new boson with properties in agreement with the Higgs boson predictions. The mass measurement of this new particle obtained from the combination of CMS and ATLAS results[38] is summarized in Fig 1.8.

The resulting mass $m_h = 125.09 \pm 0.24$ GeV has been obtained using a combination of $h \rightarrow \gamma\gamma$ and $h \rightarrow 4l$ final states which provide the best mass resolution. This result fixes the value of the last free parameter of the Brout-Englert-Higgs mechanism, as explained in Sec 1.1.2. Other channels such as

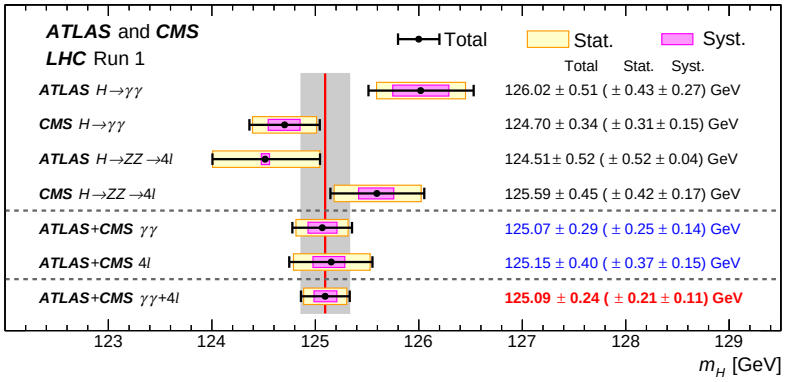

Figure 1.8: Higgs mass measurement from the combination of CMS and ATLAS results.

$h \to WW, \tau^+\tau^-, b\bar{b}$ and $\mu^+\mu^-$ have been probed, and provide a convincing agreement with the standard model predictions. The CMS and ATLAS combination note [39, 40] provides a summary of the different results. The most important results are probably the measured signal strength for the different channels and the significances of those measurements. These are reported in Tab 1.5.

Channel	ref ATLAS	ref CMS	μ ATLAS	μ CMS	σ ATLAS	σ CMS
$h \to \gamma\gamma$	[41]	[42]	$1.15^{+0.27}_{-0.25}$	$1.12^{+0.25}_{-0.23}$	5.0(4.6)	5.6(5.1)
$h \to$ ZZ	[43]	[44]	$1.51^{+0.39}_{-0.34}$	$1.05^{+0.32}_{-0.27}$	6.6(5.5)	7.0(6.8)
$h \to$ WW	[45, 46]	[47]	$1.23^{+0.23}_{-0.21}$	$0.91^{+0.24}_{-0.21}$	6.8(5.8)	4.8(5.6)
$h \to \tau^+\tau^-$	[48]	[49]	$1.41^{+0.40}_{-0.35}$	$0.89^{+0.31}_{-0.28}$	4.4(3.3)	3.4(3.7)
$h \to b\bar{b}$	[50]	[51]	$0.62^{+0.37}_{-0.36}$	$0.81^{+0.45}_{-0.42}$	1.7(2.7)	2.0(2.5)
$h \to \mu^+\mu^-$	[52]	[53]	-0.7 ± 3.6	0.8 ± 3.5		
$pp \to t\bar{t}h$	[54, 55]	[56]	$1.9^{0.8}_{-0.7}$	$2.9^{+1.0}_{-0.9}$	2.7(1.6)	3.6(1.3)

Table 1.5: ATLAS and CMS Higgs results: signal strength and significances for the different channels[39, 40].

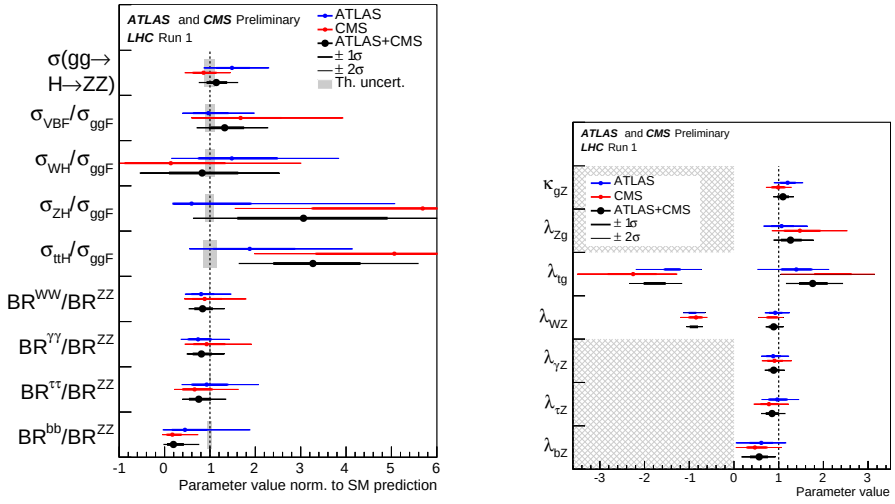

Figure 1.9: Left: Higgs boson production cross-section and branching ratios. Right: coupling modifers.

The interpretations of the measurements in terms of production cross section and branching ratios have then been computed and are presented in the left part of Fig 1.9. To reduce the systematic uncertainties and since the most precise measurements are the gluon fusion production mecanism and the $h \rightarrow 4l$ Branching ratio, those have been used as denominator.

To facilitate the interpretation, the *coupling modifiers* κ_i have been introduced in [57]. For the production or decay mode *i*, κ_i is defined as

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_i^{\mathrm{SM}}} \qquad \text{or} \qquad \kappa_i^2 = \frac{\Gamma^i}{\Gamma^i_{\mathrm{SM}}} \tag{1.45}$$

The coupling modifiers are defined for each individual couplings, including κ_g and κ_γ since new physics particles could contribute there in a different way. The results on the right part of Fig 1.9 are then presented with ratios of the coupling modifiers λ_{ij} defined as:

$$\lambda_{ij} = \frac{\kappa_i}{\kappa_j} \tag{1.46}$$

1.3.4 Phenomenology of a hierarchical 2HDM

Indirect constraints from h coupling measurements on the 2HDM

The measurement of the couplings of the newly observed boson described in 1.3.3 provides indirect constraints on the 2HDM parameters $\tan\beta$ and $\cos(\beta-\alpha)$. The CMS results are plotted in Fig 1.10 for both type I and type II [58]. Similar results have been obtained by the ATLAS collaboration [59]. The inputs used are the 7 and 8 TeV searches for h^{SM} decaying to $\gamma\gamma$, ZZ$(4l)$, WW$(2l2\nu)$, $\tau\tau$ and $b\bar{b}$. These results are very generic and are valid for any values of $\mathrm{m_A}, \mathrm{m_H} > \mathrm{m_h}$.

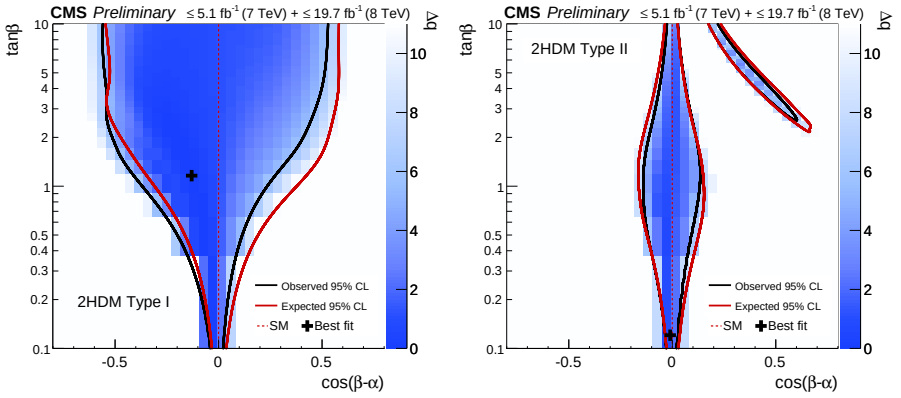

Figure 1.10: Excluded parameter space in white from h couplings measurement with the CMS detector for 2HDM type-I (left) and 2HDM type-II (right) [58].

As visible on Fig 1.10, this analysis strongly constrains $\cos(\beta-\alpha)$, in addition to the couplings to the vector bosons which are sensitive to $\cos(\beta-\alpha)$. Lower values of $\tan\beta$ in both type-I and type-II are partially excluded since the coupling to up-type quarks $\left(\text{proportional to } \frac{\cos\alpha}{\sin\beta}\right)$ would be enhanced (see Tab 1.4) Higher values of $\tan\beta$ in type-II are also partially excluded since the couplings to charged fermions proportional to $\sin\alpha/\cos\beta$ would be enhanced.

Since the coupling measurements are in agreement with the SM predictions, the alignment limit is strongly favored.

Direct searches: new scalar production mechanism

While indirect searches provide an important exclusion in the parameter space, they do not cover the full picture of the 2HDM status. For instance, we have seen that in the alignment limit, the light scalar from the 2HDM behaves exactly as the standard model Higgs boson, making this limit unreachable with only coupling measurements. It is therefore interesting to search for the processes that would be involved in this limit. Since in the alignment limit the coupling to vector bosons is negligible, only two production modes will be studied: the gluon-gluon-fusion with a top loop and the $b\bar{b}$-pair associated mechanism. Both are plotted in Fig 1.11.

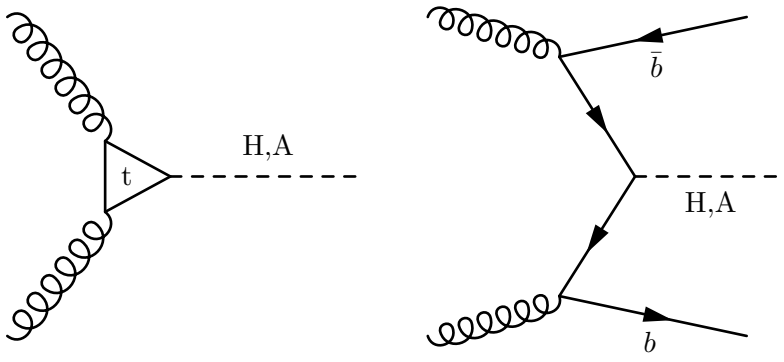

Figure 1.11: Left: Higgs boson production by gluon fusion. Right: Associate production of a Higgs boson and a $b\bar{b}$-pair.

Obviously, the cross-sections will depend on the different parameters of the model. The cross-section for the two mechanisms mentionned is plotted as a function of $\tan\beta$ in Fig 1.12. On can see that the gluon-gluon mechanism is decreasing with $\tan\beta$ since, from Tab 1.4, the coupling to the top is proportional to $\frac{\sin\alpha}{\sin\beta}$ in all types of 2HDM. On the other hand, in type-II, the coupling to the bottom quark is proportional to $\frac{\cos\alpha}{\cos\beta}$ which is therefore increasing the $b\bar{b}$-associated production as $\tan\beta$ increases.

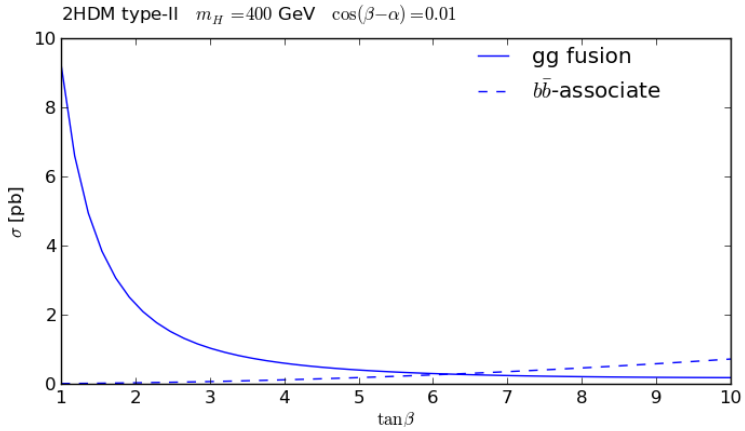

Figure 1.12: Cross-section for the production of a heavy scalar with $\mathrm{m_H} = 400$ GeV close to the alignment limit $\cos(\beta - \alpha)$ = 0.01. Two mechanisms are plotted, the gluon-gluon fusion with continuous line and the $b\bar{b}$-associated production with a dashed line. Those cross-sections are calculated at NLO in QCD and contain NNLO corrections due to the top-quark with the program SusHi [60].

Hierarchical 2HDM and exotic decays

Both CMS and ATLAS collaborations, at first, have extended their analyses searching for the Higgs boson to higher masses. Examples of such are searches for $H \to WW/ZZ$ [61] and $A \to Z\gamma$ [62]. But, in the alignment limit, decays of 2HDM scalars often involve other new (pseudo-)scalars. Those are called *exotic decays* since they involve cascades of new particles. To illustrate this statement, the branching fraction of both *A* and *H* are plotted in Fig 1.13 for the 2HDM type-I and in Fig 1.14 for the 2HDM type-II.

Some decay modes are particularly relevant in this context. Here is a non-exhaustive list of the final processes:

- $\mathrm{H/A} \to \mathrm{ZA/H} \to l^+l^-b\bar{b}/l^+l^-\tau^+\tau^-/l^+l^-t\bar{t}$ are of particular interest when $\Delta \mathrm{m}_{A-H}$ is bigger than $\mathrm{m_Z}$ [27, 28]. The $b\bar{b}$ and $\tau^+\tau^-$ final states are relevant for $\mathrm{m_A} < 2\mathrm{m_t}$. The $b\bar{b}$ final state has a higher cross-section and the $\tau\tau$ is very pure. The $t\bar{t}$ final state is dominant when $\mathrm{m_A} > 2\mathrm{m}_t$ and leads to a very clear signature.
- $\mathrm{h/H} \to \mathrm{AA} \to \mu^+\mu^-b\bar{b}/\mu^+\mu^-\tau^+\tau^-/\tau^+\tau^-\tau^+\tau^-/\mu^+\mu^-\mu^+\mu^-$ has been studied in the CMS collaboration considering the scalar as being at 125 GeV [64, 65, 66, 67, 68]. The $\mathrm{H} \to \mathrm{AA}$ is also of relevance in the Hierarchical 2HDM for $\tan\beta \neq 1$, as motivated by Figs 1.13 and 1.14.
- $A \to Zh \to llbb/ll\tau\tau$ has already been intensively studied within the CMS collaboration [69, 70]. The two final states provide very clear signatures. Unfortunately this process is strongly reduced in the alignment limit (Eq 1.43).
- $\mathrm{H} \to \mathrm{ZA} \to \mathrm{ZZh}$ is a promising channel for $\tan\beta \gtrsim 5$ and $\mathrm{m_A} < 2\mathrm{m_t}$ in type-I 2HDM, see Fig 1.13. In [71], lots of final states have been investigated. In particular $llllbb$ and $lljjbb$ provide the best limits.

Lots of final states involve two leptons and two b-jets. The searches for $\mathrm{H/A} \to \mathrm{ZA/H} \to l^+l^-b\bar{b}$ at the LHC is the main goal of this thesis. The two existing analyses at 8 and 13 TeV are presented in the third chapter.

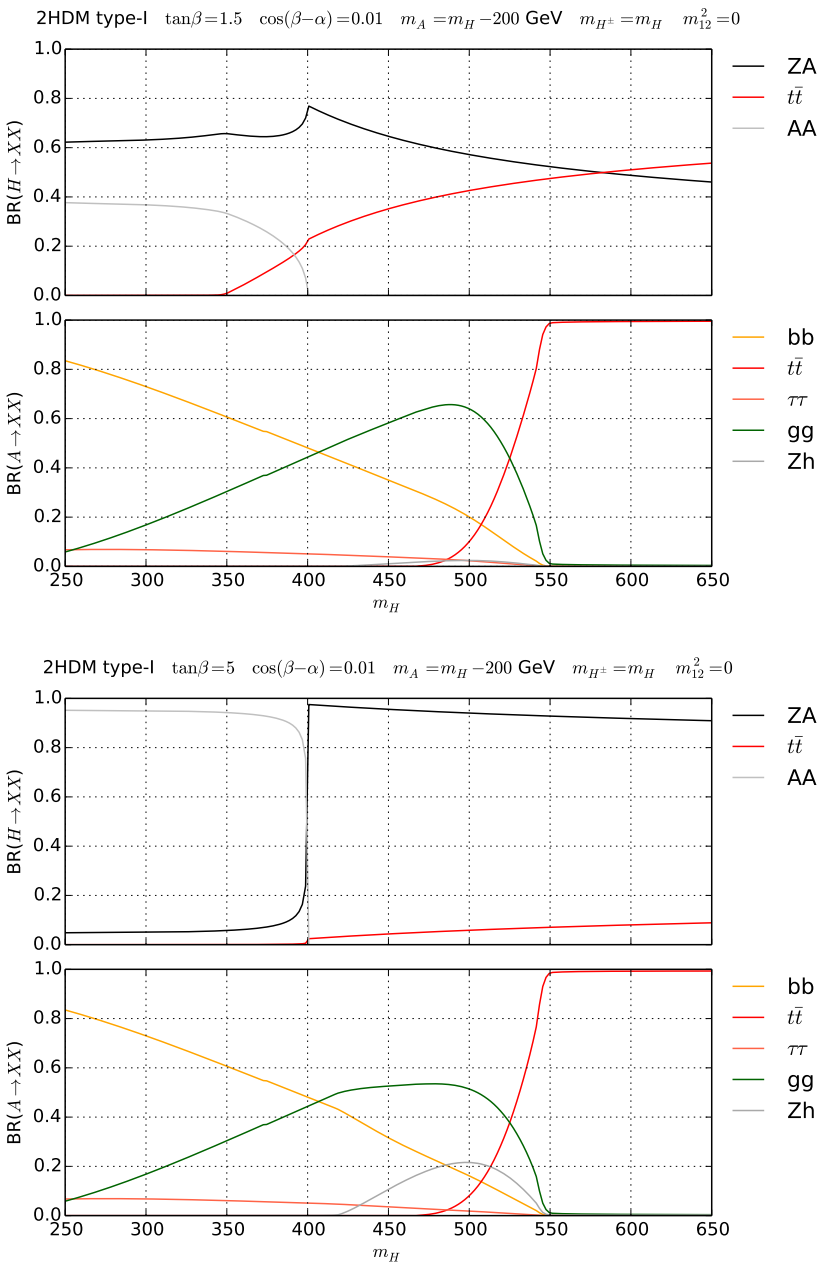

Figure 1.13: Branching fractions of the heavy scalar and the pseudoscalar in 2HDM type I, with $\mathrm{m_H} = \mathrm{m_A} + 200$ GeV, $\tan\beta = 1.5(5)$ and $\cos(\beta-\alpha) = 0.01$. Those values are obtained using 2HDMC [63].

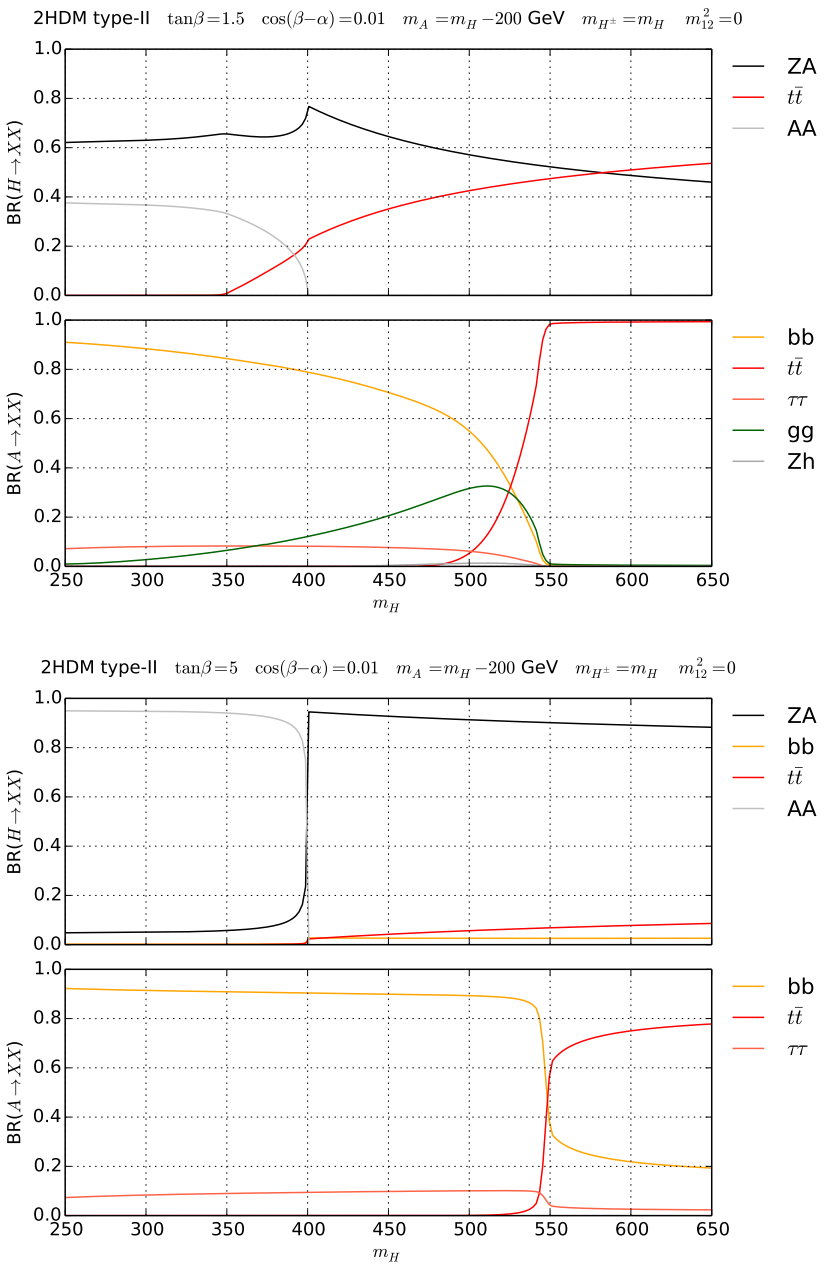

Figure 1.14: Branching fractions of the heavy scalar and the pseudoscalar in 2HDM type II, with $\mathrm{m_H} = \mathrm{m_A} + 200$ GeV, $\tan\beta = 1.5(5)$ and $\cos(\beta - \alpha) = 0.01$. Those values are obtained using 2HDMC [63].

Chapter 2

The CMS Experiment, Object Reconstruction and Simulations

After the discovery of the standard model Higgs boson, one goal of the LHC is to study the details of the scalar sector. In this context, the first chapter motivates the search for additional scalar bosons in final states containing two leptons and two b-jets. It is inspired by the 2HDM in which five degrees of freedom are to be probed. These degrees of freedom influence the masses of the new particles A and H, their width and the production cross section.

This chapter focuses on the description and simulation of the CMS detector as well as the reconstruction of leptons and b-jets. A precise simulation of the detector response is highly computing intensive, which is problematic when probing big new physics parameter spaces like the 2HDM, therefore an alternative parametric simulation framework called DELPHES is presented.

The first section focuses on the description of the CERN accelerators complex and the CMS sub-detectors. An overview of the different simulation methods used in high energy physics and an introduction to DELPHES are given in the second section. The third section describes the reconstruction of electrons, muons and b-tagged jets with the CMS detectors and comparisons with the parametric approach.

2.1 The CMS Detector at CERN

Close to Geneva, the most powerful particle accelerator has been built 100 meters under the French-Swiss border in a circular tunnel of 26.6 km of circumference. The Large Hadron Collider accelerates protons or heavy ions in both directions and collides the particles at four distinct places. Four particles detectors have been built around these points to study the high energy collisions with different goals.

- The ATLAS [72] and CMS [73] detectors are general purpose detectors. Both collaborations studies cover a wide range of physics including precise tests of the standard model (electro-weak measurement, QCD, top, Higgs, B-physics, etc.) and searches for physics beyond the standard model (super-symmetry, dark matter, extra dimensions, etc.).
- The LHCb [74] detector focuses on the physics of the b-quark by studying the particles and their decay when they are produced almost collinear to the beam.
- The ALICE [75] detector has been built specifically to study the quark-gluon plasma that is created in heavy ions collisions.

2.1.1 The LHC Accelerator Complex

At CERN, the proton accelerator complex is a succession of accelerators that progressively brings the protons to their nominal collision energy. The source of the protons is hydrogen gas from which the protons are first separated from the electrons in an electric field. The protons are then injected in a linear accelerator called LINAC2 that increases the energy of the protons to 50 MeV. The Proton Synchrotron Booster is composed of four superimposed synchrotron rings and brings further the energy up to 1.4 GeV. Next, the Proton Synchrotron, which is a 628 m circumference ring built in 1959, accelerates protons up to 25 GeV and injects them into the 6.9 km circumference Super

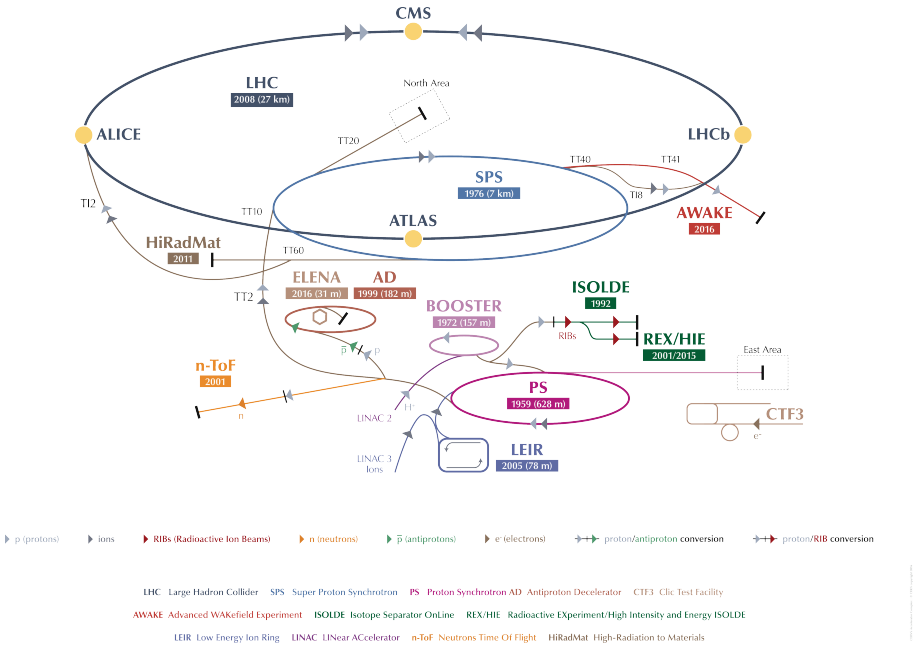

Figure 2.1: CERN's accelerator complex [77]

Proton Synchrotron. It accelerates the bunch of protons to 450 GeV and injects it into the LHC in the two directions. Fig 2.1 is a sketch of the accelerator complex and the LHC machine is described in [76].

One year after the 2008 inaugural tests, interrupted by a quench incident, the first operational run started in 2009 with 0.9 and 1.2 TeV in the center of mass. The LHC increased the beam energy up to 7 TeV in the center of mass in 2010. 45 pb^{-1} were delivered in 2010 and 6.1 fb^{-1} in 2011.

The center of mass energy has been increased further to 8 TeV in 2012 and 23.3 fb^{-1} were delivered, which ended the first run of the LHC. After a two-years long shut down (LS1), the run II of the LHC started with 13 TeV as energy in the center of mass. 4.2 fb^{-1} were delivered in 2015, the LHC aims at delivering 150 fb^{-1} during the run-II.

A second long shutdown is foreseen in 2019-2020 during which a first phase of detector upgrades will be completed. The LINAC4 is a H^- accelerator that

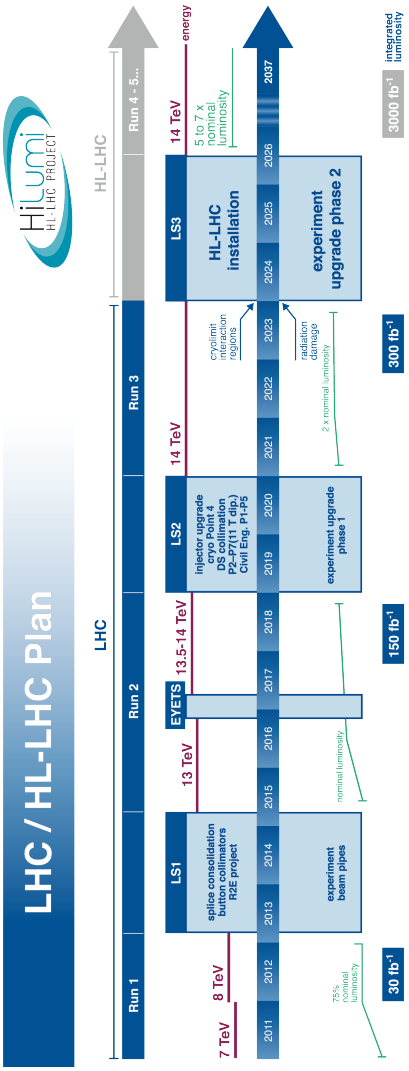

Figure 2.2: High-Luminosity LHC schedule [78].

will replace the current LINAC2. It will accelerate H^- up to 160 MeV, then remove the two electrons and inject the proton into the synchrotron booster. It is a key ingredient to increase the luminosity at the LHC. The run III is expected to start in 2021 with 14 TeV in the center of mass.

In 2024, the second phase of the upgrade will start with the installation of the High-Luminosity LHC. During the following runs, the HL-LHC is expected to deliver 3000 fb^{-1} at 14 TeV. The High-Luminosity [78] schedule is plotted in Fig 2.2.

At the LHC, several collisions per bunch-crossing occur in high luminosity conditions, most of them resulting in a small amount of activity in the surrounding detectors. Due to the elongated shape of the proton bunches constituting the beams, such *pile-up events* take place in a similarly elongated region (called the beam spot) around the nominal interaction point.

2.1.2 The CMS Detector

The *Compact Muon Solenoid* (CMS) is built inside and around a superconductive solenoid of 13 m with an inner radius of 2.95 m. It produces a quasi-homogeneous magnetic field of 3.8 T inside and about 2 T outside its volume. A sketch of the CMS detector is represented in Fig 2.3. It is a symmetric detector with an onion structure. Two kinds of detection methods are present in the CMS detector: trackers and calorimeters. The trackers are used to measure the trajectories of the charged particles in the magnetic field. They therefore provide information on the momentum of the particles. The inner tracker is placed around the beam pipe to detect all charged particles. The outer tracker is a second layer of trackers dubbed muon chambers are placed outside the solenoid to better measure the muons and to identify them. The calorimeters are used to measure the energy lost by the particles in their sensitive volume. The superconducting magnet provides the large bending power needed to measure the momenta of high energy charged particle with a sufficiently large bore to place the calorimeters inside the magnet volume.

A specific frame is attached to the CMS detector. The x and y axis are defined as the direction of the inward LHC radius and the upward vertical direction

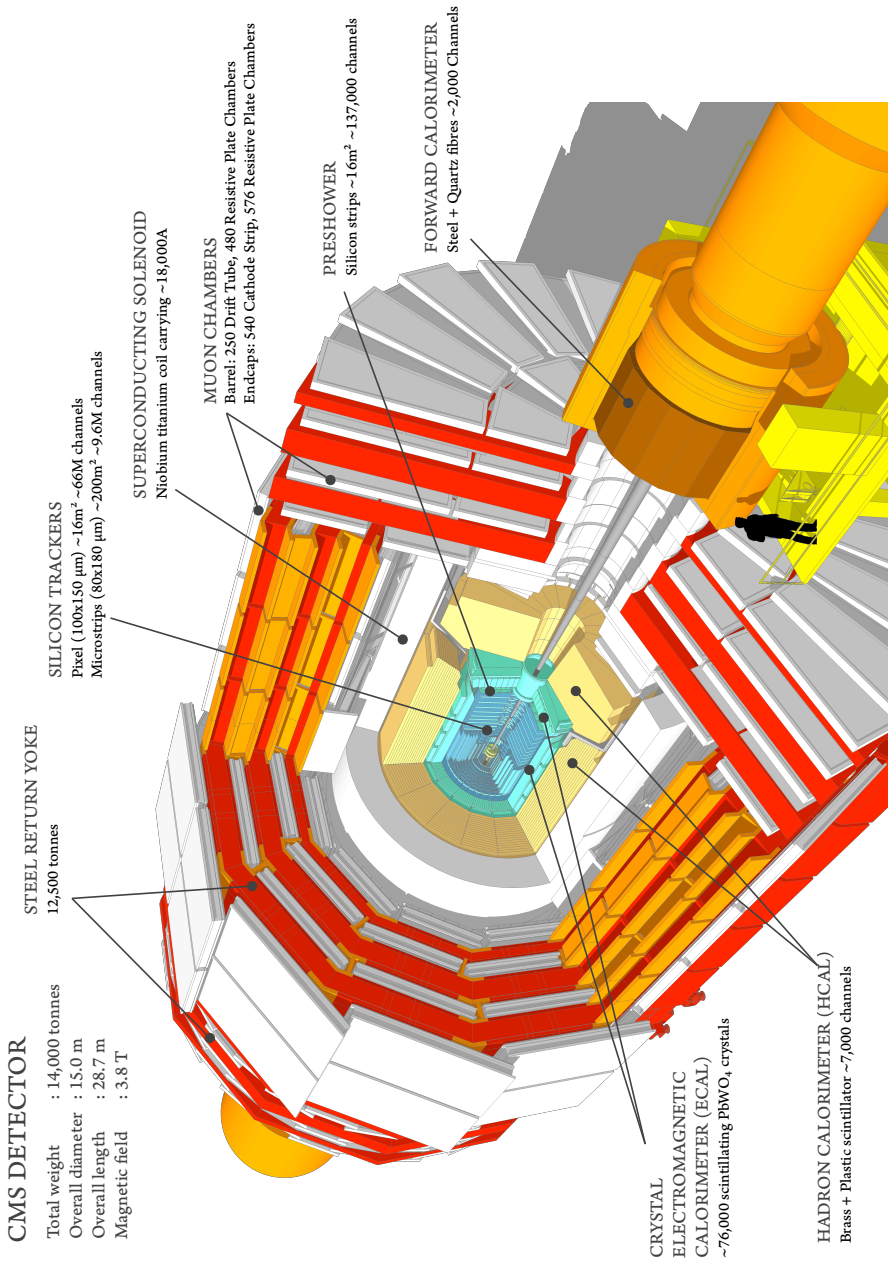

Figure 2.3: Sketch [79] of the CMS detector.

respectively. The z axis is defined as the direction of the beam pipe such that $\hat{x} \times \hat{y} = \hat{z}$ and points towards the Jura mountains. The pseudo-rapidity η defines the angle with the beam pipe and is defined as

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

where θ is the polar angle to the z axis. The azimuth angle ϕ is measured as the angle with the x axis in the $x - y$ plane and r is the radial coordinate.

Inner Tracking

The tracker is the closest detector to the interaction point. Its goal is to reconstruct tracks with a very high efficiency within the rapidity range $|\eta| < 2.5$. The geometry of the tracker is drawn in Fig 2.4. The CMS Tracker is entirely made of silicon detectors and is composed of two sub-detectors. The innermost part of the tracker is the *pixel tracker*.

- In the barrel region, it consists of three 53 cm long layers at $r = 4.4, 7.3$ and 10.2 cm.
- Two endcap disks on each side at $|z| = 34.5$ and 46.5 cm covering a radius from 6 to 15 cm.

Each layer is composed of $100\,\mu\text{m} \times 150\,\mu\text{m}$ pixels. The silicon strip tracker is the second part of the tracker and covers the region with radius between 20 and 110 cm.

- The *Tracker Inner Barrel* (TIB) is the inner part of the strip tracker. The four layers cover up to $|z| < 65$ cm with a cell size of 10 cm $\times$ $80(120)$ μm. The outermost part is the *Tracker Outer Barrel* (TOB). It consist of six layers covering $|z| < 110$ cm with a cell size of 25 cm $\times$ $120(180)$ μm. The two first layers of the TIB and of the TOB are combined with slightly tilted modules called *stereo modules* in order to improve the tracks reconstruction.

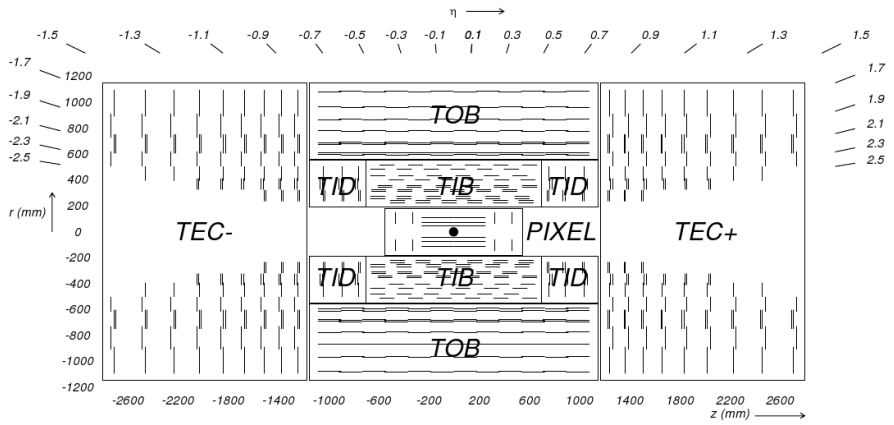

Figure 2.4: Detailed geometry of the CMS inner tracker.

- Each *Tracker End Cap* (TEC) is composed of nine vertical disks that cover the region 120 cm $< |z| < 280$ cm. The *Tracker Inner Disks* (TID) cover the region between the TEC and the TIB with three small disks.

The tracker covers the region in pseudo-rapidity $|\eta| < 2.5$.

The presence of silicon detectors implies the presence of a relatively high material quantity in the tracker volume. The material budget in unit of radiation length is plotted in Fig 2.5.

The event reconstruction in the tracker starts by identifying clusters as groups of neighboring cells with a signal significantly higher than the noise. A *hit* is a cluster with its position information and is the basic ingredient for track reconstruction. Once the track angle is known, another pass of the pixel cluster reconstruction algorithm is performed, based the cluster shape which significantly improves the resolution. The efficiency of the hit reconstruction has been measured to be higher than 99% in the pixel detector and overall 99.8% in the strip detector when excluding the defective modules. The resolution in $r\phi$ is 9.4μm and between 20 and 45μm in z.

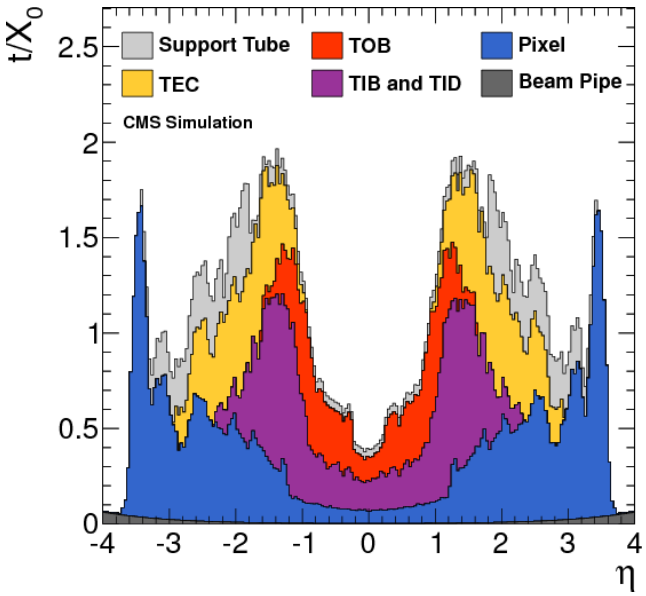

Figure 2.5: Material budget of the CMS tracker in units of radiation length χ_0

Electromagnetic Calorimeter

The Electromagnetic calorimeter (ECAL) is a hermetic homogeneous detector built of lead tungstate ($PbWO_4$) crystals. The barrel region (EB) is composed of 61200 crystals with a front face cross-section of 22×22 mm^2 and a length of 230 mm($25.8\chi_0$) placed with an inner radius of 129 cm and tilted at 3°. The endcaps (EE) are 7324 crystals with a front face of 28.6×28.6 mm^2 and a length of 220 mm ($24.7\chi_0$). It covers the rapidity range $1.479 < |\eta| < 3.0$. Two planes of lead radiator followed by silicon strips with a pitch of 1.9 mm and a radiation length of $2\,\chi_0$ and $1\,\chi_0$ are placed over much of the endcap rapidity range.

The energy resolution of the ECAL is usually parametrized by

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.2}$$

Where S is called the stochastic term, N the noise and C the constant term.

These terms have been estimated using simulation and test beams. In the barrel (endcap) region, the stochastic term is $\frac{2.7(5.7)}{\sqrt{E}}\%$, the noise term is measured to be $155(770)$ MeV and the constant term is $0.55(0.55)\%$ [80].

Hadronic Calorimeter

The hadronic calorimeter (HCAL) surrounds the ECAL system. To minimize the non-gaussian tails and hermicity of the measurement, the design maximizes the material inside the magnet. The HCAL is divided in four distinct parts. The CMS barrel and endcap sampling calorimeters use scintillator tiles as active medium combined with brass as absorber material.

- The Hadron Barrel (HB) covers the region $-1.4 < \eta < 1.4$ with 2304 towers. Each tower has a segmentation of the order $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. After crossing the ECAL, the particles cross a 9 mm scintillator. Then 15 plates of brass are separated with 3.7 mm scintillator.
- The Hadron Endcaps (HE) consist in 2304 towers and cover the pseudo-rapidity $1.3 < |\eta| < 3.0$. The segmentations vary from $\Delta\eta = 0.087$ to 0.35 and $\Delta\phi = 5^\circ$ to 10°.
- The Hadron Outer (HO) is an extra layer of scintillators located outside the magnet and covers the region $-1.26 < \eta < 1.26$ using the solenoid coil as an extra absorber. Its goal is to increase the number of interaction lengths and therefore reduce the tail in the energy resolution distribution.
- The Hadron Forward (HF) covers the pseudo-rapidity region from 3 to 5 and is exposed to huge particle fluxes. These particularly harsh conditions are the main motivation to use Cerenkov quartz fibres as active medium. The fibres are inserted in the steel absorber structure and photomultipliers are used to measure the energy deposit.

Muon Chambers

Three types of gaseous detectors are used to identify and measure muons:

- The Drift Tubes (DT) are located in the barrel region, where the neutron and muon fluxes are small and the magnetic field uniform. The four stations of DT chambers that cover the region $|\eta| < 1.2$ are interspersed among the layers of the flux return plates.
- In the endcap region, the muon flux is high and the magnetic field is large and non-uniform, therefore four stations of Cathode Strip Chambers (CSC) are used. CSCs are fast detectors with fine segmentation that cover the range $0.9 < |\eta| < 2.4$.
- In addition, some Resistive plate Chambers (RPC) are installed in the barrel and in the endcaps. These are fast gaseous detectors providing a muon trigger system parallel to those of the DTs and CSCs.

2.2 Detector Simulations

Testing the standard model requires a precise modeling of the proton-proton interaction and the detector response. These simulations are based on Monte-Carlo methods and aim at directly comparing reconstructed quantities from the data with the simulations. This is used to validate our understanding of various processes as well as to search for new physics. Indeed, simulations are often used to estimate the backgrounds and are very useful for optimizing and validating the analyses. In this thesis, simulations are also used to provide a signal model.

Currently, the chain containing the most precise modeling of the proton-proton collision and the object reconstruction are obtained with the following recipe.

1. The physics occurring at the collision point is modeled by the *event generators* described in Sec 1.3.1 with an NLO precision, re-weighted to (N)NNLO cross section.
2. The produced particles are then propagated through the *detector*. The *detector simulation* consists in the simulation of the particle-matter interaction in a very fine step using a very detailed description of the detector geometry. It includes a precise simulation of the physics processes

such as electro-magnetic showering, nuclear interactions, bremsstrahlung, photon conversions, etc. This step is currently based on GEANT 4 [81].

3. The emulation of the *digitization* process uses up-to-date models to translate the energy or charge deposits into electronic signals.

4. The *object reconstruction* is the last step and consists in applying algorithms to reconstruct objects (electrons, muons, jets, etc.) from the electronic signal information. In principle, this step is the same for both the data and the simulation. Sec 2.3 is dedicated to the description of the different objects and their reconstruction.

This procedure called *full simulation* provides the best data/simulation agreement. It is also very time consuming: the complete detector simulation, digitization and reconstruction procedure of one $t\bar{t}$-event in the ATLAS or CMS detector takes around 1-10 minutes depending on the pile-up conditions. To put this in perspective, during the second run of the LHC, 150 fb are expected to be delivered, which means approximately 140×10^6 $t\bar{t}$ events.

Speeding up the simulations It is possible to increase the speed of the simulation in many ways. The most common techniques are:

- the use of approximate detector geometries.
- using parametrization in the digitization.
- using libraries of pre-simulated showers (e.g. FrozenShowers [82])
- parametrization of the calorimeter response (e.g. GFlash [83], FastCaloSim [84], etc.)
- using *MC-truth* information to speed up object reconstruction, for example when tracking particles in central detectors.

When some of these techniques are used, the simulations are called *Fast Simulations* since the speed is typically increased by a factor 10-100. Fast simulations are mainly used by the experiments to simulate signal events when

big parameter scans are needed, such as in supersymmetry. For instance, the CMS experiment fast simulation [85] is approximately 100 times faster than the GEANT 4-based simulation.

the parametric approach Instead of speeding up the three steps separately, one can just smear the generated jets and leptons or all the particles to simulate the detector response. This method is extremely fast and provides results with precision driven by the quality of the parametrizations.

An example of a pure parametric approach is the TURBOSIM algorithm [86]. It uses full-simulated events to produce a lookup table matching one or more partons to 0 or more reconstructed objects. This method provides, by construction, the inefficiencies, the misidentification probabilities and the merging-overlap of particles. A TURBOSIM table of half a million lines with a total size of ≈ 100 MB has been produced using the H1 detector simulation [87]. The table is then interpolated and used to perform the detector simulation. The resulting simulation runs at roughly 10 ms per event (in 2006) and is in agreement with the full simulation.

The Pretty Good Simulation (PGS) and DELPHES programs are also parametric simulation. The philosophy is to first perform a very basic detector simulation with tracks and calorimeter deposits, then use these tracks and calorimeters to reconstruct the physical objects. The method is therefore a hybrid between a purely parametric and a fast simulation. PGS has been initially designed for CDF and D0 detectors at the Tevatron while the first version of DELPHES was more dedicated to the simulation of LHC experiments.

Since the time per event is reduced to 10 ms, this type of simulation has a very wide scope of applications.

- Parametric simulations are used to design an analysis and bring first estimations of the discovery potential of an analysis.
- These very fast simulations are ideal to recast existing analyses with tools such as CHECKMATE [88] and MADANALYSIS 5 [**Conte:2012fm**] [**Conte:2014zja**] [89].

- When developing detectors, one can test the impact of new geometries or resolutions on the physical results (e.g. analysis forecast for high-PU scenarios)
- It is also used a lot for pedagogical purposes, since it avoids all the complexity of the full simulation.

2.2.1 CMSSW and the CMS detector simulation

The CMS collection of software, dubbed CMSSW, aims at facilitating the development of the reconstruction and the analysis software. It is built around a modular framework, an Event Data Model (EDM) and the services needed by the simulation, the calibration and alignment and reconstruction modules [90]. An *event* is a C++ object. It is a container for all *raw* and reconstructed data for a proton-proton collision. The events based on both the detector and the Monte-Carlo data are processed with the same executable called *cmsRun*.

The Fig 2.6 is a work-flow chart of the simulation program [91] implemented in CMSSW. Events generators provide a collection of particle 4-vectors which are provided to the GEANT 4-based simulation code. Among others, the geometry of the detector, the material content and the magnetic field are provided to the simulation code and are configurable at run time.

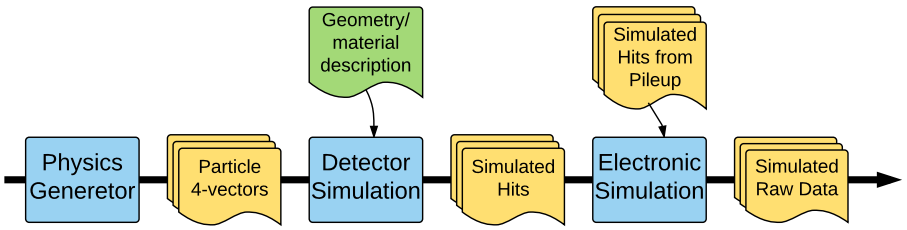

Figure 2.6: Flow chart of the CMS simulation program, inspired from [91].

The number of in-time and out-of-time pileup events are selected randomly from a poisson distribution based on the luminosity and the total inelastic cross-section. The pileup events are extracted from a collection of pre-processed GEANT 4 simulated hits from minimum bias events generated with PYTHIA6

and 8. The hard scattering event is then merged with these pileup events and processed by digitization or electronic simulation.

The complexity of the CMS simulation contrasts with the parametric approach taken in DELPHES. To illustrate this statement, here are a few facts about each sub-detector simulation:

- The tracker requires a high degree of accuracy in the active and passive component description and a precise treatment of bremsstrahlung.
- The requirement for the ECAL subdetector is an accurate description of the geometry and material budget and a complete implementation of the electromagnetic physics process.
- The HCAL simulation is very time consuming, therefore shower libraries and noise libraries are used.
- The muon physics description up to 10 TeV in GEANT 4 has been improved and validated at the LHC. It includes μ-bremsstrahlung, μ-nuclear effects and multiple scattering, which are requirement for muon detectors.

2.2.2 DELPHES 3

DELPHES3 is a modular framework written in C++ that takes as input the most common event generator output and performs a fast and realistic detector simulation of a general purpose collider detector. The project has been very successful and has lead to a publication with more than 600 citations and many conference presentations. Writing, testing and validating modules are part of the work done in this thesis.

With respect to its previous incarnation [92], the present version of DELPHES [93] includes a successul attempt to emulate the particle-flow reconstruction philosophy used in ALEPH [94] and CMS [95], based on the optimally-combined use of the information from all the subdetectors to reconstruct and identify all particles indvidually. While the aim is not to re-implement the particle-flow algorithm in all its complexity, the simplified approach adopted is particularly

suitable for the treatment of pile-up, as well as for the emulation of b and τ tagging, and is able to reproduce the jet and missing energy resolutions observed in CMS with their complete reconstruction.

DELPHES validation suite From a technical perspective, the code structure is fully modular, providing great flexibility to the user and allowing the integration of DELPHES routines in other projects. The DELPHES developments are community-based and users are strongly encouraged to participate to the improvement of the program. It is therefore important to provide an automatic procedure to test the effect of any modification to the code (e.g. new efficiency or resolution paramatrization, new energy flow algorithm, etc.). The DELPHES *particle gun* has been developed for that purpose. It uses `PYTHIA` to generate single-electron, -muons, -taus, -photons events or dijets event with a $q\bar{q}$ or a gluon-pair. The DELPHES detector simulation is performed and the output is used to produce a standard set of validation plots, including the resolution and reconstruction efficiencies for each object.

DELPHES detector simulation An example of flow-chart is plotted in Fig 2.7 where one can find an overview of the basic modules in the rectangular boxes. In a parametric simulation, it is not straightforward to divide the detector simulation part from the object reconstruction since some of them are performed simultaneously. For example, one can take a generated electron and modify its 4-momentum according the experimental resolution to produce a reconstructed electron. This step therefore reproduces the results of both the detector simulation and an object reconstruction. In this section, only the pure simulation steps are discussed.

In DELPHES, only in time pile-up events are considered in the pile-up merger module. Minimum bias interactions are extracted from a pre-generated low-Q^2 QCD sample and randomly placed along the beam axis according to a given longitudinal spread that can be set by the user. The actual number of pile-up interactions per bunch-crossing is randomly extracted from a Poisson distribution.

The particles are propagated through the tracker in the magnetic field, neglecting the interactions with the material of the tracker. Therefore, the long-lived

charged particles are propagated through the tracker volume according to a helicoidal trajectory. The long-lived neutral particles are propagated straight.

The calorimeter cells are defined in the (η, ϕ)-plane. When a particle reaches the calorimeters, its energy fraction deposited depends on its type. The fractions f_{ECAL} and f_{HCAL} correspond to the assumed deposition ratios are summarized in 2.1.

Table 2.1: Energy fraction deposited in each calorimeter for the different long-lived particles

	f_{ECAL}	f_{HCAL}
e, γ	100%	0%
μ, ν	0%	0%
K_s^0, Λ	30%	70%
charged hadrons	0%	100%

The ECAL and HCAL energy deposits are independently smeared with a log-normal distribution centered at the generated energy and with the corresponding resolution as standard deviation. The resolutions are parametrized as a function of the energy of the particle and the pseudorapidity of the reached calorimeter cell:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S(\eta)}{\sqrt{E}}\right)^2 + \left(\frac{N(\eta)}{E}\right)^2 + C(\eta)^2 \tag{2.3}$$

The ECAL parametrization is obtained from [96, 97], from which a 2D resolution is extracted.

$$\begin{aligned}
\sigma = &\quad (1 + 0.64\eta^2)\sqrt{E^2 0.008^2 + E 0.11^2 + 0.40^2} && |\eta| \leq 1.5 \\
\sigma = &\quad (2.16 + 5.6(\eta - 2)^2)\sqrt{E^2 0.008^2 + E 0.11^2 + 0.40^2} && 1.5 < |\eta| \leq 2.5 \\
\sigma = &\quad \sqrt{E^2 0.107^2 + E 2.08^2} && 2.5 < |\eta| \leq 5.0
\end{aligned}$$

The HCAL parametrization is assumed to be:

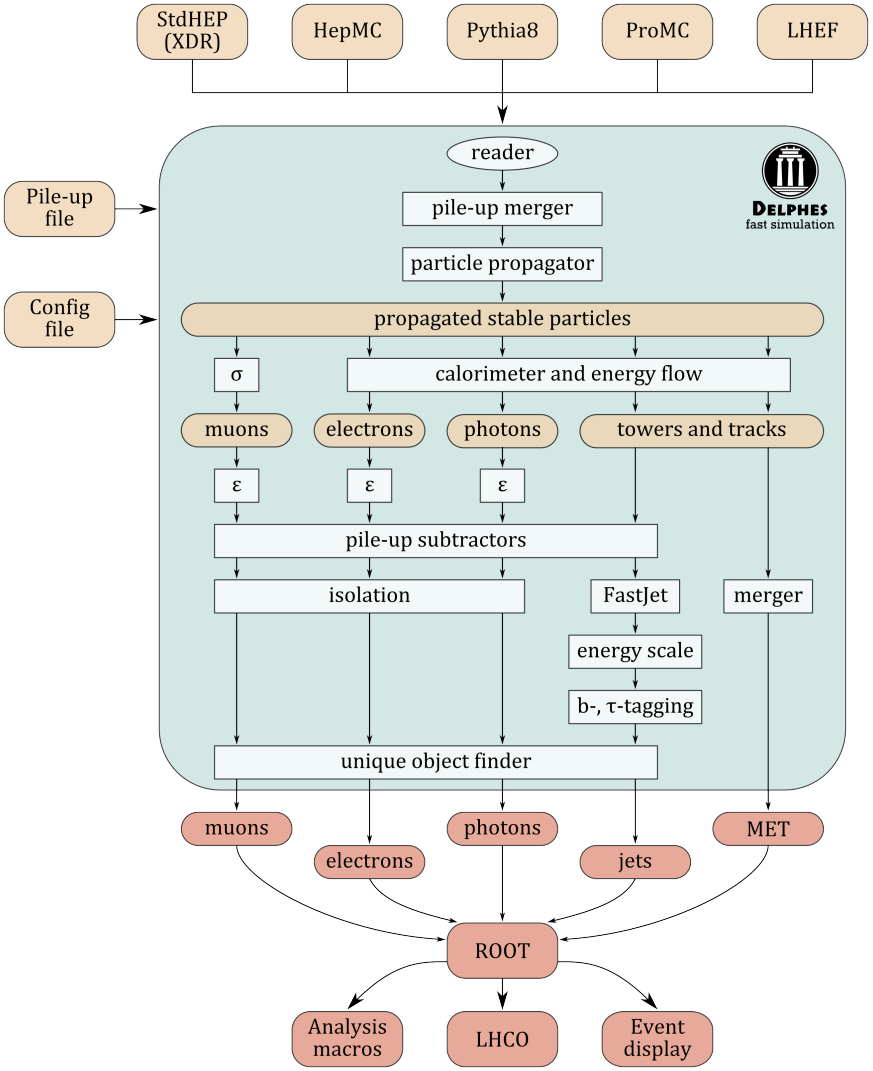

Figure 2.7: Typical work-flow chart of the DELPHES fast simulation. Event files coming from external Monte-Carlo generators are first processed by a reader stage (top). The output data are stored in a ROOT tree format (bottom) and can be analyzed and visualized with the help of the ROOT data analysis framework. The ROOT tree files can be also converted to the LHCO file format. Each step is controlled by the configuration file.

$$\begin{aligned}\sigma = & \sqrt{E^2 0.050^2 + E 1.50^2} & |\eta| \leq 3.0 \\ \sigma = & \sqrt{E^2 0.130^2 + E 2.70^2} & 3.0 < |\eta| \leq 5.0\end{aligned}$$

As a result, jets, the missing energy, isolated electrons, muons and photons, and taus can be reconstructed, as described in Sec 2.3.

2.3 Object reconstruction: from complete to parametric simulation

It is important to understand the limitations and advantages of the parametric approach adopted in DELPHES. To do so, the object reconstruction in CMS will be reviewed and compared with DELPHES with the CMS configuration card used in this thesis.

2.3.1 Tracking

The tracking is the reconstruction of the charged particles trajectories in the inner tracker using the hits positions. The algorithm developed in CMS is called *iterative tracking*. It consists in several passes of the Combinatorial Track Finder (CTF) which is in the following sequence:

- The *seed generation* using only a few hits (2 or 3) provides initial track candidates.
- The *track finding* extrapolates the seed and check if extra hits can be added to the seed. This step is based on a Kalman filter.
- The *track fitting* estimates the final track parameters using position of all hits associated to the same charged particle.
- The *track selection* selects the tracks based on quality requirement.

The basic idea is to reconstruct the easiest tracks during the first iterations: The iteration "0" is the source of most reconstructed tracks with $p_T > 0.8$ GeV and three reconstructed pixel hits. Iteration "1" aims at reconstructing tracks with only two pixel hits while the "2" recovers low-p_T tracks. The steps "3" to "5" are designed to reconstruct tracks originating outside the beam spot. Between each iteration, hits associated with high purity tracks are removed which reduce the combinatorial complexity.

Parametric tracking

The particles follow ideal trajectories in the DELPHES tracker while track reconstruction implies inefficiencies and imprecisions. The track reconstruction is mimicked by applying to these trajectories a probability to be reconstructed (an *efficiency*) as a track and a degradation of the transverse momentum (a *smearing*). The efficiency and the smearing are function of η and p_T. Since parametric reconstruction doesn't need to be applied on the data, one can use the information on the type of particles that are going through the tracker volume. The tracks are therefore divided in three different categories with trajectories originating from electrons, muons and charged hadrons and smeared accordingly.

In the case of the CMS detector, the functions are computed using the full detector simulation and standard reconstruction. When no complete simulation is available yet, tools such as TKLAYOUT [98] can be used to estimate the tracking performances for a given geometry.

The tracking efficiencies are plotted in Fig 2.8. While the efficiencies are by construction correctly taken into account, the disadvantage of the parametric approach is the difficulty to implement the *fake tracks*. One can, for instance, define a fake track as a reconstructed track that has less than three reconstructed hits corresponding to the simulated hits originating from a generated particle. The fake rate is strongly dependent on the local multiplicity of tracks and therefore on the pileup conditions. It is also linked to the non-reconstructed tracks around it, from which one or two hits are shared. It is therefore very difficult to emulate those fakes with a parametric approach.

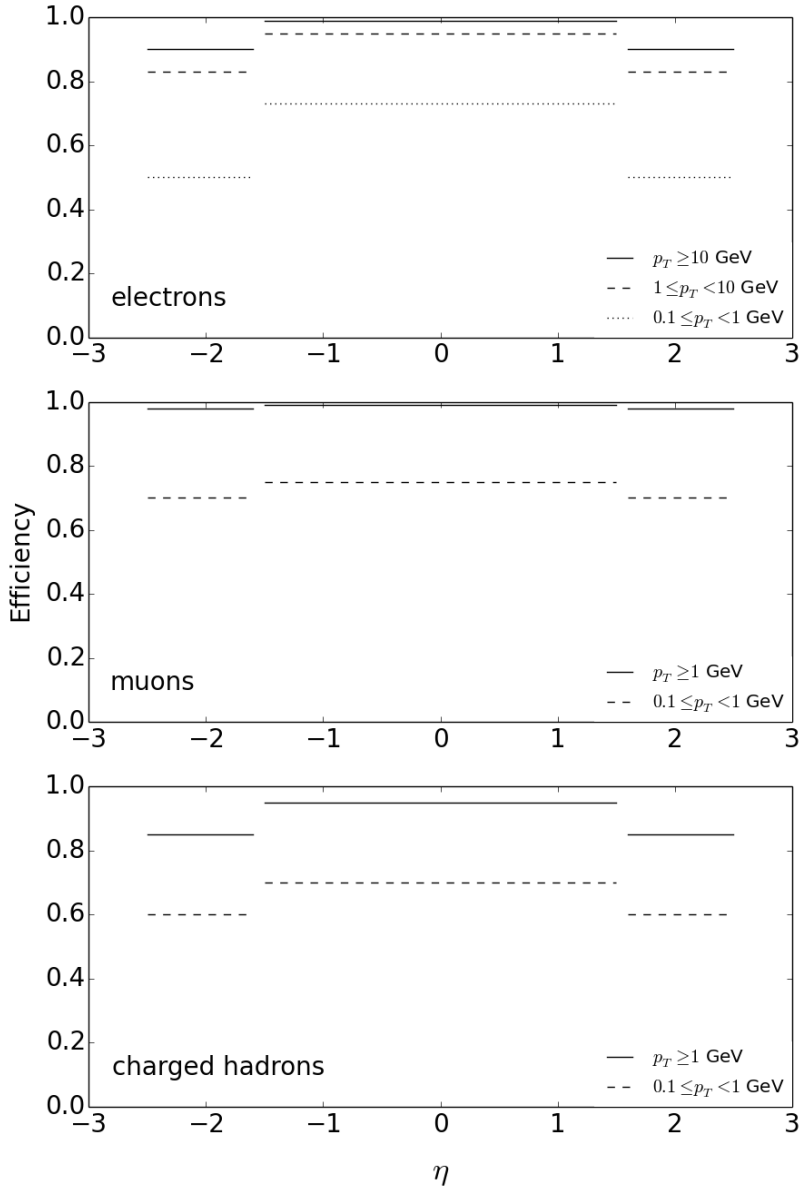

Figure 2.8: Tracking efficiencies (input) in DELPHES for the electrons(top), muons(middle) and charged hadrons(bottom).

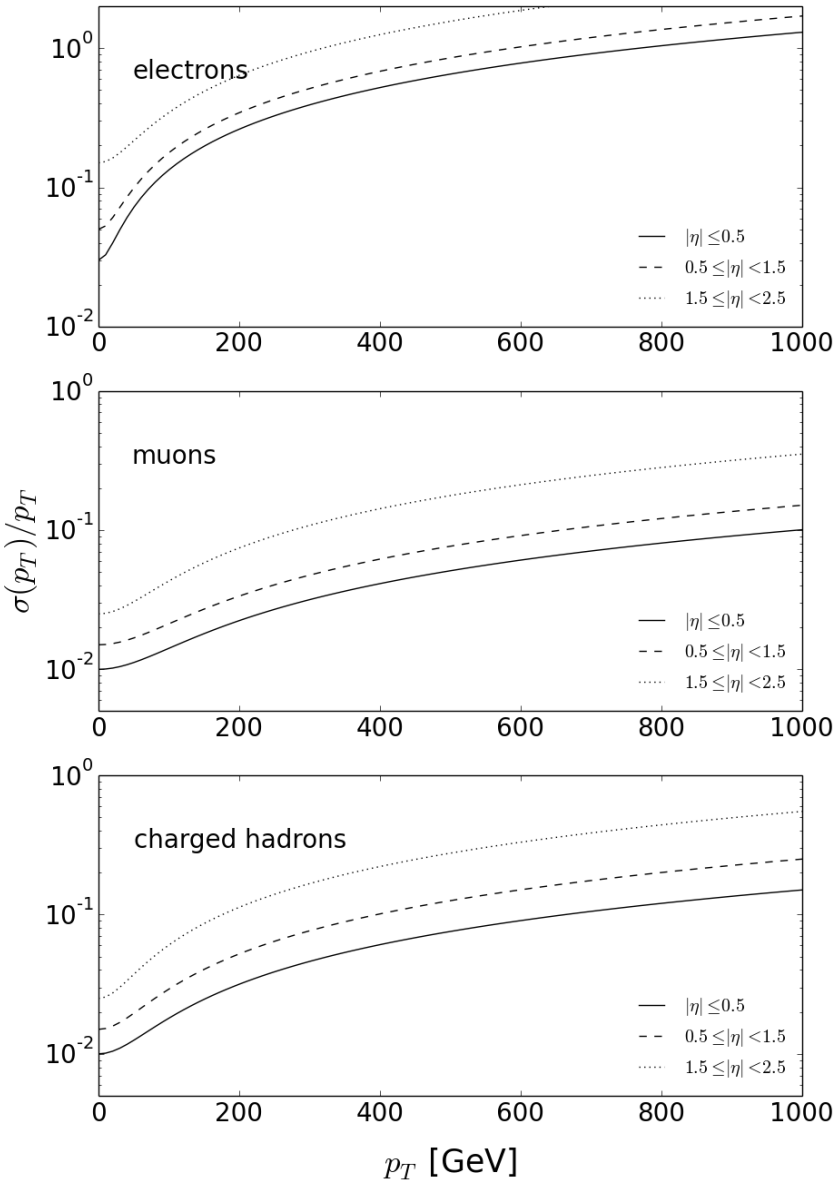

Figure 2.9: Track momentum smearing (input) in DELPHES for the electrons(top), muons(middle) and charged hadrons(bottom).

2.3.2 Calorimeter clustering

The role of the calorimeter clustering [99] is essential to detect the stable neutral particles. It is a key ingredient in the reconstruction of the electrons and the bremsstrahlung photons. It also allows to improve the charged particles measurement at high-p_T since high-p_T tracks are not measured accurately.

The clustering consists in identifiying *cluster seeds* as local maxima of energy deposits. The neighboring cells passing the threshold are then combined with the seed if they have at least one corner in common to form the *topological clusters*.

This procedure is performed in each calorimeter separately. It aims for high detection efficiency while keeping separated close energy deposits.

Parametric calorimeter clustering

In the DELPHES framework, no energy sharing is implemented between neighboring calorimeter cells. Therefore, the clustering is not needed, and not implemented in the `CMS` configuration card.

2.3.3 Linking and particle flow

In general, particles are expected to induce the creation of clusters and/or tracks. For instance high energy muons are likely to leave tracks, in the inner and outer tracker while electrons typically generate an inner track and an ECAL deposit, as illustrated in Fig 2.10. The key ingredient to identify the particles with a detector like `CMS` is to properly link the different elements. Once linked, one can combine the different measurements which are assumed to be produced by the same particle. This technique, called *particle flow* [99, 100] is heavily used within the `CMS` collaboration.

The charged particle tracks are first extrapolated to the PS (in the endcaps), the ECAL and the HCAL. The depth of the extrapolation depends on the expected longitudinal electron profile in the ECAL and to one interaction length in the

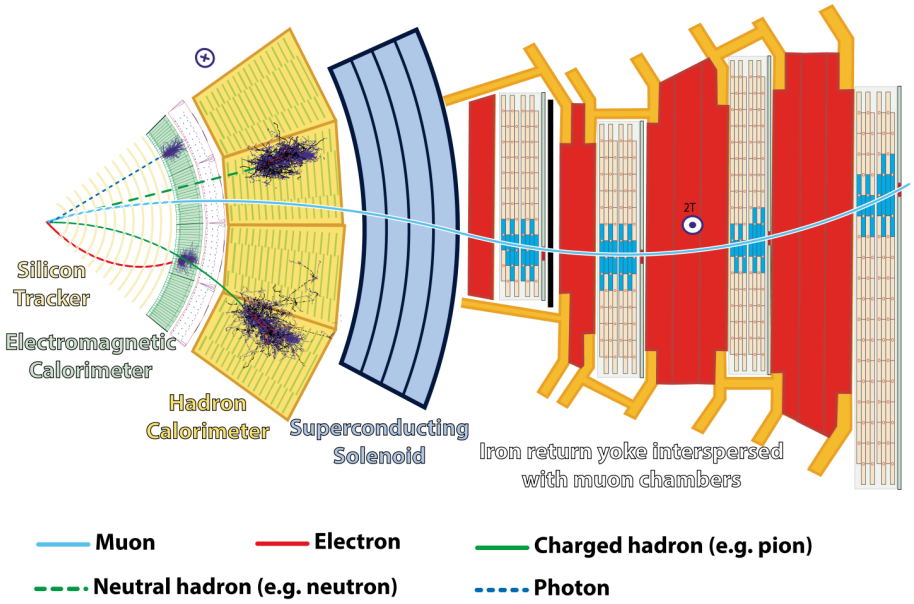

Figure 2.10: CMS slice [101]

HCAL. A link between the charged particle track and a calorimeter cluster is created if the extrapolation crosses the cluster boundaries. Links between ECAL and HCAL are also created when the ECAL cluster is within the HCAL cluster envelope. In addition, links between charged particle tracks and muon tracks, called *global muons*, are created when a global fit between the two tracks return a satisfactory χ^2.

Particle flow approach in DELPHES

A simplified version of the particle flow aiming at reproducing properly the CMS object resolutions is implemented in DELPHES. The tracker and calorimeter information is combined for each particle that interacts in both sub-detectors. In practice, when a track is reconstructed in the direction of a calorimeter cell, the energy of the track is computed assuming a massless particle and is com-

pared to the measured energy in the calorimeter. If the difference is smaller than twice the uncertainty:

$$E_{\text{calo}} - E_{\text{track}} < \sigma_E(E, \eta) \,, \tag{2.4}$$

the energies are combined to improve the energy resolution. If the calorimeter cell energy is larger than twice the track energy, a *neutral tower* is created with the *missing* energy. The case where several tracks reach the same calorimeter cell is solved using the particle truth information.

2.3.4 Pile-up subtraction

Pile-up interactions are usually identified by means of vertex reconstruction. Indeed, if such interactions occur far enough from the hard interaction, a precise vertexing algorithm is able to detect these charged tracks and remove PF candidates not associated with the primary vertex. This procedure is called *charged hadron subtraction* in CMS.

The residual pile-up contamination comes from neutral particles and charged particles matching the main primary vertex. It directly affects the performance of jets, E_T^{miss} and isolation. The neutral pile-up subtraction of jets is described in Sec 2.3.7.

DELPHES Charged pile-up subtraction In the default CMS configuration, no vertexing is performed. Actually, it would not make sense since only the transverse momentum of the tracks are smeared. This is not a limitation of the DELPHES framework but a choice when creating the CMS default configuration card which allows to save time and simplify the procedure. Therefore it is assumed that vertices corresponding to pile-up interactions occurring at a distance z from the hard scattering, such that $|\mathrm{z}| > \delta \mathrm{Z}_{\mathrm{vtx}}$ can be reconstructed. The parameter $\delta \mathrm{Z}_{\mathrm{vtx}}$ is the spatial vertex resolution of the detector. We assume that pile-up interactions occurring at a coordinate z, such that $|\mathrm{z}| < \delta \mathrm{Z}_{\mathrm{vtx}}$ cannot be disentangled from those originating from the high-Q^2 process. Therefore every charged particle originating from such vertices cannot be subtracted

from the event, while every charged particle originating from a vertex positioned at $|\mathrm{z}| > \delta \mathrm{Z}_{\mathrm{vtx}}$ can be identified as originating from pile-up, provided that the corresponding track has been reconstructed. If the particle-flow algorithm is used, the particle-flow tracks identified as originating from pile-up interactions are removed from the list of 4-vectors entering the jet clustering and the isolation procedures.

2.3.5 Electrons

The electrons are reconstructed by combining a track in the silicon detector with a cluster in the ECAL.

In CMS, electrons deposit about 97% of their energy in a 5×5 crystal array, as measured in test beams [102]. The magnetic field bends the trajectory of the electrons which on average lose 33% of their energy by radiating photons mainly along the ϕ direction. It is therefore mandatory to measure these radiations in order to precisely estimate the initial energy of the electron. The clustering of the electron energy in the ECAL is made separately in the EB and in the EE.

- The *hybrid algorithm* is used in the EB. It starts with a cluster that contains most of the energy deposited in the considered region: *the crystal seed*. Arrays of 5×1 in $\eta \times \phi$ crystals surrounding the seed in the ϕ directions are added if their energies is superior to a minimum threshold. The global cluster produced is called a *supercluster* (SC)
- The *multi-*5×5 *algorithm* is used in the EE. It starts with seeds that are the crystals with with maximal energy relative to their four direct neighbors and an energy exceeding a threshold. The energy is then collected by adding clusters of 5×5 crystals that can partly overlap around the seed within a range in $\eta^{\text{range}} = 0.07$ and $\phi^{\text{range}} = 0.3$ rad to produce SC.

Since the electron loses a large amount of energy through radiation in the tracker, the hit efficiency is reduced and it results in a poor estimation of track parameter when using the standard track Kalman Filter(KF)-based reconstruction. A dedicated tracking procedure is therefore used for the electrons. Since

this method is very computing intensive, it is initiated only when one of the two following kinds of seeds are present. The ECAL-based seeds are based on the SC energy and position. The tracker-based seeds are based on standard track reconstruction extrapolated to the ECAL and that match a SC.

Once an *electron seed* is reconstructed, they are used to start the electron-track building, followed by the track fitting.

- The track building is based on the combinatorial KF method, taking into account the energy loss modeled through a Bethe-Heitler function.
- The *Gaussian-sum filter* (GSF) has been developed for the electron track fitting. It allows for important energy loss along the trajectory.

The electron energy measurement is optimized by combining the ECAL SC energy with the track momentum. The track momentum is more precise than the ECAL SC energy for electrons with transverse momentum lower than 15 GeV and brings significant improvement up to 35 GeV. Results at 13 TeV can be found in [96, 97, 103, 104].

Electron selection and isolation The identification is needed to discriminate the prompt isolated electrons from the main background sources including photon conversion, jet misidentification and electrons coming from b- or c-decay. In order to perform this selection, many variables are available and divided into three main categories.

First, the purely calorimetric observables category includes the variables related to the shape of the shower in the ECAL. It also contains the energy fraction in the HCAL and the preshower energy. Then, the purely tracking observables are related to the tracking information, in particular the differences between the GSF and the KF fitted tracks. Last, the "ECAL and tracker" variables includes the matching (geometrical and energy-momentum) between the track and the SC. An example of selection applied to the *tight* electron identification is written in Tab 2.2.

To further reject background coming from misidentified jets or electrons from b- and c-decays, one can select only electrons with a small energy flow surrounding their trajectory. This *isolation* procedure consists in first looping

Variables	barrel cuts	endcap cuts
5×5 cluster rms $\sigma_{i\eta i\eta}$	0.0101	0.0279
$\lvert\Delta\eta\rvert$ between SC and extrapolated track	0.00926	0.00724
$\lvert\Delta\phi\rvert$ between SC and extrapolated track	0.0336	0.0918
HCAL energy behind the SC / SC energy	0.0597	0.0646
Relative isolation	0.0354	0.0646
$\frac{1}{E_{\text{SC}}} - \frac{1}{p_{\text{Track}}}$	$0.012\,\text{GeV}^{-1}$	$0.009\,99\,\text{GeV}^{-1}$
$\lvert d_0\rvert$: transverse distance to the beamspot	0.0111 cm	0.0351 cm
$\lvert d_z\rvert$: longitudinal distance to the PV	0.0466 cm	0.417 cm
Max. number of missing inner hits	2	1

Table 2.2: Cuts for the "tight" electron identification used with 2015 data. The variable $\sigma_{i\eta i\eta}$ smoothen the efficiencies across the ECAL crack.

over particle flow charged hadrons, neutral hadrons and photons candidates in a cone of $\Delta R = 0.3$ around the electron. Then the sum of the energy/momentum of those candidates is added and compared to the energy of the selected electron. Typical valued are reported in Tab 2.2. This computation takes into account the correction for the pile-up.

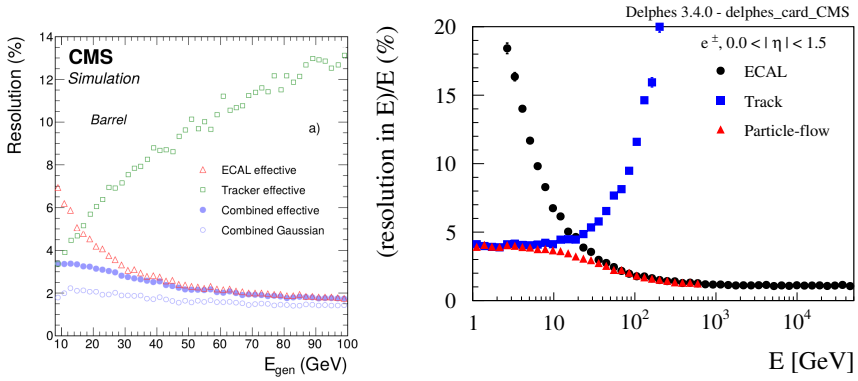

Figure 2.11: Central electron energy resolution comparison. Left: CMS performances at 8 TeV [97]. Right: Delphes electron energy resolution, using calorimeter deposits, tracks and the combination.

Electron parametric reconstruction

In DELPHES, since the material of the tracker is neglected, no bremsstrahlung photons are simulated. The electron can therefore directly be reconstructed from the energy flow candidates. The energy associated to the electron is computed by combining the track momentum and the ECAL energy deposit.

To match the observed selection efficiencies, an extra efficiency is applied with p_T- and η-dependence. One can see the difference between the track efficiencies and the reconstructed electron efficiencies in the two top plots of Fig 2.11. The impact of the *particle flow emulation* on the energy resolution is well visible.

The isolation is also modeled by summing the p_T of the energy flow charged hadrons, neutral hadrons and photons candidates. The electron is considered as isolated if this sum is smaller than 15% of its own energy.

2.3.6 Muons

The detection of muons relies mainly on the muon chambers since muons are essentially the only charged particle reaching the outer part of the detector. This detection is improved by using the inner tracker that allows very precise position and momentum measurements. This complementarity is used to define several types of reconstructed muons.

First, the *segments* (track stubs) are formed with hits from the DT and the CSC. These segments are used as seeds for the track fitting using DT, CSC and the RPC hits. These muon chamber tracks are called *stand-alone muon tracks*. Then, a track properties comparison is performed with the tracker tracks, and the global fit (inner and outer track) provides a *global muon track*. Low-p_T muons are likely to not reach the outer stations of the muon system and therefore do not trigger the creation of stand-alone muons. Therefore, all inner tracks with $p_T > 0.5$ GeV are extrapolated to the muon spectrometer. If at least one muon segment matches the extrapolated track, it is qualified as a *tracker muon*.

Fraction of valid tracker hits	0.8
χ^2/ndof	3
Kink finder	20
Segment compatibility	0.303
Or	
Fraction of valid tracker hits	0.8
Segment compatibility	0.451

Table 2.3: Requirements for a global or tracker muon to pass the *medium* identification criteria.

As for the electrons, an isolation variable is computed for each muon by summing the p_T of the energy flow charged hadrons, neutral hadrons and photons candidates in a cone around the muon trajectory.

Muon parametric reconstruction

Muon reconstruction is completely driven by the tracking in DELPHES. The existence of the muon chambers in the CMS detector is translated into better tracking performances. The reconstructed muons therefore have exactly the same behavior as the muon induced tracks as shown in Fig 2.12. As for the electrons, an extra (η, p_T)-dependent efficiency is applied to match the observed identification step, as it can be seen plots of Fig 2.12.

2.3.7 Jets

The creation of high energy quarks and gluons induces cascades of hadrons that interact with the detector, as argued in Sec 1.1.4. To reconstruct and understand hadron collider interactions, it is therefore of primary importance to reconstruct the jets efficiently. Several algorithms have been developed to reconstruct jets.

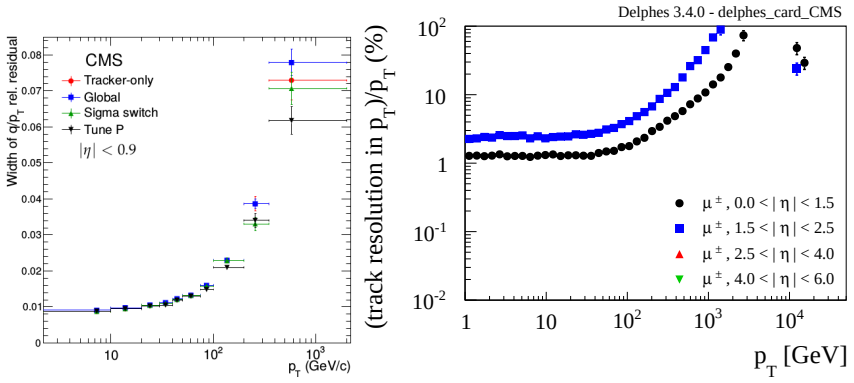

Figure 2.12: Central muon transverse momentum resolution comparison. Left: CMS performances at 8 TeV [105]. Right: Delphes muon momentum resolution.

At the LHC, the *anti-k_t* algorithm is used by both CMS and ATLAS collaborations. A distance to the beam pipe and between two four-vectors is defined as, respectively:

$$\begin{aligned} d_i &= (E_{Ti})^{2p} \\ d_{ij} &= \min\{E_{Ti}^{2p}, E_{Tj}^{2p}\}\frac{\Delta R^2(i,j)}{R^2} \end{aligned}$$

The jets are then clustered by combining the four-vectors i and j when their distance $d_{ij} < d_i$. Once d_i is the smaller distance, i is qualified as a jet and the algorithm continues iteratively.

The choice of the parameter p defines distinct algorithms, for instance:

$$\begin{aligned} p = 1: &\quad k_T \text{ algorithm} \\ p = 0: &\quad \text{Cambridge/Aachen algorithm} \\ p = -1: &\quad \text{anti-}k_T \text{ algorithm} \end{aligned}$$

The main difference between the different algorithms is the shape of the produced jets when their distance is of the order of their size. In practice, these

algorithms provide infrared and collinear safe jets with the advantage that high-p_T jets will tend to be more circular with the anti-k_T algorithm.

CMS Particle flow Jets

In CMS, jets are reconstructed with the *anti-k_t* algorithm applied on the particle flow objects. For the first and second run of the LHC, jets are reconstructed with the anti-k_T algorithm and the R distance parameter of 0.5 and 0.4 respectively.

Residual pile-up subtraction A combination of vertexing and tracking information allows the identification of contaminating charged particles from the pile-up, as seen in Sec 2.3.4, while neutral particles do not produce tracks. Therefore, neutral pile-up contamination can only be estimated *on average*. This is also true for particles originating from vertices which are too close to the hard interaction vertex to be identified as a pile-up and charged particles for which the track is not reconstructed.

The Jet Area method [106, 107] is widely used in current collider experiments. It allows the extraction of an average contamination density ρ on an event-by-event basis. The pile-up density ρ, can then be used to correct observables that are sensitive to the residual contamination, the jet energies and the isolation variable. In the presence of residual pile-up contamination, these two quantities are corrected in the following way:

$$p_{jet} \rightarrow p_{jet} - \rho \cdot A_{jet}, \tag{2.5}$$

$$I(P) \rightarrow I(P) - \frac{\rho \cdot \pi R^2}{p_T(P)}, \tag{2.6}$$

where A_{jet} is the jet area estimated via the FASTJET Package, and R is the diameter of the isolation cone.

In CMS, a slightly modified version of the jet area, dubbed *hybrid jet area method*, is developed. It improves the performances by considering a more complex parametrization of the correction including a η-dependence [108].

The Jet Energy Correction is mainly a simulation-based correction applied to the reconstructed jet to correct the bias due to the reconstruction. First, the particle-level jets are matched to the closest reconstructed jets and a correction factor is extracted in η- and p_T- bins. This factor is then applied to modify the reconstructed jets momentum. A residual correction is extracted from data using γ+jets events.

Particle-Flow-like jets in DELPHES

In DELPHES, the energy flow objects are created and therefore available for jet clustering. The jet algorithm uses the tracks and towers as input and produces jets with any of the algorithms available in the FASTJET suite. This method produces realistic anti-k_t jet collections, depending on the track and calorimeter parametrization. It has the advantage to have some predictive behavior. For instance, if one changes the calorimeter size or resolution, the energy resolution will change accordingly.

The neutral pile-up subtraction is, as in the full simulation, based on a η-dependent jet area method provided in FASTJET. Finally, the reconstructed jet energy is corrected to remove the bias from the particle level jet energy.

Similar results are observed when comparing CMS and DELPHES jet energy resolution, as drawn on Fig 2.13. An improvement of jet resolution up to 1 TeV is observed thanks to the particle flow approach. These particle flow jets are reconstructed with a resolution of 15% at low p_T and decrease to approximatively 5% at very high p_T.

2.3.8 B-tagging

The identification of jets that result from the hadronization of heavy flavour quarks, typically b or c quarks, is important in high energy collider experiments. In particular, the decay of the Higgs boson to a $\mathrm{b\bar{b}}$ pair or the the decay of the top to bW are important motivations. B hadrons have a typical decay length of $500\,\mu$m, inducing typical trajectory of a few mm in the detector. Here are some examples:

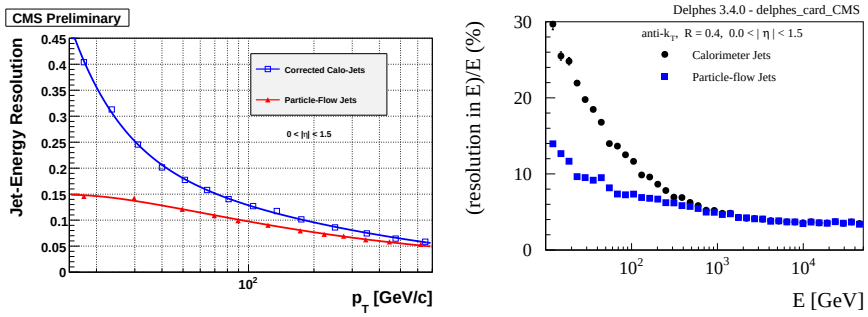

Figure 2.13: Jet energy resolution for corrected calorimeter jets and particle flow jets in the barrel region. CMS results on the left are obtained with $R = 0.5$ [95], while DELPHES on the right plot uses 0.4.

hadron	$c\tau$
$B^{\pm}$	491.1 µm
B^0	455.7 µm
B_s^0	452.7 µm
Λ_b^0	439.5 µm

The charged particles originating from the B hadron decays induces the reconstruction of several tracks slightly displaced from the interaction point. The identification of the jets originating from a b quark is based on the reconstructions of displaced secondary vertices built from these charged hadrons tracks.

CMS secondary vertex reconstruction The *Inclusive Vertex Finder* is used to produce a collection of reconstructed vertices. The algorithm takes as input the tracks passing a very basic quality selection and which are not too far displaced in the z-direction from the primary vertex. These tracks are processed with a clustering algorithm. It loops over all displaced seed tracks and checks which other tracks are close to it to construct clusters. Each cluster is then fitted and the output of the algorithm is a collection of non-independent secondary vertices.

It is followed by the Vertex Merger, then the Vertex Arbitrator and the Vertex Merger again.

The *Vertex Merger* loops over all the reconstructed clusters and checks if they share a significant number of tracks with another cluster. If it is the case and if the distance between the two vertices is small compared to the uncertainty, the two vertices are considered as merged and one of the two is removed from the list.

The *Vertex Arbitrator* is applied to each secondary vertex. It loops over all the associated tracks and checks if they are more compatible with the primary vertex or with the secondary vertex. If at least two tracks remain, the vertex is re-fitted.

A second pass of the *Vertex Merger* is then applied to get rid of tracks that are associated with several secondaries.

Combined Secondary Vertices Several different algorithms are used to tag jets in CMS. The *Combined Secondary Vertex* (CSVv2) is a very efficient algorithm that takes advantage of the very high performances of the tracker and combines the vertices information with extra tracks and jet information. Examples of track information are the impact parameter significance in the transverse plane and in the three dimensions. The jet kinematic is also used and the three kinds of information are combined using a boosted decision tree. Three working points are available: loose, medium and tight corresponding to the purity of approximately 0.1, 0.01 and 0.001 respectively.

DELPHES b-tagging possibilities

In DELPHES a purely parametric approach based on the truth information is available as well as a track counting algorithm.

Parametric The algorithm for b jet identification proceeds as follows: the jet becomes a potential b jet if a generated b is found within some distance $\Delta R = \sqrt{(\eta^{\text{jet}} - \eta^{\mathrm{b}})^2 + (\phi^{\text{jet}} - \phi^{\mathrm{b}})^2}$ of the jet axis. The probability to be identified as a b depends on user-defined parameterizations of the efficiency. The user can also specify a mistagging probability parameterization, that is, the probability that a jet originating from a particle other than a b is wrongly

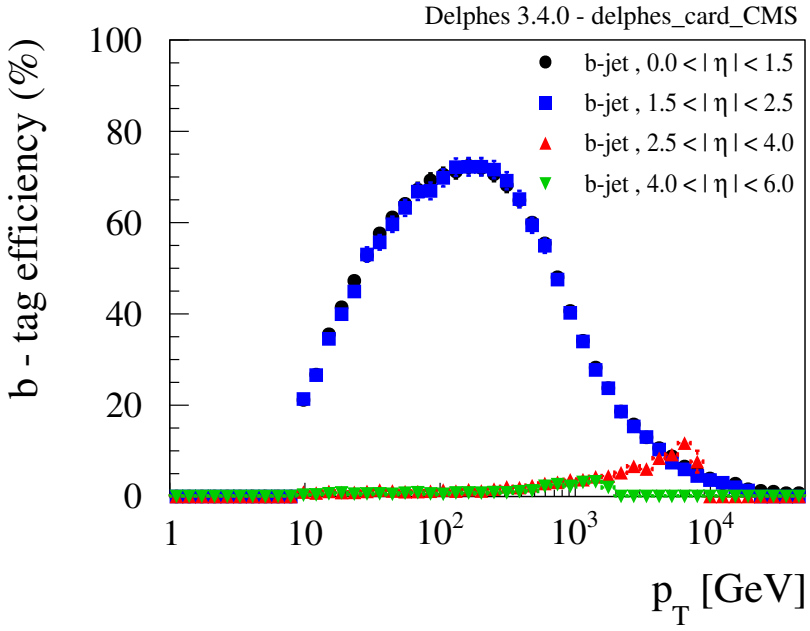

Figure 2.14: Parametric b-tagging efficiencies for different η region. The b-tagging efficiency used in DELPHES corresponds to the medium working point of the CMS CSVv2 algorithm.

identified. Modularity allows the user to use several b-tagging algorithms for the same jet collection and to easily implement other tagging algorithms, possibly involving an analysis of the jet constituents.

The parametric b-tagging efficiency obtained using the default CMS card corresponds to the typical efficiency obtained for the medium working point of the CSVv2 algorithm, see Fig 2.3.8.

Track Counting When the track counting b-tagging is selected, the impact parameter of each track inside all reconstructed jets is smeared according to a Gaussian distribution. The width of the Gaussian distribution is defined by the user and can depend on the p_T and η of the corresponding track. The *Track Counting* is then applied on each jet and requires a minimum number of tracks

(typically 2 or 3) with a high impact parameter significance for the jet to be tagged. While less precise than the parametric approach, this method provides the predictive power needed for a future detector design when a full simulation is not yet available.

2.3.9 Trigger

When comparing the cross-section of the processes of interest with the total cross section in Fig 1.6, it is obvious that only a tiny proportion of the events are "interesting" such as those involving the Higgs or a weak boson. Also, with groups of protons colliding every 25 ns, its impossible to read and save every events, given the quantity of information in one single event.

A *trigger system* has been developed to select the potentially interesting events and reduce the rate of stored events to less than 1 kHz.

The CMS triggering system is divided in two levels.

The Level 1 Trigger is an extremely fast and automatic procedure that selects around 100 kHz of data using very simple criteria. It is based on a combination of local information to reconstruct simple objects. The decision is taken in approximately $3\mu s$ during which the data is kept in memory buffers.

The High Level Trigger (HLT) is a computer farm selecting events interesting for physics analysis based on particular requirements on one or more particle candidates. Some conditions are of general interest such as the *single-electron trigger*, and some are dedicated to specific analyses, for example the *3-jets-with-MET trigger*. This second step reduces the number of events recorded to less than 1kHz.

Two triggers are particularly relevant for this work: the di-muon and the di-electron triggers. In 2015, they recorded events with two reconstructed muons (electrons) with transverse momentum higher than 17 (17) and 7 (12) GeV for the leading and subleading muon (electron) respectively.

Chapter 3

Search for new scalar bosons in the llbb final state

The search for the production of a heavy Higgs boson often involves leptons and b-jets in the final state. In particular, it has been shown in the first chapter that the search for a heavy scalar decaying into a Z-boson and a pseudo-scalar A in the final state with two leptons and two b-jets is very relevant. As shown in Fig 1.12, the main production mode of the heavy scalar is the gluon fusion process. The corresponding Feynman diagram is drawn in Fig 3.1.

The following sections will be motivated by this specific process. It is nevertheless important to keep in mind the other processes arising from extensions of the scalar sector that provide the same final state: A $\rightarrow Z(l^+l^-)H(b\bar{b})$, h $\rightarrow Z(l^+l^-)A(b\bar{b})$ and A $\rightarrow Z(l^+l^-)h(b\bar{b})$. One could also think of the specific process H $\rightarrow Z(l^+l^-)Z(b\bar{b})$. The resonant and non-resonant Higgs pair productions $(H \rightarrow)h(b\bar{b})h(l^+\nu_l l^-\bar{\nu_l})$ induce similar final state but induce some extra missing-E_T due to the presence of neutrinos in the final state.

In order to be sensitive for the three hierarchies sketched in Fig 3.2 one needs to search for both $H \rightarrow ZA$ and $A \rightarrow ZH$. In practice, the search will therefore be based only on spin-independent variables, allowing to easily recast results for one process to the other. The im2HDM-motivated $H \rightarrow ZA$ process is the

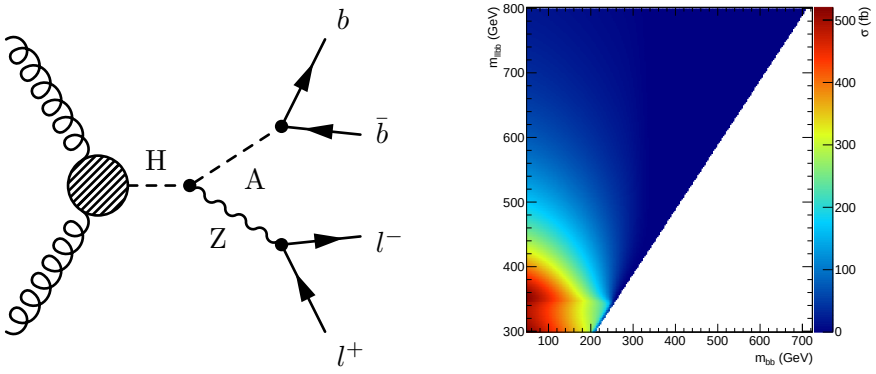

Figure 3.1: Left: Feynman diagram of the production of a heavy Higgs boson, decaying to Z and A in the final state with two leptons and two b-quarks. Right: Corresponding production cross-section in the 2HDM type-II, with $\tan\beta = 1.5$ and $\cos(\beta - \alpha) = 0.01$.

best choice since it offers a more complete parameter space: the pseudo-scalar mass is not bound by the Higgs boson mass, as illustrated in Fig 3.2.

As seen in Fig 1.12, the associate production of a heavy scalar and two b-quarks leads at high values of $\tan\beta$. There is currently no specific search for this final state which consist in two extra forward b-jets, but the analyses are also sensitive to this production mode, since no vetos are applied on extra jets or b-jets.

This search has been performed for the first time with the 2012 data at 8 TeV. In this analysis, we have tested the complete 2D $\mathrm{m_A} - \mathrm{m_H}$ map and made many use of the DELPHES detector simulation. A second iteration has followed using 2015 data at 13 TeV, where we have tested several mass point with a similar analysis strategy. In the first and second sections, the signal process and the standard model processes involving final states with two leptons and two b-jets are reviewed. The sections three to four are the details of the analysis: selections, strategy and systematic error estimation. Since these are the newest results, only the 13 TeV analysis will be described these sections. In the fifth section, applications of fast simulations described in the second chapter to the analysis are commented. These are mainly personnal results obtained in the

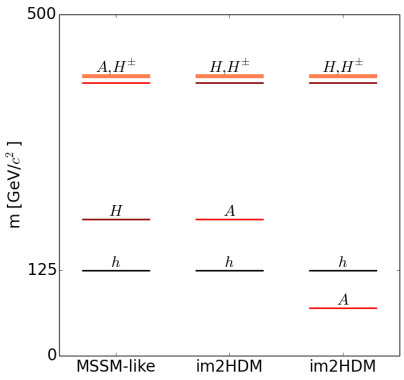

Figure 3.2: Three different possible hierarchies from the 2HDM where the light scalar mass is fixed at the discovered Higgs boson mass.

context of the 8 TeV search. The results of both analysis are presented in the sixth section.

3.1 The signal

The signal samples are produced using MADGRAPH5 interfaced with the parton shower generator PYTHIA8.2. The parton distribution function is NNPDF2.3 LO four flavors, and the factorization and renormalization scales are estimated dynamically. The NNLO cross-section for the process $pp \to H$ is calculated with SUSHI and used to normalize the Monte Carlo samples. The mass hypotheses considered in the 13 TeV (henceforth denoted $\mathcal{H}^{\mathrm{m_H}}_{\mathrm{m_A}}$) search are:

- $\mathcal{H}^{300}_{50}$, $\mathcal{H}^{300}_{100}$, $\mathcal{H}^{300}_{200}$,
- $\mathcal{H}^{500}_{50}$, $\mathcal{H}^{500}_{100}$, $\mathcal{H}^{500}_{200}$, $\mathcal{H}^{500}_{300}$, $\mathcal{H}^{500}_{400}$,
- $\mathcal{H}^{800}_{50}$, $\mathcal{H}^{800}_{100}$, $\mathcal{H}^{800}_{200}$, $\mathcal{H}^{800}_{400}$, $\mathcal{H}^{800}_{700}$.

The lightest scalar boson mass is fixed at 125 GeV. Following the constraints from the standard model Higgs measurements described in Sec 1.3.4, the samples have been generated close to the alignment limit. The chosen values for

$\tan\beta$ and $\cos(\beta-\alpha)$ are 1.5 and 0.01 in a type-II 2HDM. As illustrated in Fig 3.2, the masses of the charged Higgs bosons are kept equal to the mass of the heaviest neutral boson to preserve the degeneracy $\mathrm{m_{H^\pm}} \approx \max(\mathrm{m_H}, \mathrm{m_A})$. The value of the soft $\mathcal{Z}_2$-symmetry breaking is set to $\mathrm{m_{12}^2} = \mathrm{m_{H^\pm}}^2 \tan\beta/(1+\tan^2\beta)$, inspired by the MSSM parametrization. This choice is arbitrary and not very relevant since it affects only the scalar self-interaction allowing only secondary order impact: if the $\mathrm{H} \to \mathrm{hh}$ branching fraction is enhanced, the $\mathrm{H} \to \mathrm{ZA}$ process rate decreases.

3.2 The backgrounds

The typical signal signature is two b-jets and two same flavor, opposite signs leptons. Several processes of the standard model, called *backgrounds*, will produce signatures in the detector very close to the one of our signal. On the one hand, the *irreducible* backgrounds are the processes that involve exactly the same final state object. This is the case of the $ZZ \to l^+l^-b\bar{b}$ for instance. On the other hand, the *reducible* backgrounds are the processes that provide a different final state that possibly fakes the final state object.

3.2.1 Drell-Yann plus jets

The production of one Z/γ-boson in addition to at least two jets will be the most important background. At leading order, two production mechanisms are possible, the *gluon-gluon* case where a gluon from one proton interacts with a gluon from the other proton and the *quark-antiquark* case where a quark from one proton interacts with an anti-quark from the other proton.

The $gg \to Z(l^+l^-)jj$ has the biggest cross-section thanks to the high gluon pdf, while the $q\bar{q} \to Z(l^+l^-)jj$ is reduced by the parton density function of the antiquark. These two processes are irreducible when two b-jets are in the final state, as illustrated in the Feynman diagrams in Fig 3.3. When no or only one b-quark is produced in the acceptance, the background is reducible and will depend on the b-tagging performances.

The simulated sample used in the analysis is generated using MADGRAPH_aMC@NLO and PYTHIA8. Two samples have been generated: one with a generation cut of $\mathrm{m}_{ll} > 50$ GeV yielding a cross-section of 6025.2 pb at 13 TeV; the other with a generation cut of $10 < \mathrm{m}_{ll} < 50$ GeV yielding a cross-section of 18610 pb at 13 TeV.

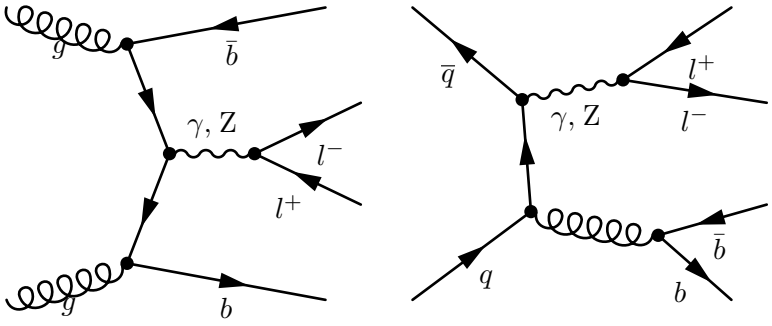

Figure 3.3: Example of leading order diagrams of the $gg \to l^+l^-b\bar{b}$ process.

3.2.2 The top-pair background

The production of a top-antitop pair also provides two leptons when both $\mathrm{W}^{\pm}$ decay leptonically. There are different possibilities to produce two same flavor opposite sign electrons or muons. The obvious one is when both W decay to electrons or muons and the less obvious one is when at least one of the W decays to a τ-lepton. The τ can itself decay leptonically, creating an extra possibility to observe two same flavor leptons.

A POWHEG [109] sample of $pp \to t\bar{t} \to b\bar{b}l^+l^-\nu_l\overline{\nu_l}$ is used at 13 TeV. Its cross-section computed at NNLO is $87.31^{+2.08}_{-3.07} \pm 3.68$ pb.

3.2.3 ZZ and Zh

The production of two Z-bosons arises with a much lower rate because of the extra weak couplings. A sample of $ZZ \to l^+l^-q\bar{q}$ is generated with MADGRAPH_aMC@NLO and a cross-section of 3.22 pb. Interestingly, the asso-

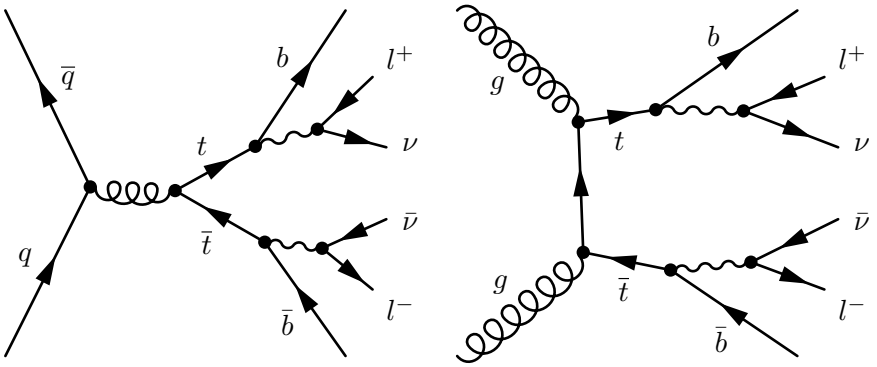

Figure 3.4: $t\bar{t}$ Process, leading diagrams.

ciate production of a Higgs boson with a Z where the Higgs particle decays to two b-quarks is a similar background with its cross-section computed at NLO of 0.87 pb. Leading order diagrams of these two processes are drawn on Fig 3.5.

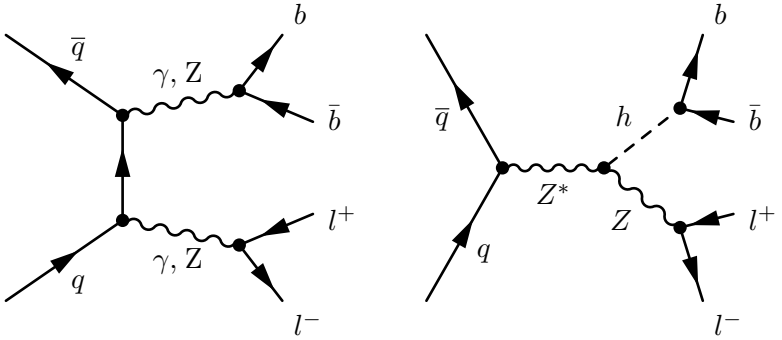

Figure 3.5: Leading order ZZ and Zh process diagram.

3.2.4 Single top production

The production of only one top quark is also a source of background. The most important production mode is the associate production of a (anti)top quark and a W boson, see Fig 3.6. The production cross section is 72.2 pb for an approximated NLO sample generated in the five flavor scheme. It means that the

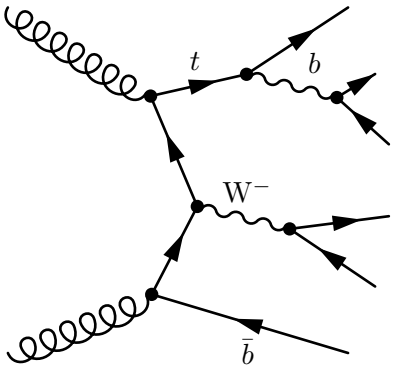

Figure 3.6: Single top production in addition to a W boson, leading diagram. The production of the antitop is explained by a similar diagram, where the direction of the arrows are changed.

b-quark is a possible initial state and its pdf is considered instead of considering a gluon splitting into a b-quark pair.

There are two other mechanisms of single top production with a W in the t- and s-channel. Samples are generated with respectively 70.69 and 3.36 pb with an NLO precision. These two modes are considered but provide only one lepton in the final state and are therefor heavily reduced when selecting signal events.

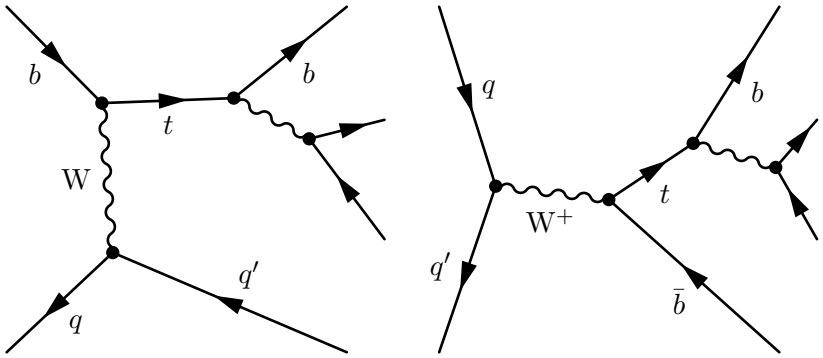

Figure 3.7: Feynman diagram of the production of a single top with a W boson in the t-channel on the left and s-channel on the right.

3.2.5 W plus heavy flavor jets

The production of one W-boson in addition to at least two jets is also a background. The simulated sample used in the analysis is generated using MADGRAPH_aMC@NLO and PYTHIA8. An inclusive sample with all W+jets has been generated and the cross-section computed at NNLO is 61536.7 pb at 13 TeV.

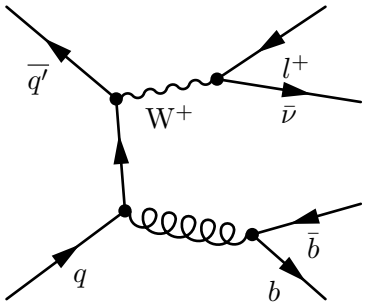

Figure 3.8: Example of leading order diagrams of the $gg \to Wb\bar{b}$ process.

3.3 Event Selection, Background Control and MC-Corrections

The online selection of the events is based on the dilepton triggers which require two leptons passing the identification and isolation criteria. In addition, a sufficiently high transverse momentum is required. The di-electron(muon) trigger requires at least 17 and 12(8) GeV respectively for the leading electron(muon) and subleading electron(muon).

3.3.1 Pile-Up Reweighting

The MC samples are produced before the data taking. Therefore, the primary vertex multiplicity distribution does not reproduce the observed distribution from the data. An event-by-event weight is applied to each simulated event to

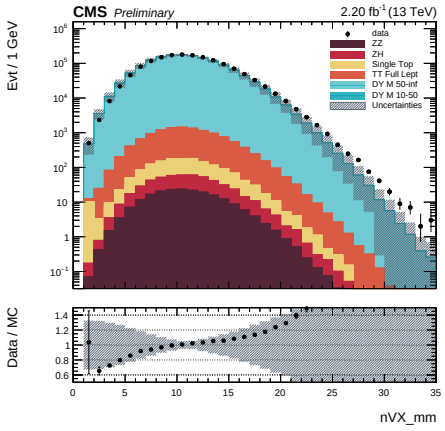

Figure 3.9: Distribution of the number of reconstructed vertices. The events are selected when at least two muons are reconstructed, the grey area represent the associated systematic error.

match the data distribution derived from the instantaneous luminosity for each bunch-crossing. The data/MC agreement in the primary interaction multiplicity is shown in Fig 3.9.

3.3.2 Muons

The first event selection step of this analysis is the requirement of two same flavor leptons. We use all events passing that very simple selection criterion to check the primary vertices and the leptons reconstruction.

The two global or tracker muons are selected by requiring $p_T > 20$ GeV, $|\eta| < 2.4$ and some additional track- and muon-quality requirements defined in Tab 2.3. This selection is driven by the double muon trigger acceptance and spatial boundaries of the CMS muon chambers. In addition, the muons must be well isolated. The sum of the charged neutral hadrons, photons and PU (corrected with the $\Delta\beta$ correction) contribution, combined as

$$I^{\mu}_{\text{rel}} = \left[\sum p_T^{\text{ch.had.}} + \max(0, \sum E_T^{\text{neut.had.}} + \sum E_T^{\gamma} - 0.5 \sum p_T^{\text{ch,PU}})\right] / p_T^{\mu}$$

must not exceed 0.25 in a cone with radius of 0.4 around the muon. The $\Delta\beta$ correction consists in subtracting one half neutral PU quantity for every charged PU quantity found in the cone.

Each selected muon must lie within a cone of $\Delta R = 0.3$ from one of the object that has fired the DoubleMuon trigger, with a relative p_T difference of 0.5.

Some basic kinematic distributions are plotted in Fig 3.10 for the leading and sub-leading muons. The isolation variable is plotted in Fig 3.11.

The MC muon momentum is corrected to remove the effects from the detector misalignment or errors in the magnetic field. This method is called *Rochester correction* and induces a very good agreement of the di-muon invariant mass distribution with the data (see bottom of Fig 3.10). This method improves also the di-lepton transverse momentum plotted on the bottom right part of Fig 3.10. A slight disagreement is still visible in the data-MC ratio at low transverse momentum of the dilepton system.

The efficiency of triggering, identification and isolation are extracted from the data using a *Tag and Probe* technique [110] and the scale factors extracted are applied on the simulated events.

3.3.3 Electrons

The two electrons are selected by requiring $p_T > 20$ GeV, $|\eta| < 2.5$ (excluding the ECAL gap at $1.566 < |\eta| < 1.442$). Additional quality requirements are applied. These are based on the supercluster quality, the energy deposited in the HCAL, the track distance to the primary vertex and the energy matching between the track and the ECAL deposits. The identification corresponds to the selection in Tab 2.2 and typically yields an efficiency close to 70% on the electrons from Z boson decay. This selection is driven by the double electron/gamma trigger acceptance and spatial boundaries of the electromagnetic calorimeter.

The selection also includes the isolation in a cone with radius $\Delta R < 0.3$. As for the muons, the contributions of the charged hadrons, neutral hadrons,

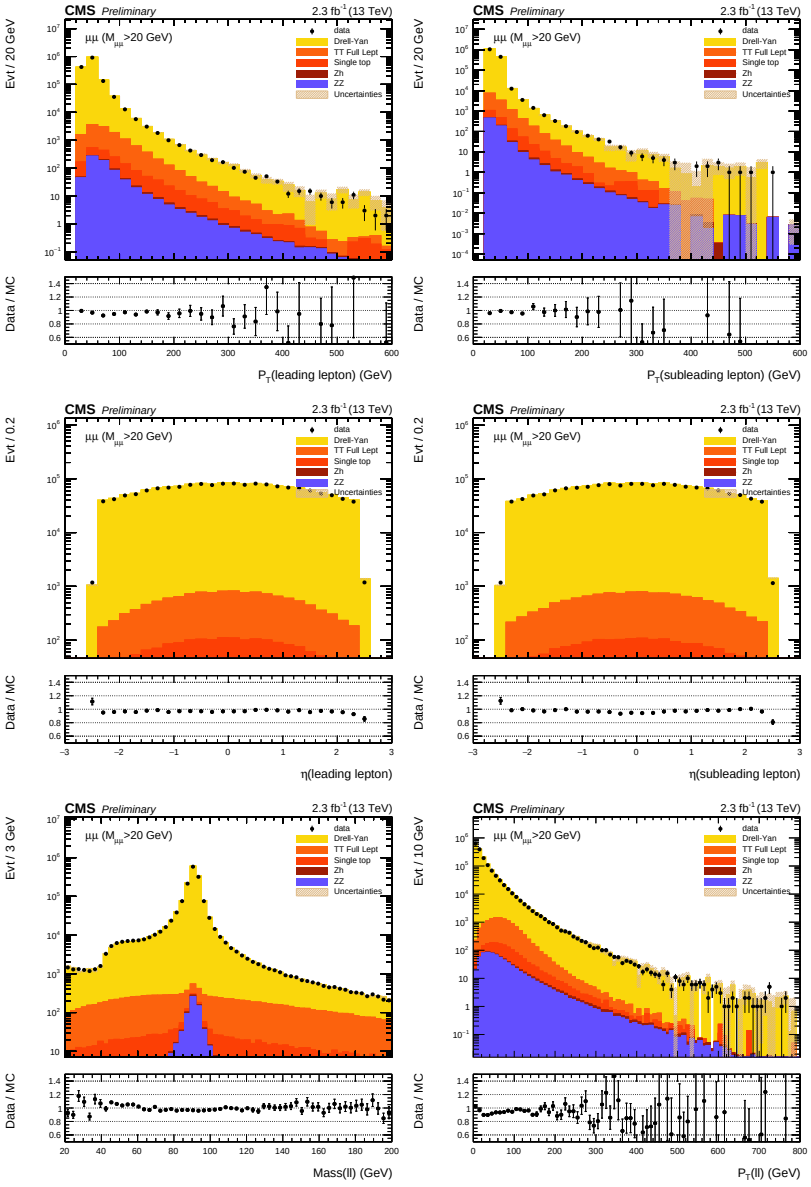

Figure 3.10: Top left(right): transverse momentum of the (sub-)leading muon. Center left(right): pseudo-rapidity of the (sub-)leading muon. Bottom left: di-muon invariant mass. Bottom right: di-muon transverse momentum.

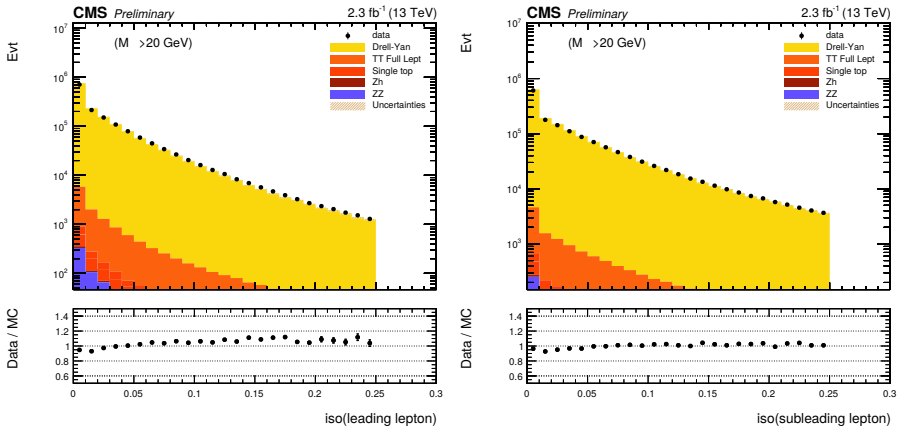

Figure 3.11: Distribution of the leading(left) and sub-leading(right) muon isolation variable.

photons and PU are used. The relative isolation is set to $I_{\text{rel}} < 0.035$ in the barrel region and 0.064 in the endcap region.

Some kinematic distributions are plotted in Fig 3.12 for the leading and subleading electrons. The isolation variables are plotted in Fig 3.13.

The electron energy correction was not yet available when performing this search. One can see that the di-electron mass distribution is slightly shifted, producing a wave in the data-MC ratio plot around the Z-mass on the bottom-left plot of Fig 3.12. This effect doesn't affect significantly the results since the m_{ll} variable is used only for background normalization.

As for the muons, a *Tag and Probe* technique is used to extract scale factors that are applied to the simulated events.

3.3.4 Jets

The jets are required to satisfy $p_T > 30$ GeV and $|\eta| < 2.4$ to be selected. Jets are selected requiring less than 90% of the EM and HCAL deposits to come from neutral deposits and composed of at least two objects. In addition, tight

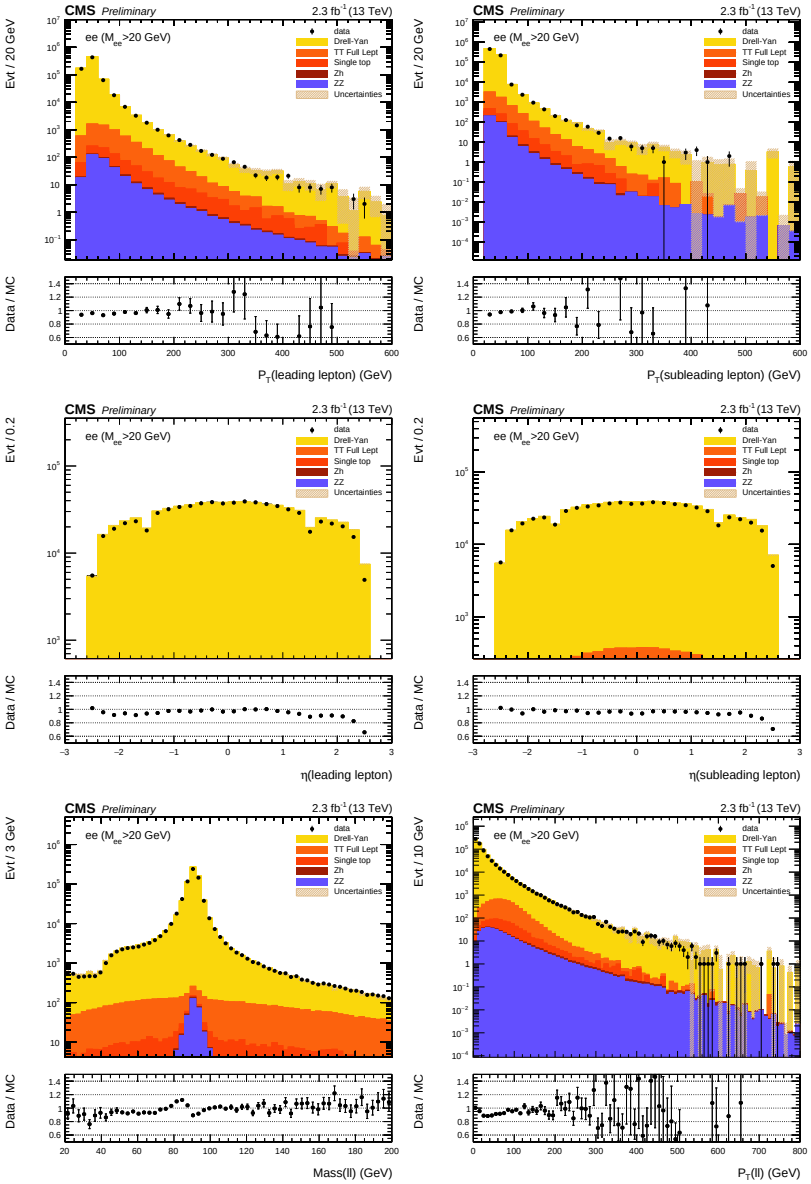

Figure 3.12: Top left(right): transverse momentum of the (sub-)leading electron. Center left(right): pseudo-rapidity of the (sub-)leading electron. Bottom left: di-electron invariant mass. Bottom right: di-electron transverse momentum.

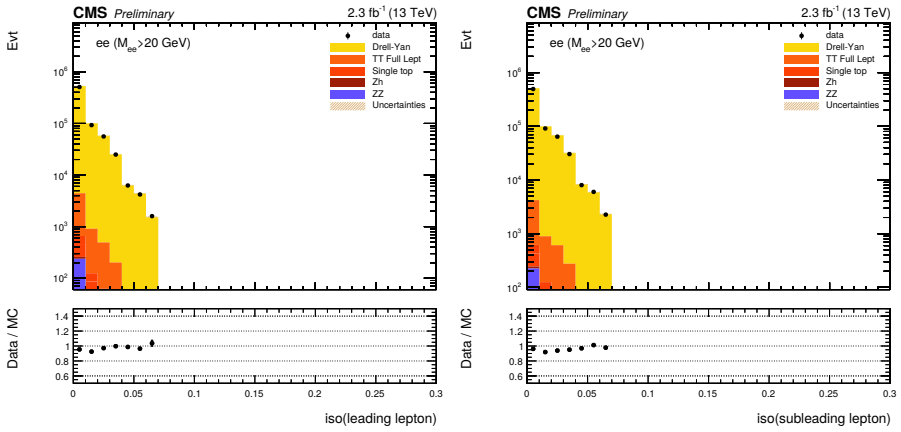

Figure 3.13: Distribution of the leading(left) and sub-leading(right) electron isolation variable.

jets are required to have an EM fraction of less than 99% and a charged hadron fraction as well as the multiplicity greater than zero. The leading and sub-leading jet transverse momentum and η distributions are plotted in Fig 3.14. Only events with two selected leptons and at least two selected jets enter this distribution.

The b-jets are selected by requiring the medium working point of the CSVv2 b-tagging algorithm described in Sec 2.3.8. The CSVv2 discriminant is plotted on the bottom of Fig 3.14 before applying the b-tagging scale factors. The medium working point corresponds to a misidentification probability of light-parton jets (jets originating from gluons or u, d and s quarks) equal to 1%. By default, the medium working point is considered, which is justified by a couple of reasons. First, the presence of light or c-jets tagged as b-jets by the CSVv2 algorithm is small: the $t\bar{t}$ background provides two genuine physical b-jets per events, and it is known from previous studies that the proportion of $Z + c$/light does not contribute to more than 15% of the yields. Therefore, there is no reason to enforce the b-jet purity, which would degrade the signal efficiency significantly. Second, using a tighter working point may lead to a larger sensitivity to modeling issues especially at the level of the parton shower (fragmentation, hadron decay). The loose working point is not considered

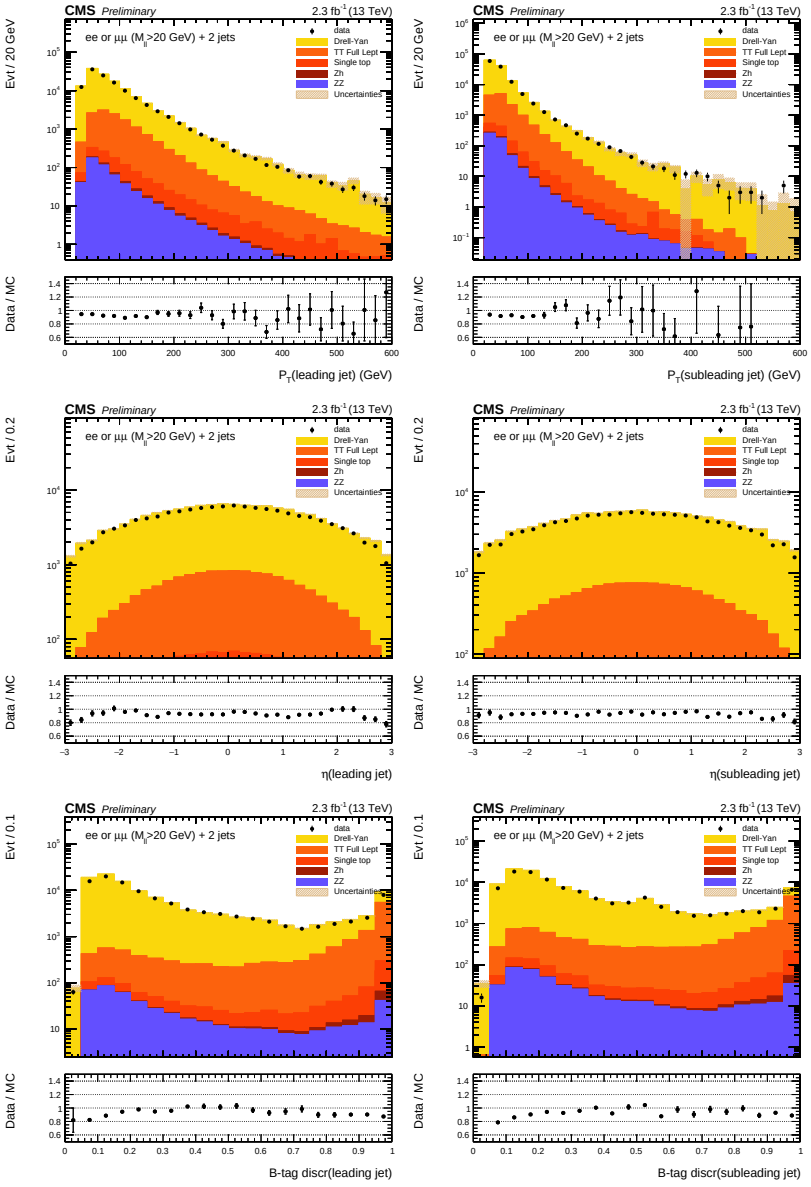

Figure 3.14: Top left(right): transverse momentum of the (sub-)leading jet. Center left(right): pseudo-rapidity of the (sub-)leading jet. Bottom left(right): (sub-)leading jet CSVv2 discriminant.

since the fake rate from Drell-Yan process is more problematic to control. The medium working point is therefore the most reasonable option.

The efficiency of the CSVv2 b-tagging has been shown to be different in data and simulations. Scaling factors are therefore used to correct the prediction from simulation. The baseline selection of events with two leptons and two b-jets is the last step used to check that backgrounds are well understood.

Table 3.1: Selection yields for basic selections aiming at testing the background modelling. The selection $\mu\mu$ and ee requires two same flavor leptons with $\mathrm{m}_{ll} > 20$ GeV. Two additional jets are required in $lljj$ selections and two b-tagged jets are required in the $llbb$ step.

Cat.	Drell-Yan	Single top	TT Full Lept	ZZ	Zh	Tot. MC	Data	Data/MC
$\mu\mu$	1564063.2 ± 104018.8	1047.1 ± 10.2	11079.0 ± 609.7	673.9 ± 1.0	57.8 ± 0.1	1576921.0 ± 104020.6	1530064	0.97 ± 0.06
$\mu\mu jj$	77996.7 ± 5203.8	444.3 ± 6.7	8440.9 ± 464.6	344.0 ± 0.8	38.6 ± 0.1	87264.6 ± 5224.5	81252	0.93 ± 0.06
$\mu\mu bb$	897.8 ± 75.7	77.1 ± 2.7	2641.0 ± 145.6	16.6 ± 0.2	11.9 ± 0.1	3644.5 ± 164.1	3385	0.93 ± 0.04
ee	723098.4 ± 48094.7	532.6 ± 7.3	5231.1 ± 288.1	322.1 ± 0.7	28.2 ± 0.1	729212.3 ± 48095.5	695462	0.95 ± 0.06
$eejj$	37497.3 ± 2510.5	228.9 ± 4.8	3954.4 ± 217.9	168.0 ± 0.5	19.0 ± 0.1	41867.6 ± 2520.0	38801	0.93 ± 0.06
$eebb$	439.3 ± 43.3	38.5 ± 1.9	1219.3 ± 67.4	8.2 ± 0.1	5.9 ± 0.0	1711.2 ± 80.2	1585	0.93 ± 0.04

Table 3.2: Selection yields for the selection of one Z candidate and two b-tagged jets. The yields are not scaled by the fit value (pre-fit), all the considered systematics are included in the uncertainties.

Cat.	Drell-Yan	Single Top	TT Full Lept	ZZ	Zh	Tot. MC	Data	Data/MC
$Z(\mu\mu)bb$	731.4 ± 128.1	31.9 ± 7.1	1175.8 ± 79.4	16.5 ± 1.8	12.0 ± 1.1	1967.7 ± 206.7	1877	0.95 ± 0.10
$Z(ee)bb$	384.2 ± 88.2	13.5 ± 3.2	525.0 ± 51.8	8.2 ± 1.2	5.9 ± 0.8	936.7 ± 141.3	859	0.92 ± 0.14

Table 3.3: Selection yields for the selection of one Z candidate and two b-tagged jets. The yields are scaled by the fit value (post-fit), the post-fit systematics are included in the uncertainties.

Cat.	Drell-Yan	Single Top	TT Full Lept	ZZ	Zh	Tot. MC	Data	Data/MC
$Z(\mu\mu)bb$	701.9 ± 52.0	31.9 ± 2.7	1100.8 ± 53.1	16.5 ± 0.9	12.0 ± 0.5	1863.2 ± 74.4	1877	1.01 ± 0.04
$Z(ee)bb$	368.7 ± 31.7	13.5 ± 1.4	491.5 ± 23.9	8.2 ± 0.5	5.9 ± 0.3	887.7 ± 39.7	859	0.97 ± 0.04

3.4 Analysis strategy

Due to the potentially large differences between the background and signal levels (and assuming that there is more background than signal), this analysis requires a method to control the background level accurately. The option of controlling the background level using the prediction from theory only cannot be considered here. The recent progress in terms of precision of event collision modeling through the use of higher-order matrix-element calculations and parton shower tunes, allow in general a better predictability in terms of normalization and/or differential cross-section for the standard model processes. But, there are still various physics cases where a large disagreement between the data and the predictions from theory is visible, even on "simple" quantities (number of jets, etc.).

In this work, the chosen strategy, while being conceptually simple, certainly fits the needs considering the available statistics. What is proposed is the following: for each tested couple of mass, a signal region is defined by a rectangle in $[m_{bb}, m_{llbb}]$ plane, the size of the sides being defined by three times the width of the signal peak (different for m_{bb} and m_{llbb}). This ensures to recover nearly 100% of the signal distribution. More details are discussed in Sec 3.5.4.

The control region is defined by the rest of the $[m_{bb}, m_{llbb}]$ plane, excluding the rectangle defined for the signal, as illustrated in Fig 3.15.

For each tested couple of masses, a fit is then performed in the control region to estimate the different backgrounds. The backgrounds are fitted using a histogram of the invariant mass of the dilepton system in the control region, which allows to efficiently separate the DY (dominated by the associated production of a Z boson and a pair of b-quarks) and the $t\bar{t}$ components. The MC simulations are used to obtain the ratio of number of events in the signal and in the control region for each background source. The expected background in the signal region is interpolated using this information in addition to the control region fit. The rescaling factor, together with the resulting statistical and systematic uncertainty, is propagated onto the signal region. The comparison between the data and the predicted background can be used for a statistical

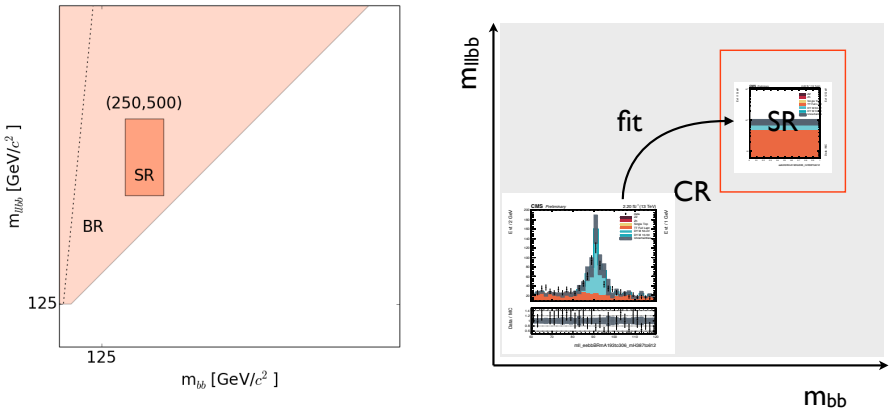

Figure 3.15: Left: Illustration of the signal and background region definition. Right: Illustration of the fitting procedure.

interpretation either in terms of the agreement between the data with the SM prediction, or with a particular signal hypothesis.

3.4.1 Systematic Uncertainties

For each hypothesis considered in $\mathcal{P}_{HA}$, a fit is performed in the corresponding CR to match appropriately the background processes to the data. This fit takes into account different sources of systematic uncertainties affecting the normalization or the shape.

- The luminosity uncertainty is set to 2.7%. It affects the normalization of each backgrounds.
- The uncertainty on the lepton reconstruction and efficiency is obtained by varying the scale factors applied on the theory predictions by one standard deviation. This uncertainty ranges from 1 to 4% depending on the p_T and η conditions.
- The uncertainties on the jet energy correction(resolution) are evaluated by shifting the correction factor up and down by one standard devia-

tion with respect to the default correction factors. The uncertainty associated to the correction ranges from 6% at low p_T to 1.5% around 100 GeV [111].

- The uncertainty associated with the correction for the b-tagging efficiency is obtained by varying up and down by one standard deviation. The b-tag scale factor (SF) variation is taken to be correlated with the c-mistag SF variation, while the light SF variation are supposed uncorrelated. Uncertainties of about 4% are associated to b-tagging efficiencies [112].

- The Drell-Yann and the $t\bar{t}$ normalization uncertainties are taken from the NNLO predictions.

- The scales uncertainties on the theory predictions are obtained by varying the factorization and renormalization scale by a factor two, independently, and ignoring the extremal variations: $\{\mu_F, \mu_R\} = \{/2, \times 2\}$ and $\{\times 2, /2\}$.

- The uncertainty coming from the PDF is also taken into account. A set of Monte Carlo replicas is generated by NNPDF taking into account the correlated systematics from the different measurements described in Sec 1.3.1. The uncertainty is obtained by running on these hundred replicas and taking as error the RMS of the variation.

For $\mathrm{m_H} = 300$ GeV, the b-tagging remains the source of the largest experimental systematic uncertainty while counting the number of events in the signal region across the whole $\mathrm{m_A}$ range, oscillating between 6 and 9%. Both JEC and JER uncertainties remain small, fluctuating around one percent. The distribution of the invariant mass built from the two b-tagged jets and from the two b-tagged jets and leptons from the Z boson decay is shown in Fig. 3.16.

The rescaling factor obtained from the fit, together with the resulting statistical and systematic uncertainty, is propagated onto the signal region, and the comparison between the data and the predicted background can be used for a statistical interpretation either in terms of the agreement between the data with the SM prediction, or with a particular signal hypothesis. Some distributions

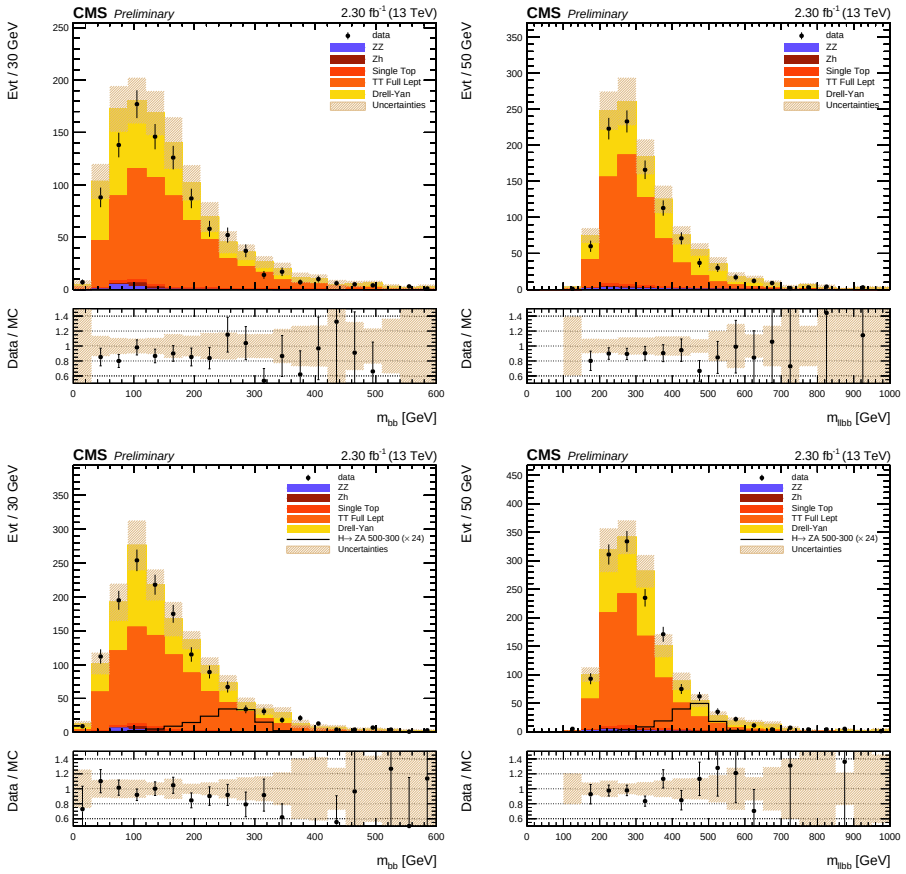

Figure 3.16: Pre-fit distribution of the invariant mass build from the two b-tagged jets (left) and the two b-tagged jets and leptons from the Z boson decay (right), for the ee-channel (top) and the $\mu\mu$-channel (bottom). The shaded band shows the quadratic sum of the systematic uncertainties, including the MC statistical uncertainties, and the error bars show the data statistical uncertainties. On the ratio plots, the shaded band shows the effect of the systematic uncertainties and the error bars show the data and MC statistical uncertainties added quadratically.

obtained by applying the rescaling factor are drawn in Fig 3.17. The top left figure of Fig 3.17 is particularly interesting since it shows that both the DY-like and the $t\bar{t}$-like backgrounds are well scaled.

This method allows to strongly constrain the different systematics and therefore reduce their impacts on the final results. For instance, the b-tagging uncertainty degrades the 300 GeV-100 GeV limit of about one percent and the JEC impacts the limit of less than half a percent. The effect is very visible while comparing the uncertainty bands in Fig 3.16 and Fig 3.17.

3.4.2 Statistical method

Following the discussion and notations from [113], the likelihood function is the product of the Poisson probabilities on each bin:

$$L(\mu, \vec{\theta}) = \prod_{i=1}^{N} \text{Pois}(n_i | b_i(\vec{\theta}) + \mu s_i(\vec{\theta})) \cdot \prod_{p} f_p(a_p|\theta_p) \tag{3.1}$$

where n_i is the observed rate, b_i is the background expected rate and s_i is the signal expected rate. The parameter μ is defined as the signal strength modifier. When $\mu = 0$, the likelihood obtained corresponds to the background only case, and when $\mu = 1$ it correspond to the reference signal model. Here, the refence signal model has a cross section of 1 fb. $\vec{\theta}$ contains all the nuisance parameter terms and the $f_p(a_p|\theta_p)$ are the constraints terms.

Shape uncertainties Since the number of nuisance parameter is high, the multi-dimentional modelling of the nuisance parameter is very complex. To simplify the problem, what is usually done is to have a nominal setting of the parameters α^0 and then individually vary each parameter 'up' and 'down' by a reasonable amont. For each sample, one has to interpolate and extrapolate these predictions to produce the parametrized variation in rate and shape. The shapes are interpolated quadratically for shifts below 1σ and linearly beyond. When building the likelihood, each shape uncertainty is associated to a nuisance parameter taken from a unit gaussian distribution.

In this analysis, the histograms and the up and down variations of the histograms for each non-constant systematic are produced (b-tagging, jet energy correction and resolution, parton density functions, lepton identifications, scales (μ_F and μ_R).

Rate uncertainties The functional form of the constraints are taken to be a log-normal distribution with the computed uncertainty used as standard deviation for the normalization uncertainties. For instance, log-normal uncertainties are used for the background cross-sections and for the luminosity.

Test statistic, p-value and $\mathrm{CL_S}$ In this analysis, the procedure to establish a model exclusion is based on a frequentist significance test using the likelihood ratio as statistic:

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

where $\hat{\theta}$ and $\hat{\mu}$ are the values of the parameters that maximize the likelihood function $L(\mu, \vec{\theta})$. $\hat{\hat{\theta}}$ is the value of $\vec{\theta}$ that maximizes the likelihood function with μ fixed. It is often called *the profiled value of* θ.

The p-value associated with the hypothesis of a signal strenght μ is written:

$$p_\mu = \int_{q_{\mu \mathrm{obs}}}^{\infty} f(q|\mu, \hat{\hat{\theta}}(\mu, \mathrm{obs}))dq_\mu \tag{3.3}$$

where $q_{\mu \mathrm{obs}}$ is the value of the test statistic for the observed data and f is the pdf of the test statistic q_μ. The most general method, called "toy Monte Carlo", is to perform the integral by generating pseudo-experiment. Here, the distributions are assumed to take on their asymptototic form. Therefore, the *Higgs combine tool* is used to compute these upper limit on the cross-section in the asymptotic limit [114]. The p-value has the feature that for a downward fluctuation, the upper limit of the confidence interval can be arbitrarily small. Therefore, in particle physics, a modified frequentist approach called $\mathrm{CL_S}$ is often used [115]. It is defined as a ratio of p-values

$$p'_\mu = \frac{p_\mu}{1 - p_b} \,, \tag{3.4}$$

where p_b is the p-value associated with the background only hypothesis is given by:

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

The CL_S upper limit μ_up at 95% is obtained by solving $p'_{\mu_\text{up}} = 5\%$. With the choice on the definition of μ, the obtained μ_up provides directly the excluded cross-section.

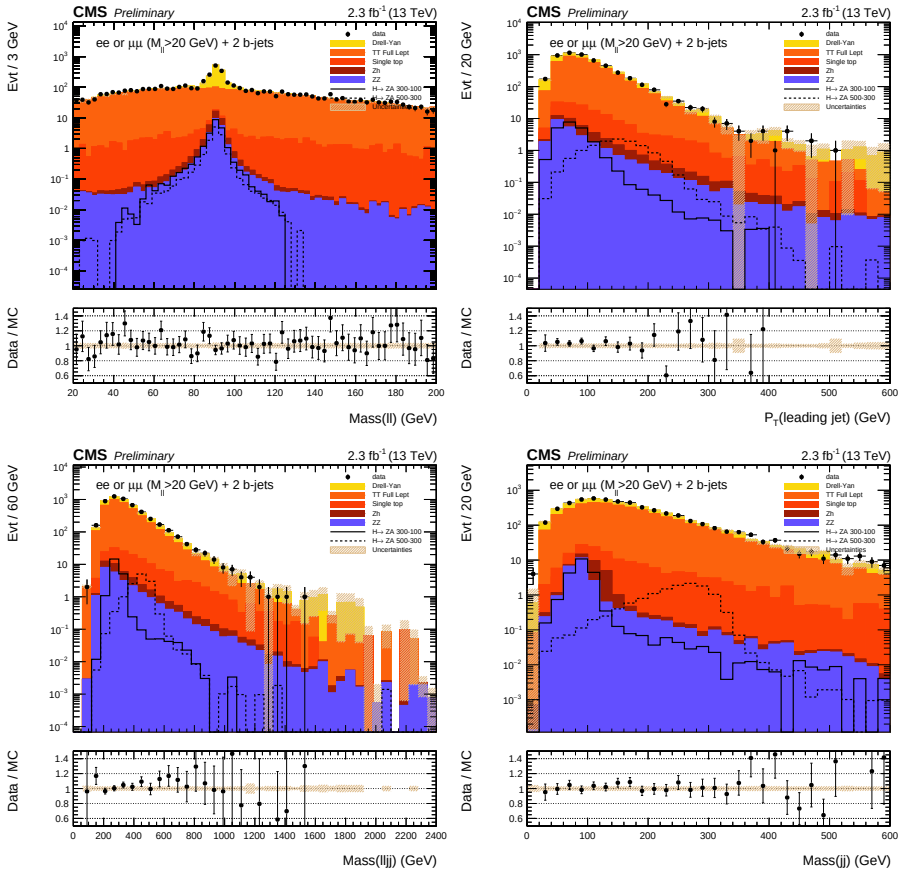

Figure 3.17: Post-fit distributions after the selection of events with two same flavor leptons and two b-tagged jets. Top left: invariant mass of the di-lepton system. Top right: transverse momentum of the leading b-tagged jet. Bottom: Post-fit distributions of the di-jets invariant mass (left) and the four-body mass (right) after the selection of events with two same flavor leptons and two b-tagged jets.

3.5 Going back: 8 TeV analysis

Before looking at the results, let's discuss the first dedicated search for the $\mathrm{H} \to \mathrm{ZA} \to l^+l^-b\bar{b}$ with the 8 TeV data from the LHC since it is also a work realized as part of this thesis. Since the principles are similar to the 13 TeV analysis, the accent will be put on the few differences first. Then, each aspect of the analysis performed with the DELPHES framework will be reviewed.

3.5.1 Some differences

The basic selection was similar at 8 TeV. One extra selection was applied on the E_T^{miss}-significance where it was required to be less than 10 in order to further suppress background events originating from the $t\bar{t}$ processes. The E_T^{miss}-significance [116] is an event-by-event assessment of the likelihood that the E_T^{miss} is consistent with zero, given the content of the reconstructed event and the experimental resolutions.

The background estimation is the main difference between the two analyses (8 vs. 13 TeV). The method used at 8 TeV has been originally designed for a standard model measurement: the Z plus heavy flavor jets cross-section. It was then mandatory to efficiently distinguish the other main processes, namely $t\bar{t}$ and Z plus light jets from the Z plus heavy flavor processes. The DY categorization was done as follow:

- The Zbb was the category with events containing two b-originated jets and no extra-jets.
- The Zbx contained the events with three jets including two b-originated jets and events with two jets of which one is originating from b-hadrons.
- The Zxx was the category with two or three reconstructed light jets.

To perform the background normalization, the three DY contributions and the $t\bar{t}$ contribution were left free to float in a single two-dimensional fit to the shape of the following observables:

- The product of the CSV discriminants of both selected jets was designed to discriminate the Zbx and Zxx categories versus the Zbb and $t\bar{t}$.
- The invariant mass of the lepton pair discriminated the three DY contributions versus the $t\bar{t}$ events.

In this analysis, the backgrounds were therefore estimated once and for all, assuming that the signal is small enough to be negligible at this stage.

3.5.2 DY Mismodeling

A deviation between the data and the prediction has been observed in the distribution of the m_{llbb} mass. The magnitude of the deviation ranges up to 20% in some region after the background fit. Several investigations have been performed and it was shown that the discrepancy was also present with looser selection: one Z candidate and two jets. These discrepancies were due to the fact that the DY samples were generated only at LO, since big samples are needed for this analysis.

Given the huge amount of resources needed to produce an NLO sample with a decent number of events, the strategy is to apply a re-weighting technique to the existing MADGRAPH 5 reconstructed sample already in use for this search. A comparison between the LO MADGRAPH 5 sample and a MADGRAPH_aMC@NLO NLO sample has been performed and has shown a disagreement comparable to what is seen in the data. The comparison has been made after reconstructing both generated samples: MADGRAPH 5 + PYTHIA6 and MADGRAPH_aMC@NLO + PYTHIA8 with DELPHES.

Even though they are related, the NLO effects are expected to be different for the Z+jets and for the Z+heavy flavor events. Therefore, it was mandatory to separate the re-weighting according to the originating partons or hadrons of the jets in the events. The categorization used is the same as in the background estimation: Zxx, Zbx and Zbb. The NLO/LO ratio is shown in Fig 3.18 in which the jet flavor is distinguished (light or heavy). The ratios have then been fitted using a third-order polynomial function. These functions are the ones that have been used to re-weight the LO MADGRAPH 5 sample. The method

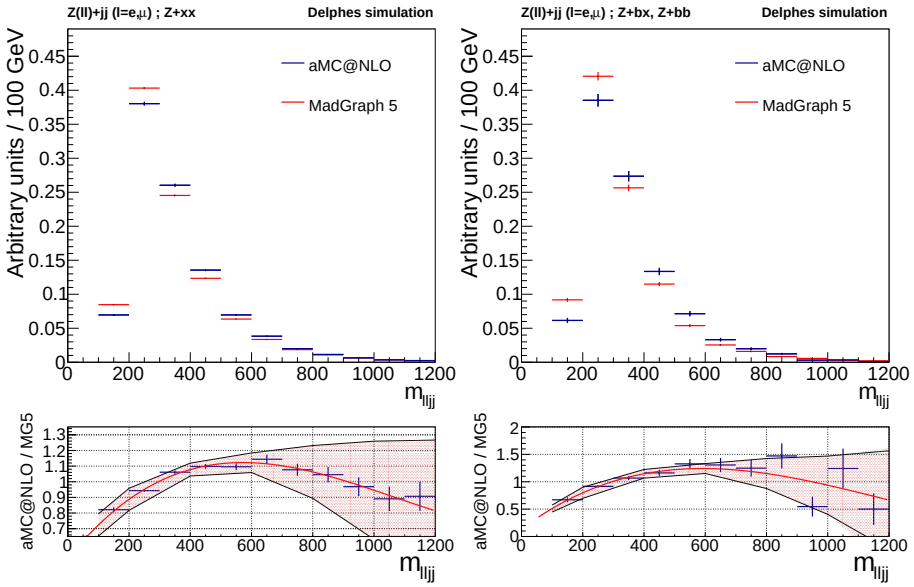

Figure 3.18: Comparison between LO sample MADGRAPH + PYTHIA6 and the NLO sample MADGRAPH_aMC@NLO + PYTHIA8 for the four-body masses (two leptons and two light jets on the left, two leptons and two b-tagged jets on the right). Both generated samples have been reconstructed using DELPHES simulation.

has shown to improve the data/MC agreement of the m_{llbb} without impacting negatively the other variables.

A systematic uncertainty has been assigned to this re-weighting. In addition to the four parameter estimation, their uncertainty and the covariance matrix are derived as a result of the fit performed. A set of curves is generated according to the uncertainties and taking into account the correlation of the parameters. This is done keeping the same normalization such that only the shape is affected.

The curves corresponding to the $\pm 1\sigma$ variation of the parameters are shown on the ratio plots of Fig 3.18. The final effect on the results has been checked and turned out to be a major systematic ranging from 10% at low m_{llbb} to 25-30% at high values.

3.5.3 Many signal samples

For each bin in $[\mathrm{m_A} - \mathrm{m_H}]$, an estimation of the expected quantity of signal events passing the selection is needed. Therefore, more than 300 signal samples were generated using MADGRAPH 5 and PYTHIA with masses corresponding to the center of the different rectangular signal regions. While only twelve of these were produced using the CMS full simulation, the DELPHES fast simulation as been used on each sample. DELPHES has shown to reproduce realistically the performance of the CMS detector 10.000 times faster than the full simulation of the CMS detector. Nevertheless, the tuning of the parameters, efficiencies and smearing is mandatory to go beyond the typical 10% agreement obtained using public CMS informations. In the tuning procedure, we worked on the main reconstruction and selection effect of this analysis: the tracking efficiency, the electron and muon reconstruction efficiency and isolation, the B-tagging parametrization.

The b-tagging efficiency depends on the final state particles and momentum. Twelve fully simulated samples were therefore used to extract the p_T and η-dependence of the b-tagging for this particular topology. The b-tagging efficiency has been implemented in DELPHES.

Once all the mentioned DELPHES tuning was performed, the variables of interest such as $\mathrm{m_{bb}}$ and $\mathrm{m_{llbb}}$ have been compared with the full simulation. These distributions are shown in Fig 3.19 for two specific mass choices. The agreement observed is satisfactory in the light of the many simplifications described in the second chapter.

The "only goal" of these samples is to provide the efficiency for a sample to pass the basic selection and rectangular cut centered at the mass where it has been generated. So, this efficiency is the important quantity to compare with the full-simulation results and it was found that the agreement was within 10%, see Tab 3.4. The very low mass comparison suffers from an important statistical uncertainty coming from the fully simulated sample, since the efficiency is very small. In addition, efficiencies for this signal are particularly difficult to model since the leptons and b-jets tranverse momentum are very low and close to the selection criterias. A complete 2D map is extracted from the ratios $\frac{\epsilon_{\mathrm{CMS}}}{\epsilon_{\mathrm{DELPHES}}}$. The interpolation is performed using the Delauney triangulation and

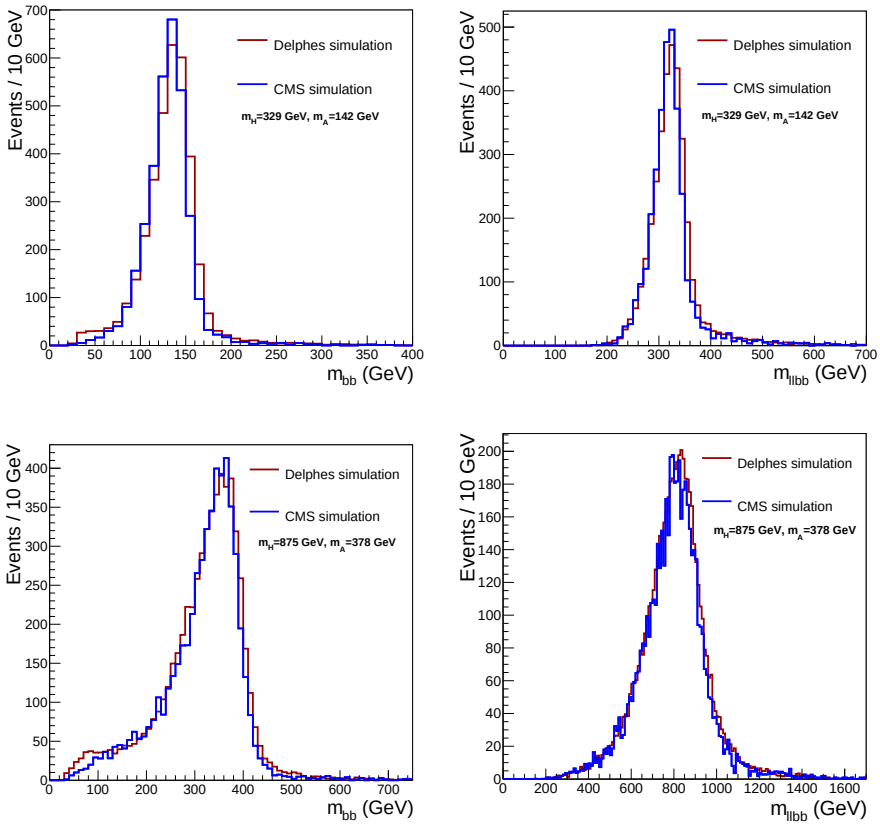

Figure 3.19: DELPHES simulation in comparison with the standard CMSSW for two benchmark mass points for the $\mathrm{m_{bb}}$ (left) and $\mathrm{m_{llbb}}$ (right).

m_A	m_H [GeV]				
[GeV]	142	200	329	575	875
30-35	41 ± 24		83 ± 10		
50		84 ± 5			
70				96 ± 2	91 ± 4
90		80 ± 3			
142			93 ± 2		93 ± 2
378				89 ± 2	91 ± 1
575					89 ± 2
761					85 ± 2

Table 3.4: The ratio $\frac{\epsilon_{\text{CMS}}}{\epsilon_{\text{DELPHES}}}(\%)$ of the efficiencies between DELPHES and the full simulation for a few representative benchmarks.

the map is also extrapolated to cover the entire parameter space. This map is then used to multiply the DELPHES based efficiencies, which provides a better modeling, and in particular allows to recover the full-simulation efficiencies for the dedicated points.

A systematic uncertainty is derived from the interpolations and assigned to the final results. When the ratio was extrapolated, the uncertainty was set to 10%.

3.5.4 Analysis design: bin size

The choice of the bin size is a key ingredient of this search. On the one hand, the rectangle size has to be chosen to maximize the significance of signal events inside the selection cuts. On the other hand, since the search has to be easily re-casted, the bin size must be as simple as possible and not depend on the 2HDM parameters (which change the width of the scalar resonances).

The choice made is to take the bin size as three standard deviations of the expected width of each resonance. The experimental resolution on the di-jet and the four body masses is assumed to be 15% of the reconstructed mass. To validate this assumption, the DELPHES samples have been used to compute

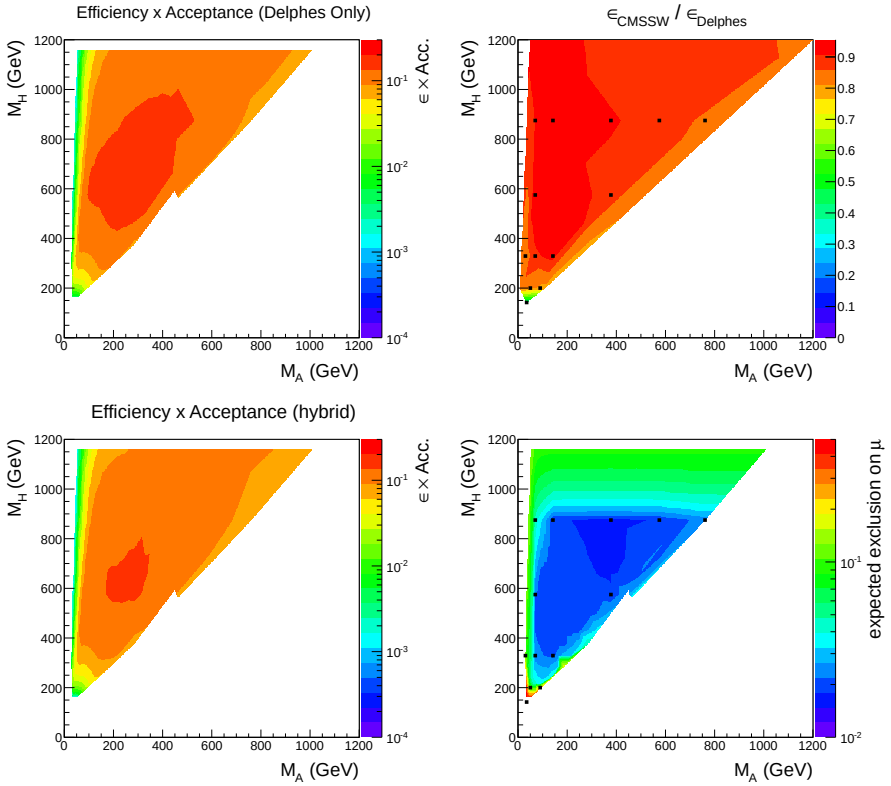

Figure 3.20: Top left: Map of the efficiency $\times$ acceptance for the signal mass points using DELPHES. Top right: Map resulting from the interpolation of the efficiencies ratio reported in Tab 3.4. Bottom left: Hybrid efficiency map obtained as a product of the "DELPHES efficiency" (top left plot) and the efficiency ratio (top right plot). Bottom right: Systematic uncertainty assigned on the hybrid signal x acceptance in the $[\mathrm{m_A}, \mathrm{m_H}]$ plane.

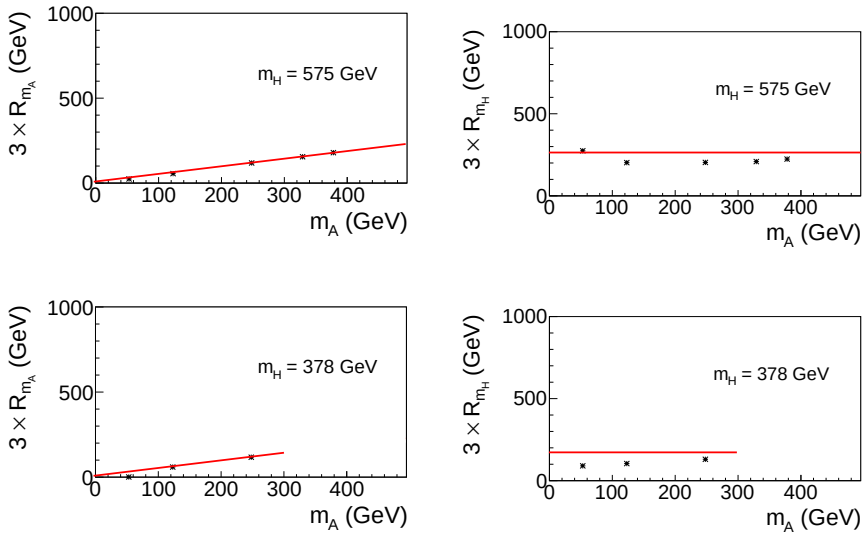

Figure 3.21: Left: di-jet mass resolution as a function of the pseudo-scalar mass for $\mathrm{m_H} = 575(378)$ GeV in the top(bottom), the 45% line is drawn in red. Right: four-body mass resolution as a function of the pseudo-scalar mass for $\mathrm{m_H} = 575(378)$ GeV in the top(bottom), the 45% line is drawn in red.

the standard deviation of the signal peak in function of the masses it has been generated with. The Fig 3.21 shows the result for two choices of $\mathrm{m_H}$: 378 and 575 GeV. The left plots are the di-jet mass resolution as a function of the pseudo-scalar mass and show a very good agreement with the $3 \times 15\%$ (drawn in red) choice. The right plots are the four-body mass resolution as a function of the pseudo-scalar mass. The 15% lines provide a good order of magnitude, and are higher than the experimental resolution, allowing to not lose signal events. In fact, it is expected, since the two leptons have a much better resolution than jets. Combining the four objects therefore tends to reduce the σ_M/M ratio and the reduction depends on the generated masses.

So, for simplicity, the $3 \times 15\%$ *of the mass* has been chosen in these analyses for both $\mathrm{m_{bb}}$ and $\mathrm{m_{llbb}}$.

The natural width of the 2HDM heavy scalar is plotted in Fig 3.22 for comparison. The natural width reaches the same order of magnitude as the experimental resolution only for high $\mathrm{m_H}$ and low $\mathrm{m_A}$. Using signal samples generated with this model guarantees that the results are valid for typical choices of 2HDM with masses up to 1 TeV. Nevertheless, computing the excluded cross section for models with higher width in that region requires the computation of a new efficiency map as in Fig 3.20.

3.6 Results at 8 and 13 TeV

3.6.1 Limits on cross-section

At 13 TeV, in addition to the 1- and 2-σ CLs, the observed and expected limits are displayed in Fig 3.23 for several $\mathrm{m_A}$ hypotheses and $\mathrm{m_H}$ being 300 GeV (top left), 500 GeV (top right) and 800 GeV (bottom left). For $\mathrm{m_H} = 300$ GeV, the expected limit on $\sigma \times BR$ ranges from $300\mathrm{fb}$ to $1000\mathrm{fb}$. The limits are lower, around $150\mathrm{fb}$, for $\mathrm{m_H} = 500$ and 800 GeV since backgrounds decrease (Fig 3.17). The sensitivity gets worse for the point $\mathrm{m_H}, \mathrm{m_A} = 800$ GeV, 50 GeV because the typical ΔR distance between the two the b-tagged jets

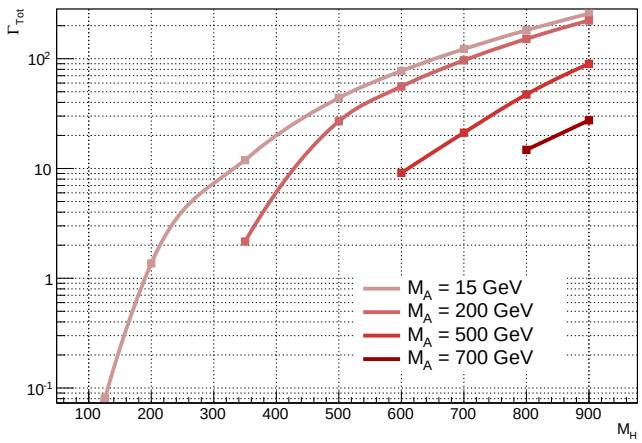

Figure 3.22: Natural width of the heavy scalar in the 2HDM type-II, with $\tan\beta = 1.5$ and $\cos(\beta - \alpha) = 0.01$

become smaller than the jet size. The observed limit is in agreement with the expected limit for each $\mathcal{H}^{\mathrm{m_H}}_{\mathrm{m_A}}$.

At 8 TeV, the complete map is computed thanks to the use of hundreds of DELPHES samples. The bottom right plot of Fig 3.23 shows the observed upper exclusion limits on the cross section of the $\mathrm{H} \rightarrow \mathrm{Z}(ll)\mathrm{A}(bb)$. Since no parity dependent variables were used, the limit has also been interpreted in terms of the process $\mathrm{A} \rightarrow \mathrm{Z}(ll)\mathrm{H}(bb)$. Typically, a cross section around 10-100 fb is excluded. Two excesses with local significance above 2σ have been found, these are discussed in detail in A.Caudron thesis [117]. As for the 13 TeV results, the excluded cross section is higher around 300 GeV-100 GeV, which is the region with the highest background, and in the triangular region where $\mathrm{m_A}$ is lower than $\frac{\mathrm{m_H}}{10}$ in which the b-tagged jets are entering the collinear regime.

Following the discussion in Sec 3.4, the limit is valid for any model in which the scalar and pseudo-scalar width are small compared to the bin size, or to the used model.

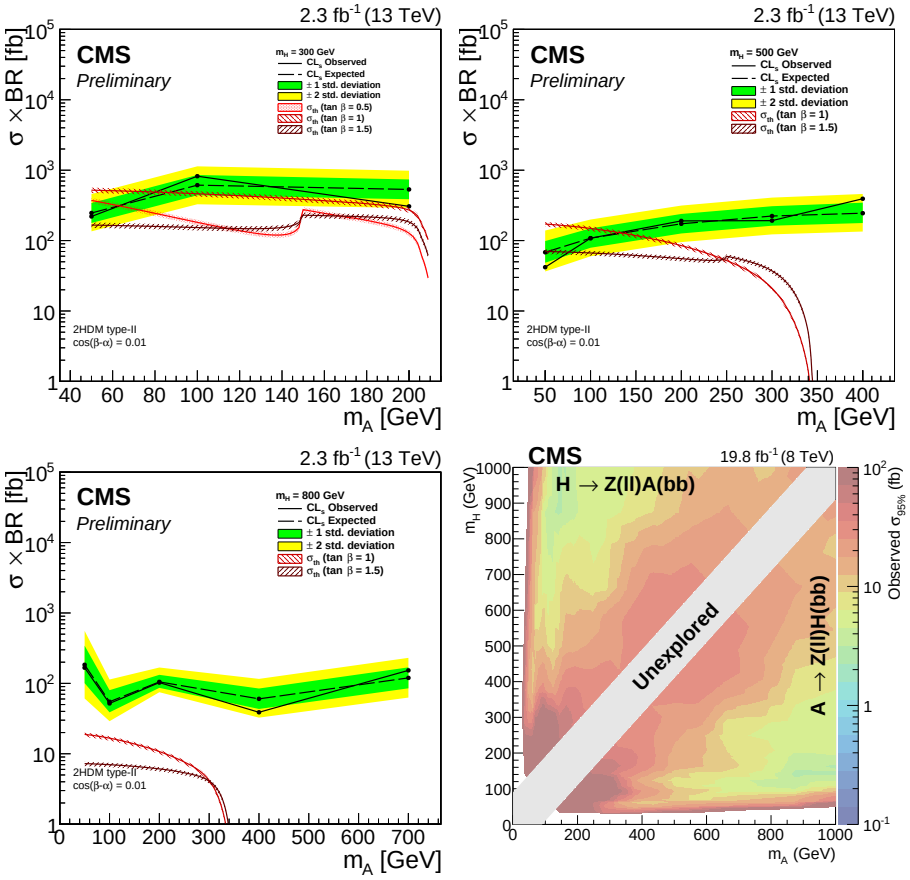

Figure 3.23: Limit on the $\sigma \times \mathrm{BR}(\mathrm{H} \to \mathrm{ZA}) \times \mathrm{BR}(\mathrm{Z} \to ll) \times \mathrm{BR}(\mathrm{A} \to b\bar{b})$. At 13 TeV for $\mathrm{m_H} = 300$ (top left), 500 (top right), 800 (bottom left) GeV where the predicted production cross sections of $pp \to \mathrm{H} \to \mathrm{ZA} \to l^+l^-b\bar{b}$ are also shown for different values of $\tan\beta$. At 8 TeV in the bottom right in the plane $[\mathrm{m_A}, \mathrm{m_H}]$.

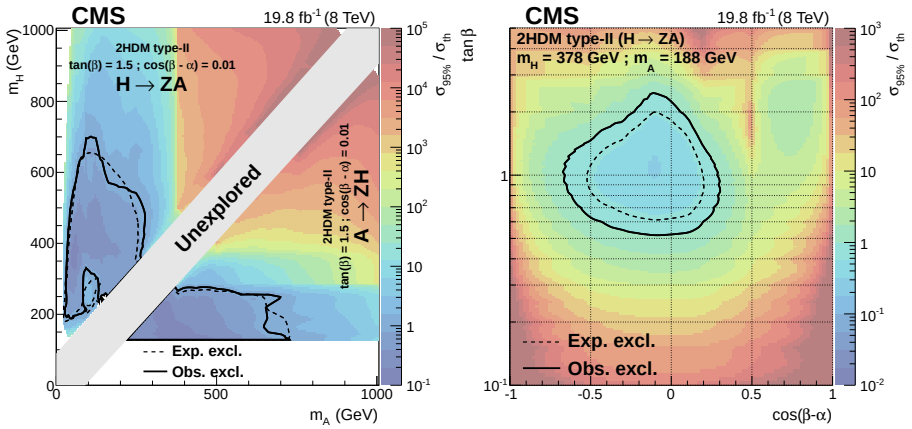

Figure 3.24: Observed limit on the signal strength $\mu = \sigma_{95\%}/\sigma_{\text{th}}$ after combining results from the $llbb$ and $ll\tau\tau$ final states. The dashed contour shows the region expected to be excluded. The solid contour shows the region excluded by the data. Left: Limits are shown in the plane $[\mathrm{m_A}, \mathrm{m_H}]$ for the considered 2HDM benchmark. Right: Limits are shown in the plane $[\cos(\beta-\alpha),\tan\beta\,]$ for the signal masses of $\mathrm{m_H} = 378$ GeV and $\mathrm{m_A} = 188$ GeV.

3.6.2 2HDM interpretation of 8 TeV results

The 8 TeV limits on the cross-section have been combined with the similar $ll\tau\tau$ analysis [118] and interpreted as a limit on a signal strength defined as the ratio between the excluded cross-section at 95% confidence level and the 2HDM benchmarks theoretical cross-section. The signal strength is interpolated on the left part of Fig 3.24 for any masses of the heavy scalar in the range $[140, 1000]$ and of the pseudo-scalar in the range $[10, 1000]$.

In the MSSM-like mass hierarchy, this model is excluded from approximately 200 to 600 GeV for the pseudo-scalar mass. The inverted hierarchy case is more subtle. In the left part of the plot, the pseudo-scalar is very light. A small ratio $\frac{\mathrm{m_A}}{\mathrm{m_H}}$ implies a boosted topology, which mean that the two b-quarks induce only one jet. These events are therefore not selected and the analysis is not sensitive in that region, which explains the shape of the not excluded

region. There is also a not excluded island around $[\mathrm{m_A}, \mathrm{m_H}] = [100, 250]$ corresponding to the highest background region.

To illustrate the statement that this analysis is motivated by the alignment limit, the mass point: $[\mathrm{m_A}, \mathrm{m_H}] = [188\ \mathrm{GeV}, 378\ \mathrm{GeV}]$ is selected to plot the signal strength in the [$\cos(\beta-\alpha)$-$\tan\beta$] plane into the right of Fig 3.24. The data allows to exclude a region between $\tan\beta = 0.5$ and $\tan\beta = 2$ in the alignment limit. A lower $\tan\beta$ induces a small $\mathrm{A} \to b\bar{b}$ branching fraction, while a high $\tan\beta$ implies a smaller $gg \to H$ production cross-section.

Since only three masses have been tested with the 2015 data, some theoretical cross section lines have been drawn on the three plots of Fig 3.23, allowing an easy interpretation.

3.6.3 2HDM results combination

Several analyses with different final states are complementary to study the 2HDM. Therefore summary plots with the different analysis performed with the CMS detector have been produced in [58]. The choice has been to fix the value of $\cos(\beta-\alpha)$ at 0.1, since it is close to the constraints obtained from the Higgs coupling measurements (see Fig 1.10). It is also assumed that the mass difference between the scalar and the pseudo-scalar is 100 GeV.

To ensure a stable vacuum and perturbativity, the soft-$\mathcal{Z}_2$ symmetry breaking parameter is taken to be:

$$\mathrm{m_{12}} = \max(1 - \tan\beta^{-2}, 0) \cdot \frac{1}{2}\sin(2\beta)(\mathrm{m_A}^2 + \lambda_5 v^2) \tag{3.6}$$

according to [119, 58].

The cross section and the partial decay width have been computed using SusHi and 2HDMC for the 2HDM type-I and type-II. A table with all parameters is given in Tab 3.5.

The exclusions are plotted on the left(right) of Fig 3.25 for the type-I(type-II).

The h(125) results are directly taken from the indirect constraints from h measurements [40]. These are explained in Sec 1.3.4 and is visible in Fig 1.10.

Table 3.5: Parameters of the 2HDM type-I and II used in this section. The 2-dimensional scan is performed in $\mathrm{m_H}$-$\tan\beta$.

Parameter	Value
$\mathrm{m_h}$	125.09 GeV
$\mathrm{m_A}$	$\mathrm{m_H} + 100$ GeV
$\mathrm{m_{H^\pm}}$	$\mathrm{m_H} + 100$ GeV
$\cos(\beta - \alpha)$	0.1
$\mathrm{m_{12}^2}$	$\max(1 - \tan\beta^{-2}, 0) \cdot \frac{1}{2}\sin(2\beta)(\mathrm{m_A}^2 + \lambda_5 v^2)$

The $\mathrm{A} \to \mathrm{ZH}$ analyses are helping when the pseudo-scalar mass is smaller than twice the top mass. This is translated in a limit around $\mathrm{m_H} = 250$ GeV that can be observed in both type-I and type-II plots. In both models, the production cross-section decreases with $\tan\beta$. It is excluded up to $\tan\beta \approx 1 - 1.5(2)$ in the type-I(type-II).

The $\mathrm{A/H/h} \to \tau\tau$ [120] is, as expected from Fig 1.13 and 1.14, a very constraining final state. In particular, the $\tau\tau$ final state tackles more the low-$\tan\beta$ case in the type-I and the high-$\tan\beta$ in the type-II.

The $\mathrm{H} \to \mathrm{WW/ZZ}$ analyses [61] are helpful when the heavy scalar is heavier than twice the W-mass and lighter than twice the top mass. We note here the fact that this analysis loses its relevance when getting closer to the alignment limit.

3.7 Going beyond

The first conclusion from the 2015 data analysis is that the standard model describes the observed data extremely well, as confirmed in the many results presented in the ICHEP 2016 conference. The tools used such as the NLO event generators coupled with parton shower and the full detector simulation provide a very accurate modeling of the proton-proton interactions. In addition, the LHC is (at the time of writing the thesis) providing data with the expected rate and more than 37 fb^{-1} have been collected in 2016.

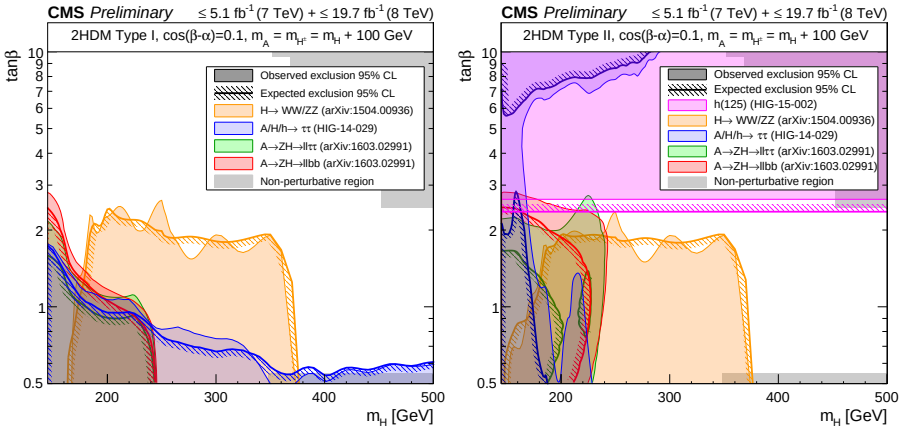

Figure 3.25: 95% CL exclusion contour, in two 2HDM scenarios of type-I on the left and type-II on the right. These limits are obtained by several analyses performed with the run-I dataset. The colored lines with hatches indicate the regions that were expected to be excluded based on the null-hypothesis assumption of a SM-like Higgs sector. The gray regions correspond to non-perturbative or unstable scenarios of 2HDM.

Taking these facts into account, as well as the many arguments in the analyses' descriptions, leads to several ideas for future developments.

Background estimation will therefore be a central question. Indeed, using MC doesn't induce significant systematic error as far as there more simulated events than observed events. Once this condition is not fulfilled, it is important to estimate the background using the data only. Different choices are possible, for instance defining a signal region in the $[m_{bb}, m_{llbb}]$ plane and estimate the background using a region surrounding the signal region.

Many signal samples are needed to make a precise 2D scan in $\mathrm{m_A} - \mathrm{m_H}$. A combination of fully simulated and parametric samples has shown to efficiently provide a complete map at a decent cost in computing-time.

Cut optimization Tilted elliptic selection could drastically improve the analysis performances. For a given value of $\mathrm{m_A}$ and $\mathrm{m_H}$, the $\mathrm{m_{bb}}$ and $\mathrm{m_{llbb}}$ will be correlated since they are built from the two same b-jets. The correlation is dependent on $\mathrm{m_A}$ and $\mathrm{m_H}$ and a tilted elliptical selection would take the correlation into account. The relation between $\mathrm{m_A}$, $\mathrm{m_H}$ and the ellipse angle is under study.

The boosted regime defined as the region where $\mathrm{m_A} < \mathrm{m_H}/10$ is not probed with the current analysis. Indeed, in this region, the ΔR distance between two jets originating from the pseudo-scalar is smaller than the jet size used in the analysis. Such events would therefore not be selected when requiring two b-tagged jets. The key ingredient is the requirement of one *fat-jet* with a ΔR-cone of 0.8 for instance. The signal fat-jets are particular since they are induced by two b-quarks. So, the fat-jet will contain two prongs and both *sub-jets* are originating from b-quarks. The exploration of the analysis optimization to tackle this parameter space have been studied by Liliya Milenska in her master's thesis [121].

Conclusion

I started my PhD thesis in 2012, just after the discovery of a new scalar boson, at the end of the first run of the LHC. The second run of the LHC was foreseen in 2015 with two main goals: The study of the properties of the newly discovered boson and the complete exploration of the TeV scale. *Is it really the Higgs boson?* is the question that motivated me during the first months. To address it, we started working on tools to measure the spin and parity of this new particle using the Matrix-Element-Method in three different final states ($ZZ \to llll$, $WW \to ll\nu\nu$ and $Zh \to llbb$). At that time, a simplified Matrix-Element-Method was used in the ZZ analysis and we observed no significant gain by applying a more general method. Soon, the ATLAS and CMS released the first official results that favored the scalar hypothesis and excluded the spin-2 hypothesis. After more than one year of research, I rapidly gained interest in the second goal: searching for new physics. To me, the discovery of a first fundamental scalar particle strongly motivates the searches for new spin-0 particles. Indeed, while the existence of the electroweak symmetry breaking mechanism is now a piece of evidence, the scalar particle content could not be limited to the Higgs boson. In fact, extra scalar bosons arise naturally when extending the standard model of particle physics which describes beautifully the experimental observations but still contains many theoretical difficulties.

The 2HDM is a very generic extension of the scalar sector in which an extra Higgs doublets is added to the standard model. Many existing indirect and direct searches are constraining the 2HDM, including the measurement of the coupling of the Higgs to the other standard model particles. With the inter-

esting parameter choice called *alignment limit*, in which the 2HDM lightest scalar behaves as the standard model Higgs boson, the 2HDM survives the existing observations. In this limit, the processes $pp \to \mathrm{H} \to \mathrm{ZA} \to l^+l^-b\bar{b}$ or $pp \to \mathrm{A} \to \mathrm{ZH} \to l^+l^-b\bar{b}$, depending on the considered mass hierarchy, are strongly enhanced. The main work of thesis is the search for such processes with the CMS detector using data taken during the first run and the first year of the second run of the LHC. The LHC has delivered more than 23 fb^{-1} of proton-proton collisions in 2012 with 8 TeV in the center of mass and $4.2\ \mathrm{fb}^{-1}$ at 13 TeV in 2015.

Having in mind the large variety of 2HDM and more complex models that could involve such processes, the analysis strategy has been chosen to be easily re-casted or reinterpreted. The adopted strategy provides model independent results, which allow to exclude a cross section of $10\ \mathrm{fb}$ for a large fraction of the parameter space since no significant deviation from the standard model expectation has been found. The analysis allows to exclude parts of the parameter space in the alignment region, which is unreachable using only the Higgs boson properties.

In 2016, the LHC has delivered more than 36 fb^{-1} of data with 13 TeV in the center of mass. It is the first time in history that such an amount of data is taken and the complete second run is expected to contain 150 fb^{-1}. This amount of data is a motivation to reduce the dependence on the backgrounds simulation. Several improvements described in the third chapter are therefore foreseen and include a set of optimizations of the analysis aiming at improving its sensitivity while keeping the model independent approach. In fact, the analysis performed in this work is not only sensitive to both $\mathrm{H} \to \mathrm{ZA}$ and $\mathrm{A} \to \mathrm{ZH}$ processes, but also for any resonance decaying into a Z and another resonance. Therefore, it covers as well processes like $\mathrm{A} \to \mathrm{Zh}$, $\mathrm{h} \to \mathrm{ZA}$ or a new resonance decaying to a pair of Z bosons. The goal of the planned improvements is to reach a competitive sensitivity also for such particular processes.

The $\mathrm{H} \to \mathrm{ZA}$ process is not the only exotic decay in hierarchical 2HDM: the process $\mathrm{H} \to \mathrm{AA}$ is also potentially important. While the $\mathrm{h} \to \mathrm{AA}$ process has been intensively studied already with the CMS detector, the case where the initial resonance mass is different from 125 GeV hasn't been studied yet. The final state with two b-jets and two muons is potentially very interesting and

if an excess is found in the $H \to ZA$ analysis, this is the place to look for a related excess and a better interpretation of an eventual new particle. Extending the current analysis to di-muon resonances with a mass different from the Z-mass could therefore improve strongly the sensitivity to many extensions of the scalar sector.

Alongside this, we have provided the community with a modular parametric detector simulation widely used and ideal for recasting. The current DELPHES version is able to tackle events as complex as the foreseen 200 pileup events and apply state-of-the-art pileup subtraction. The emulation of the particle flow has been also significantly improved. The program has shown to realistically reproduce the performances of the CMS detector at the LHC. In addition to many other applications, DELPHES appeared as a useful solution to perform the many detector simulations of 2HDM processes events needed to cover the full 2HDM A and H masses spectrum as well as for establishing first physics prospect for the HL-LHC.

To summarize, we have explored a new region in the 2HDM parameter space by searching for new resonances with exotic decays $H \to ZA$ or $A \to ZH$. No evidences of new physics has been found with the data collected in 2012 and 2015, we have therefore partially excluded the parameter space. The next steps are the optimization and the generalization of the analysis and aiming at gaining in sensitivity and increasing the number of studied processes.

Bibliography

[1] S. Kanesawa and S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields. IV — Case of Interacting Electromagnetic and Meson Fields". In: *Progress of Theoretical Physics* 3 (Jan. 1948), pp. 1–13.

[2] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". In: *Physical Review* 74 (Nov. 1948), pp. 1439–1461. DOI: 10.1103/PhysRev.74.1439.

[3] J. Schwinger. "On Quantum-Electrodynamics and the Magnetic Moment of the Electron". In: *Physical Review* 73 (Feb. 1948), pp. 416–417. DOI: 10.1103/PhysRev.73.416.

[4] F. J. Dyson. "The Radiation Theories of Tomonaga, Schwinger and Feynman". In: *Phys. Rev.* 75 (3 Feb. 1949), pp. 486–502. DOI: 10.1103/PhysRev.75.486. URL: http://link.aps.org/doi/10.1103/PhysRev.75.486.

[5] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". In: *Physical Review* 76 (Sept. 1949), pp. 769–789. DOI: 10.1103/PhysRev.76.769.

[6] R. P. Feynman. "Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction". In: *Physical Review* 80 (Nov. 1950), pp. 440–457. DOI: 10.1103/PhysRev.80.440.

[7] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: http://link.aps.org/doi/10.1103/PhysRev.96.191.

[8] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[9] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13.9 (Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: http://prl.aps.org/export/PRL/v13/i9/p321_1?type=bibtex.

[10] Peter W. Higgs. "Broken Symmetries And The Masses Of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13.20 (Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: http://prl.aps.org/export/PRL/v13/i20/p585_1?type=bibtex.

[12] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[13] A. Salam. "Weak and electromagnetic interactions". In: (1969). Proc. of the 8th Nobel Symposium on 'Elementary particle theory, relativistic groups and analyticity', Stockholm, Sweden, 1968, edited by N. Svartholm, p.367-377.

[14] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev.* D2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[15] D. J. Gross and F. Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[16] D. J. Gross and F. Wilczek. "Asymptotically Free Gauge Theories. I". In: *Physical Review D* 8 (Nov. 1973), pp. 3633–3652. DOI: 10.1103/PhysRevD.8.3633.

[17] H. David Politzer. "Asymptotic freedom: An approach to strong interactions". In: *Physics Reports* 14 (Nov. 1974), pp. 129–180. DOI: 10.1016/0370-1573(74)90014-3.

[18] *The economist, Daily chart: Worth the wait*. http://www.economist.com/blogs/graphicdetail/2012/07/daily-chart-1. Accessed: 2016-07-08.

[19] M.E. Peskin and D.V. Schroeder. *An Introduction to Quantum Field Theory*. Advanced book classics. Addison-Wesley Publishing Company, 1995. ISBN: 9780201503975. URL: https://books.google.be/books?id=i35LALN0GosC.

[20] Gerard 't Hooft. "Gauge theory and renormalization". In: *Proceedings, International Conference on History of Original Ideas and Basic Discoveries in Particle Physics: Erice, Italy, July 29-August 4, 1994*. 1994. arXiv: hep-th/9410038 [hep-th].

[21] James M. Cline et al. "Update on scalar singlet dark matter". In: *Phys. Rev.* D88 (2013). [Erratum: Phys. Rev.D92,no.3,039906(2015)], p. 055025. DOI: 10.1103/PhysRevD.92.039906,10.1103/PhysRevD.88.055025. arXiv: 1306.4710 [hep-ph].

[22] Giuseppe Degrassi et al. "Higgs mass and vacuum stability in the Standard Model at NNLO". In: *JHEP* 08 (2012), p. 098. DOI: 10.1007/JHEP08(2012)098. arXiv: 1205.6497 [hep-ph].

[23] T. D. Lee. "A Theory of Spontaneous T Violation". In: *Phys. Rev.* D8 (1973). [,516(1973)], pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226.

[24] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].

[25] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[26] J. -M. Gerard and M. Herquet. "A Twisted custodial symmetry in the two-Higgs-doublet model". In: *Phys. Rev. Lett.* 98 (2007), p. 251802. DOI: 10.1103/PhysRevLett.98.251802. arXiv: hep-ph/0703051 [HEP-PH].

[27] Simon de Visscher et al. "Unconventional phenomenology of a minimal two-Higgs-doublet model". In: *JHEP* 08 (2009), p. 042. DOI: 10.1088/1126-6708/2009/08/042. arXiv: 0904.0705 [hep-ph].

[28] G. C. Dorsch et al. "Hierarchical vs Degenerate 2HDM: The LHC Run 1 Legacy at the Onset of Run 2". In: (2016). arXiv: 1601.04545 [hep-ph].

[29] G. C. Dorsch et al. "Echoes of the Electroweak Phase Transition: Discovering a second Higgs doublet through $A_0 \to ZH_0$". In: *Phys. Rev. Lett.* 113.21 (2014), p. 211802. DOI: 10.1103/PhysRevLett.113.211802. arXiv: 1405.5537 [hep-ph].

[30] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[31] Jun Gao et al. "CT10 next-to-next-to-leading order global analysis of QCD". In: *Phys. Rev.* D89.3 (2014), p. 033009. DOI: 10.1103/PhysRevD.89.033009. arXiv: 1302.6246 [hep-ph].

[32] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J.* C75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[33] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[34] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[35] S. Dittmaier et al. “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”. In: (2011). DOI: 10.5170/CERN-2011-002. arXiv: 1101.0593 [hep-ph].

[36] Charalampos Anastasiou et al. “Higgs Boson Gluon-Fusion Production in QCD at Three Loops”. In: *Phys. Rev. Lett.* 114 (2015), p. 212001. DOI: 10.1103/PhysRevLett.114.212001. arXiv: 1503.06056 [hep-ph].

[37] W.J. Stirling. *Standard Model cross sections as a function of collider energy, with 125 GeV Higgs and including 33 TeV HE LHC*. private communication. http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html.

[38] ATLAS and CMS Collaborations. “Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”. In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].

[39] CMS Collaboration. “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”. In: *Eur. Phys. J.* C75.5 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].

[40] CMS Collaboration. *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV*. Tech. rep. CMS-PAS-HIG-15-002. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2053103.

[41] ATLAS Collaboration. “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 detector”. In: *Phys. Rev.* D90.11 (2014), p. 112015. DOI: 10.1103/PhysRevD.90.112015. arXiv: 1408.7084 [hep-ex].

[42] CMS Collaboration. “Observation of the diphoton decay of the Higgs boson and measurement of its properties”. In: *Eur. Phys. J.* C74.10

(2014), p. 3076. DOI: 10.1140/epjc/s10052-014-3076-z. arXiv: 1407.0558 [hep-ex].

[43] ATLAS Collaboration. "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". In: *Phys. Rev.* D91.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.91.012006. arXiv: 1408.5191 [hep-ex].

[44] CMS Collaboration. "Measurement of the properties of a Higgs boson in the four-lepton final state". In: *Phys. Rev.* D89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[45] ATLAS Collaboration. "Observation and measurement of Higgs boson decays to WW* with the ATLAS detector". In: *Phys. Rev.* D92.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.92.012006. arXiv: 1412.2641 [hep-ex].

[46] ATLAS Collaboration. "Study of (W/Z)H production and Higgs boson couplings using $H \to WW^*$ decays with the ATLAS detector". In: *JHEP* 08 (2015), p. 137. DOI: 10.1007/JHEP08(2015)137. arXiv: 1506.06641 [hep-ex].

[47] CMS Collaboration. "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states". In: *JHEP* 01 (2014), p. 096. DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].

[48] ATLAS Collaboration. "Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector". In: *JHEP* 04 (2015), p. 117. DOI: 10.1007/JHEP04(2015)117. arXiv: 1501.04943 [hep-ex].

[49] CMS Collaboration. "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons". In: *JHEP* 05 (2014), p. 104. DOI: 10.1007/JHEP05(2014)104. arXiv: 1401.5041 [hep-ex].

[50] ATLAS Collaboration. "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector". In: *JHEP* 01 (2015), p. 069. DOI: 10.1007/JHEP01(2015)069. arXiv: 1409.6212 [hep-ex].

[51] CMS Collaboration. "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks". In: *Phys. Rev.* D89.1 (2014), p. 012003. DOI: 10.1103/PhysRevD.89.012003. arXiv: 1310.3687 [hep-ex].

[52] ATLAS Collaboration. "Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector". In: *Phys. Lett.* B738 (2014), pp. 68–86. DOI: 10.1016/j.physletb.2014.09.008. arXiv: 1406.7663 [hep-ex].

[53] CMS Collaboration. "Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC". In: *Phys. Lett.* B744 (2015), pp. 184–207. DOI: 10.1016/j.physletb.2015.03.048. arXiv: 1410.6679 [hep-ex].

[54] ATLAS Collaboration. "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector". In: *Eur. Phys. J.* C75.7 (2015), p. 349. DOI: 10.1140/epjc/s10052-015-3543-1. arXiv: 1503.05066 [hep-ex].

[55] ATLAS Collaboration. "Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector". In: *Phys. Lett.* B740 (2015), pp. 222–242. DOI: 10.1016/j.physletb.2014.11.049. arXiv: 1409.3122 [hep-ex].

[56] CMS Collaboration. "Search for the associated production of the Higgs boson with a top-quark pair". In: *JHEP* 09 (2014). [Erratum: JHEP10,106(2014)], p. 087. DOI: 10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106. arXiv: 1408.1682 [hep-ex].

[57] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". In: (2013). Ed. by S Heinemeyer et al. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].

[58] CMS Collaboration. *Summary results of high mass BSM Higgs searches using CMS run-I data*. Tech. rep. CMS-PAS-HIG-16-007. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2142432.

[59] ATLAS Collaboration. "Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector". In: *JHEP* 11 (2015), p. 206. DOI: 10.1007/JHEP11(2015)206. arXiv: 1509.00672 [hep-ex].

[60] 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". In: *Comput. Phys. Commun.* 184 (2013), pp. 1605–1617. DOI: 10.1016/j.cpc.2013.02.006. arXiv: 1212.3249 [hep-ph].

[61] CMS Collaboration. "Search for a Higgs Boson in the Mass Range from 145 to 1000 GeV Decaying to a Pair of W or Z Bosons". In: *JHEP* 10 (2015), p. 144. DOI: 10.1007/JHEP10(2015)144. arXiv: 1504.00936 [hep-ex].

[62] CMS Collaboration. *Search for scalar resonances in the 200-500 GeV mass range decaying into a Z and a photon in pp collisions at* $\sqrt{s} = 8$ TeV. Tech. rep. CMS-PAS-HIG-14-031. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/1997885.

[63] David Eriksson, Johan Rathsman, and Oscar Stal. "2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual". In: *Comput. Phys. Commun.* 181 (2010), pp. 189–205. DOI: 10.1016/j.cpc.2009.09.011. arXiv: 0902.0851 [hep-ph].

[64] CMS Collaboration. *Search for exotic decays of the Higgs boson to a pair of new light bosons with two muon and two b jets in final states*. Tech. rep. CMS-PAS-HIG-14-041. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2135985.

[65] CMS Collaboration. *Search for the exotic decay of the Higgs boson to two light pseudoscalar bosons with two taus and two muons in the final state at* $\sqrt{s} = 8$ *TeV*. Tech. rep. CMS-PAS-HIG-15-011. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2128149.

[66] CMS Collaboration. "Search for a very light NMSSM Higgs boson produced in decays of the 125 GeV scalar boson and decaying into τ leptons in pp collisions at $\sqrt{s} = 8$ TeV". In: *JHEP* 01 (2016),

p. 079. DOI: 10.1007/JHEP01(2016)079. arXiv: 1510.06534 [hep-ex].

[67] CMS Collaboration. *Search for Higgs Decays to New Light Bosons in Boosted Tau Final States*. Tech. rep. CMS-PAS-HIG-14-022. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2058768.

[68] CMS Collaboration. "A search for pair production of new light bosons decaying into muons". In: *Phys. Lett.* B752 (2016), pp. 146–168. DOI: 10.1016/j.physletb.2015.10.067. arXiv: 1506.00424 [hep-ex].

[69] CMS Collaboration. "Search for a pseudoscalar boson decaying into a Z boson and the 125 GeV Higgs boson in $l^+l^-b\bar{b}$ final states". In: *Phys. Lett.* B748 (2015), pp. 221–243. DOI: 10.1016/j.physletb.2015.07.010. arXiv: 1504.04710 [hep-ex].

[70] CMS Collaboration. "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$". In: *Phys. Lett.* B755 (2016), pp. 217–244. DOI: 10.1016/j.physletb.2016.01.056. arXiv: 1510.01181 [hep-ex].

[71] Christina Gao et al. "Searching for Additional Higgs Bosons via Higgs Cascades". In: (2016). arXiv: 1604.03108 [hep-ph].

[72] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

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

[74] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[75] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

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

[77] Cinzia De Melis. "The CERN accelerator complex. Complexe des accélérateurs du CERN". In: (July 2016). General Photo. URL: https://cds.cern.ch/record/2197559.

[78] A Barachetti, L Rossi, and A Szeberenyi. "Final Project Report: Deliverable D1.14". In: (Jan. 2016). URL: https://cds.cern.ch/record/2120851.

[79] Tai Sakuma and Thomas McCauley. "Detector and Event Visualization with SketchUp at the CMS Experiment". In: *J. Phys. Conf. Ser.* 513 (2014), p. 022032. DOI: 10.1088/1742-6596/513/2/022032. arXiv: 1311.4942 [physics.ins-det].

[80] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/349375.

[81] John Allison et al. "Geant4 developments and applications". In: *IEEE Trans. Nucl. Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.

[82] E. Barberio et al. "Fast simulation of electromagnetic showers in the ATLAS calorimeter: Frozen showers". In: *J. Phys. Conf. Ser.* 160 (2009), p. 012082. DOI: 10.1088/1742-6596/160/1/012082.

[83] Rahmat Rahmat and Rob Kroeger. "HF GFlash". In: *Phys. Procedia* 37 (2012), pp. 340–346. DOI: 10.1016/j.phpro.2012.02.385.

[84] ATLAS Collaboration et al. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*. Tech. rep. ATL-PHYS-PUB-2010-013. Geneva: CERN, Oct. 2010. URL: https://cds.cern.ch/record/1300517.

[85] Andrea Giammanco. "The Fast Simulation of the CMS Experiment". In: *Journal of Physics: Conference Series* 513.2 (2014), p. 022012. URL: http://stacks.iop.org/1742-6596/513/i=2/a=022012.

[86] Bruce Knuteson. "Systematic analysis of high-energy collider data". In: *Nucl. Instrum. Meth.* A534 (2004), pp. 7–14. DOI: 10.1016/j.nima.2004.07.050. arXiv: hep-ex/0402029 [hep-ex].

[87] Sascha Caron and Bruce Knuteson. "QUAERO and H1: An Interface to high-p(T) HERA event data". In: *Eur. Phys. J.* C53 (2008), pp. 167–175. DOI: 10.1140/epjc/s10052-007-0468-3. arXiv: hep-ph/0612201 [hep-ph].

[88] Manuel Drees et al. "CheckMATE: Confronting your Favourite New Physics Model with LHC Data". In: *Comput. Phys. Commun.* 187 (2015), pp. 227–265. DOI: 10.1016/j.cpc.2014.10.018. arXiv: 1312.2591 [hep-ph].

[89] B. Dumont et al. "Toward a public analysis database for LHC new physics searches using MADANALYSIS 5". In: *Eur. Phys. J.* C75.2 (2015), p. 56. DOI: 10.1140/epjc/s10052-014-3242-3. arXiv: 1407.3278 [hep-ph].

[90] CMS Collaboration. *CMSSW Application Framework*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework.

[91] Sunanda Banerjee. *CMS Simulation Software*. Tech. rep. CMS-CR-2012-146. Geneva: CERN, June 2012. URL: http://cds.cern.ch/record/1457820.

[92] S. Ovyn, X. Rouby, and V. Lemaitre. "DELPHES, a framework for fast simulation of a generic collider experiment". In: (2009). arXiv: 0903.2225 [hep-ph].

[93] J. de Favereau et al. "DELPHES 3, A modular framework for fast simulation of a generic collider experiment". In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex].

[94] D. Buskulic et al. "Performance of the ALEPH detector at LEP". In: *Nucl. Instrum. Meth.* A360 (1995), pp. 481–506. DOI: 10.1016/0168-9002(95)00138-7.

[95] CMS Collaboration. "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". In: (2009).

[96] CMS Collaboration. "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JINST* 8 (2013). [JINST8,9009(2013)], P09009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[97] CMS Collaboration. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[98] G Bianchi. "tkLayout: a design tool for innovative silicon tracking detectors". In: *Journal of Instrumentation* 9.03 (2014), p. C03054. URL: http://stacks.iop.org/1748-0221/9/i=03/a=C03054.

[99] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. CMS-PAS-PFT-09-001. 2009. Geneva: CERN, Apr. 2009. URL: https://cds.cern.ch/record/1194487.

[100] Florian Beaudette. "The CMS Particle Flow Algorithm". In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, pp. 295–304. arXiv: 1401.8155 [hep-ex]. URL: https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf.

[101] David Barney. "CMS Detector Slice". CMS Collection. Jan. 2016. URL: https://cds.cern.ch/record/2120661.

[102] CMS Collaboration. "CMS Physics: Technical Design Report". In: Technical Design Report CMS (2006). URL: http://cds.cern.ch/record/922757.

[103] CMS Collaboration. "2015 ECAL detector performance plots". In: (Dec. 2015). https://cds.cern.ch/record/2114735.

[104] CMS Collaboration. "Electron and Photon performance using data collected by CMS at $\sqrt{s}$ = 13 TeV and 25ns". In: (Dec. 2015). URL: https://cds.cern.ch/record/2118397.

[105] 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. arXiv: 1206.4071 [physics.ins-det].

[106] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Catchment Area of Jets". In: *JHEP* 04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].

[107] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[108] The CMS collaboration. "Determination of jet energy calibration and transverse momentum resolution in CMS". In: *Journal of Instrumentation* 6.11 (2011), P11002. URL: http://stacks.iop.org/1748-0221/6/i=11/a=P11002.

[109] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. "Hadronic top-quark pair-production with one jet and parton showering". In: *JHEP* 01 (2012), p. 137. DOI: 10.1007/JHEP01(2012)137. arXiv: 1110.5251 [hep-ph].

[110] CMS Collaboration. "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 01 (2011), p. 080. DOI: 10.1007/JHEP01(2011)080. arXiv: 1012.2466 [hep-ex].

[111] CMS Collaboration. "Jet energy scale and resolution performances with 13TeV data". In: (June 2016). URL: http://cds.cern.ch/record/2160347.

[112] CMS Collaboration. *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Tech. rep. CMS-PAS-BTV-15-001. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2138504.

[113] Kyle Cranmer. "Practical Statistics for the LHC". In: *Proceedings, 2011 European School of High-Energy Physics (ESHEP 2011): Cheile Gradistei, Romania, September 7-20, 2011*. [,247(2015)]. 2015, pp. 267–308. DOI: 10.5170/CERN-2015-001.247, 10.5170/CERN-2014-003.267. arXiv: 1503.07622 [physics.data-an]. URL: https://inspirehep.net/record/1356277/files/arXiv:1503.07622.pdf.

[114] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[115] A L Read. "Modified frequentist analysis of search results (the CL_s method)". In: CERN-OPEN-2000-205 (2000). URL: https://cds.cern.ch/record/451614.

[116] CMS Collaboration. "Missing transverse energy performance of the CMS detector". In: *Journal of Instrumentation* 6 (Sept. 2011), p. 9001. DOI: 10.1088/1748-0221/6/09/P09001. arXiv: 1106.5048 [physics.ins-det].

[117] Adrien Caudron. "The final state with two b-jets and two leptons at the LHC as a probe of the scalar sector". cp3.irmp.ucl.ac.be/upload/theses/phd/acaudron.pdf. Université Catholique de Louvain, 2016.

[118] CMS Collaboration. "Search for neutral resonances decaying into a Z boson and a pair of b jets or tau leptons". In: *Phys. Lett.* B759 (2016), pp. 369–394. DOI: 10.1016/j.physletb.2016.05.087. arXiv: 1603.02991 [hep-ex].

[119] Howard E. Haber and Oscar Stal. "New LHC benchmarks for the $\mathcal{CP}$-conserving two-Higgs-doublet model". In: *Eur. Phys. J.* C75.10 (2015). [Erratum: Eur. Phys. J.C76,no.6,312(2016)], p. 491. DOI: 10.1140/epjc/s10052-015-3697-x, 10.1140/epjc/s10052-016-4151-4. arXiv: 1507.04281 [hep-ph].

[120] CMS Collaboration. *Search for additional neutral Higgs bosons decaying to a pair of tau leptons in pp collisions at $\sqrt{s} = 7$ and 8 TeV*. Tech. rep. CMS-PAS-HIG-14-029. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2041463.

[121] Liliya Milenska. "Study of the associated production of a Z boson and a pseudoscalar A in a boosted topology in the context of a two-Higgs-doublet model". MA thesis. Université Catholique de Louvain, Juin 2016.