

Università degli Studi di Siena

DIPARTIMENTO DI SCIENZE FISICHE, DELLA TERRA E DELL'AMBIENTE

SEZIONE DI FISICA



UNIVERSITÀ
DI SIENA
1240

Search for a heavy scalar boson decaying into a pair of Standard-Model-like Higgs bosons, in the final state $b\bar{b}\tau^+\tau^-$ with both τ decaying hadronically, at CMS experiment

Tesi di Dottorato di Ricerca in Fisica Sperimentale
In partial fulfillment of the requirements for
the Ph.D. Thesis in Experimental Physics

Candidato:
Maria Teresa Grippo

Supervisor:
Prof.ssa Maria Agnese Ciocci

Ciclo di dottorato XXVIII

Contents

Contents	3
Abstract	7
Introduction	9
1 Theory and motivation	11
1.1 The Standard Model	11
1.1.1 Fundamental particles of the Standard Model	11
1.1.2 Standard Model as gauge theory	13
1.1.3 The Brout–Englert–Higgs mechanism	14
1.1.4 Higgs physics at LHC	18
1.2 Beyond the Standard Model Higgs	23
1.2.1 The Supersymmetry and the Minimal Supersymmetry Model	26
1.2.2 The Two–Higgs–Doublet Model	29
1.3 The $H \rightarrow hh$ process at LHC	32
1.4 Physics motivations and challenges for the search of $H \rightarrow hh \rightarrow$ $bb\tau_h\tau_h$ channel	37
2 The CMS experiment at LHC	39
2.1 The LHC machine	39
2.1.1 LHC complex and acceleration scheme	39
2.1.2 LHC milestones	41
2.2 The CMS detector	42
2.2.1 Coordinate System and Useful Variables	45
2.2.2 Tracking System	46
2.2.3 Electromagnetic calorimeter	48
2.2.4 Hadron calorimeter	51
2.2.5 The magnet	53
2.2.6 Muon System	53
2.2.7 Trigger system	56
2.2.8 Computing system	58
3 Event reconstruction	61
3.1 Reconstruction at detector level	61

3.1.1	Track reconstruction	61
3.1.2	Primary vertex	64
3.1.3	Energy Reconstruction in the Calorimeter	67
3.2	Physical Objects	68
3.2.1	Particle-Flow Algorithm	68
3.2.2	Electrons	71
3.2.3	Muons	73
3.2.4	Electron and Muon Isolation	75
3.2.5	Jets	77
3.2.6	b jets	80
3.2.7	Missing Transverse Energy	84
3.2.8	τ leptons	86
4	Event Selection and modelling	91
4.1	Signal and Background Definition	91
4.2	Data samples	93
4.3	Trigger selection	95
4.4	Sample composition modelling	96
4.4.1	Monte Carlo samples	97
4.5	Event Selection	99
4.5.1	$h \rightarrow \tau_h \tau_h$ selection	100
4.5.2	$h \rightarrow bb$ selection	101
4.5.3	$H \rightarrow hh \rightarrow bb\tau\tau$ selection	105
4.5.4	$H \rightarrow hh$ reconstruction	108
4.6	Data driven Background modelling	112
4.6.1	$Z \rightarrow \tau\tau$	113
4.6.2	QCD	115
5	Results	119
5.1	Systematic uncertainties	119
5.1.1	Normalization uncertainties	121
5.1.2	Shape uncertainties	125
5.2	Signal extraction procedure	125
5.3	Results	129
5.4	SM Higgs background contribution	130
5.5	Model independent limits	133
5.6	Model dependent limits	138
	Conclusion	141
	A Luminosity Measurement	143
	B Monte Carlo simulation	145

C Data and MC corrections	147
C.1 Hadronic τ efficiency for $\tau_h\tau_h$ trigger	147
C.2 τ_h energy scale corrections	148
C.3 \cancel{E}_T resolution and response	148
C.4 b tagging efficiency and mistag rate	149
D Di-tau mass reconstruction – SVFit algorithm	151
Bibliography	155

Abstract

This thesis presents the search for a heavy scalar boson H , decaying in two Standard Model like Higgs bosons h with a mass around 125 GeV. The final state consists of a pair of b jets and a pair of τ , where both τ leptons decay into hadrons plus a tau neutrino.

The search uses proton-proton collisions data collected by the Compact Muon Solenoid (CMS) experiment during 2012, at a centre of mass energy of 8 TeV corresponding to an integrated luminosity of 19.7 fb^{-1} .

The results have shown no evidence for a heavy boson H and are compatible with the Standard Model expectations. Upper limits are set at 95% confidence level on the production cross-section times the branching fraction of the heavy boson H , in the final state $hh \rightarrow bb\tau\tau$, for a H mass in the range $260 \leq M_H \leq 350 \text{ GeV}$.

The results are also interpreted in the context of benchmark scenarios of the Minimal Supersymmetric Extension of the Standard Model (MSSM) and of the Two-Higgs-Doublet Model (2HDM). Exclusion limits are set in the two dimensional parameter space of these two models.

Introduction

The discovery of a scalar boson with a mass around 125 GeV¹ [1, 2] at the Large Hadron Collider (LHC) [3] by the ATLAS and CMS collaborations [4, 5], was the starting point for a new era of measurements in particle physics. Is the discovered boson the Standard Model (SM) Higgs boson [6, 7]? Latest results [8, 9] confirm that the discovered particle is compatible, within the achieved precision, with the SM Higgs boson.

On the other hand the SM theory has some critical points: it does not include the gravitational force, it does not account for neutrinos masses, in contrast with the experimental evidence [10, 11], it does not also explain the baryogenesis [12], the dark matter and the dark energy [13, 14]. In addition, in the SM the *naturalness* and the *hierarchy problem* [15] are not solved. These issues lead to introduce new theories, generally called *Beyond Standard Model* (BSM) theories. The two models on which this thesis will be focused are the Minimal Supersymmetry Model (MSSM) [16–18] and the Two–Higgs–Doublet Models (2HDM) [19], two possible extensions of the SM. They introduce two Higgs doublets and therefore the symmetry breaking mechanism produces five physical particles for the Higgs sector: a light and a heavy CP–even neutral Higgs bosons, h and H , a CP–odd neutral Higgs boson A , and two charged Higgs bosons H^\pm .

In the context of these two models and in specific scenarios, the SM–like Higgs boson discovered can be the light Higgs boson h , which is produced through the decay of the heavy Higgs boson H , $H \rightarrow hh$. For a H mass in the range $260 \leq M_H \leq 350$ GeV, this decay is the most sensitive for the search of an additional Higgs doublet [20–23]. The data collected by CMS in 2012 open the possibility to investigate this decay mode.

The final state investigated in this thesis, in which one h decays into a pair of τ leptons and the other h decays into b quarks, is very appealing, thanks to the high branching fraction of the $h \rightarrow bb^2$ ($B = 57.5 \pm 1.9\%$ [24]) and the quite clean and efficient signature of the h decaying into a τ pair ($B = 6.3 \pm 0.36\%$ [24]). The choice of this final state is a compromise between the decay modes, $H \rightarrow hh \rightarrow bbbb$, which has the highest branching fraction but an overwhelming

¹In this thesis the system of natural units is used with $\hbar = c = 1$, where $\hbar = h/2\pi = 6.58211889(26) \cdot 10^{-22} \text{MeVs}$ e $c = 299792458 \text{ms}^{-1}$.

²For simplicity throughout the paper it is not specified the charge of the leptons and the particle-antiparticle nature of the quarks: $\tau\tau$ corresponds to $\tau^+\tau^-$, for leptons $\ell\ell$ is equal to $\ell^+\ell^-$, bb stays for $b\bar{b}$ and in general qq stays for $q\bar{q}$.

QCD background, and $H \rightarrow hh \rightarrow bb\gamma\gamma$, which has a very clean signature but a very low branching fraction ($B(h \rightarrow \gamma\gamma) = 0.228 \pm 0.011\%$). Given the potential of discovery or exclusion of new BSM physics, searches for the decays $H \rightarrow hh$ have been already performed by the ATLAS [25–27] and CMS Collaborations [28–33] in several final states. In this thesis, for the first time at CMS, the $bb\tau_h\tau_h$ final state is investigated, where τ_h indicates a τ which decays into hadrons plus a tau neutrino. The decay $h \rightarrow \tau_h\tau_h$, the *fully-hadronic* decay mode, was chosen because it constitutes the 42% of the di-tau decay modes and because of the very good performance of the τ_h reconstruction and identification provided by CMS experiment. In the following, the different $hh \rightarrow bb\tau\tau$ final states will be named according to the tau pair decay modes: respectively, the *fully hadronic channel*, $\tau_h\tau_h$, the *semileptonic channel*, $\tau_l\tau_h$, and the *fully leptonic channel*, $\tau_l\tau_l$.

The SM theory, the Higgs boson’s discovery and properties, and these two BSM theories will be described in the first chapter. The physics motivations for the $H \rightarrow hh \rightarrow bb\tau\tau$ process will be highlighted at the end of the same chapter.

In the second chapter the LHC machine will be introduced and the Compact Muon Solenoid (CMS) experiment will be described, focusing the attention on its components, useful to reconstruct the hunted final objects (τ_h , muons, electrons). The reconstruction and identification algorithms used to obtain the high level objects will be described in the third chapter.

In the fourth chapter, starting from the reconstructed objects, the analysis strategy will be described, focusing the attention on the signal identification and on the modelling of the background processes.

In the fifth chapter the signal extraction procedure will be described and the final results as well as the exclusion limits will be presented.

The results of this analysis are combined with the CMS results for the process $H \rightarrow hh \rightarrow bb\tau_l\tau_h$, the *semileptonic* channel, and are also combined with the results coming from the process $A \rightarrow Zh \rightarrow \ell\ell\tau\tau$, for both theoretical models in specific benchmark scenarios.

Finally, in the conclusion paragraph the comparison of our exclusion limits for $H \rightarrow hh$ channel with the results obtained by other analysis in CMS and in ATLAS experiments will be reported, showing the sensitivity of the obtained results and the future perspectives of this search.

The results of the analysis described in this thesis are published in Physics Letters B journal [34] by the CMS collaboration. I presented the results of the $H \rightarrow hh \rightarrow bb\tau\tau$ process at “HH2015: Higgs Hunting 2015”, at Orsay (France), 30 Jul-1 Aug 2015, on behalf of the CMS collaboration. I also presented both the analyses, $H \rightarrow hh \rightarrow bb\tau\tau$ and $A \rightarrow Zh \rightarrow \ell\ell\tau\tau$, at the “LP2015: XXVII International Symposium on Lepton Photon Interactions at High Energies”, 17-22 Aug 2015, hosted in Ljubljana (Slovenia). I am also the first author of the proceeding, that will be published on the Proceedings of Science. This work is partially funded by MIUR Italy, under contract 2012Z23ERZ of PRIN 2012, “H-TEAM: Trigger, Elettronica Avanzata e Metodi innovativi per misure di precisione nel settore dell’ Higgs ad LHC”.

Chapter 1

Theory and motivation

This chapter describes the theoretical framework for the search performed in this thesis. It begins with a brief overview of the Standard Model (SM), our current best understanding of elementary particles physics, followed by a brief description of the reasons why extending the theory Beyond the SM (BSM) is necessary. Two popular BSM models are considered in this thesis that can resolve the pending issues left by the SM: the Minimal Supersymmetry Model (MSSM) and the Two-Higgs-Doublet-Model (2HDM). The discovered scalar neutral boson h can be interpreted by these models, in some benchmark scenarios with specific parameters, as a particle with the same properties of the SM Higgs boson, which can be produced in pair by the decay of the heavy neutral scalar Higgs boson H , also predicted by these models and observable with the current available data at LHC. This chapter provides an outline of these models, focusing the attention on the electroweak sector and on the theoretical and experimental motivations for the search of $H \rightarrow hh \rightarrow bb\tau\tau$, the main argument of this thesis.

1.1 The Standard Model

1.1.1 Fundamental particles of the Standard Model

The Standard Model (SM) theory [6, 7] of the particle physics describes the microscopic world in term of its constituents and its interactions. It is an elegant theoretical formulation of three of the fundamental interactions¹, the electromagnetic, weak nuclear and strong nuclear interactions, and of the particles that compose the matter.

¹The description of the gravitational force cannot be included because it is 40 orders weaker than the other three forces. It is expected to become relevant at the so called *Planck scale*, of the order of 10^{19} GeV and in the context of the *Grand Unification Theory* [6].

The matter is made up of elementary particles with spin 1/2 (*fermions*) and divided in *leptons* and *quarks* families. The fermions interact among themselves through the *bosons*, particles of spin 1 that are the mediators of the fundamental forces described in the SM.

The elementary building blocks of matter, the fermions, are composed of six leptons and six quarks, organized in *generations* according to the mass, starting from the lightest and most stable particles. For both families there are so doublets of leptons and quarks in each generation, and the same goes for antiparticles². The doublet of leptons is composed of a charged massive lepton and a neutral particle, called *neutrino*, assumed massless in the SM theory³; the leptons interact through electroweak forces and not through the strong interaction. The same organization characterizes the quarks doublets: they can be of six *flavours*, are massive, have a fractional electric charge and interact through strong and electroweak interactions. Quarks have an additional charge, the color charge, which exists in three states “red” (R), “green” (G) and “blue” (B). Quarks don’t exist in nature as free particles but they combine in “white” color charge states, called *hadrons*, divided in *mesons*, a quark–antiquark couple, and in *barions*, made of three quarks. Properties of leptons and quarks are summarized in Table 1.1.

	Generation I		Generation II		Generation III		
Family	Particle	Mass	Particle	Mass	Particle	Mass	Charge
Leptons	ν_e	< 2 eV	ν_μ	< 0.19 MeV	ν_τ	< 18.2 MeV	0
	e	0.511 MeV	μ	105.7 MeV	τ	1.777 GeV	-1
Quarks	u	2.3 MeV	c	1.27 GeV	t	173.2 GeV	+2/3
	d	4.8 MeV	s	95 MeV	b	4.18 GeV	-1/3

Table 1.1: Leptons and quarks: some properties [35].

The interactions among elementary fermions is mediated, as already stated, through *bosons*. The photon, γ , is the mediator of the electromagnetic interactions for all fermions except the neutrinos, is massless and electrically neutral. Eight *gluons* are the mediators of the strong nuclear interaction among quarks, are massless, electrically neutral and have a color charge; the W^\pm and Z are the carriers of the weak nuclear interactions for all the fermions. Properties of bosons are summarized in Table 1.2.

²Antiparticles are particles with the same mass, spin and lifetime of its partner but opposite charge.

³By the observation of the neutrino flavour oscillation the different neutrino flavors have different masses, although these masses have been shown to be very tiny by the experiments [10, 11].

Force	Particle	Charge	Mass
Electromagnetic	γ	0	$< 1 \times 10^{-18}$ eV
	W^\pm	± 1	80.39 GeV
Weak	Z	0	91.19 GeV
	g	0	0 (theoretical)

Table 1.2: The SM bosons mediators: some properties [35].

1.1.2 Standard Model as gauge theory

The SM is a quantum field theory and its mathematical formulation is based on the construction of a local gauge invariant Lagrangian [6]⁴ under the *non-abelian* group $SU(3)_C \times SU(2)_L \times U(1)_Y$.

The $SU(3)_C$ is the symmetry group which describes the strong interactions between quarks and gluons. The theory was introduced by Yang and Mills [36] and included in the so called Quantum Chromodynamics (QCD). It contains 8 generators, the Gell–mann matrices, which correspond to eight physical gauge fields, the gluons. The non–abelian property of the group allows to have interactions among gluons, which are the color carriers and, as a consequence of the gauge invariance, they are massless. Further details about QCD can be found in [6, 7].

The $SU(2)_L \times U(1)_Y$ group describes the unified electromagnetic and weak interactions (EWK), as theorized by Glashow, Weinberg and Salam [37–39], where the Quantum Electrodynamics (QED), represented by the $U(1)_{em}$ group, is included. The generators of the groups are respectively the three isospin weak operators, $\vec{T} \equiv (T_1, T_2, T_3)$, and the weak hypercharge Y . The gauge invariance allows to associate the $\vec{W}_\mu \equiv (W_\mu^1, W_\mu^2, W_\mu^3)$ gauge bosons to the isospin operators and B_μ gauge boson to the hypercharge operator. The fermions interact with gauge bosons, which, as a consequence of the non–abelian property, interact also with each other. Moreover the third component of the isospin (T_3) and hypercharge are related to the electric charge Q ($Q = T_3 + Y/2$). The electric charge is the generator of the abelian $U(1)_{em}$ group, the gauge boson A_μ , representing the photon, is associated to this. The A_μ boson is massless as a consequence of the gauge invariance, and it is not self–interacting because of the abelian property of the group.

The electroweak theory is *chiral* [7], which means that the *left-handed* and *right-handed* fermion components transform in a different way under the local gauge transformations. The *left-handed* fermions form an isospin doublet with which the weak forces interact, instead the *right-handed* form a singlet which is invariant under $SU(2)_L$ transformations. They can be written as

⁴A system is considered *symmetric* if it remains unchanged, despite of the changes of its properties, under a *global* transformation, that does not depend on the space–time coordinates, or under a *local* transformation, which depends on space–time coordinates.

$$\begin{aligned}\psi_L &= \begin{pmatrix} \nu_i \\ l_i \end{pmatrix}_L, \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L \\ \psi_R &= l_R, u_R, d_R,\end{aligned}\tag{1.1}$$

ψ usually represents the fermion, the index i refers to the generation: ν are the neutrinos, which have only the left-handed component because they are massless in the SM and the right component has not been yet observed; l are the charged leptons; u the up-type quarks (u, c, t), and d the down-type quarks (d, s, b), that have both also the right-handed components. Moreover the gauge fields \vec{W}_μ and B^μ do not represent the physical mediators of the electroweak interactions, which are indeed a combination of the gauge fields:

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}, Z_\mu = \frac{g_1 W_\mu^3 - g_2 B_\mu}{\sqrt{g_2^2 + g_1^2}}, A_\mu = \frac{g_1 W_\mu^3 + g_2 B_\mu}{\sqrt{g_2^2 + g_1^2}}\tag{1.2}$$

g_1 and g_2 correspond to the electroweak couplings, related to the electromagnetic coupling e through the Weinberg angle θ_W (experimentally measured [40]), as reported in the following equation:

$$e = g_1 \sin \theta_W = g_2 \cos \theta_W\tag{1.3}$$

The resulting EWK Lagrangian describes all the interactions between the gauge fields and the fermions, but crucially does not contain mass terms for the weak force bosons or any of the fermions. Attempts to add gauge boson mass terms of the form $-m_W^2 W_\mu W^\mu$ or $-m_Z^2 Z_\mu Z^\mu$ or fermion mass terms break the gauge invariance of the Lagrangian. On the other side this contradicts the experimental results: the Z and W^\pm bosons were discovered by UA1 and UA2 experiments in 1983 at CERN [41–44] with a mass of around 91.2 GeV and 80.4 GeV respectively, also the fermions have a definite mass, as reported in [35]. The only particle massless is the photon and this property has been confirmed experimentally so far [35]. In order to confirm the experimental results the spontaneous breaking of the electroweak symmetry was theorized and formulated in the Brout–Englert–Higgs mechanism in the SM.

1.1.3 The Brout–Englert–Higgs mechanism

In order to give a mass to the fermions and to the weak bosons, the existence of a scalar boson, the **Higgs boson**, was predicted by Brout, Englert and Higgs in the 60's and theorized in the Brout–Englert–Higgs mechanism [45–47]. The mechanism is based on the spontaneous breaking of the electroweak symmetry. The spontaneous symmetry breaking has the important feature that the Lagrangian is

invariant under the symmetry but the ground state of the theory not [7], which results to be degenerate and therefore it is necessary to break the symmetry choosing one eigenstate. In the electroweak theory the group $SU(2)_L \times U(1)_Y$ is spontaneously broken in the $U(1)_{em}$, related to the QED that must stay an exact symmetry. For the Goldstone theorem [7], three massless Goldstone bosons appear, which are absorbed by three out of four gauge bosons, giving mass to the vector bosons and keeping massless the photon.

The spontaneous symmetry breaking [48] can be obtained introducing a complex scalar field ϕ which attains a non-zero vacuum expectation value (vev) and that, once added to the electroweak Lagrangian, preserves the gauge invariance. It is an isospin doublet with hypercharge equal one and a linear combination of four real scalar fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.4)$$

The additional invariant term to the EWK Lagrangian due to the scalar field part can be written:

$$\mathcal{L}_S = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (1.5)$$

$$\text{with } D_\mu = \partial_\mu - ig_2 T_i W_\mu^i - i\frac{g_1}{2} B_\mu, \quad (1.6)$$

where $V(\phi)$ is the potential which describes the evolution of the ϕ field and depends on λ and μ^2 . The parameter λ is needed to be > 0 to make the potential bounded from below, instead the parameter μ^2 is considered < 0 . In this way the ground state is not uniquely identified but the minima lie on a hypersphere equal to:

$$|\phi|^2 = -\frac{\mu^2}{2\lambda}, \quad (1.7)$$

whose trend can be observed in Fig. 1.1.

The vev is defined as $v = \sqrt{-\mu^2/\lambda}$ and a particular minimum is chosen:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.8)$$

This choice ensures both the electroweak symmetry breaking with the consequent generation of massive vector bosons and the invariance under the $U(1)_{em}$ symmetry group, with the consequence that the photon still remains massless. The expansion of the ϕ field around the vacuum state $\langle \phi \rangle$ introduces four scalar fields: $\theta_{1,2,3}$ and

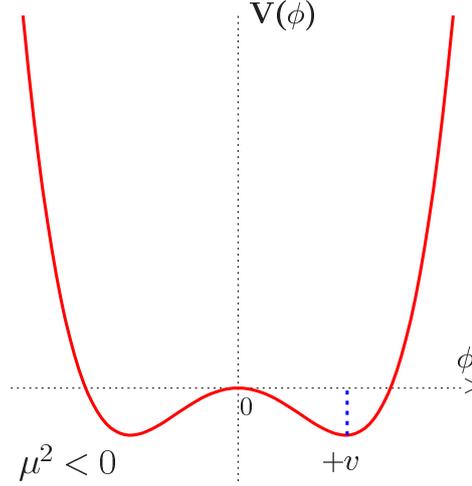


Figure 1.1: The potential $V(\phi)$ in the case $\mu^2 < 0$. [48].

$h(x)$ and the field is written as:

$$\phi(x) = e^{i\vec{\tau} \cdot \vec{\theta}(x)/v} \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix}. \quad (1.9)$$

The fields $\theta_{1,2,3}$ represent the massless Goldstone bosons, obtained from the electroweak symmetry breaking, that can be re-absorbed thanks to a gauge boson transformation into the physical gauge bosons as their longitudinal component and to give them mass. Therefore the ϕ field can be rewritten as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (1.10)$$

where the scalar and neutral Higgs field $h(x)$ can be recognized. The massive gauge vector bosons can be obtained replacing the vacuum expectation value of the ϕ field in the Lagrangian 1.5 and using the relations of the vector bosons 1.2. The resulting massive terms are:

$$M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu + \frac{1}{2} M_A^2 A_\mu A^\mu; \quad (1.11)$$

the W and Z bosons have acquired masses, while the photon, the generator of the $U(1)_{em}$ symmetry, to grant the $U(1)_{em}$ symmetry, remains massless as it should be.

$$M_W = \frac{1}{2} v g_2, \quad M_Z = \frac{1}{2} v \sqrt{g_1^2 + g_2^2} \text{ and } M_A = 0. \quad (1.12)$$

From the masses and couplings 1.22 equations, the following relations are obtained:

$$\frac{M_W}{M_Z} = \cos \theta_W \quad (1.13)$$

$$\rho \equiv \frac{M_W^2}{M_Z^2 \cos^2 \theta_W} = 1 \quad (1.14)$$

where the first allows to determine the Weinberg angle, the second represents the relative strength of nuclear weak interactions with neutral and charged current and has been verified experimentally [49].

The same doublet ϕ can be introduced in the electroweak Lagrangian to generate the fermion masses [48] and using the same procedure described for the gauge bosons masses, the fermion masses emerge as constant terms in front of the $\bar{\psi}_L \psi_R$ component and are identified with the fermion masses

$$m_e = \frac{\lambda_e v}{\sqrt{2}}, \quad m_u = \frac{\lambda_u v}{\sqrt{2}} \quad \text{and} \quad m_d = \frac{\lambda_d v}{\sqrt{2}}. \quad (1.15)$$

Also for these terms the gauge invariance is preserved and the $U(1)_{em}$ and $SU(3)_C$ symmetries are unbroken. The masses of the fermions are parameters in the theory and are not predicted by the SM, the input is given from experimental values.

Moreover it is possible to determine the value of v , using the experimental value of the Fermi constant G_F , evaluated from the muon decay and related to the W boson mass in the following way:

$$\frac{G_F}{\sqrt{2}} = \frac{g_2^2}{8M_W^2} \quad (1.16)$$

Joining the 1.12 and 1.16 equations, the resulting value of v is

$$v^2 = \frac{1}{\sqrt{2}G_F} \approx (246 \text{ GeV})^2. \quad (1.17)$$

Finally the potential $V(\phi)$ can be rewritten using the 1.10 equation:

$$V(\phi) = -\lambda v^2 h^2(x) - \lambda v h^3(x) - \frac{1}{4} \lambda h^4(x), \quad (1.18)$$

where the second and the third terms represent respectively the triple and quartic self-couplings of the Higgs boson; the first term instead represents the Higgs mass, equal to

$$M_h^2 = 2\lambda v^2 = -2\mu^2, \quad (1.19)$$

which is not predicted by the theory despite the fact that v is fixed.

Using equations 1.18 and 1.19, the SM Higgs self-couplings are equal to⁵:

$$g_{h^3} = 3\frac{M_{h^2}}{v} \text{ and } g_{h^4} = 3\frac{M_{h^2}}{v^2}. \quad (1.20)$$

The couplings to gauge bosons and fermions arise when calculating their masses. Indeed from the Lagrangian terms that represent the mass terms

$$\mathcal{L}_{M_V} \sim M_V^2 \left(1 + \frac{h}{v}\right)^2, \quad \mathcal{L}_{m_f} \sim m_f \left(1 + \frac{h}{v}\right) \quad (1.21)$$

one obtains also the Higgs boson couplings to gauge bosons and the Yukawa couplings to fermions⁶:

$$g_{hff} = \frac{m_f}{v}, \quad g_{hVV} = -2\frac{M_V^2}{v}, \quad g_{hhVV} = -2\frac{M_V^2}{v^2}. \quad (1.22)$$

In Fig. 1.2 the Feynman diagrams of the SM Higgs interactions are reported. The direct couplings to gauge bosons and fermions are proportional to their mass and the discovery of the Higgs boson (that will be briefly described in section 1.1.4) provided a confirmation of these parameters and as a consequence of the experimental values of the masses.

1.1.4 Higgs physics at LHC

The Large Hadron Collider (LHC) at CERN (whose description will be given in the next chapter) has been built with the purpose of testing the SM theory, and in particular to discover and study the characteristics of the SM Higgs boson, and also to discover new physics BSM. In the next sections the production mechanisms, the decay processes and the discovery of the SM Higgs boson by CMS and ATLAS experiments will be treated.

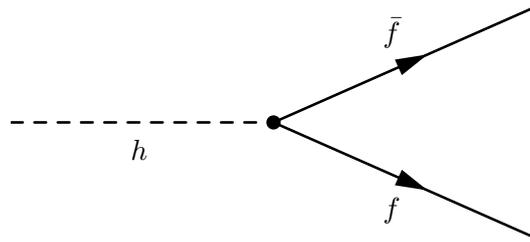
Production mechanisms and decay channels

The SM Higgs boson mass is not predicted by the theory so its production cross-section and branching fraction⁷ are calculated as function of its mass. In Fig. 1.3 the theoretical cross-section values are reported as function of M_h at the centre of mass energy $\sqrt{s} = 8$ TeV at which LHC proton-proton collisions occurred during

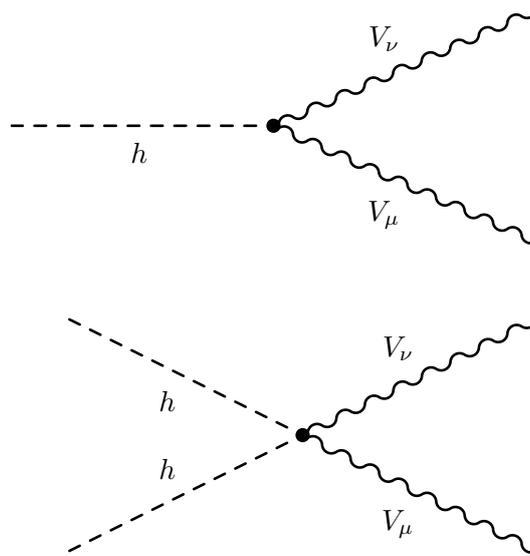
⁵ g_{h^3} and g_{h^4} are multiplied by i in the reference [48]

⁶Also in this case the couplings are multiplied by i in the reference [48]

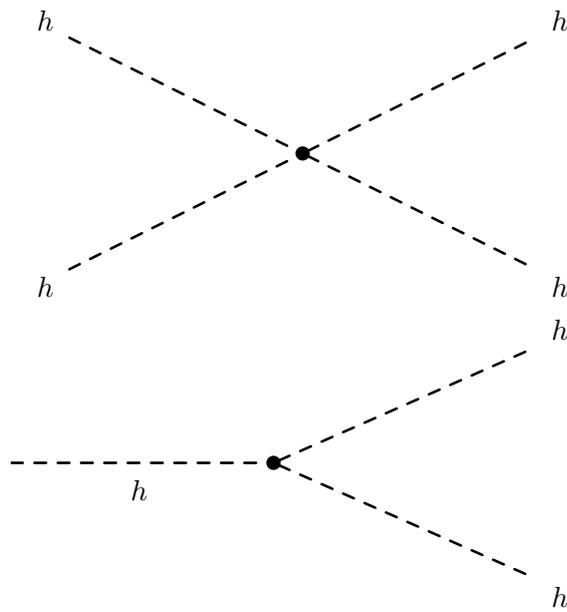
⁷The cross-section is defined in a simple way as the probability of a particular process occurs or some particles are produced; the branching fraction is the probability that a particle decays in a specific channel.



(a) SM Higgs - fermions interaction.



(b) SM Higgs - Vector bosons interactions.



(c) SM Higgs - Self interactions

Figure 1.2: Feynman diagrams of the SM Higgs interactions.

the 2012. The data collected during that period were analysed in this thesis. They are calculated at *leading order* (LO) considering the QCD and electroweak corrections [50] at *next-to-leading-order* (NLO).

As it can be observed, the most probable production mechanism is the *gluon fusion* $gg \rightarrow h$ (Fig. 1.4a), mediated by a loop of quark (the Higgs does not couple directly to the gluons because they are massless), and dominated by the loop of top quark which is the heaviest quark. The second process is the *vector boson fusion* (VBF) $qq \rightarrow V^*V^* \rightarrow qqh$ with $V = W, Z$, relevant in the search of the Higgs boson despite its cross-section is one order less than the gluon fusion. An important feature of the VBF production (Fig. 1.4b) is that the two jets are highly energetic and have the tendency to be produced along the beam line, with a high pseudorapidity⁸ ($|\eta| > 4$) and high invariant mass ($M_{jj} \gtrsim 700$ GeV). This peculiarity allows to discriminate better the signal from background. Despite their lower cross-sections the associated production with vector bosons (Fig. 1.4c) and with heavy quarks (Fig. 1.4d) can be used for the SM Higgs discovery especially in the low mass region. The first process, also called *Higgsstrahlung*, is quite recognisable thanks to the presence of one or two isolated leptons coming from the vector boson. The production in association with heavy quarks, beauty or top quarks, is quite challenging process because of the presence of a lot of hadronic jets in the event, but it is crucial for directly measuring the Higgs Yukawa couplings to t/b quarks.

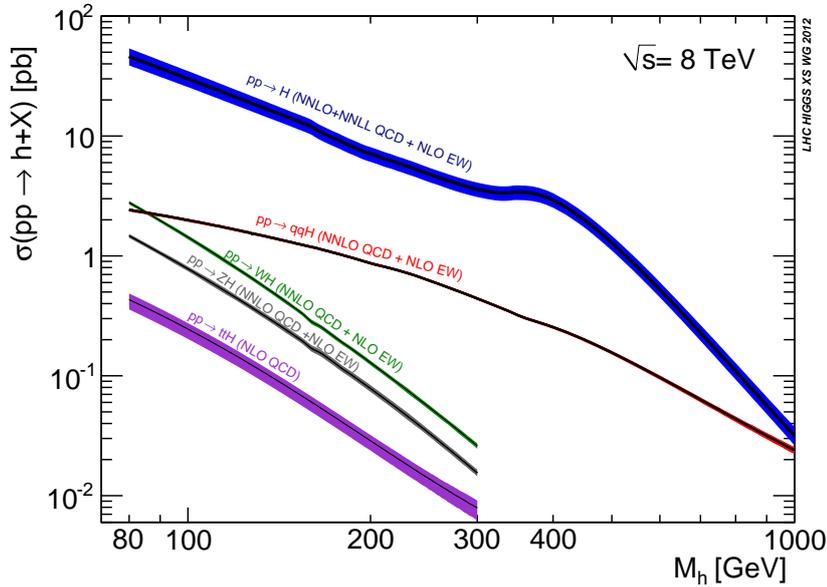


Figure 1.3: The SM Higgs boson production cross-sections as a function of M_h [50] (in this figure H corresponds to h).

The mass of the Higgs boson is crucial also in the calculation of the branching fractions of the Higgs [24], as can be seen in Fig. 1.5, where the theoretical *branching fractions* are reported as function of the Higgs mass. The SM theory provides

⁸The pseudorapidity will be defined in the section 2.2.1.

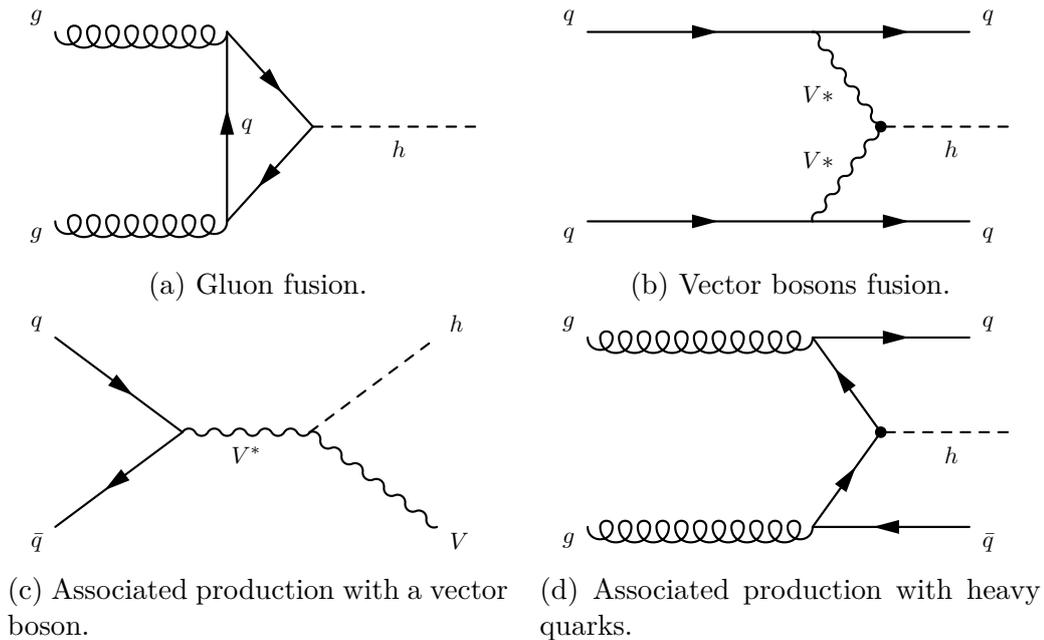


Figure 1.4: Feynman diagrams of the SM Higgs production mechanisms.

a mean lifetime for the Higgs boson $\approx 10^{-22}$ s, so the only way to detect the SM Higgs particle is through its decay products. The theoretical total decay width Γ_h is the inverse of the lifetime and from this and the partial decay width of a particle into X, it is possible to calculate the branching fraction B:

$$B(h \rightarrow X) = \frac{\Gamma(h \rightarrow X)}{\Gamma_h}. \quad (1.23)$$

The partial decay widths are directly proportional to the couplings and therefore to the masses of the particles in which the Higgs particle decays, and to the Higgs boson mass itself, with the tendency to decay in the heaviest particles.

According to the mass region some decay channels are more favoured than others. In the low mass region, $M_h \leq 130$ GeV, the dominant decay is $h \rightarrow bb$ ($B \approx 57\%$), which is quite challenging to identify due to a huge presence of jets in a hadronic environment but at the same time very important to study the coupling of Higgs with fermions. The decays $h \rightarrow gg$ through a top quark loop can compete with the $h \rightarrow \tau\tau$ and $h \rightarrow cc$ decays. The decay $h \rightarrow \gamma\gamma$ is highly suppressed in this region despite its clean signature, because it is produced through a W boson loop. In the high mass region the dominant decays are into W or Z bosons pair. The decay into real vector bosons is possible if $M_h > 2M_V$, below this kinematic threshold it is possible that the Higgs boson decays into one or two virtual gauge bosons, decaying into fermions.

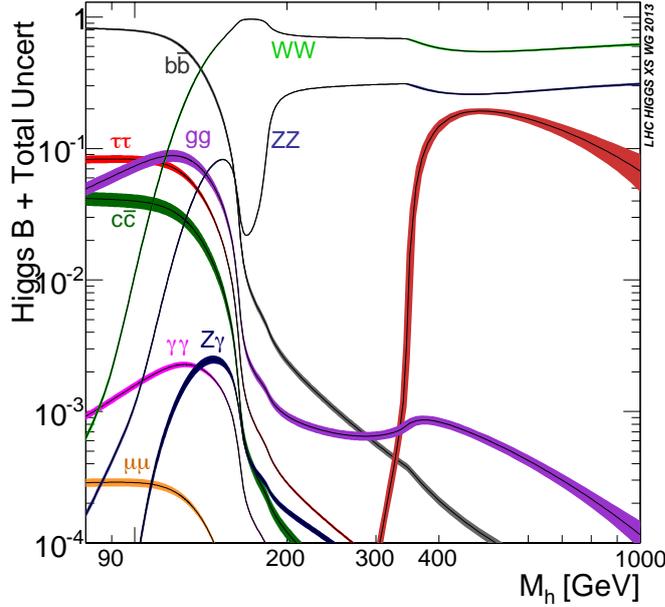


Figure 1.5: Theoretical branching fractions as a function of M_h , where the theoretical uncertainties are reported as color bands [24].

SM Higgs boson: experimental searches and discovery

The searches of the SM Higgs boson started at Large Electron–Positron collider (LEP) [51], that operated at CERN from 1989 to 2000 with centre of mass energies \sqrt{s} between 90 and 209 GeV. The process more probable at this collider was the Higgsstrahlung in association with Z boson and the final states with $b\bar{b}$ and $\tau\tau$ pairs were explored. Direct and indirect searches were performed but without any evidence of the presence of the SM Higgs boson. A lower bound of $M_h = 114.4$ GeV was established, at the 95% confidence level [52]. Other searches were performed at Tevatron [53], proton–antiproton collider at Fermilab, by the CDF and D0 experiments at $\sqrt{s} = 1.96$ TeV. The studied channel was the associated production of a Higgs boson with a W or Z boson and the decay of the Higgs boson to a $b\bar{b}$ pair. An excess of events was observed in the data compared with the background predictions in the mass range between 120 and 135 GeV, with a significance of around 3 standard deviations, which can be interpreted as the evidence of a particle compatible with the SM Higgs boson [54].

The discovery of a new particle with mass around 125 GeV was announced the 4th July 2012 separately by the ATLAS [4] and CMS [5] collaborations, after having analysed the data collected by both experiments at $\sqrt{s} = 7$ TeV, around 5 fb^{-1} , and at $\sqrt{s} = 8$ TeV, around $5\text{--}6 \text{ fb}^{-1}$. Despite the low branching fractions the discovery of the Higgs boson was performed in the $h \rightarrow \gamma\gamma$ and $h \rightarrow ZZ$ decays for their clean signature, giving an excess of events of 5 standard deviations with respect to the background [1,2]. After the end of the Run I at LHC at the beginning of 2013 the data collected at 8 TeV were increased up to an integrated luminosity of $\approx 20 \text{ fb}^{-1}$ so it was possible to study some properties of this new particle. The

combined measurements of ATLAS and CMS experiments for the mass gave the value $M_h = 125.09 \pm 0.21$ (stat) ± 0.11 (syst) GeV (Fig. 1.6) [8].

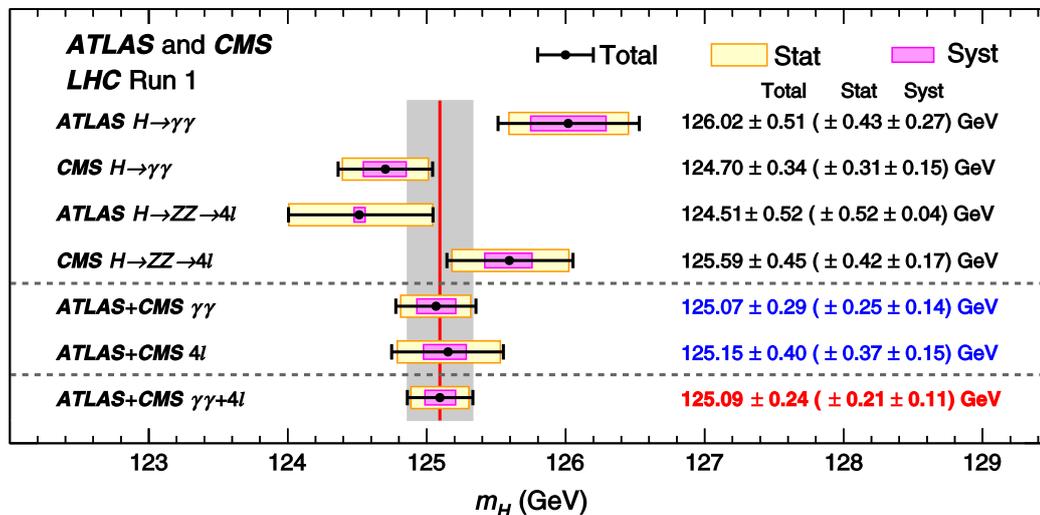


Figure 1.6: Summary of Higgs boson mass measurements from the individual analyses of ATLAS and CMS and from the combined analysis [8]. The systematic (narrower, magenta-shaded bands), statistical (wider, yellow-shaded bands), and total (black error bars) uncertainties are indicated. The (red) vertical line and corresponding (gray) shaded column indicate the central value and the total uncertainty of the combined measurement, respectively.

Further studies conducted by ATLAS and CMS confirmed that the 125 GeV particle discovered is compatible with the SM Higgs boson, indeed the spin and parity were confirmed to be equal to 0^+ by both ATLAS [55] and CMS [56] searches. The Higgs width predicted by the SM, Γ_h , is almost equal to 4 MeV, quite low with respect to the experimental mass resolution, so upper limits of $\Gamma_h \lesssim 22$ MeV have been set at the 95% of confidence level by both experiments [57, 58]. Finally the signal strength measured combining the results of both experiments is $\mu = 1.09_{-0.10}^{+0.11}$ and also the couplings to fermions and vector bosons were confirmed to be compatible with the SM predictions, as shown in Fig. 1.7 [9].

Recently the combination of the two experiments have shown that the significances of the VBF production process and of the $h \rightarrow \tau\tau$ decay are at the level of 5.4 and 5.5 standard deviations, respectively [9].

1.2 Beyond the Standard Model Higgs

Despite its success in the description of the strong and electroweak interactions and of the Higgs mechanism and properties, the SM theory presents unsolved issues and seems to be valid only at the energies explored by the present hadron colliders, the TeV scale [18]. The SM theory problems are the following:

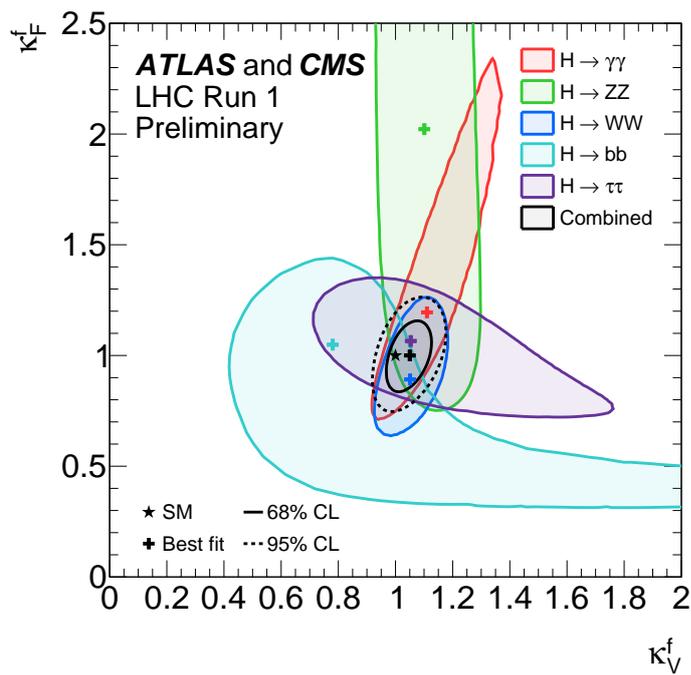


Figure 1.7: Negative log-likelihood contours of the relative to the SM predictions Higgs couplings to fermions (k_F^f) versus the relative to the SM predictions Higgs couplings to bosons (k_V^f) for the combined ATLAS and CMS measurements for the individual decay channels (identified by the symbol f) as well as for their global combination (k_F versus k_V shown in black), assuming that all coupling modifiers are positive [9].

- it does not include the gravitational force
- it assumes massless neutrinos, which on the contrary have been observed with non-zero masses [10, 11]
- it does not also explain the baryogenesis [12], the dark matter and the dark energy [13, 14]
- it is based on the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$ which does not provide the unification of the gauge coupling constants, the so called *unification problem*

A *Grand Unified Theory* (GUT), solving the *unification problem* beyond the TeV scale, for example up to the Planck scale $M_P \approx 10^{19}$ GeV, is expected to exist [59]. Furthermore in the SM two important issues are the *naturalness* and the *hierarchy problem* [15]. The first issue is related to the fact that the radiative corrections to the Higgs boson mass squared are quadratically divergent in the cut-off scale Λ , which represents the energy scale at which the theory remains still valid and consistent, and beyond this scale New Physics should appear. If Λ is chosen to be the GUT scale, the mass of the Higgs particle will prefer to be close to the very high scale unless an unnatural fine adjustment of parameters is performed, and in order to cancel the divergences, to lie in the range of the electroweak symmetry breaking scale, $v \sim 250$ GeV. The second issue is related to the reason why the cut-off parameter Λ is much greater than v and the forces are spread along different orders of magnitude.

These problems are not solved in the context of the SM but many *Beyond Standard Model* (BSM) theories have this purpose, besides wanting to fill the gap between the electroweak and the Planck scale with new particles.

Last but not least: is the discovered scalar boson [1, 2, 8, 9] the SM Higgs boson? The way to test this hypothesis is to proceed with the precision measurements of the SM parameters in the Higgs sector, including the Higgs mass and couplings to the fermions and to the other bosons, and, more important, to directly measure the Higgs self-coupling ($g_h^3 = 3 \frac{M_h^2}{v}$, from equation 1.20) by studying the double-Higgs production modes [60]. However, data collected so far at LHC by ATLAS and CMS (approximately 25 fb^{-1} in total per experiment during Run I) are insensitive to the self-coupling in the SM [61, 62], because of the expected small signal rates and the large backgrounds. On the other hand, the current measurements do not exclude the possibility that the discovered Higgs boson belongs to the Higgs sector predicted by models beyond the SM (BSM). As discussed in the next sections, the Higgs sector of some of these models, requires two Higgs doublets [18, 19], and one of the neutral heavy Higgs of these models can decay into two lighter “SM-like” Higgs bosons with the same mass and properties of the discovered scalar boson, with a signal rate, depending on the model parameters, significantly enhanced with respect to the SM double Higgs production. Furthermore, for these models the invariant mass of the two SM-like Higgs, is another handle to discriminate the signal of the decaying Higgs. The search for a resonant pair of

“SM-like” Higgs boson is a probe for New Physics Beyond the Standard Model, since the current achieved statistics at LHC can be sufficient to cover some part of the parameter spaces of these BSM models, not yet excluded from the other measurements [18,59]. This is the main motivation for the search presented in this thesis. Two BSM models, the Minimal Supersymmetry Model (MSSM) [17,18] and the Two-Higgs-Doublet-Model (2HDM) [19] will be used to interpret the Higgs boson observed at LHC (section 1.1.4). In this section the few key properties of the Higgs sector of the MSSM and of the 2HDM will be described, which motivated the experimental data analysis of this thesis. A detailed description of these models can be found in [16,18] and in [19].

1.2.1 The Supersymmetry and the Minimal Supersymmetry Model

The Minimal Supersymmetry Model (MSSM) [18] is the simplest extension to SM in the supersymmetry context based on the group $SU(3)_C \times SU(2)_L \times U(1)_Y$ and on the R-parity conservation.

The supersymmetry (SUSY) [16] is a symmetry that relates and transforms the bosons into fermions and viceversa. They are organized in superfields of two types: the simplest case is a *scalar* superfield, containing a complex scalar field and a two-component Weyl fermionic field; the second is a combination of spin-1/2 *gauginos* and spin-1 gauge bosons, called *vector* supermultiplet. Each supermultiplet is composed of both fermion and boson states, which are commonly known as *superpartners* of each other, and contains an equal number of fermion and boson degrees of freedom. The presence of superpartners cures the quadratic divergences of the Higgs mass squared and if the symmetry was exact the masses of the SM particles and their superpartners would be equal. The only freedom present in this theory is the superpotential which gives form to the scalar field potential and the Yukawa interactions between fermion and scalar fields [18]. Unfortunately the SUSY cannot be an exact symmetry since no fundamental scalar particles have been observed with the same mass of known fermions. Therefore SUSY has to be broken in a way that the hierarchy problem is not reintroduced and that the gauge invariance and the renormalizability of the theory are preserved. This can be done introducing by hand terms that break the symmetry and give the SUSY-breaking mechanism. This leads to a low-energy effective SUSY theory, called the Minimal Supersymmetry Model (MSSM).

SUSY then implies that the spin-1 gauge bosons and their spin-1/2 partners, the *gauginos* (Table 1.3), are in vector supermultiplets. There are only three generations of spin-1/2 quarks and leptons, which belong to *chiral* superfields together with their spin-0 superpartners, the squarks and sleptons (Table 1.3). Finally two chiral superfields \hat{H}_1, \hat{H}_2 with hypercharges equal to -1 and +1 respectively are needed to delete the quadratic divergences, and their scalar components H_1 and H_2 (Table 1.3) give masses to fermions while the two doublets fields lead to five Higgs particles, as will be described later.

The R-parity is defined as

$$R_P = (-1)^{2s+3B+L} \quad (1.24)$$

where s , B and L are the spin, barion and lepton quantum numbers respectively. The R-parity is equal to +1 for the SM particles and -1 for their supersymmetric partners. Its conservation leads to the fact that the SUSY particles are produced in pairs and that the lightest SUSY particle is stable. The symmetry breaking is realized adding terms to the Lagrangian without introducing other divergences, called soft-SUSY-breaking, introducing a huge number of unknown parameters. These parameters are reduced making assumptions like no CP-violation and no flavour-changing neutral currents (FCNC)⁹ at three level. Further details can be found in [16, 18].

Superfields	Quantum numbers			Particle contents	
	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	SM particles	Superpartners
Superpartners of gauge bosons					
\hat{G}_a	8	1	0	G_a^μ	\tilde{G}_a
\hat{W}_a	1	3	0	W_a^μ	\tilde{W}_a
\hat{B}_a	1	1	0	B^μ	\tilde{B}
Superpartners of fermions					
\hat{Q}	3	2	1/3	(u_L, d_L)	$(\tilde{u}_L, \tilde{d}_L)$
\hat{U}^C	$\bar{3}$	1	-4/3	\bar{u}_R	\tilde{u}_R^*
\hat{D}^C	$\bar{3}$	1	2/3	\bar{d}_R	\tilde{d}_R^*
\hat{L}	1	2	-1	(ν_L, e_L)	$(\tilde{\nu}_L, \tilde{e}_L)$
\hat{E}^C	1	1	2	\bar{e}_R	\tilde{e}_R^*
Superpartners of Higgs bosons					
\hat{H}_1	1	2	-1	H_1, \tilde{H}_1	
\hat{H}_2	1	2	1	H_2, \tilde{H}_2	

Table 1.3: The list of the MSSM superfields composed by the SM particles and their superpartners, and their quantum numbers. The MSSM Higgs sector is composed by the two superfields \hat{H}_1 and \hat{H}_2 [18].

Let's focus then the attention on the Higgs sector in the MSSM. As already said two doublets of complex scalar fields, H_1 and H_2 , are necessary to break the electroweak symmetry and to delete divergencies:

$$H_1 = \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix} \text{ with } Y = -1, \quad H_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} \text{ with } Y = +1 \quad (1.25)$$

⁹The flavour-changing neutral currents (FCNCs) represent the transition of a fermion which changes its flavour and not the charge. These processes are suppressed at tree-level in the SM and as well at higher orders by the Glashow-Iliopoulos-Maiani (GIM) mechanism [63].

The expression of the scalar potential (see eq. 1.18 for the SM scalar potential) in this case is given by:

$$\begin{aligned}
V_S = & (|\mu|^2 + m_{H_1}^2) |H_1|^2 + \\
& (|\mu|^2 + m_{H_2}^2) |H_2|^2 - \\
& \mu B \epsilon_{ij} (H_1^i H_2^j + \text{h.c.}) + \\
& \frac{g_2^2 + g_1^2}{8} (|H_1^0|^2 + |H_1^-|^2 - |H_2^0|^2 - |H_2^+|^2)^2 + \\
& \frac{g_2^2}{2} |H_1^{-*} H_1^0 + H_2^{0*} H_2^+|^2
\end{aligned} \tag{1.26}$$

where μ is the MSSM equivalent strength parameter used in the SM Higgs description and g_1 and g_2 are defined in 1.1.2. This potential provides a spontaneous symmetry breaking at its minima, therefore leading the two Higgs doublets to vacuum expectation values:

$$\langle H_1^0 \rangle = \frac{v_1}{\sqrt{2}}, \quad \langle H_2^0 \rangle = \frac{v_2}{\sqrt{2}}, \tag{1.27}$$

where the two vacuum expectation values are related to the SM one by

$$(v_1^2 + v_2^2) = v^2 = \frac{4M_Z^2}{(g_1^2 + g_2^2)} = (246 \text{ GeV})^2 \tag{1.28}$$

An important parameter can be defined to describe the MSSM phase space:

$$\tan \beta = \frac{v_2}{v_1} = \frac{v \sin \beta}{v \cos \beta} \tag{1.29}$$

In order to obtain the Higgs physical fields and their masses, the two Higgs doublets have to be rewritten around the vacuum values and using real and imaginary parts. The real parts represent the CP-even Higgs bosons and the imaginary parts correspond to the CP-odd Higgs and the Goldstone bosons. The resulting fields are:

- The two CP-odd scalar fields:

$$\begin{pmatrix} G^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} \text{Im}(H_1^0) \\ \text{Im}(H_2^0) \end{pmatrix} \tag{1.30}$$

- The four charged scalar fields:

$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} H_1^\pm \\ H_2^\pm \end{pmatrix} \quad (1.31)$$

- The two CP–even scalar fields:

$$\begin{pmatrix} H \\ h \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \text{Re}(H_1^0) - v_1 \\ \text{Re}(H_2^0) - v_2 \end{pmatrix} \quad (1.32)$$

where α , the mixing angle between the two CP–even scalar Higgs fields, is equal to:

$$\alpha = \frac{1}{2} \arctan \left(\tan 2\beta \frac{M_A^2 + M_Z^2}{M_A^2 - M_Z^2} \right), \quad -\frac{\pi}{2} \leq \alpha \leq 0 \quad (1.33)$$

The Goldstone bosons result to be massless and become the longitudinal modes of the Z and W^\pm massive vector bosons. The resulting five Higgs physical states, a neutral pseudoscalar A, two neutral scalars H, h, and two charged scalars H^\pm , are massive and their masses are related in the following way:

$$\begin{aligned} M_{H^\pm}^2 &= M_A^2 + M_W^2, \\ M_{H,h}^2 &= \frac{1}{2} \left[M_A^2 + M_Z^2 \mp \sqrt{(M_A^2 + M_Z^2)^2 - 4M_A^2 M_Z^2 \cos^2 2\beta} \right]. \end{aligned} \quad (1.34)$$

The supersymmetric structure has strong constraints on Higgs spectrum. Only two parameters are considered free at tree level, $\tan\beta$ and M_A , and there is a strong hierarchy on the mass spectrum:

- $M_H > \max(M_A, M_Z)$
- $M_{H^\pm} > M_W$
- $M_h \leq \min(M_A, M_Z) \cdot |\cos 2\beta| \leq M_Z$

These parameters will be used in section 1.3 in the context of a specific MSSM scenario and to interpret the results of the search conducted in this thesis in section 5.6.

1.2.2 The Two–Higgs–Doublet Model

Another possible extension of the SM are the Two–Higgs–Doublet–Models (2HDM) [18, 19]. In these models two scalar doublets provide masses to other particles by the means of the Higgs mechanism solving, under some hypothesis, several anomalies

of the SM like for instance the supersymmetry. The most general scalar potential, V_S , assuming two complex doublets of $SU(2)_L$ with hypercharge $Y = +1$, with scalar fields ϕ_1 and ϕ_2 equal to

$$\phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad \phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_5 + i\phi_6 \\ \phi_7 + i\phi_8 \end{pmatrix} \quad (1.35)$$

contains 14 parameters and it can be expressed assuming a CP–noninvariant theory, the gauge invariance and a correct breaking of $SU(2)_L \times U(1)_Y$ by

$$\begin{aligned} V_S = & m_{11}^2 \phi_1^\dagger \phi_1 + m_{22}^2 \phi_2^\dagger \phi_2 - m_{12}^2 (\phi_1^\dagger \phi_2 + \phi_2^\dagger \phi_1) + \\ & \frac{\lambda_1}{2} (\phi_1^\dagger \phi_1)^2 + \frac{\lambda_2}{2} (\phi_2^\dagger \phi_2)^2 + \\ & \lambda_3 \phi_1^\dagger \phi_1 \phi_2^\dagger \phi_2 + \lambda_4 \phi_1^\dagger \phi_2 \phi_2^\dagger \phi_1 + \\ & \left\{ \frac{\lambda_5}{2} (\phi_1^\dagger \phi_2)^2 + [\lambda_6 (\phi_1^\dagger \phi_1) + \lambda_7 (\phi_2^\dagger \phi_2)] \phi_1^\dagger \phi_2 + \text{h.c.} \right\} \end{aligned} \quad (1.36)$$

The 14 coefficients are all non–zero and real, except for m_{12}^2 and $\lambda_{5,6,7}$ that are complex and lead to a possible violation of CP. The expression of the scalar potential is quite complicated from a phenomenological point of view. However, to construct models that could be tested in modern experiments, several reasonable simplifications might be applied [19]:

- CP-symmetry is not spontaneously broken and is conserved in the Higgs sector.
- Tree level flavour-changing neutral currents (FCNC) are absent. Introduced symmetries, which ensure absence of FCNC, will also eliminate quadratic terms (see equation 1.36), where one of the doublets appears an odd number of times.

Under these hypotheses, $\lambda_6 = \lambda_7 = 0$. Similarly to the SM (see section 1.1.3) the minimization of this potential provides the vacuum expectation values of the two scalar fields:

$$\langle \phi_1 \rangle = \frac{v_1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \langle \phi_2 \rangle = \frac{v_2}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (1.37)$$

where v_1 and v_2 are the same vacuum expectation values described in the MSSM section (1.2.1). In order to obtain the Higgs physical fields, the fields have to be decomposed, as already described for the SM and the MSSM theories, obtaining

eight fields:

$$\phi_a = \begin{pmatrix} \phi_a^+ \\ (v_a + \rho_a + i\eta_a)/\sqrt{2} \end{pmatrix}, \quad a = 1, 2 \quad (1.38)$$

Three of them, the Goldstone bosons, are massless and give mass to the Z and W^\pm vector bosons. The other five fields are the physical fields, the same of the MSSM, H^\pm , A, h, H. After expanding ϕ_a (eq. 1.38) into the 2HDM potential, considering the simplifications mentioned above, the squared mass matrices for ϕ_a^+ , ρ_a and η_a fields might be obtained. Mixing angles that diagonalizes these matrices are the two most important 2HDM parameters: the angle α that diagonalizes the neutral scalars matrix and the angle β ($\tan \beta = v_2/v_1$) that diagonalizes the pseudoscalars and the charged scalars matrices. By rotating the basis we can express the neutral physical fields as function of ρ_a and η_a using the definitions of α and β , obtaining:

- the physical pseudoscalar field $A = \eta_1 \sin \beta - \eta_2 \cos \beta$
- The physical scalar fields, the lighter $h = \rho_1 \sin \alpha - \rho_2 \cos \alpha$ and the heavier $H = -\rho_1 \cos \alpha - \rho_2 \sin \alpha$

The SM Higgs boson field would be then:

$$H^{SM} = \rho_1 \cos \beta + \rho_2 \sin \beta = h \sin(\alpha - \beta) - H \cos(\alpha - \beta) \quad (1.39)$$

The assumption of absence of the FCNC is guaranteed by the Glashow–Weinberg condition that all the fermions of a given representation receive their masses through renormalizable Yukawa couplings to a single Higgs doublet [23]. This condition is satisfied by four assignments, where by convention up–type quarks are always taken to couple to ϕ_2 doublet:

- **Type I:** all fermions couple to one doublet, the ϕ_2
- **Type II:** the up–type quarks couple to ϕ_2 doublet, the down–type and leptons couple to ϕ_1
- **Type 3 - lepton–specific:** quarks couple to ϕ_2 doublet and leptons ϕ_1
- **Type 4 - flipped:** up–type quarks and leptons couple to ϕ_2 doublet, the down–type quarks to ϕ_1

Under all these assumptions, the free parameters left in 2HDM are then: M_A , M_H , M_h , M_{H^\pm} , m_{12}^2 , $\tan \beta$ and $\cos(\beta - \alpha)$ [19]. The interactions of the Higgs fields with vector bosons and fermions are determined by β and α , so they are the only parameters which determine the phenomenology of 2HDM.

In this thesis the 2HDM Type II¹⁰, like the MSSM, has been chosen to interpret the process $H \rightarrow hh$, since this process is enhanced in a specific benchmark scenario, as will be explained in the following section. Moreover, this model has the structure of the supersymmetric model, indeed it includes the MSSM. Obviously there are differences between the two models and it is important to highlight some of them: for example the 2HDM Type II doesn't have an upper bound for the lightest Higgs boson which is an important characteristics of the MSSM; in addition the parameters α , the scalar and the pseudoscalar masses are arbitrary, while in the MSSM the same quantities have constraints (respectively eq. 1.33 and 1.34).

1.3 The $H \rightarrow hh$ process at LHC

In the following sections the process $H \rightarrow hh$ at LHC and the main motivations of the search conducted in this thesis will be described in the context of MSSM and 2HDM models. Both theories, in order to introduce the observed scalar boson, require that the CP-even neutral scalar h has the same couplings and properties of the SM Higgs. This assumption is supported by recent results, where the hypothesis that the heavy BSM H is the discovered new boson is strongly disfavoured [64, 65]. Last but not least, this process has an enhanced sensitivity in the parameter region with low $\tan\beta$, not yet excluded by direct searches of the MSSM Higgs boson, described in [65–69], and accessible with the data collected so far by the ATLAS and CMS experiments at the LHC working at a centre of mass energy $\sqrt{s} = 8$ TeV. The theoretical predictions about cross-sections and branching fractions reported in the following, are evaluated at 8 TeV.

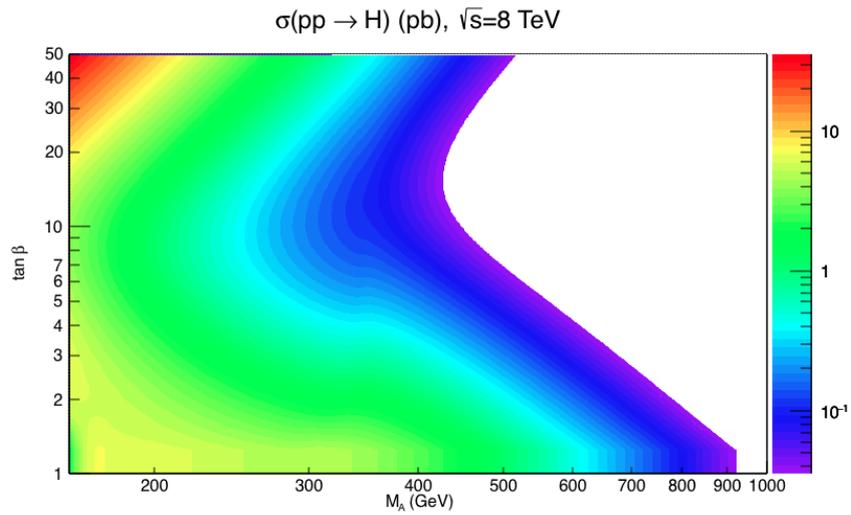
MSSM low- $\tan\beta$ -high scenario

The discovery of a boson with a mass of around 125 GeV supports the MSSM h Higgs, since in this model the lightest CP-even h boson is predicted with a mass less than ≈ 130 GeV, with an uncertainty of 3 GeV [20–22]. The $H \rightarrow hh$ process can be used to probe the MSSM in the $[M_A, \tan\beta]$ plane and in particular the low $\tan\beta$ region ($\tan\beta \lesssim 3$).

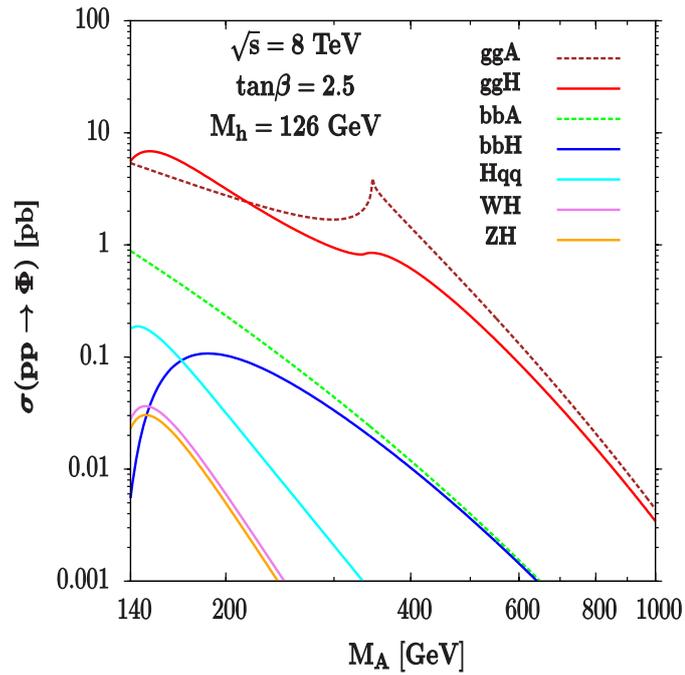
In Fig. 1.8a the total H production cross-section in the parameter space $[M_A, \tan\beta]$ is shown. Recent measurements by ATLAS [66, 67] and CMS [65, 68] excluded the so called high $\tan\beta$ region [21], leaving still room for the search of this process in the low $\tan\beta$ region, not yet excluded. Furthermore, in this region the dominant production process is the gluon fusion, $gg \rightarrow H$, (red line of Fig. 1.8b).

An important aspect for the prediction of the H branching fractions, is that the couplings to fermions and bosons are all described in terms of the mixing angles α and β [21, 22].

¹⁰In following 2HDM, if not otherwise specified, will indicate 2HDM Type II.



(a)



(b)

Figure 1.8: (a) MSSM production cross-sections of the neutral heavy Higgs boson H in the $[M_A, \tan\beta]$ MSSM plane with the constraint $M_h = 125$ GeV [21]. (b) MSSM production cross-sections of the neutral Higgs bosons as a function of M_A with $M_h = 126$ GeV and $\tan\beta = 2.5$ [22]. The simulation assumes a centre of mass energy of 8 TeV for LHC.

In particular, the trilinear Hhh coupling can be written:

$$g_{Hhh} \propto 2 \sin 2\alpha \sin(\beta + \alpha) - \cos 2\alpha \cos(\beta + \alpha) \quad (1.40)$$

For a SM-like h , α tends to $\beta - \pi/2$ (the so called *decoupling limit* [22]), as a consequence, the $H \rightarrow hh$ decay is highly suppressed at high $\tan \beta$, but dominant at low $\tan \beta$, thanks to the coupling g_{Hhh} (eq. 1.40).

In figure 1.9a the $B(H \rightarrow hh)$ in the $[M_A, \tan \beta]$ plane is reported, showing that the highest value of this branching fraction is predicted by the model in the low $\tan \beta$ region. In addition, fixing $\tan \beta$ equal to 2.5, the $H \rightarrow hh$ is the favored process when the H mass is in the range $2M_h \leq M_H \leq 2M_t$, as shown in Fig. 1.9b.

In this thesis the results of the search of $H \rightarrow hh$ process (section 5.6) will be interpreted in terms of the so called “low- $\tan \beta$ -high” scenario [70], which provides the production cross-sections and branching fractions for a grid of values of M_A and $\tan \beta$, with $1 \leq \tan \beta \leq 4$.

2HDM Type II scenario

The 2HDM Type II also assumes that the lightest Higgs h is the SM-like Higgs boson with $M_H \sim M_A \sim M_{H^\pm}$ [23]. Under these constraints, the couplings of the h are SM-like, realizing the so called *alignment limit*, where $\cos(\beta - \alpha) = 0$. Also in this case, the couplings of the Higgses of this model to fermions and bosons are parameterized in terms of the mixing angles α and β [19].

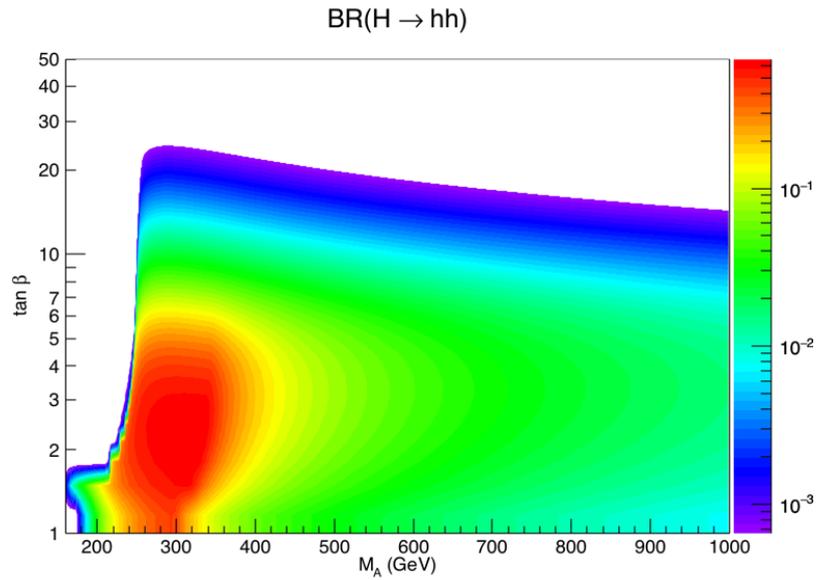
In particular, the Hhh coupling is given by:

$$g_{Hhh} \propto \frac{\cos(\beta - \alpha)}{v} M_H^2 \quad (1.41)$$

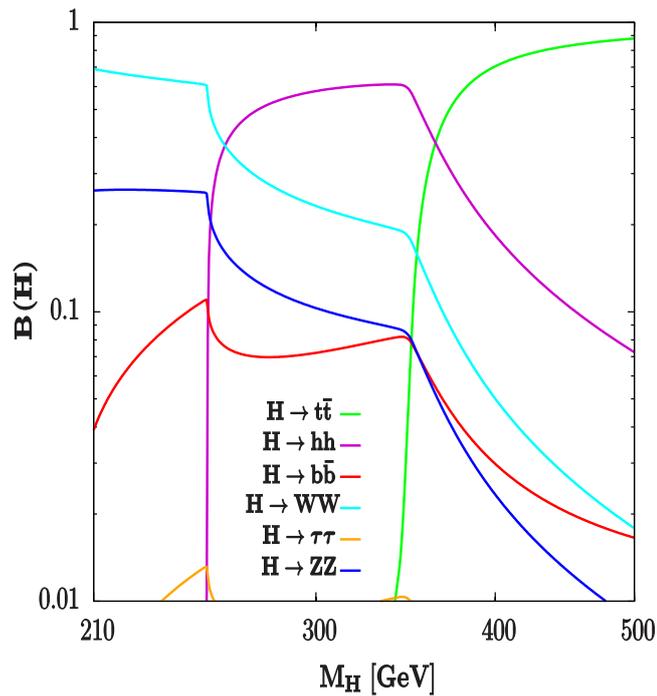
As a consequence, in the alignment limit this decay is suppressed, while in the approximation $\cos(\beta - \alpha) \approx 0$, near the alignment limit, the dominant decay is $H \rightarrow hh$, being for this process the decay width, $\Gamma(H \rightarrow hh)$, proportional to M_H^3/v^2 in the region $2M_h \leq M_H \leq 2M_t$.

The inclusive production cross-section of H boson times its branching fractions, $\sigma \cdot B(H \rightarrow X)$, is reported in Fig. 1.10a. At low $\tan \beta$, near the alignment limit, the dominant contribution is the $H \rightarrow hh$ and the leading production process from [19] is the gluon fusion. On the other hand, for high values of $\tan \beta$ the $H \rightarrow hh$ process is suppressed, as shown in Fig. 1.10b.

In this thesis the results of the search of $H \rightarrow hh$ will be interpreted in terms of the so called “2HDM Type II” scenario which provides the production cross-sections and branching fractions for a grid of values in the $[\cos(\beta - \alpha), \tan \beta]$ plane, with fixed masses for Higgses, $M_h = 125$ GeV (near the alignment limit), $M_H = M_A = M_{H^\pm}$ and $m_{12}^2 = M_A^2 \tan \beta / (1 + \tan^2 \beta)$.



(a)



(b)

Figure 1.9: (a) MSSM branching fractions of the neutral heavy Higgs boson into two h bosons in the $[M_A, \tan \beta]$ plane with the constraint $M_h = 125$ GeV [21]. (b) Possible branching fractions of the neutral heavy MSSM Higgs boson as a function of M_H with $M_h = 126$ GeV and $\tan \beta = 2.5$ [22]. The simulation assumes a centre of mass energy of 8 TeV for LHC.

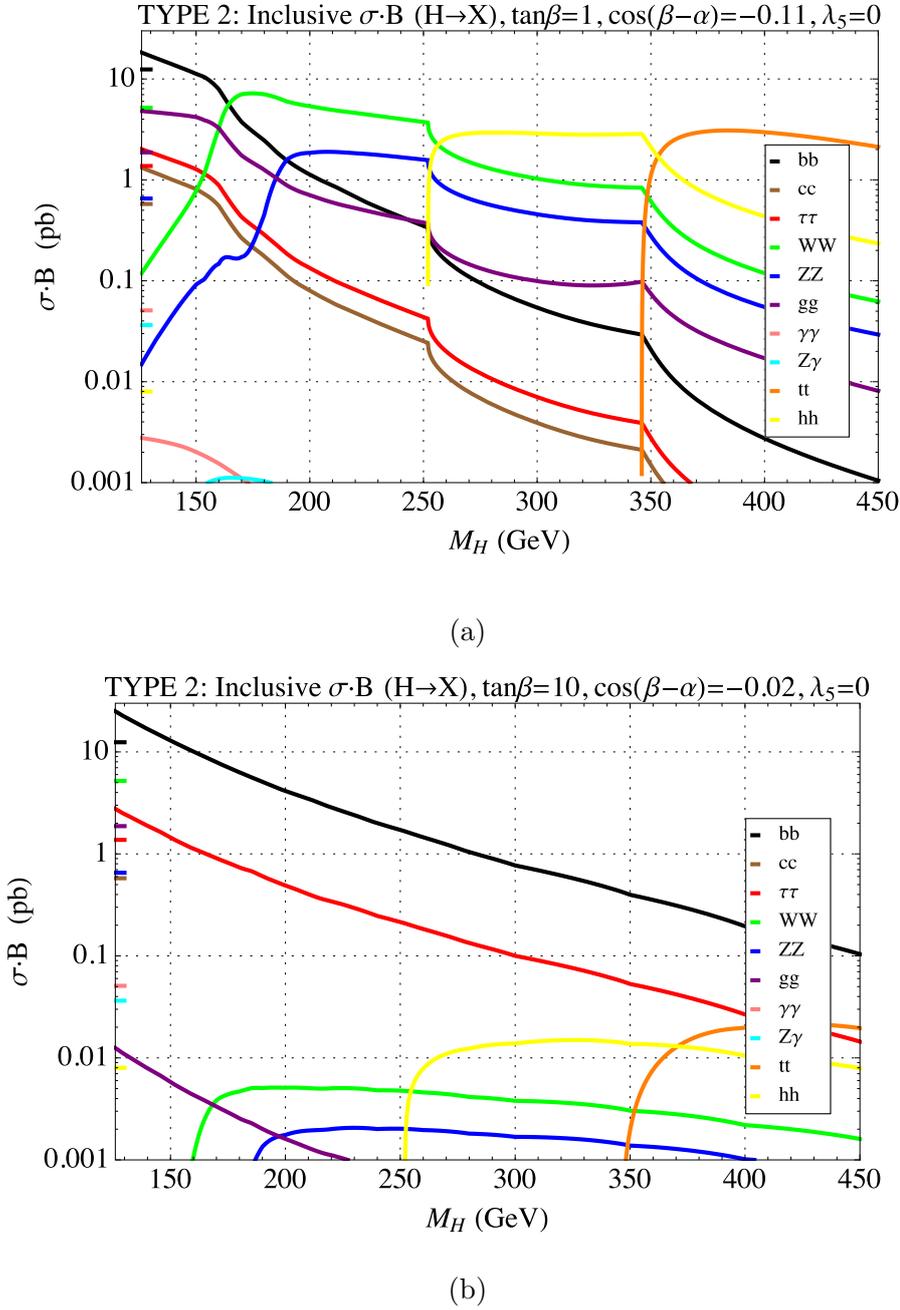


Figure 1.10: Cross-section times branching fraction $\sigma \cdot B(H \rightarrow X)$ to available final states in units of pb for 8 TeV proton-proton collisions for the non-SM-like scalar Higgs boson as a function of M_H for 2HDM Type II with $\tan\beta = 1$, $\cos(\beta - \alpha) = -0.11$ (a) and with $\tan\beta = 10$, $\cos(\beta - \alpha) = -0.02$ in (b), with $\lambda_{5,6,7} = 0$ and $M_A = M_H$ for both scenarios [23].

1.4 Physics motivations and challenges for the search of $H \rightarrow hh \rightarrow bb\tau_h\tau_h$ channel

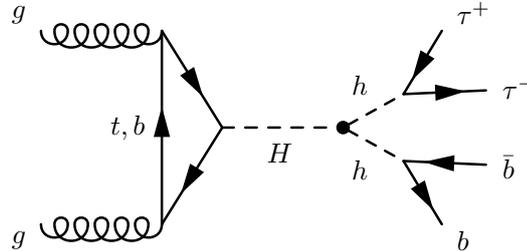


Figure 1.11: Feynman diagram of the process $gg \rightarrow H \rightarrow hh \rightarrow bb\tau\tau$.

The process studied in this thesis is the production of a heavy Higgs boson decaying in two 125 GeV Higgs bosons in $bb\tau\tau$ final state (Fig. 1.11), in the context of MSSM and 2HDM.

Main motivation of this search is that these models predict, for a well defined parameter region (previous sections), an enhanced production rate for the process $gg \rightarrow H$ and meanwhile the favoured decay mode is $H \rightarrow hh$ with h being the SM-like Higgs boson. All these characteristics make the search for double Higgs production very appealing at LHC with the data collected in 2012, in order to probe for new physics beyond the SM. On the other hand, several decay modes of the h pair produced in the process $gg \rightarrow H \rightarrow hh$ can be investigated. In the context of looking for two 125 GeV Higgs bosons, the final state in which one h decays into a pair of τ leptons and the other h decays into b quarks is very appealing, thanks to the high branching fraction of the $h \rightarrow bb$ ($B = 57.5 \pm 1.9\%$ [24]) and the quite clean and efficient signature of the h decaying into a τ pair ($B = 6.3 \pm 0.36\%$ [24]). Moreover, the most recent results of the evidence greater than 3σ for the 125 GeV Higgs boson decaying into τ leptons [71, 72] and its combination with $h \rightarrow bb$ searches [73, 74] gave stronger evidence of the 125 Higgs boson coupling to the fermions [75]. The choice of this final state is a compromise between the decay modes, $H \rightarrow hh \rightarrow bbbb$, which has the highest branching fraction but an overwhelming QCD background, and $H \rightarrow hh \rightarrow bb\gamma\gamma$, which has a very clean signature but a very low branching fraction ($B(h \rightarrow \gamma\gamma) = 0.228 \pm 0.011\%$). Given the potential of discovery or exclusion of new physics beyond SM, searches for the decays $H \rightarrow hh$ have been performed by the ATLAS [25–27] and CMS Collaborations [28–33] in di-photon, multilepton and bb final states. The $H \rightarrow hh \rightarrow bbbb$ and $H \rightarrow hh \rightarrow \gamma\gamma bb$ final states were already studied in other CMS analyses [29–33].

In this thesis the $\tau_h\tau_h$ final state, where τ_h indicates a τ which decays into hadrons plus a tau neutrino, was investigated. The decay $h \rightarrow \tau_h\tau_h$, called *fully-hadronic* channel, was chosen thanks to the very good performance of the τ_h reconstruction and identification provided by CMS experiment. In CMS the τ_h is reconstructed as a jet with well defined properties (section 3.2.8). Last but not least the *fully-*

hadronic channel constitutes the 42% of the di-tau decay modes. This decay channel is complementary to the $\tau_l\tau_h$, the *semileptonic* channel ($\approx 46\%$), where τ_l stays for a τ lepton which decays into muon or electron, and to the $\tau_l\tau_l$, the *fully-leptonic* channel ($\approx 12\%$), where both taus decay into muons or electrons.

The expected signature for $H \rightarrow hh \rightarrow bb\tau_h\tau_h$ is then characterized by the presence of two b jets and 2 τ jets (Chapter 3). Unfortunately this final state can be reproduced by different processes in which jets from quark or gluon are present. These are the multiple interactions in the proton-proton collisions and SM processes which mimic the searched signal process. In Fig. 1.12 the cross-sections of the SM processes investigated so far by CMS are shown. This figure reveals that the SM double Higgs production is quite prohibitive (with a cross-section of the order of 10 fb), while the process $gg \rightarrow H \rightarrow hh \rightarrow bb\tau\tau$ in the context of the MSSM and 2HDM, despite the low cross-section (order of 1.5 pb), can be still achievable. The dominant background of this search is the multijets process that with the SM background processes, are the main challenge for this search, having cross-section orders of magnitude greater than the studied process. These backgrounds and their modelling will be described in Chapter 4.

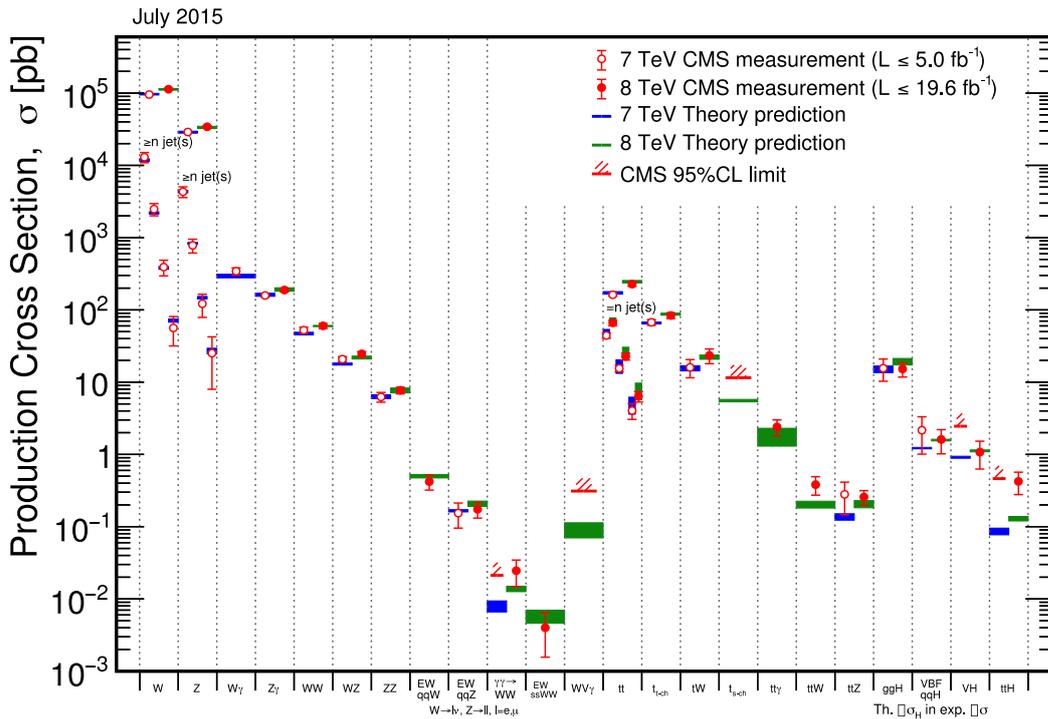


Figure 1.12: The most recent CMS cross-section measurements (and upper limits) of the SM processes compared to their theoretical prediction. The horizontal blu (7 TeV) and green (8 TeV) lines indicate the predicted cross-sections while the points mark the measured values. The production cross-sections of processes measured at CMS span several orders of magnitude [76].

Chapter 2

The CMS experiment at LHC

The data analysed in this thesis were collected by the Compact Muon Solenoid (CMS) detector [5] at the Large Hadron Collider (LHC) [3] at CERN. This chapter provides a general description of the experimental framework, the accelerator system, the detector, the trigger and the infrastructure, involved in producing and recording the data.

2.1 The LHC machine

The LHC is a circular proton-proton collider situated at CERN near Geneva in Switzerland. It is the last stage of the CERN accelerator complex shown in Fig. 2.1. Main purpose of this collider is to reach collision energies never achieved before at other hadron colliders, in order to probe the Standard Model theory, to explore the TeV scale and looking for New Physics.

It has been installed in the 27-km long tunnel built for the Large Electron Positron collider (LEP) [51], dismissed in 2000. The design of the machine provides a beam acceleration up to 7 TeV, reaching a centre of mass energy of 14 TeV and an instantaneous luminosity $\mathcal{L}(t)$ of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Moreover heavy ions Pb-Pb collisions are possible, in which the beam energy is 2.7 TeV, achieving an instantaneous luminosity of $\mathcal{L} = 10^{27} \text{ cm}^{-2} \text{ s}^{-1}$.

2.1.1 LHC complex and acceleration scheme

In Fig. 2.1, a schematic overview of the complex acceleration system of the LHC machine is reported. The protons are produced from hydrogen ionization and are pre-accelerated before entering in the LHC machine. The first acceleration is provided by the linear accelerator LINAC2 up to 50 MeV. Then the beam is injected in the Proton Synchrotron Booster (PBS) in which the protons are

accelerated up to 1.4 GeV. Subsequently the beam energy arrives to 26 GeV in the Proton Synchrotron (PS) where the protons are separated in bunches with 25 ns time-separation. Finally the bunches are transferred to the Super Proton Synchrotron (SPS), accelerated up to 450 GeV and then injected to LHC. The LHC is the last stage of a complex system of accelerators. It is composed of two rings in which injected protons circulate in opposite directions, with separated magnetic fields and vacuum chambers in the main arcs and with common parts at the interaction points. The original design provides that protons are collected in 2808 *bunches* composed of 10^{11} protons each and with a time separation of 25 ns. The tunnel is not a perfect ring but it is composed of eight arcs and eight straight sections. In the arcs are present dipole magnets and focusing quadrupoles. The dipole magnets (in total 1232) for bending the protons operate at 1.9 K reaching a maximum of 8.33 T, thanks to their superconductive characteristics. In each straight section there are superconductive radio-frequency cavities which are tuned to oscillate at 400 MHz.

Inside LHC the protons are accelerated up to 7 TeV where, thanks to the radio-frequency cavities, they keep their energy.

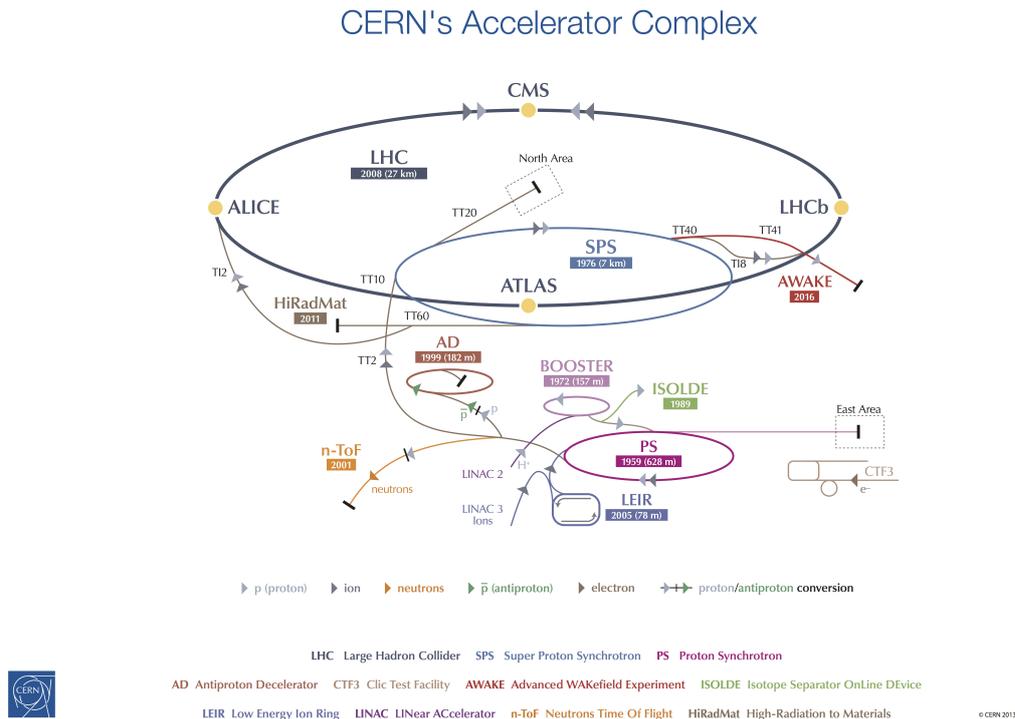


Figure 2.1: A schematic representation of the CERN accelerator complex [77].

In the LHC tunnel, four different experiments are located in the interaction points: ALICE (A Large Ion Collider Experiment) [78], aiming to study the quark-gluon plasma characteristics generated in the heavy-ion collisions; ATLAS (A Toroidal LHC ApparatuS) [4] and CMS (Compact Muon Solenoid) [5], general-purpose detectors, located in the points of highest luminosity and diametrically opposite

in the straight sections, designed to study the nature of electroweak symmetry breaking and discover new physics up to the TeV scale; LHCb (Large Hadron Collider beauty) [79], dedicated to study the rare decays of b hadrons and to perform precision measurements of CP violation.

Luminosity One of the key parameters that characterizes a collider performance is the machine instantaneous luminosity (\mathcal{L}), since the production rate for events of a given physical process is proportional to it. In particular at LHC:

$$\nu_{LHC\ process} = \sigma_{LHC\ process} \cdot \mathcal{L} \quad (2.1)$$

where $\nu_{LHC\ process}$ and $\sigma_{LHC\ process}$ are respectively the rate and the production cross-section of a process at the LHC. \mathcal{L} is a function of the beams parameters and for beams composed by Gaussian distributed bunches with equal amount of protons per bunch, it can be written as:

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n\beta^*} F \quad (2.2)$$

In eq. 2.2 γ is the Lorentz factor of the protons in the beam, f is the bunch frequency, k_B is the number of bunches per beam, N_p is the number of proton per bunch, ϵ_n is the normalized transverse emittance, β^* is the beta function at the interaction point and F is a reduction factor due to a non- π intersecting angle of the beams.

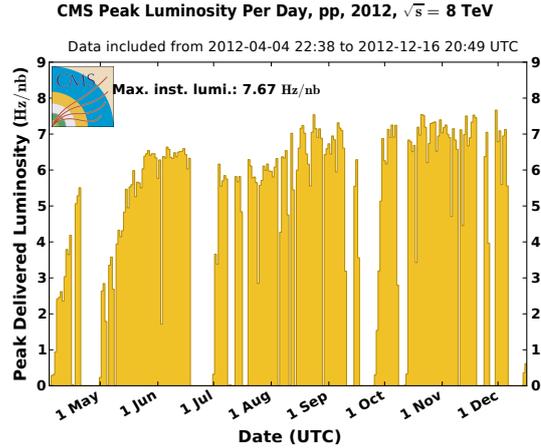
2.1.2 LHC milestones

The LHC adventure started in 2010 after some technical problems which delayed the data-taking of almost one year, operating at a centre of mass energy $\sqrt{s} = 7$ TeV and delivering at the end of 2010 a total integrated luminosity of 44 pb^{-1} and 6 fb^{-1} at the end of 2011. In the 2012 the beam energy was increased to 4 TeV and with a bunch spacing of 50 ns per each. At the end of the data-taking period, called Run I, the total integrated luminosity delivered to CMS was 23.3 fb^{-1} out of which 21.79 fb^{-1} was recorded. During this period the maximal instantaneous luminosity registered at the CMS collision point was $7.67 \times 10^{33}\text{ cm}^{-2}\text{ s}^{-1}$, while the LHC design peak luminosity is $10^{34}\text{ cm}^{-2}\text{ s}^{-1}$.

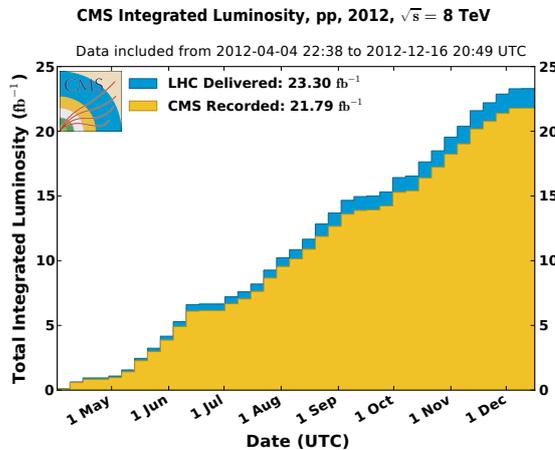
The analysis presented in this thesis uses the data collected during the Run I: instantaneous luminosity and cumulative luminosity at CMS collision point for this period are shown in Fig. 2.2.

The total inelastic cross-section measured for p-p collisions during Run I, as reported in [80], is about 75 mb. Despite the high luminosity reached by the LHC during the Run I, the high rate of collisions caused the overlap of several interactions in the same bunch crossing, the *pile-up* effect, degrading the performance

of the CMS detector. For Run I an average of 21 proton-proton interactions was reconstructed per bunch crossing, as shown in Fig. 2.3, where the multiple interactions arise from the same bunch crossing and from different bunch crossings (*out-of-time pile-up*). At the analysis level, in order to discriminate the primary hard collision corresponding to an interesting event from the pile-up events, the most energetic proton-proton collision per event is selected and is referred as the *primary hard* interaction, while the other collisions in the event are called *pile-up* interactions.



(a)



(b)

Figure 2.2: LHC performance at $\sqrt{s} = 8$ TeV in terms of (a) instantaneous luminosity delivered and (b) integrated luminosity in 2012. The blue curve corresponds to the delivered integrated luminosity while yellow curve corresponds to the data recorded under stable beam conditions [81].

2.2 The CMS detector

The CMS experiment [5] has been designed to study the p-p collisions at the LHC with the purpose of testing the Standard Model (SM) theory, which is culminated

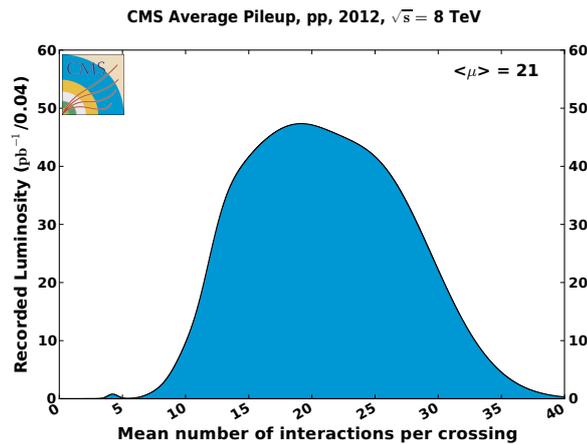


Figure 2.3: Mean number of interactions per bunch crossing collected in 2012 [81].

with the discovery of a scalar boson compatible with the SM Higgs boson [1, 2] and in order to explore the TeV scale.

The CMS detector has a cylindrical symmetry around the beam axis, composed of a central part, called *barrel*, and enclosed by two disks, called *endcaps*. It is quite *compact*, as the structure reveal: it has a diameter of 14.6 m, a length of 21.6 m and a weight of 12500 tons, and its coverage is hermetic up to $\eta = 5$. In Fig. 2.4 a schematic view of the CMS detector is shown. Looking at Fig. 2.5 and starting from the collision region, the following subdetectors can be found:

- an **inner tracking system**, composed of silicon sensors equipped with pixels and strips, which gives a very good vertex reconstruction and a good resolution for the particle momenta
- an **electromagnetic calorimeter (ECAL)**, which gives an accurate measurement of the electrons and photons energies, thanks to its granularity and wide coverage
- a **hadron calorimeter (HCAL)**, which covers hermetically the collision zone and, thanks to its sampling structure, provides a very good reconstruction of jets and of the missing transverse energy
- a solenoidal **superconducting magnet**, long 13 m and with a radius of 3 m, quite big to host the inner tracking and the calorimeter systems. It provides an almost uniform magnetic field of 3.8 T along z , allowing the measurement of the charged particles momenta in the inner tracking system.
- a **muonic system** interspersed between the iron plates of the return yoke. The complete coverage and a redundant system of four muon chambers provide a good identification of the muons and a good momentum resolution also at very high energies

The search presented in this thesis uses all the CMS subdetectors. In the next sections a short description of the CMS subdetectors and of the trigger system is given. More details can be found in [5].

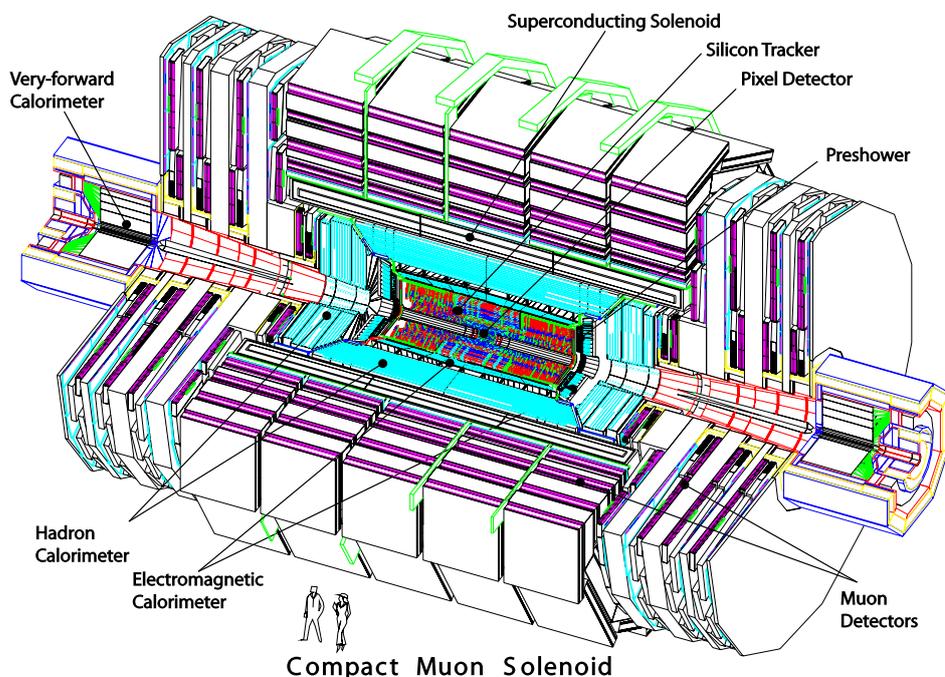


Figure 2.4: Schematic view of CMS detector and its subdetectors [5].

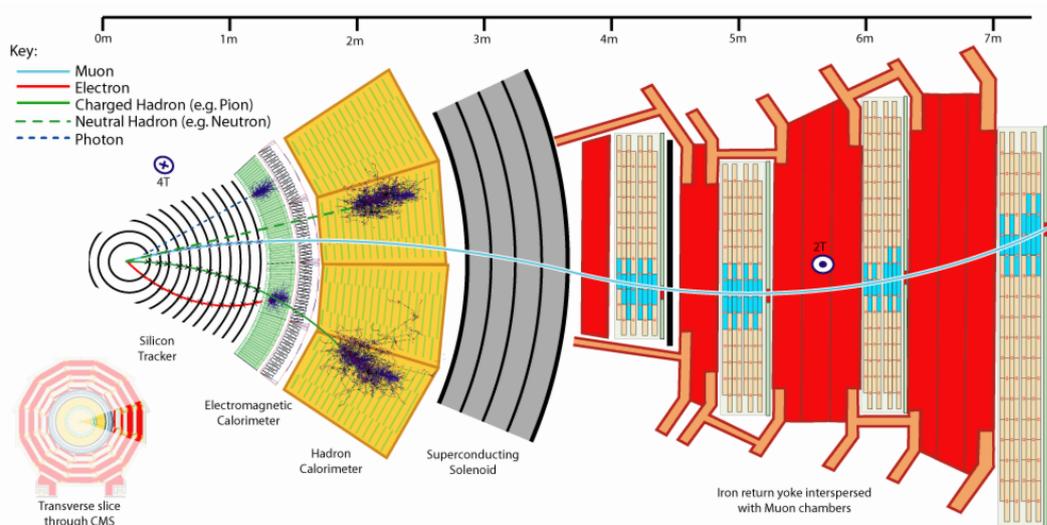


Figure 2.5: Transverse view of CMS detector in the barrel region, where detectable particles signatures are also highlighted.

2.2.1 Coordinate System and Useful Variables

The CMS coordinate system has its origin in the nominal interaction point, located in the center of the detector. The z axis is along the beam line and points toward the Jura mountains from LHC P5, where CMS is located. The x axis is horizontal and points toward the center of the LHC. The y axis is vertical and points upward. Thanks to its geometric symmetry, it is natural to define a cylindrical coordinate set (r, ϕ, θ) . Using the cartesian coordinate set (x, y, z) , r is the distance from the beam axis in the transverse plane $r = \sqrt{x^2 + y^2}$; $\phi = \tan^{-1} \frac{y}{x}$ is the azimuthal angle measured in the $x - y$ plane in the range $[-\pi, +\pi]$; $\theta = \tan^{-1} \frac{\sqrt{x^2 + y^2}}{z}$, is the polar angle measured from z axis in the range $[0, +\pi]$. Instead of the polar angle θ , the *pseudorapidity* is used in high energy physics (HEP) and in this thesis. The pseudorapidity (η) is defined as follows:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (2.3)$$

in the range $[-\infty, +\infty]$. The pseudorapidity is an approximation at high energies of the *rapidity* defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_L}{E - p_L} \right) \quad (2.4)$$

where p_L is the longitudinal momentum along the beam line. The *rapidity* is important because the difference of rapidities of two particles is a relativistic invariant for a Lorentz boost along the z axis and at the same time it is strongly dependent on kinematic variables. The same is true for the *pseudorapidity*. Another important quantity, the number of particles produced during a collision in a $d\eta$ interval ($dN/d\eta$) is invariant along the z axis. Other useful variables used in HEP and in this thesis are

- the transverse momentum $p_T = p \cdot \sin \theta = \sqrt{p_x^2 + p_y^2 + p_z^2} \cdot \sin \theta$, where p_x , p_y and p_z are the projections of the momentum of a particle \mathbf{p} respectively along the x , y and z axis
- Another useful variable is the Lorentz invariant separation, ΔR , between two particles, defined in terms of their difference in pseudorapidity, $\Delta\eta$, and in azimuthal angle, $\Delta\phi$, as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (2.5)$$

ΔR is used in this thesis to define a z -boost invariant cone opened around a single particle direction or detector position, in order to study nearby detector activity.

Another important element is the *beam spot*. The beam spot is the luminous

region near to the coordinates origin, where two proton beams collide. It is wide $16 \mu\text{m}$ along x axis, $100 \mu\text{m}$ along y, and it is long around 15 cm along z direction. The beam spot is the origin for all proton-proton interaction products and its location plays an important role during the event reconstruction, as described in Chapter 3.

2.2.2 Tracking System

The CMS tracking system [82] is the innermost CMS subdetector. It is designed to provide precise measurements of charged particle trajectories and momentum up to $\eta \approx 5$, as well as the positions of the primary and secondary vertices, which are crucial for the subsequent physical objects reconstruction. At the design operation conditions, the average amount of particles produced by proton-proton interactions in one bunch crossing is approximately 1000. To achieve a good precision in such high flux environment, the tracking system is composed by high granularity silicon detectors. The almost uniform magnetic field of 3.8 T, produced by the superconducting solenoid, described in section 2.2.5, allows to measure the transverse momentum of charged particles by bending their trajectories. The chosen tracker design allows to detect with a good efficiency and accuracy the trajectories of all charged particles with $1 < p_{\text{T}} < 100 \text{ GeV}$ produced in the collisions with $|\eta| < 2.5$. The layout of the inner tracking system is shown in figure 2.6. It is composed by the pixel detector and the strip detector described in the following.

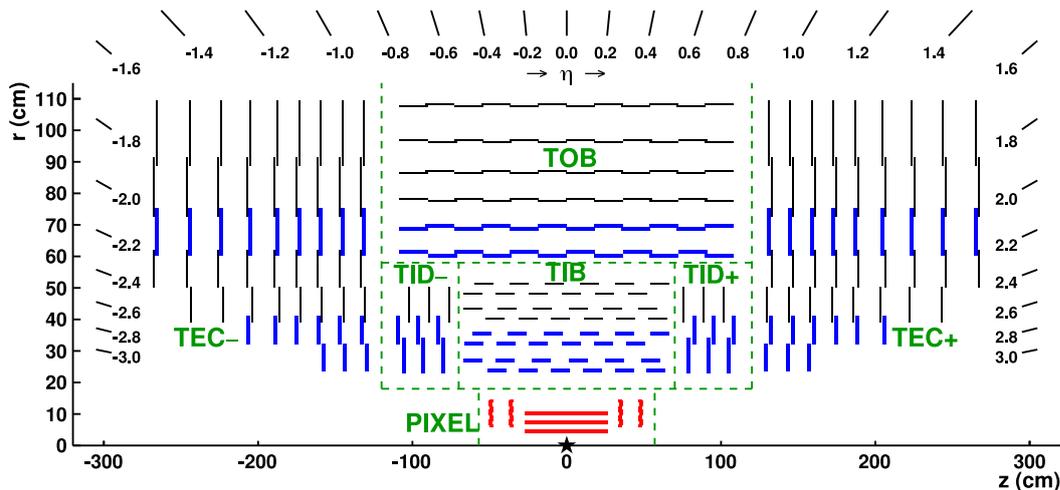


Figure 2.6: Schematic cross-section through the CMS tracker in the $r - z$ plane, where the interaction point is highlighted with a star and the main components are written: Pixel Detector (PIXEL), Tracker Inner barrel (TIB), Tracker Outer barrel (TOB), Tracker Inner Disk (TID) and Tracker endcap (TEC) [83].

Pixel Detector

The innermost part of the tracker is composed by the silicon pixel detectors (PIXEL), while the outer parts are using silicon strip detectors. Each of these parts is divided in barrel and endcap regions. The PIXEL (Fig. 2.6) includes three cylindrical barrel layers and two endcap disks on each side [82]. The barrel layers are 53 cm long with the average radii of 4.4, 7.3 and 10.2 cm. The endcap disks are located at $|z|$ equal to 34.5 and 46.5 cm extending in radius from 6 to 15 cm, covering a region up to $|\eta| < 2.5$. Both barrel and endcap are using 285 μm thick silicon modules with pixel size $100 \times 150 \mu\text{m}^2$. The pixel modules are arranged into half-ladders in the barrel part and into blades in the endcap. The total number of modules in the pixel barrel and in the endcaps is 768 and 672, respectively, which represents the overall area of approximately 1 m^2 with about $6.6 \cdot 10^7$ readout channels, organized in about 16000 modules of 52 columns and 80 rows. The endcap disks are assembled in a *turbine*-like geometry, where the blades are each rotated by 20° around their radial symmetry axis. This specific geometry has been chosen to profit of the Lorentz force in order to achieve a good spatial resolution in the η region, covered by the endcaps. A charged particle that traverses a module of PIXEL, produces a signal on different nearby pixels, which are combined in a hit. The high pixel granularity allows to obtain a three-dimensional measurement of the hit position, which is a key component for the precise vertex reconstruction. By calculating weighted average through the charge collected within a hit and by taking into account corrections due to the Lorentz drift of the collected electrons, the resulting hit position resolution of PIXEL is approximately $10 \mu\text{m}$ in the transverse direction and $20\text{--}40 \mu\text{m}$ in the longitudinal direction, depending on η .

Strip Detector

In the outer region ($20 < r < 110 \text{ cm}$), where the flux of particles is reduced, a strip silicon detector is used. The strip tracker is divided in four parts (Fig. 2.6): Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker Inner Disk (TID) and Tracker End Cap (TEC). The inner parts of the strip tracker, TIB and TID, are located in the radial region $20\text{--}55 \text{ cm}$ with $|z| < 124 \text{ cm}$ and are composed of 4 layers and 3×2 disks, respectively. The TIB and TID use 320 μm thick micro-strip silicon sensors with the pitch size varying for different layers in the range $[80,120] \mu\text{m}$ for TIB and $[100,141] \mu\text{m}$ for TID. The resulting TIB spatial resolution is $23\text{--}35 \mu\text{m}$ in the transverse direction. The outer strip tracker comprises 6 TOB layers, which occupy $55 < r < 116 \text{ cm}$ with $|z| < 110 \text{ cm}$, and 9×2 TEC disks, which occupy the region $124 < |z| < 282 \text{ cm}$, extending the η region covered to $\eta = 5$. In the TOB and 3 outer TEC disks 500 μm thick strip silicon sensors are used to improve signal-over-noise ratio, while in the 4 inner TEC disks 320 μm sensors are used, same as TIB and TID. The pitch size varies from 122 to 183 μm for the TOB and from 97 to 184 μm for the TEC. The resulting TOB spacial resolution is $35\text{--}53 \mu\text{m}$ in the transverse direction. In the first two

layers of TIB and TOB, and in the transition region between different parts of the tracker, as shown in figure 2.6, double sided modules are used. These are back-to-back modules with 0.1 rad “stereo” rotated strips. Such configuration allows to simultaneously measure transverse and longitudinal hit position. Longitudinal resolution for the “stereo” modules is $230 \mu\text{m}$ for the TIB and $530 \mu\text{m}$ for the TOB. The inner tracker covers the region $|\eta| < 5$, providing up to 10 high resolution measurements for charged particles with $|\eta| < 2.4$.

While the tracker provides key information for trigger (being a fast detector) and to reconstruct physical objects, it also adds significant amount of material budget. The total material budget added by the tracker is shown in Fig. 2.7 and, depending on η , it is equivalent to $\approx 0.4 - 2$ radiation length, X_0 [35], or $\approx 0.14 - 0.56$ nuclear interaction length, λ_I [35].

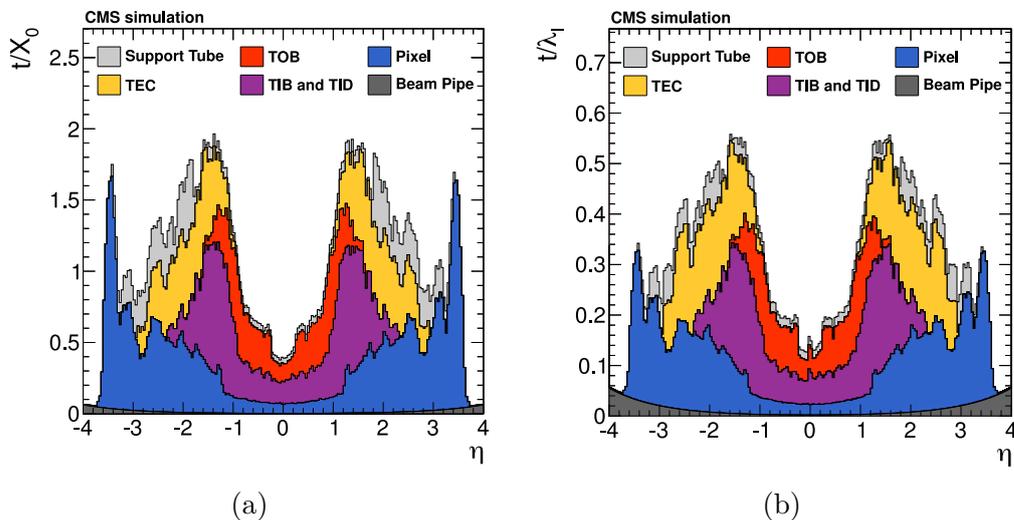


Figure 2.7: Total thickness of the tracker material traversed by a particle produced at the nominal interaction point, as a function of pseudorapidity η , expressed in units of radiation length X_0 [35] (a) and in units of interaction length λ_I [35] (b), showing the total material budget of each of the tracker subdetectors, together with contributions from the beam pipe and from the support tube that surrounds the tracker [83].

2.2.3 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) [84] has been designed to measure at best the energy and the direction of the electrons and of the photons. The design was optimized for the detection of the process $h \rightarrow \gamma\gamma$, which was considered as one of the “golden” channels for the Higgs boson discovery. Therefore, in order to obtain a good resolution for invariant mass of the γ pair, both energy and angular resolutions of the calorimeter are important. The high angular resolution it is also important to separate two closely located photons to suppress $\pi^0 \rightarrow \gamma\gamma$ background.

ECAL, located after the inner tracker, is a homogeneous and hermetic calorimeter composed by lead tungstate (PbWO_4) crystals with an array of photodetectors placed after a preshower detector. The choice of the PbWO_4 is appropriate for the LHC requirements. The high density, $\rho = 8.28 \text{ g/cm}^3$, the short radiation length, $X_0 = 0.89 \text{ cm}$ and the small Molière radius¹, $R_M = 2.2 \text{ cm}$ made possible to construct a compact calorimeter with high granularity. The crystal has also a fast time-response (emitting the 80% of light in 25 ns) and it is radiation hard (up to 10 Mrad). Due to low light yield (30 γ/MeV) it is important to use fast photodetectors with a high gain and suitable for operating in a strong magnetic field, like avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps.

It is separated in the barrel (EB) and endcap (EE) sections, as in Fig. 2.8. The barrel part (EB), with an internal radius of 129 cm, covers a region up to $|\eta| < 1.479$. The 61200 crystals of EB have a truncated pyramid shape which varies with η and create a $\eta - \phi$ grid. Each crystal is long 230 mm with a section towards the interaction point wide $22 \times 22 \text{ mm}^2$, with a granularity equal to 0.0174×0.0174 in $\Delta\eta \times \Delta\phi$, corresponding to an average of $25.8 X_0$.

The endcap part (EE) covers a region $1.479 < |\eta| < 3.0$ and it is distant 314 cm from the nominal interaction point. Each endcap is divided in two parts with a D shape, named “Dees”. The crystals are all identical and have a front area wide $28.62 \times 28.62 \text{ mm}^2$, the back area wide $30 \times 30 \text{ mm}^2$ and they are long 220 mm ($24.7 X_0$). A *preshower* sampling calorimeter, located in front of the EB calorimeter, provides a good separation between π^0 and γ . It is made by two alternate layers of lead and silicon strip detectors. The preshower has a total thickness of 20 cm, corresponding to about $3 X_0$, and covers the region between $1.653 < |\eta| < 2.6$.

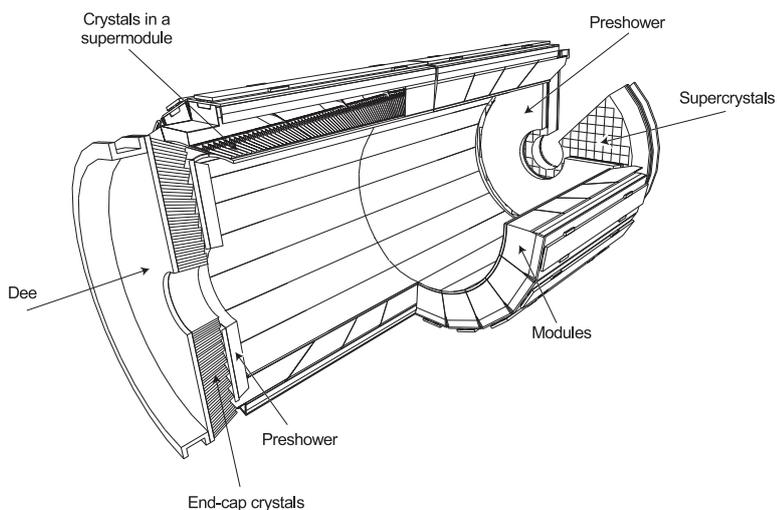
The energy resolution for ECAL up to energies of 500 GeV has been parameterized as follows:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_N}{E}\right)^2 + c^2 \quad (\text{with } E \text{ in GeV}), \quad (2.6)$$

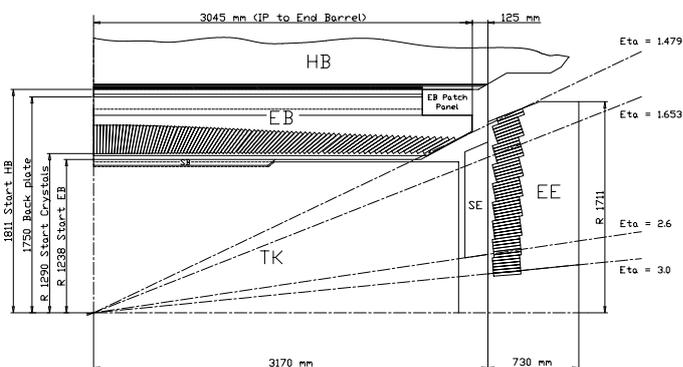
where a is a stochastic term that includes the statistical fluctuations of the shower containment as well as a contribution from the photostatistics; σ_N is the noise term due to electronic noise and *pile-up* energy; the non uniformity of the light collected by the calorimeter, the intercalibration errors and the loss of energy from the back side of the crystal contribute to the constant term c , which has to be kept lower than 0.55% to profit of the excellent stochastic term provided by PbWO_4 crystal, relevant for the Higgs searches.

Different contributions to the energy resolution are shown in Fig. 2.9. The parameterization of the energy resolution used to obtain Fig. 2.9 has been obtained in a test-beam [84], and the parameters values are reported in Table 2.1.

¹The Molière radius R_M is the radius of a cylinder containing the 90% of the electromagnetic showers in the lateral direction. It sets the transverse shower size and gives the lateral deflection of critical energy electrons after traversing one radiation length [85].



(a)



(b)

Figure 2.8: CMS electromagnetic calorimeter: 3D view (a) [5] and part of the longitudinal section (b) [84].

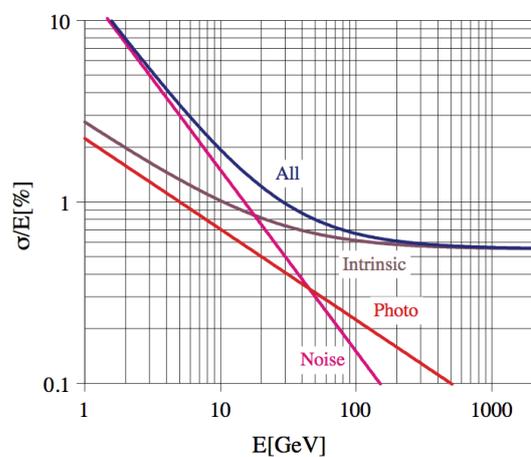


Figure 2.9: Different contributions to the energy resolution of the PbWO_4 calorimeter [84].

<i>Region</i>	a [GeV ^{1/2}]	σ_N [GeV]	c [%]
Barrel	0.027	0.155	0.55
Endcap	0.057	0.770	0.55

Table 2.1: ECAL energy resolution contributions in barrel and endcap. The unit of measurement for the energy E is GeV [84].

2.2.4 Hadron calorimeter

Surrounding the ECAL there is the hadron calorimeter (HCAL) [86], designed to detect and measure the energy of strongly interacting particles. It is a sampling calorimeter with alternating layers of absorber and scintillator material.

The longitudinal scheme of HCAL is shown in Fig. 2.10. It is composed of four parts: *Hadron Barrel* (HB), *Hadron Endcap* (HE), *Hadron Outer* (HO) and *Hadron Forward* (HF). HB and HE are inside the solenoid magnet, HO is outside the magnet and completes the structure of the calorimeter in the barrel, HF covers the forward region up to $\eta \approx 5.2$.

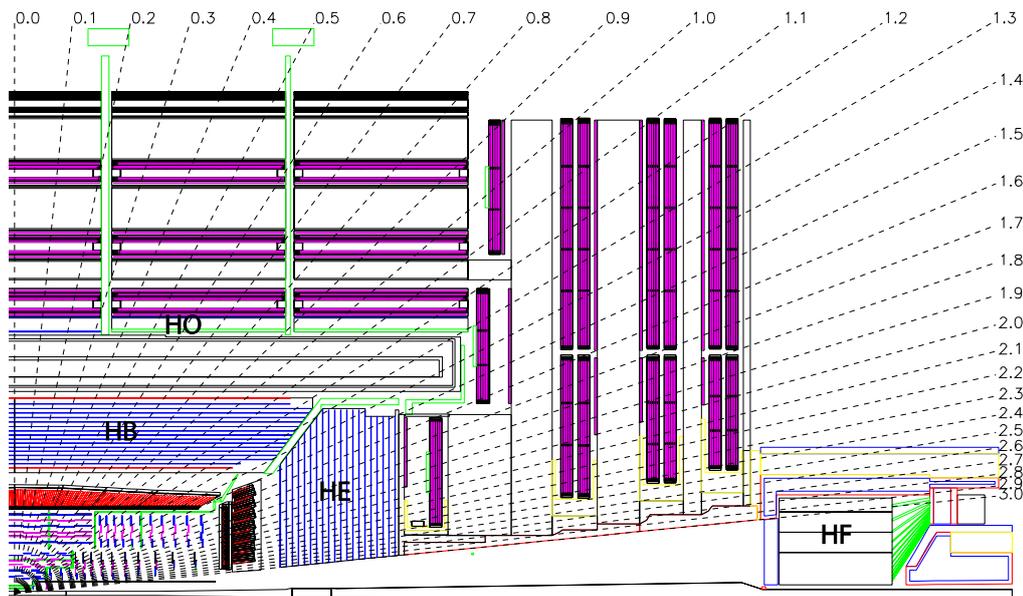


Figure 2.10: Layout of one quadrant of the hadron calorimeter in the $r - z$ plane. The HB, HE, HO and HF components of the detector are highlighted. The tracker and ECAL sub-detectors are visible in the inner part of the detector and the muon chambers are visible in the outer part [5].

Hadron Barrel HB is a sampling calorimeter which covers the region $|\eta| < 1.3$ from $r = 1.77$ m up to $r = 2.95$ m. The barrel is divided in two parts made of 18 identical wedges, each of which is segmented in four azimuthal angle

(ϕ) sectors. The plastic scintillator is segmented and readout in $\eta - \phi$ towers granularity 0.087×0.087 , thus equivalent to the area of a 5×5 array of the ECAL crystals. The active medium in the scintillator is composed of 70000 tiles, connected to wavelength shifter fibres that collect the emitted light. The tiles placed in the same ϕ region are grouped in scintillating units, called tray. The absorber consists of a front steel plate, 40 mm thick, 8 brass plates 50.5 mm thick plus other 6 brass plates 56.5 mm thick and finally a 75 mm thick back steel plate. The brass has been chosen because it maximises the radiation length, it is easy to model and it is not a ferromagnetic material. The effective thickness in units of nuclear absorption length, λ_I , seen by a hadron entering in the barrel grows as $1/\sin\theta$, reaching $10.6 \lambda_I$ at $|\eta| = 1.3$.

Hadron Endcap HE is a sampling calorimeter as well and covers a region $1.3 < |\eta| < 3$. It is divided in two sections: eight internal towers with a $\eta - \phi$ granularity 0.175×0.175 , and five external towers with a granularity 0.087×0.087 . Its hermetic structure consists of 79 mm thick brass plates with 9 mm gaps to accommodate the scintillators.

Hadron Outer HB does not provide sufficient hadron shower containment for high energy hadrons, then a HO calorimeter has been built. It covers the same η region as the HB, is placed outside the solenoid coil and it is used to detect the tails of high energy hadron showers. It contains 10 mm thick scintillators and is divided in five η regions, called ring, each covering 2.5 m along z . The central ring has two layers of scintillators (at radial distances of 3.82 m and 4.07 m, respectively) divided by a 19.5 cm thick piece of iron (the tail catcher iron). All other rings have a single scintillator layer at a radial distance of 4.07 m. The additional absorber material of the HO corresponds to an interaction length of about $1.4\lambda_I/\sin\theta$. HO allows to increase the minimal interaction length of the HCAL at $\eta = 0$ of the HCAL to $11.8 \lambda_I$.

Hadron Forward HF has a cylindrical structure with an external radius of 130 cm. It covers a region $3 < |\eta| < 5.2$ and it is placed outside the muon chambers. It is a sampling calorimeter, whose absorber is made of steel cylindrical plates for a total of 1.65 m thickness along z and the active part uses radiation hard quartz fibers. The fibers run in a longitudinal direction and are organized to form towers with $0.175 \times 0.175 \eta - \phi$ granularity. This configuration has been chosen in order to obtain a detector which resists to the high flux of particles in this region, resulting in a radiation dose that approaches the 100 Mrad/year for the LHC operation luminosity design. Because the HF provides good precision for energy measurements in the forward region, it has been used for the CMS luminosity measurements, as described in appendix A.

Energy Resolution HCAL has been designed to provide a good energy resolution for hadronic jets (section 3.2.5). In addition its hermetic design and shower

containment are driven by the need for a precise measurement of the missing transverse energy (section 3.2.7) per event and to minimize the muon misidentification strongly reducing the punch-through hadrons in the muon chambers. Based on a test beam results, the energy resolution of the central HCAL parts can be fit with a parameterization similar to the one used for ECAL and shown in equation 2.6. In the case of HCAL, the contribution from the noise σ_N is negligible. The resolution is measured using a test beam with charged pions in the energy range 5–300 GeV and is found to be approximately [87]:

$$\frac{\sigma}{E} = \frac{84.7\%[\text{GeV}^{1/2}]}{\sqrt{E}} + 7.4\% \quad (2.7)$$

2.2.5 The magnet

The superconductive magnet [88] is the central device of CMS around which the entire experiment has been built with the aim of bending particles coming from LHC collisions. This allows to measure with high precision the transverse momentum of charged particles. Thanks to its high magnetic field, 3.8 T, a momentum resolution better than 10% can be reached at 1 TeV and high performances of the muonic system can be kept. The CMS magnetic field is a solenoid made of 4 layers of NbTi superconductor, long 12.9 m, with an inner diameter of 5.9 m. Inside the magnet the calorimeters and the tracker system are located. The magnet is composed of 2168 turns cooled at the temperature of -268.5°C ; a nominal current of 19 kA circulates in the cables and a total energy of 2.6 GJ is stored in the magnetic system. The lines of the field are closed in the iron-yoke, thick 1.55 m in the barrel and 1.45 m in the endcap, where the muon system (section 2.2.6) is housed. The magnetic field in the iron yoke is 2 T and points in a direction opposite to the direction of the field inside the coil. Then high momentum muons are bended in two opposite directions in the tracking and in the muon systems (see muon track example of Fig 2.5). This improves their momentum resolution.

2.2.6 Muon System

The particles produced by the interaction and subsequent decays are absorbed by the system described above with a very high probability. The most common particle that escapes the calorimetric system is the muon. For that reason, the muon system is the most outer part of the CMS detector. The CMS muon system [89, 90] is designed to efficiently identify muons, to precisely measure their momentum and to be used for triggering on events with muons. It utilizes gaseous detectors, exploiting different technologies, positioned outside of the solenoid and covering the range $|\eta| < 2.4$.

It has a cylindrical section along the barrel and two endcap planar regions. It is composed of three kinds of gaseous particle detectors, as shown in Fig. 2.11.

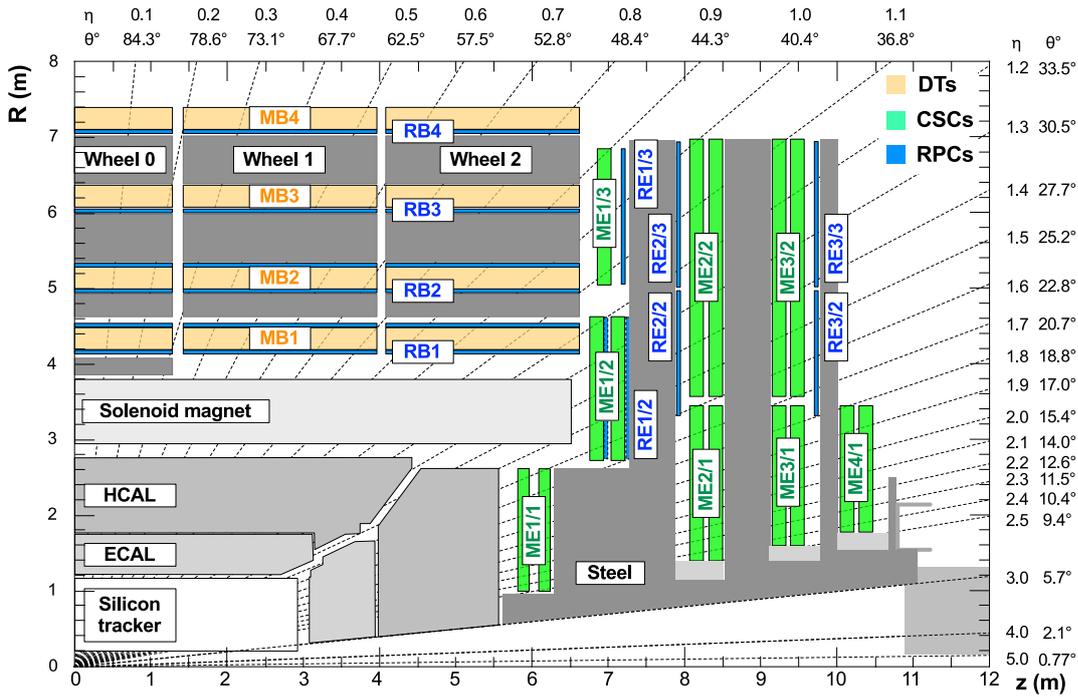


Figure 2.11: Layout of one quadrant of the muon system in the $r - z$ plane. The positions of the DT, CSC and RPC chambers are highlighted [90].

Drift Tubes Chambers (DT) are used in the barrel region $|\eta| < 1.2$ because of the low rate for background and the almost uniform magnetic field. The chambers (Fig 2.12) are divided into 12 ϕ -segments per wheel, forming 4 stations (MB1, . . . , MB4, with MB standing for Muon Barrel) at different radii interspersed between plates of the magnet flux-return yoke. Each station consists of 8 layers of tubes measuring the position in the bending plane and 4 layers in the longitudinal plane (except for MB4). In the fourth station the radial separation among chambers is increased to achieve the best angular resolution. The basic element of a DT, the drift cell, has a transverse size of $42 \times 13 \text{ mm}^2$. The gas in each cell is a mixture of argon (85%) and carbon dioxide (15%), providing a saturated drift velocity of about $55 \mu\text{m/ns}$. The maximum drift time is almost 400 ns. Drift cells in a layer are staggered of a half-cell with respect to their neighbor layer to avoid any dead hit and allowing to solve the left-right ambiguity.

Cathode Strip Chambers (CSC) have been installed in both endcaps ($0.9 < |\eta| < 2.4$ region) since they are suitable to resist to the high particle flux and to ensure best performance in the non uniform magnetic field thanks to their fine segmentation, their radiation hardness and fast-time response. The chambers are divided in four stations (named ME1, . . . , ME4), placed perpendicular to the beam and interspersed with iron layers for the flux-return yoke. There are 468 CSC all arranged in a way that they overlap along ϕ , avoiding dead zones. The CSC are wedge-shaped multi-wire proportional chambers containing six gas layers with cathode strips running radially outward to measure hits in the $r - \phi$ plane, and

anode wires running perpendicular to measure η . The anodes also provide good time resolution, which is used to correctly identify the bunch-crossing from which a muon originates.

Resistive Plate Chambers (RPCs) augment the CSC cover in the region $|\eta| < 1.6$. RPCs are constructed from parallel anode and cathode plates with a gas gap in between operating in avalanche regime with a high intrinsic time resolution of ≈ 2 ns. Muon ionisation is detected by arrays of metallic strips that run parallel to the beam axis. The 4 RPC stations in the barrel region and the 3 RPC stations in the endcaps region provide additional capabilities for triggering on high p_T tracks, due to the fast RPC response time, and extra information for tracks ambiguity resolution.

The spatial and time resolutions of the muon system were measured using proton–proton collisions data collected by the CMS in 2010 at $\sqrt{s} = 7$ TeV [90]. The average DT chamber resolution are around $80\text{--}120\ \mu\text{m}$ and $130\text{--}390\ \mu\text{m}$, respectively in $r - \phi$ and $r - z$ directions. The spatial resolution of CSC varies from 58 to $136\ \mu\text{m}$ in the strips local coordinates, corresponding to $r - \phi$ global coordinates. For the RPC chambers the spatial resolution varies from 0.8 to 1.3 cm, similarly in the strip local coordinates. The time resolution achievable was 3 ns or better per chamber for all 3 systems. This technology allows then fast access to hit information from the entire muon detector for use in a hardware muon trigger and to correctly identify the bunch crossing from which a muon originates.

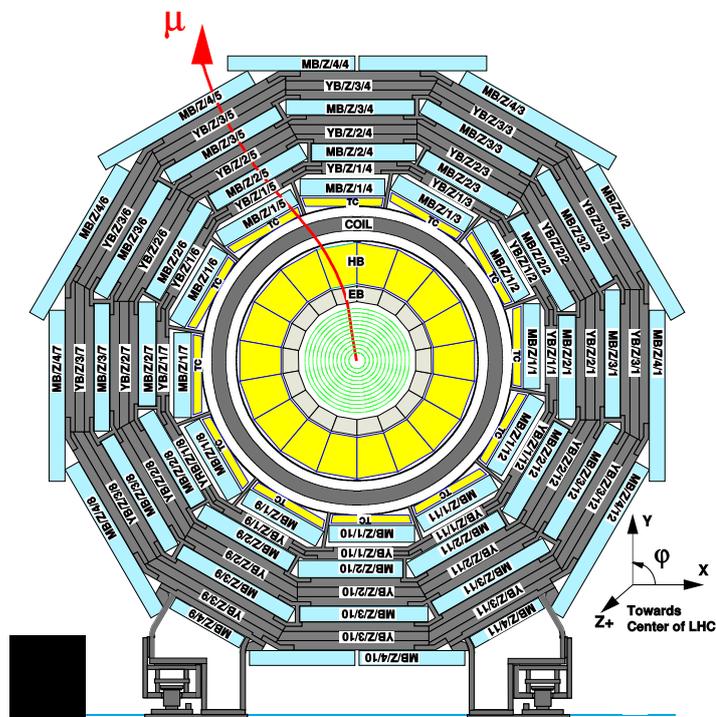


Figure 2.12: Schematic representation of the CMS barrel muon drift chambers (light blue) showing the different layers of the CMS detector in the $r - \phi$ plane [5].

2.2.7 Trigger system

At the design luminosity ($\mathcal{L}(t) = 10^{34} \text{cm}^{-2} \text{s}^{-1}$), the LHC bunch crossing frequency is 40 MHz and for 25 ns spaced beams this corresponds to a rate of 10^9 events per second from proton proton collisions. The interaction rate is orders of magnitude higher than the maximum rate (100 Hz) that the data acquisition system can handle. Furthermore, the majority of collisions are not of interest. This leads to implement a trigger system that preselects events online and decides if the corresponding event information has to be kept or discarded. The trigger system has to be very efficient for the interesting events (the signal events), typically several orders of magnitudes less frequent than the background events.

The CMS trigger provides this rate reduction in two steps: the Level-1 Trigger (L1) [91], that reduces the rate up to 100 KHz, and the High Level Trigger (HLT) [92] designed to reduce the rate up to $O(10^2)$ Hz. The L1 trigger is hardware based and uses custom electronics to implement in hardware the events selection using information from the calorimeters and muon systems only [91]. The HLT one is a processors farm several thousand CPU cores. Events accepted by the L1 trigger are read out to the high-level trigger using the complete detector information, including input from the tracker.

L1 Trigger The design latency for L1 trigger decision of $3.2 \mu\text{s}$ makes impossible to use the entire detector output at the L1 and imposes serious restrictions on the complexity of the trigger algorithms. At this stage, only information from the calorimeters and the muon system is available and can be used to implement the L1 selections. Due to a such short latency, the event processing is pipelined at each sub-detector in front-end electronics.

L1 is composed by local, regional and global components (Fig. 2.13). During the local step, trigger primitives for each sub-detector are generated. The calorimeter trigger primitive combines energies deposited in ECAL - HCAL towers (5 crystals in ECAL correspond to the granularity of HCAL) to obtain the trigger tower. The muon trigger primitives are composed by track segments, for the CSC and ϕ -projection in the DT, or hit patterns, for the RPC and η -projection in the DT. The local information is then combined by the Regional Triggers, where calorimeter electron/photon candidates are defined and combined DT/CSC tracks are assigned with physical parameters (like for instance p_T). In the next step, Global Calorimeter Trigger (GCT) and Global Muon Trigger (GMT) determines L1 trigger objects based on the combined sub-detectors information. The GCT trigger objects include jets and τ -jets, the missing transverse energies, isolated and non-isolated e/γ candidates. Then, at the final step, the L1 Global Trigger (GT) decides to accept or to reject an event based on GCT and GMT trigger objects. The decision to accept is taken if the event satisfies all the requirements of at least one of the GT algorithms. The GT may execute in parallel up to ≈ 130 algorithms, including both basic algorithms, which consist on some simple p_T or E_T thresholds to a single object, and complex algorithms based on topological

selections. The result of each algorithm is represented by one bit, which indicates if an event passed the algorithm requirements or not. If L1 accept decision is made, the entire detector information is readout and is passed to the event builder network (EB), to generate a global event and then it is transferred to the HLT. The sustainable rate of the EB is 1 Tb/s.

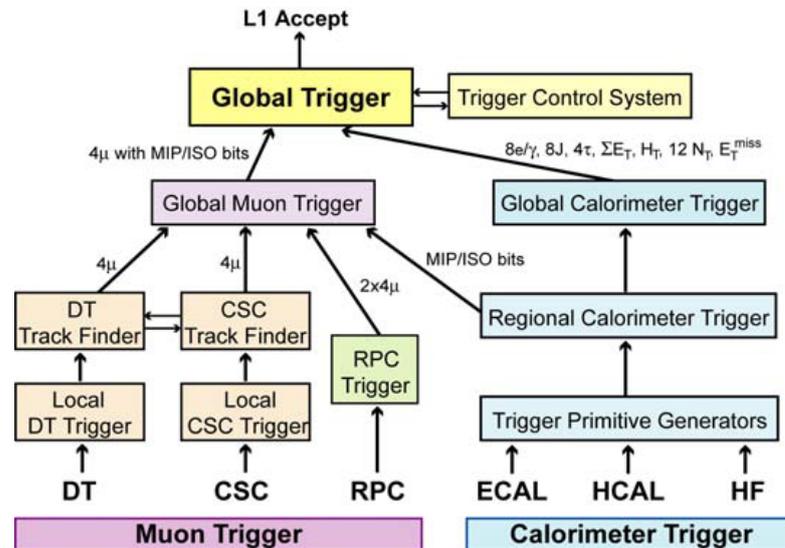


Figure 2.13: Functional block diagram of the CMS L1 trigger [5].

High Level Trigger The HLT is implemented in software. The CMS software (CMSSW) used for HLT and offline reconstruction is open-source and its source code is available online [93]. This software is executed in the Event Filter Farm (EVF) composed by a farm of several thousand CPU cores processing the complete detector information. The HLT event selection algorithms are similar to the ones of the offline reconstruction. Its architecture is designed to reduce the average CPU time needed to process an event, by rejecting events that will not pass the acceptance criteria in the earliest possible stages. For that reason, the sequence of HLT algorithms runs starting from the less CPU consuming steps.

The HLT uses “virtual levels”, where the information from each subdetector is used and faster *off-line* algorithms are applied using at each level a more refined and complete information. The virtual levels are: Level 1, where the event has fired a particular L1 path or a combination of L1 paths; Level 2, where only the information of calorimeters and muon detectors are considered in order to identify the objects; Level 2.5, here the information from the pixel sensor detector is added in order to reconstruct the tracks and the primary vertex; Level 3, where the full track reconstruction is available and the rate is reduced requiring the presence of physical objects (jets, electrons, muons, taus, ...), reconstructed using the Particle Flow algorithm, described in the next chapter.

The HLT is completely summarized in a *trigger path*, corresponding to a set of requirements that an event has to fulfill at L1 and at HLT to be kept. The CMS

trigger system implements a big amount of HLT trigger paths, as described in [92]. Since information from all sub-detectors is available at the HLT level, by combining all the measurements of the various subsystems it is possible to efficiently record, at the same time, events characterized by different signatures. Indeed, the data collected by CMS triggers can be used to study the properties of the Higgs boson, top quark, weak bosons, as well as the decays of b and c hadrons, τ leptons, and more generally SM processes, or to search evidence of physics beyond the SM. In addition to various physics requirements forming a trigger path, the CMS trigger system is able to randomly discard events. This feature of the trigger is called prescaling: to not saturate the memory, when a trigger path has a lower threshold than the expected, in order to face the event rate, a prescale factor is applied to the trigger; this is done in order to collect data samples of these trigger paths.

The data analyzed in this thesis were collected during Run I, with a bunch spacing of 50 ns corresponding to a bunch crossing frequency of 20 MHz. The trigger paths used in this analysis require two isolated central high p_T hadronic τ candidates, are not prescaled and will be briefly described in section 4.3.

2.2.8 Computing system

Events accepted by the HLT have to be stored and reprocessed to be easily used. Real data and simulated samples are stored in a structure called World LHC Computing Grid, made of different computing centres spread all over the world and linked with a high-speed network (Fig. 2.14) [94].

The fundamental concept is the *event*, that refers both to the signals produced by p-p collisions and detected in CMS and to the simulated events. In CMS there are three formats: RAW, RECO and AOD. The RAW format is referred to data collected by the CMS detector and that have fired the trigger (average size of 1.5 MB/event). The RECO data are instead data on which reconstruction and identification algorithms are applied in order to promote them to high-level physical objects (average size of 0.5 MB/event). The last step are the AOD data, which are filtered RECO data in order to obtain more accesible objects (average size of 100 KB/event). CMS uses also a format called PAT (Physics Analysis Toolkit), in which it can be chosen what to store from AOD or RECO data in the PAT [95].

For MC simulations the corresponding data are called GEN-SIM-DIGI (average size of 2 MB/event). GEN stays for generation of physical processes, using complex simulation tools, like PYTHIA [96], POWHEG [97] and MADGRAPH [98]. SIM stays for Simulation, where the interaction radiation-matter after the passage of a particle in all CMS subdetectors and their responses are simulated using the software GEANT-4 [99]. Finally DIGI stays for Digitization, that is the simulation of the electronic answer. Further details of a simulation process will be given in the section B.

The world-wide dissemination of data has shown the necessity to have a common

infrastructure of data storage. The data coming from HLT and from on-line CMS acquisition (RAW), are stored in Tier-0 (T0), placed at CERN, that organizes the collected data in different groups, called *Primary Dataset*, according to the trigger path fired during the acquisition. The *Primary Dataset* used in this work requires τ leptons candidates and will be described in section 4.2. The T0 converts the RAW data in RECO and AOD (Analysis Object Data). These are then transferred to the seven computing centers Tier-1 (T1). The T1 reconstructs, calibrates and skims the RAW DATA and stores the MC samples produced in the Tier-2 (T2) and diffuses them. The T2 are user accessible computing centres. They are 160 in total and are used for the MC production and for data analyses.

Besides the common infrastructure in CMS, it is also used a common software framework, called *CMSSW*, an object-oriented framework written in C++ and Python, fully customizable for any computer architecture and it is constantly developed in order to keep high-performances and flexibility [93].

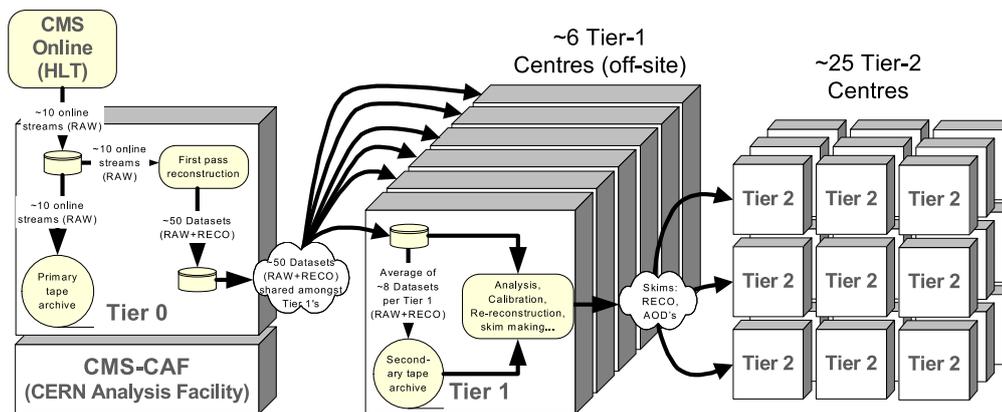


Figure 2.14: CMS Data Flow Scheme [94].

Chapter 3

Event reconstruction

In order to analyze the signals collected by the CMS subdetectors in the proton–proton collisions, a suitable reconstruction must be performed. This chapter describes the algorithms used to reconstruct the high level detector objects (such as tracks, calorimeter clusters and vertices) and then how from these the physical objects (muons, electrons, taus, jets, missing energy. . .) used by the analysis are reconstructed.

3.1 Reconstruction at detector level

The detector signals are used to reconstruct in the event the high level objects, which include the *tracks*, the trajectories of the charged particles identified as sequences of *hits* in the tracking detectors, and the collections of towers in which energy is deposited from particles, the *calorimeter clusters*. The reconstruction of tracks allows to identify and reconstruct the vertices of the proton–proton multiple interactions and to identify the hard collision vertex among them, the *primary vertex*.

In the following sections a short description of the algorithms used to reconstruct the high level objects is given.

3.1.1 Track reconstruction

Track reconstruction is a complex algorithm to estimate the trajectory and the momentum of the charged particles passing through the inner tracking system (section 2.2.2). Inside the tracker region, thanks to the almost uniform magnetic field, the trajectory of a charged particle can be parameterized by a helix. The five parameters used by CMS to determine track trajectory are: d_0 , z_0 , ϕ , $\cot\theta$ and the transverse momentum (p_T) of the track. p_T is defined at the point of

coordinates (x_0, y_0, z_0) , the closest to the nominal beam axis ($z = 0$) point of the track trajectory; θ and ϕ are, respectively, the polar and the azimuthal angles of the track momentum vector at the impact point; d_0 is the impact parameter of the track in the transverse plane and is defined through the coordinates of the impact point, $d_0 = -y_0 \cos \phi + x_0 \sin \phi$.

As explained in the previous chapter, charged particles leave small charge depositions as they pass through the inner tracking system in the pixel and in the strip detectors. The first step of the reconstruction is a *local reconstruction*, where the signals collected by the pixels and strips of the silicon detectors above specified thresholds are grouped together realizing *clusters*. The spacial positions of the charge deposits called *hits* are reconstructed from clusters using specific algorithms for each subdetector [83], with a hit reconstruction efficiency of around 99%.

Using these hits, pattern recognition algorithms can reconstruct the original trajectories of the particle, associating a group of hits to a track and measuring the five parameters of the helix that best match to the observed hits. CMS uses the Combinatorial Track Finder (CTF) [100], an extension of the Kalman Filter (KF) [101] algorithm, to perform the pattern recognition and the track fitting. It is an *iterative tracking* procedure where the first iterations search for tracks with high p_T coming from the primary vertex. Once found, these tracks are removed from the following iterations to reduce the combinatorics and simplify the subsequent iterations with a complex topology (like lower p_T or with displaced vertex).

Each iteration is decomposed in four steps [100]:

- *Track Seed Generation*: in this step track seeds candidates, consisting of two or three hits in the pixel detector, are used to give estimates of the track parameters. The seeds are constructed from the inner part of the tracker (pixel detector) and the tracks are built outwards, in order to keep high reconstruction efficiency and to facilitate the reconstruction of low-momentum tracks.
- *Track finding*: the track finding is based on the Kalman filter method. It starts from the trajectory seed and adds the hits of the outer layers updating the track parameters and their uncertainties in each layer, as well as the material crossed. This procedure is repeated until the final tracker layer is reached. Only a limited number of track candidates are accepted on the basis of number of hits and minimum χ^2 criteria. For these tracks a search in the inward direction is then performed to find new hits, and using them to update the trajectory parameters.

To avoid duplicates of the real tracks, an algorithm, the trajectory cleaner, is applied: it keeps, among the tracks sharing more than the 19% of hits, the track with the highest number of hits. If both tracks have the same number of hits, the track with the lowest χ^2 value is retained.

- *Final Track Fitting and Smoothing:* the collection of hits provided during the track-finding step is refitted to remove any bias and complete the information about the trajectory. The fit is performed using a Kalman filter (filtering) and a Runge–Kutta propagator [102] (smoothing) to obtain optimal track parameters by taking into account material effects and accommodating possible inhomogeneities in the magnetic field. Starting from the innermost hits of the track outwards and iteratively, the track trajectory is sequentially updated with new valid hits, as well as its position uncertainty. To extrapolate the trajectory from one hit to the next and to improve the precision, the Runge–Kutta [102] method is used. Then hits incorrectly associated to the track, *outliers*, are removed by applying requirements on the distance between track candidate and hit or the probability that a pixel hit is consistent with the track. This procedure is repeated many times until no more outliers are found, filtering and smoothing the track every time.
- *Track selection:* The first three steps create a significant amount of *fake* tracks, which are not associated with a charged particle. To remove fake tracks, tracks are selected by applying requirements based on the number of layers with hits associated to the track, the track quality (χ^2 of the final fit) and on the compatibility that the track is originating by one of the reconstructed vertices, primary plus pile-up (section 3.1.2).

Performances The track reconstruction performance can be expressed in term of tracking efficiency (the fraction of simulated charged particles that can be associated with corresponding reconstructed tracks), fake rate (fraction of reconstructed tracks that are not associated with any simulated particle) and track parameters resolution. The performances have been evaluated using $t\bar{t}$ simulated events and real data corresponding to 2011 LHC data-taking period.

The average track reconstruction efficiency for promptly-produced charged particles with transverse momenta of $p_T > 0.9$ GeV is 94% for $|\eta| < 0.9$ and 85% for $0.9 < |\eta| < 2.5$. The inefficiency is caused mainly by nuclear interactions produced by the hadrons in the tracker material. For isolated muons with $1 < p_T < 100$ GeV, the tracking efficiency is $> 99\%$ over the full tracker acceptance, and does not depend on p_T (Fig. 3.1a), resulting in a negligible fake rate.

In Fig. 3.1c it is reported the measured tracking efficiency both in real data and simulation, showing an efficiency $> 99\%$. The particular behaviour is caused by inactive modules and the misalignment of the tracker, and it is well reproduced by simulation.

The resolution of the track parameters is estimated calculating the track residuals defined as the difference between the reconstructed and the true (from MC simulation) track parameters. In Fig. 3.2 the resolutions of transverse impact parameter (Fig. 3.2a), of the longitudinal impact parameter (Fig. 3.2b) and of the transverse momentum (Fig. 3.2c) are reported. In every bin, the resolution is defined as the half-width of the interval in the distribution of track residuals. For an isolated

muons of $p_T = 100$ GeV produced with $|\eta| < 1.4$, the resolutions are approximately 2.8% in p_T , and respectively, 10 μm and 30 μm in the transverse (d_0) and longitudinal impact (z_0) parameters.

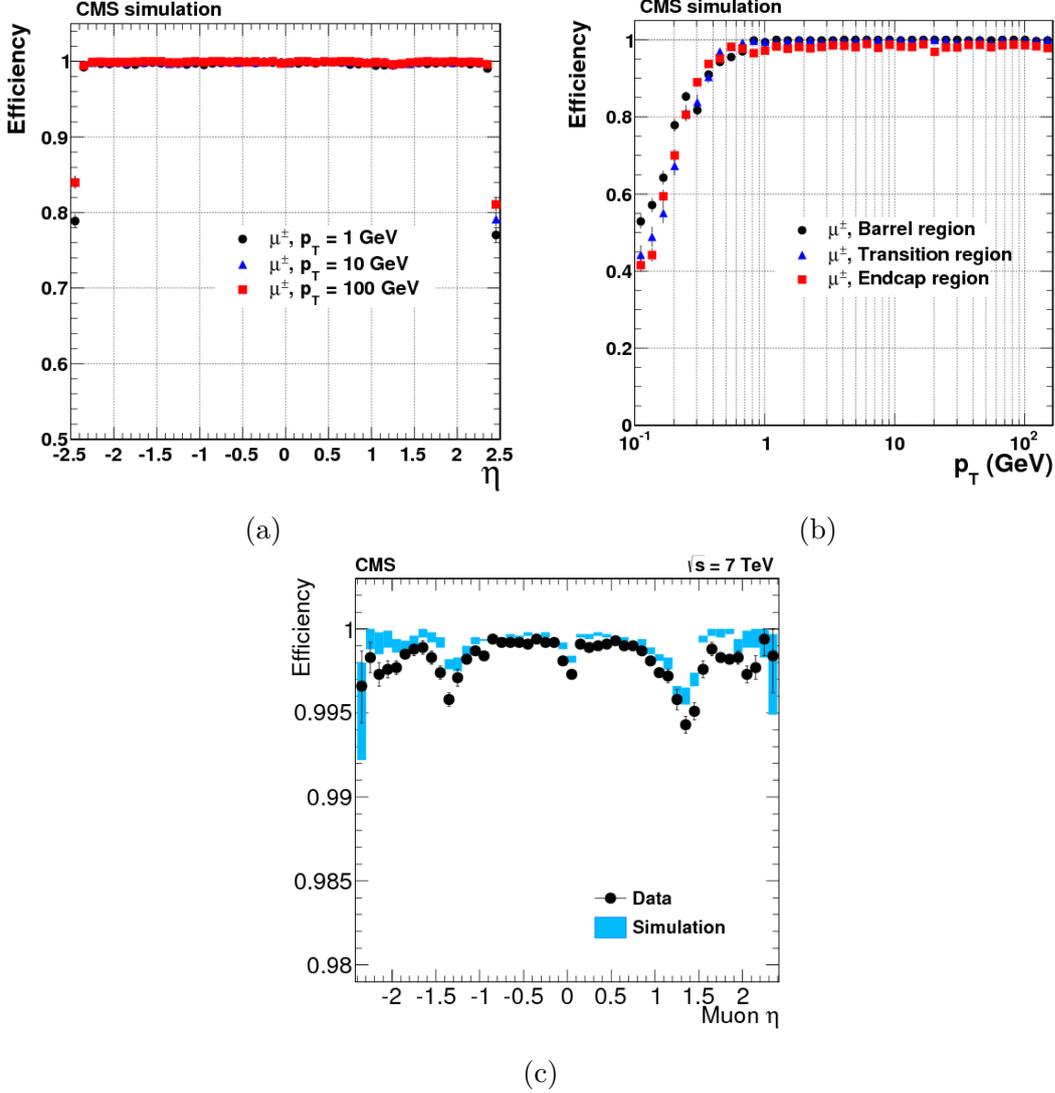


Figure 3.1: Track reconstruction *efficiencies* for *single, isolated muons* satisfying the *high-purity* quality requirements. Results are shown as a function of η (a) for $p_T = 1, 10$, and 100 GeV. They are also shown as a function of p_T (b), for η intervals of 0–0.9 (barrel), 0.9–1.4 (transition) and 1.4–2.5 (endcap), respectively. In (c) the tracking efficiency is reported measured with a tag-and-probe technique [103], for muons from Z decays, as a function of the muon η for data (black dots) and simulation (blue bands) [83].

3.1.2 Primary vertex

For the data recorded by CMS in 2012 the number of proton-proton interactions (number of reconstructed vertices) per bunch crossing were reported in figure 2.3 and its average value is 21. The probability that more than one proton-proton

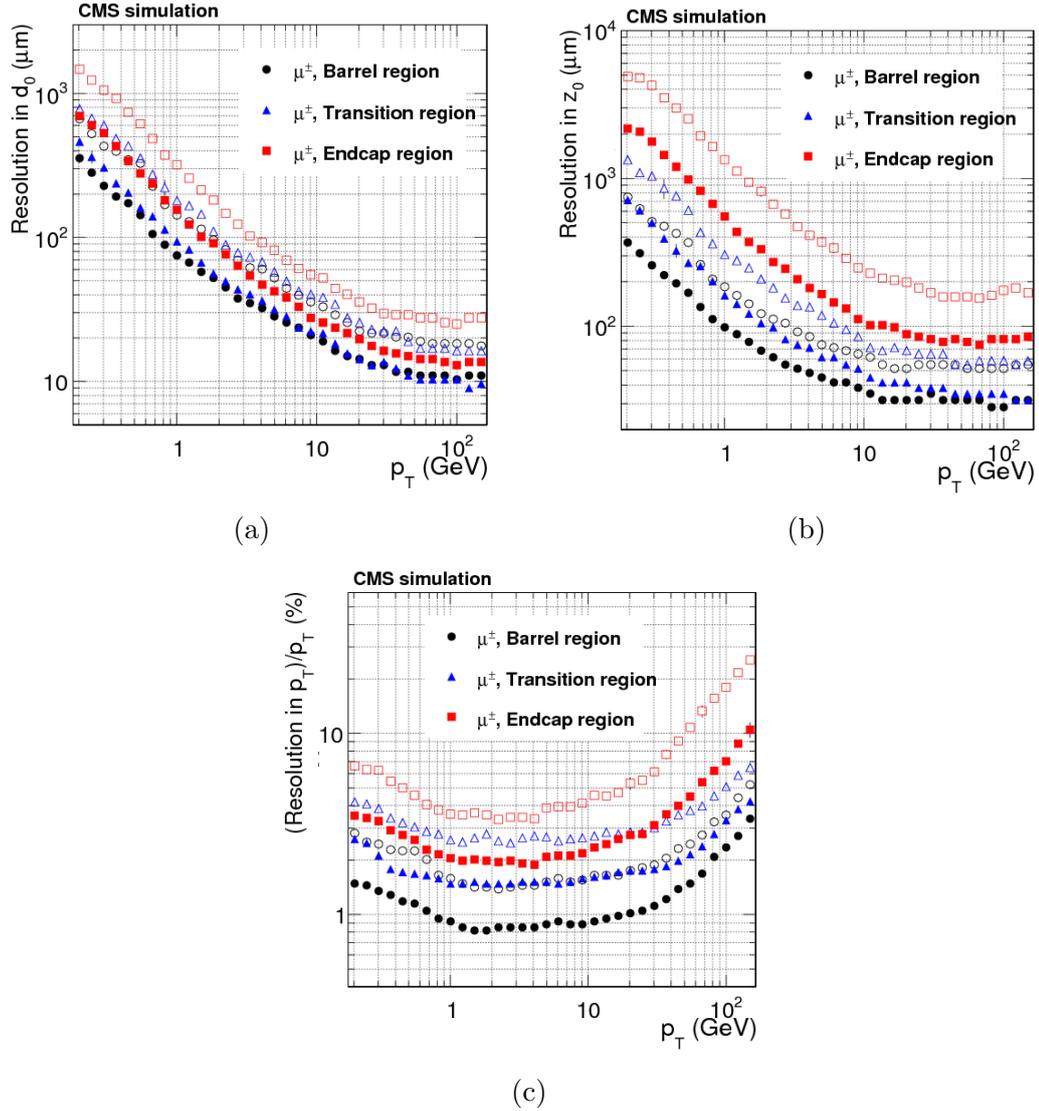


Figure 3.2: Resolutions as a function of p_T for single, isolated muons for η intervals of 0–0.9 (barrel), 0.9–1.4 (transition) and 1.4–2.5 (endcap), respectively. From left to right: transverse (a) and longitudinal (b) impact parameters, and p_T (c). For each bin in p_T , the solid (open) symbols correspond to the half-width for 68% (90%) intervals centered on the mode of the distribution in residuals [83].

interaction could produce an interesting process is usually considered to be negligible. On the other hand, the precise reconstruction of the positions of all p–p interactions in an event allows to separate the primary hard (high p_T) interaction products and the objects originated by pile–up collisions. Furthermore, the accurate location of the primary vertex, in combination with other information, e.g. the secondary vertex position, plays an important role in the determination of physical objects properties, like for instance b jets. The location and the uncertainty of the vertices in a p–p interaction are obtained using the reconstructed tracks [83]. This process is realized in three steps:

1. Tracks are preselected to be consistent with the primary interaction region, the so called *beam spot*, which is the centre of a three–dimensional profile of the luminous region where the collisions occur.
2. The selected tracks are split into several clusters based on a possible common vertex of origin, using the *Deterministic Annealing* (DA) algorithm [104], which uses as input the z coordinate of the tracks at the point of the closest approach to the beam spot.

3. For each cluster, which contains at least two tracks, the *Adaptive Vertex Fitter* [105] is used to determine the best fit of the three–dimensional vertex position as well as the fit quality. In the fit to each track is assigned a probability (p_i) for originating at this vertex. These probabilities are used to determine a quantity related to the quality of the fit, the number of degrees of freedom for the fit (n_{dof}), defined as: $n_{dof} = -3 + 2 \sum_{i=1}^{\#tracks} p_i$. At the analysis level further requirements will be applied in order to improve the selection of the primary hard interaction vertex: $n_{dof} > 4$; the distance in the z direction from the vertex to the nominal interaction point must be smaller than 24 cm; the corresponding distance in the transverse plane ($d = \sqrt{x^2 + y^2}$) must be smaller than 2 cm.

Due to the presence of events of pile–up, it is considered as primary hard interaction vertex, the collision vertex which maximizes the sum of the squared p_T of all tracks associated with each vertex, $\sum p_T^2$. While the other vertices are considered pile–up vertices.

The resolution of a reconstructed primary vertex strongly depends on the number of tracks considered in the fit and on their p_T . The primary vertex resolution in x and z are represented in Fig. 3.3, respectively Fig. 3.3a and 3.3b, for minimum–bias¹ and jets (defined in section 3.2.5) enriched events. The resolution in y is almost identical to that in x, and is therefore omitted. The resolution on the vertex position is better in jet enriched events, reaching in both x and z dimensions a resolution between 10–20 μm for 15 tracks. In Fig. 3.3c the primary vertex reconstruction efficiency for data and MC samples is shown: the primary

¹Minimum–bias events are the most frequent QCD events produced in the inelastic p–p collisions. They are characterized by the presence of low p_T jets (soft jets).

vertex reconstruction efficiency also depends on the number of tracks, showing an efficiency greater than 99% with more than three tracks [83].

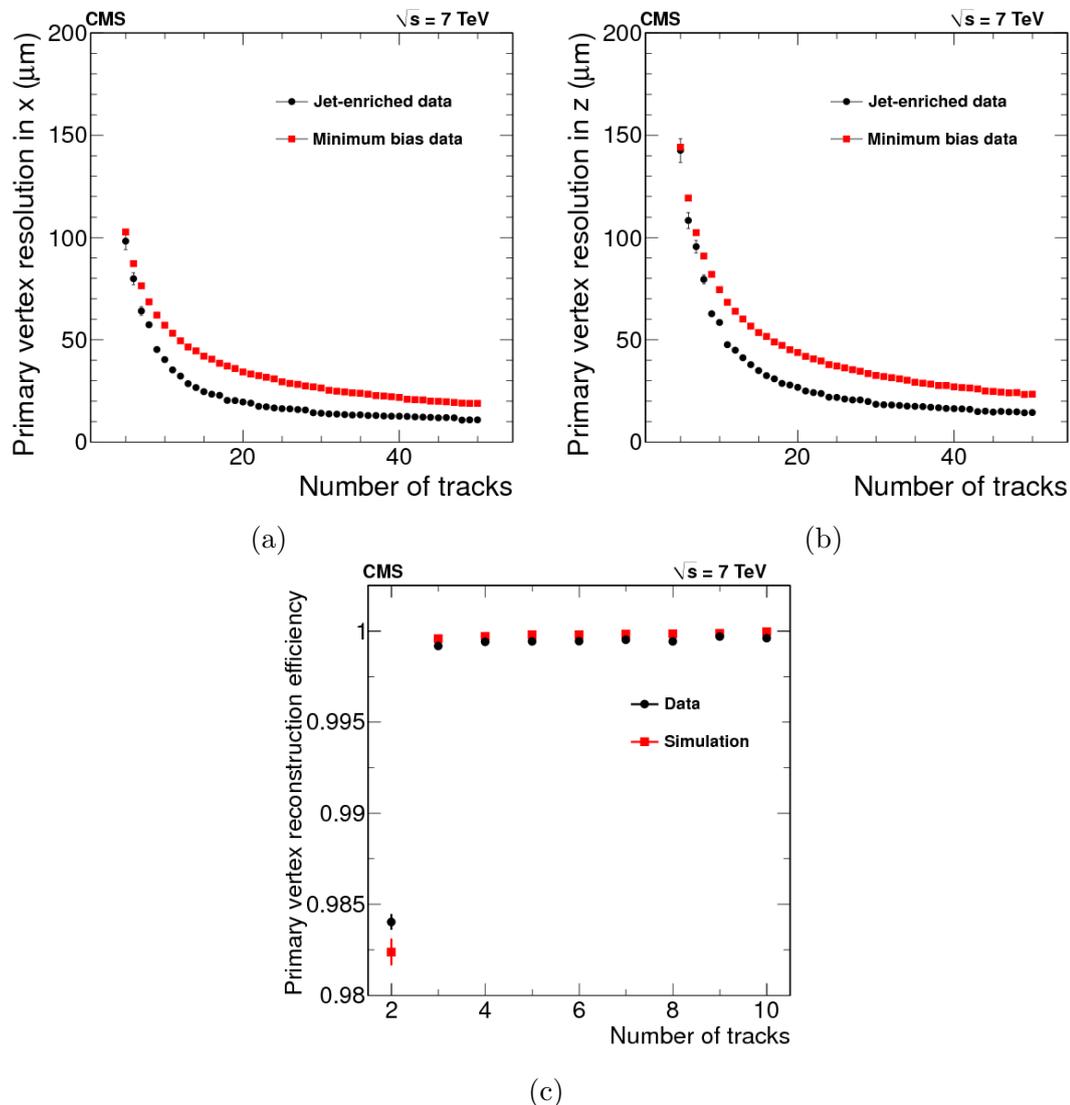


Figure 3.3: Primary-vertex resolution in x (a) and in z (b) as a function of the number of tracks at the fitted vertex, for two kinds of events. Primary vertex reconstruction efficiency (c) as a function of the number of tracks in a cluster, measured in minimum-bias data and in MC simulation [83].

3.1.3 Energy Reconstruction in the Calorimeter

The energy of a particle is collected by ECAL and HCAL systems as described in sections 2.2.3 and 2.2.4 respectively. In ECAL the energy of electron and photon showers is collected in several crystals. The 97% of energy of a single electron is collected by 5×5 crystals [95]. This configuration results to give very good performances. The not complete containment is caused by the presence of material in front of the CMS calorimeter resulting in electron bremsstrahlung, photon conversions, and spread of the energy along ϕ because of the high magnetic field.

In order to collect the entire deposited energy by showers, *clustering algorithms* are used, where the energy of nearby ECAL cells above a threshold, around a cell with a maximum energy deposit, are summed together, creating a *cluster*. To recover the spread of energy along ϕ the algorithms build cluster of clusters, called *supercluster* (SC). A detailed description of superclustering methods will be given in the description of electron reconstruction (section 3.2.2). The SC ECAL is summed to the energy deposited in individual HCAL cells, that are arranged in a *tower* pattern in $\eta - \phi$ space, projected to the nominal interaction point and with the same layout of the crystals. The energy associated with a tower is calculated as the sum of all contributing readout cells which pass specific requirements. There are in total 4176 towers and they are used as input to several jet clustering algorithms, where the input towers are treated as massless particles, with the energy given by the tower energy, and the direction is defined by the unit vector pointing from the nominal interaction point to the center of the tower itself.

3.2 Physical Objects

The basic objects described in the previous sections are the necessary elements to build more complex objects, with the goal of representing a particle, like an electron, a muon, a jet etc. In the following sections the algorithms used to link the tracks and the calorimetric information and to reconstruct high-level objects will be described.

3.2.1 Particle-Flow Algorithm

The CMS Particle-Flow (PF) algorithm [106–108] uses information from all the CMS subdetectors to reconstruct individual stable particles, like muons, electrons, photons, charged hadrons and neutral hadrons, in a crowded particles environment. The particles reconstructed by the PF are then used to reconstruct jets, b jets, hadronic τ decays, the missing transverse energy and to calculate the isolation of the physical objects (described respectively in sections 3.2.5, 3.2.6, 3.2.8, 3.2.7 and 3.2.4).

A stable particle traversing the detector usually produces a charged track in the tracker if it is charged, a muon track in the muon system if it is a muon, and calorimeter clusters in ECAL and in HCAL, respectively in the case of an electron or photon and a neutral or charged hadron. The PF algorithm combines together the so called *building elements*, charged particle tracks reconstructed by the inner tracker, calorimeter clusters and muon tracks using a linking algorithm to reconstruct and identify all the stable particles produced in the event.

Building elements Tracks in the tracker are reconstructed with the iterative sequence described in section 3.1.1. The standalone muon track reconstruction [95] in the muon system starts by building track segments in the DTs and CSCs using pattern recognition and linear fit techniques. After that, track segments and

RPC clusters are combined and refitted by a Kalman filter forming *standalone muon* tracks. A more detailed description of the reconstruction of a standalone muon track is reported in section 3.2.3. PF improves the clustering algorithm to reconstruct clusters. In order to have a high detection efficiency for low p_T particles and to split close energy deposits in the calorimeters, a specific clustering algorithm has been designed for the PF algorithm. This cluster algorithm starts from the calorimetric reconstructed clusters, the *cluster seeds*, considering the clusters with a maximum local energy. Then *topological clusters* are built, grouping the neighbours cells with energy above a specific threshold. After collecting all adjacent cells, one topological cluster may contain more than one seed. Therefore, in the last algorithm step, topological clusters are split by sharing the energy deposit of each cell between the PF clusters, whose amount is equal to the number of seeds. The fine calorimeters granularity is the key component that allows to efficiently split topological clusters into *PF clusters*.

Linking Algorithm and PF Blocks Charged particle tracks, calorimeter clusters and muon tracks are linked together into *PF blocks*. Each block links group of elements from two sub-detectors using a criterion to link these elements and, depending on the composition, the block is then interpreted as a particle of a particular type. In linking between tracks and calorimeter clusters, the track trajectory is extrapolated from the outermost tracker layer to points within PS, ECAL and HCAL volumes. If these fall within the boundaries of a cluster then the track and cluster are linked. To incorporate reconstruction uncertainties, gaps between calorimeter cells and cracks between calorimeter modules, cluster envelope of the track can be enlarged by up to the size of one cell. Additional links between charged particle tracks and calorimeter clusters are established to take into account Bremsstrahlung by electrons. In this case, tangents to the tracks are extrapolated to the ECAL from the intersection points between the track and each of the tracker layers, if the extrapolated tangent position is within the boundaries of a cluster, as defined above, then the cluster is linked to the track as a potential Bremsstrahlung photon. A link between the ECAL and HCAL clusters is created if the position of the ECAL cluster, which has higher granularity, is within the HCAL cluster envelope. Similar logic is followed to link the PS and ECAL clusters. A link between inner tracker tracks and muon system tracks is referred as a *global muon*. It is established if the χ^2 of the global fit of the two tracks is less than χ_{max}^2 . If more than one combination of tracks satisfies this criterion, then the combination with the smallest χ^2 is chosen.

Reconstruction and Identification In the final step the PF algorithm reconstructs and identifies a set of particles using the *block* of elements. PF particles are created in several iterations. Blocks assigned to a PF particle are excluded in the subsequent iterations.

A schematic view of the PF algo principle is reported in Fig. 3.4.

The reconstruction and identification of a particle or an ensemble of particles (hadrons) from each block proceed as follows:

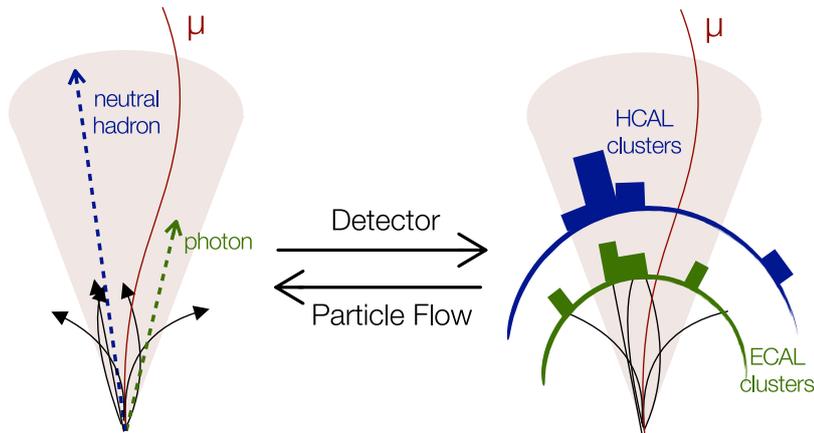


Figure 3.4: Schematic view of the PF algo principle.

- Each global muon is promoted to *PF muon*, if its transverse momentum is compatible with the muon momentum reconstructed using only the inner tracker information within three standard deviations. The corresponding track is then removed from the block. The expected energy loss in the calorimeters is also subtracted from the associated clusters. More details about the CMS muon reconstruction and identification at analysis level will be given in the section 3.2.3.
- *PF electrons* are identified in two steps. In the first step, each track is submitted to a pre-identification step considering the tracker as a pre-shower, since real electrons can produce short tracks when losing energy by Bremsstrahlung in the tracker layers, on their way towards ECAL. The pre-identified electron tracks are then refitted with a Gaussian-Sum Filter (GSF) [109]. In the second step, a multivariate discriminator assesses the compatibility with the linked ECAL clusters. This provides discrimination against charged hadrons. If the candidate passes, a *PF electron* is formed from the GSF track and linked to ECAL clusters, which are also removed from the block. More details about the CMS electron reconstruction and identification will be given in the section 3.2.2.
- Tracks classified neither electrons nor muons result in a charged hadron. The mass assigned is the pion mass and the momentum is from the track fit. If the energy of the linked calorimeter clusters is compatible with this momentum, within uncertainties, the candidate momentum is updated to a weighted average between the track and cluster measurements. However, if the cluster energy exceeds the track momentum by more than the expected resolution, this excess is interpreted as the presence of overlapping neutral particles. If the excess is larger than the total linked ECAL energy, a *PF photon* is created, with any remainder becoming a neutral hadron.
- The ECAL and HCAL clusters not linked to any track are identified as *PF photons* and *PF neutral hadrons* respectively.

3.2.2 Electrons

A distinguishable feature of the electron reconstruction is a significant amount of energy that an electron can emit via Bremsstrahlung before reaching the ECAL when interacting with the material of the tracker. This feature requires a special approach in order to correctly reconstruct the electron inside the tracker. Meanwhile, the photons produced can also convert to e^+e^- pairs before reaching the ECAL. This means that the energy deposited in the ECAL can be spread out in the ϕ direction. Hence, also in this case, dedicated algorithms are required to combine the energy deposits from both the initial electron and the bremsstrahlung products. In this section a detailed description of the algorithms is given, which are used to cluster energy in the electromagnetic calorimeter, to reconstruct electron trajectories in the tracker and to identify electrons [110].

Reconstruction

In order to achieve optimal performance, the electron reconstruction uses standalone approach combined with the PF algorithm [111]. The latter has been described in section 3.2.1. In the standalone approach, to recover all the energy of the photons emitted through bremsstrahlung, two *superclustering algorithms* (mentioned in section 3.1.3) are used: the “hybrid” algorithm, used in the ECAL barrel (Fig. 3.5), and the “multi- 5×5 ” algorithm, for the ECAL endcaps [111].

Hybrid algorithm collects the energy of the shower in a small window along η and an extended one in ϕ in the EB. The algorithm starts from crystals with energy deposit above E_T^{min2} and considers as a *seed* the crystal with the highest energy deposit. Then structures of 5×1 crystals in $\eta \times \phi$ are added in both directions of ϕ around the seed crystal, if they have an energy greater than a specific array threshold in order to create a final global cluster, called *supercluster* (SC).

Multi- 5×5 algorithm is used in the EE, where crystals are not arranged in an $\eta \times \phi$ geometry. The seed crystal is chosen as the crystal which exceeds a minimum energy threshold and having the highest energy deposit among its neighbours. Other 5×5 crystals around it are added into a *supercluster* by adding to them 5×5 neighboring crystal arrays with possible overlaps and, if the total energy of the nearby clusters is above the threshold, the clusters are grouped into the SC. The SC energy is calculated summing the energies of all the clusters belonging to it, exceeding a minimum pre-defined energy.

The position of the SC resulting from the different algorithms is the energy-weighted mean of the cluster positions. This represents the impact point in the ECAL of a non-radiating electron with energy equal to the supercluster energy.

The radiative losses for electrons in the tracker material forbid the reconstruc-

² E_T is the transverse energy, defined as $E_T = E \cdot \sin\theta$, where E is the deposit of energy in the calorimeter cell and θ is the polar angle position of the cell.

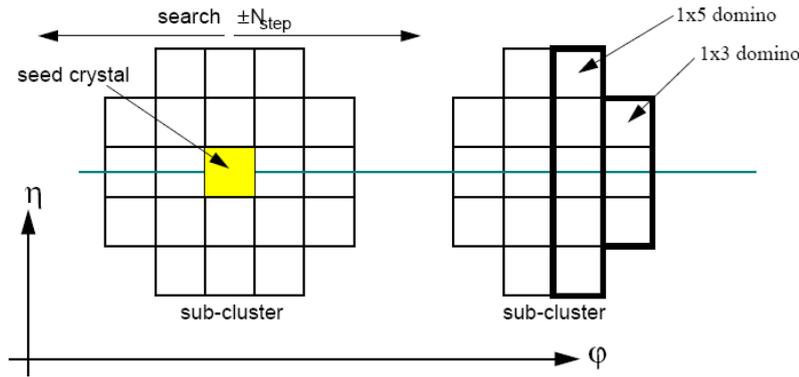


Figure 3.5: Electron clustering Hybrid algorithm [112].

tion of *electron tracks* with Kalman filter [101] procedure due to the huge time consuming. A specific tracking technique is used to reconstruct *electron tracks*. The technique starts from a seeding procedure in order to find two or three hits in the tracker. Two algorithms are used and combined together to improve the reconstruction efficiency and reduce the misidentification probability: the first starts from a SC in ECAL and tries to estimate the trajectory in the first layers of the tracker, the second starts from reconstructed tracks that are extrapolated towards ECAL and matched to an SC, both combining the information with a candidate vertex. For ECAL seeding the tracker hits are compared with SC hits within windows in ϕ and z . For tracker seeding the KF algorithm is used when the bremsstrahlung effects can be neglected, because it can't follow the curvature of the trajectory caused by the bremsstrahlung, and here the KF track is matched with a PF cluster in ECAL respecting specific requirements [111]. In this case the GSF [109] is used approximating the energy loss to a mixture of Gaussian distributions and building tracks up to ECAL³. For each GSF track several PF clusters can be grouped together, in order to collect the electron arriving at the ECAL surface and the radiated photons (see section 3.2.1). The PF cluster corresponding to the electron at the ECAL surface is the one matched to the track at the exit of the tracker. Most of the bremsstrahlung photons are recovered extrapolating for each tracker layer (where most of the material is concentrated) a straight line to the ECAL, tangent to the electron track, adding each matching PF cluster to the electron PF cluster. Converted photons missed by this procedure can be recovered by a dedicated MVA algorithm which uses displaced KF tracks to kinematically associate them with the PF clusters. In addition, for ECAL-seeded isolated electrons, any PF clusters matched geometrically with the hybrid or multi- 5×5 SC is also added to the PF electron cluster.

The reconstruction efficiency of the combined standalone and PF approaches, is $> 90\%$ for electrons from Z/γ data samples and simulated events, for a transverse energy of the electron (E_T^{SC}) greater than 10 GeV, as shown in Fig. 3.6 [111].

³The fraction of energy lost through bremsstrahlung is estimated using the momentum at the point of closest approach to the beam spot, p_{in} , and the momentum extrapolated to the surface of the ECAL from the track at the exit of the tracker, p_{out} , and is defined as $f_{brem} = [p_{in} - p_{out}] / p_{in}$.

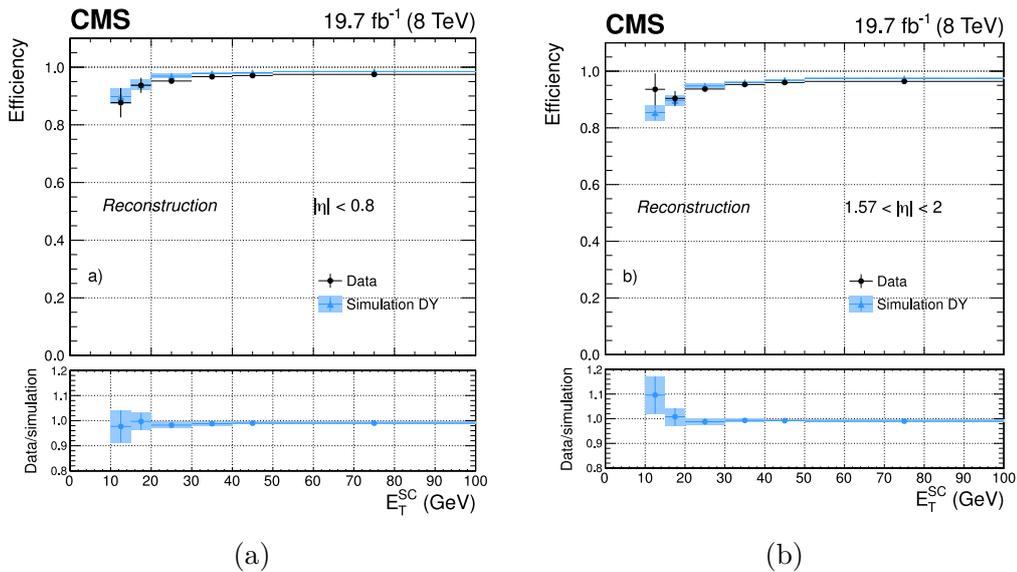


Figure 3.6: Electron reconstruction efficiency measured in dielectron events in data (dots) and Drell Yan simulation (triangles), as a function of the electron E_T^{SC} for $|\eta| < 0.8$ (a), and for $1.57 < |\eta| < 2$ (b). The bottom panels show the corresponding data-to-simulation scale factors [111].

Identification

In order to separate a good electron from jets and to increase the identification sensitivity, the identification of the electrons is based on a multivariate Boosted Decision Tree (BDT) discriminator [113]. The discriminator uses a set of most discriminating variables exploiting the track–cluster matching, the associated SC substructure and its shape, the kinematic observables as well as the fraction of energy loss through bremsstrahlung (f_{brem}) [111].

It was trained in two bins of p_T and three bins of η , for genuine electrons from $Z/\gamma^* \rightarrow ee$, using data and simulated events, and for misidentified electrons reconstructed in $W + \text{jets}$ events in data.

The BDT output is a real value in the interval $[-1;1]$. True electrons have tendency to obtain higher output values, while the fake electrons obtain lower output values with a bigger probability. The comparison of the BDT performance for electrons from Z decays and for misidentified electrons allowed to determine two working points (WP) that have been used in this thesis for the BDT–based electron identification, summarized in Table 3.1. Each WP defines acceptance threshold for the BDT output as function of p_T and η of an electron.

3.2.3 Muons

Muons with $p_T > 5$ GeV typically leave hits in the inner tracker then travel through the CMS calorimeters with minimal energy deposit and leave hits in the muon chambers. Thus muons can be reconstructed like tracks in two independent

Working Point	Electron p_T	BDT output threshold		
		$ \eta < 0.8$	$0.8 \leq \eta < 1.479$	$1.479 \leq \eta $
Loose ID	$p_T \leq 20$ GeV	0.925	0.915	0.965
	$p_T > 20$ GeV	0.905	0.955	0.975
Tight ID	$p_T > 20$ GeV	0.925	0.975	0.985

Table 3.1: Definition of *loose* and *tight* working points for the BDT electron identification as a function of electron p_T and η . To pass the WP identification criterion, BDT output should be greater than the threshold value.

places - the inner tracker and the muon chambers. This greatly improves the ability to isolate muons from hadronic activity.

Initially, muon tracks are reconstructed independently in the inner tracker, *tracker muon*, and in the muon system, *standalone muon* [95, 114].

Based on these objects, three reconstruction approaches are used:

- **Standalone muon reconstruction** uses only muon chambers taking the track segments from local reconstruction and from the three muon chambers. Starting from the segments of innermost chambers the muon trajectories are built using a Kalman-filter technique [101]. The Kalman filter is then applied backward from the outermost surfaces; the track parameters are defined in the innermost layers and the track is extrapolated to the nominal interaction point, setting a constraint to the vertex.
- **Tracker muon reconstruction** starts from the tracks reconstructed in the inner tracker, considering all the tracks with $p_T > 0.5$ GeV and $|\vec{p}| > 2.5$ GeV, and extrapolates them to the muon system. If an extrapolated track position matches within the uncertainties to at least one muon segment, the corresponding track is considered as a *tracker muon* track. The muon momentum resolution can be greatly improved by utilizing hits in both the muon system and the tracker using the Global Muon reconstruction.
- **Global Muon reconstruction** starts with a standalone muon track and searches for a compatible tracker track. A global-muon track is fitted combining hits from the tracker track and standalone-muon track, using a Kalman filter technique to take into account the expected energy loss within the magnet and support structure; in the fit, the muon track is constrained to originate from the beamspot.

The Global Muon reconstruction provides better momentum resolution for the tracks with $p_T > 200$ GeV, while the Tracker Muon reconstruction is more efficient for the low momentum tracks with $p_T < 5$ GeV.

Identification

In these thesis two muon identification techniques are used both using the muon reconstructed as a Global Muon: the PF Muon and the Tight Muon [106,114].

- *Particle-Flow Muon Selection*: the Particle-Flow algorithm identifies the muon as described in section 3.2.1. A specific selection is applied to the muon reconstructed as Global and Tracker Muons. The selection is optimized in order to identify muons within jets with high efficiency, while maintaining a low rate for the misidentification of charged hadrons as muons.
- *Tight Muon Selection*: the candidate is required to be reconstructed as a Global Muon and a PF Muon, with a χ^2/ndof of the global track fit smaller than 10, with at least one muon chamber hit; the tracker track must be matched in at least two muon stations, have at least one hit in the pixel detector and hits in at least 5 tracking layers, with a transverse impact parameter $|d_{xy}| < 2$ mm with respect to the primary vertex (Table 3.2).

Working Point	<i>Tight Muon</i>
Algorithm	Global & PF Muon
Strip hits	> 5
Pixel hits	≥ 1
Muon chambers with segments	≥ 2
Global fit	$\chi_{dof}^2 < 10$
Hits in muon system	≥ 1
Transverse impact parameter	$ d_{xy} < 2$ mm

Table 3.2: Tight Muon selection.

In Fig. 3.7 the muon reconstruction and identification efficiencies are reported respectively for the PF muon, (a) and (b), and for the Tight Muon selections, (c) and (d), showing a very good agreement between data and MC simulation; further details can be found in [114].

3.2.4 Electron and Muon Isolation

In this thesis we are interested in electrons and muons *isolated*, which means that there is a small activity of particles around the lepton direction in the tracker and in the calorimeter. The variable used to define the lepton isolation, I_{rel} , is a function of sums of p_{T} of PF charged candidates, $p_{\text{T}}^{\text{charged}}$, PF neutral hadrons, $p_{\text{T}}^{\text{neutral-had}}$, PF photons, p_{T}^{γ} , and pile-up corrections, $\Delta\beta$ [111,114]:

$$I_{\text{rel}} = \frac{\sum p_{\text{T}}^{\text{charged}} + \max\left(\sum p_{\text{T}}^{\text{neutral-had}} + \sum p_{\text{T}}^{\gamma} - \Delta\beta, 0\right)}{p_{\text{T}}^l}, \quad (3.1)$$

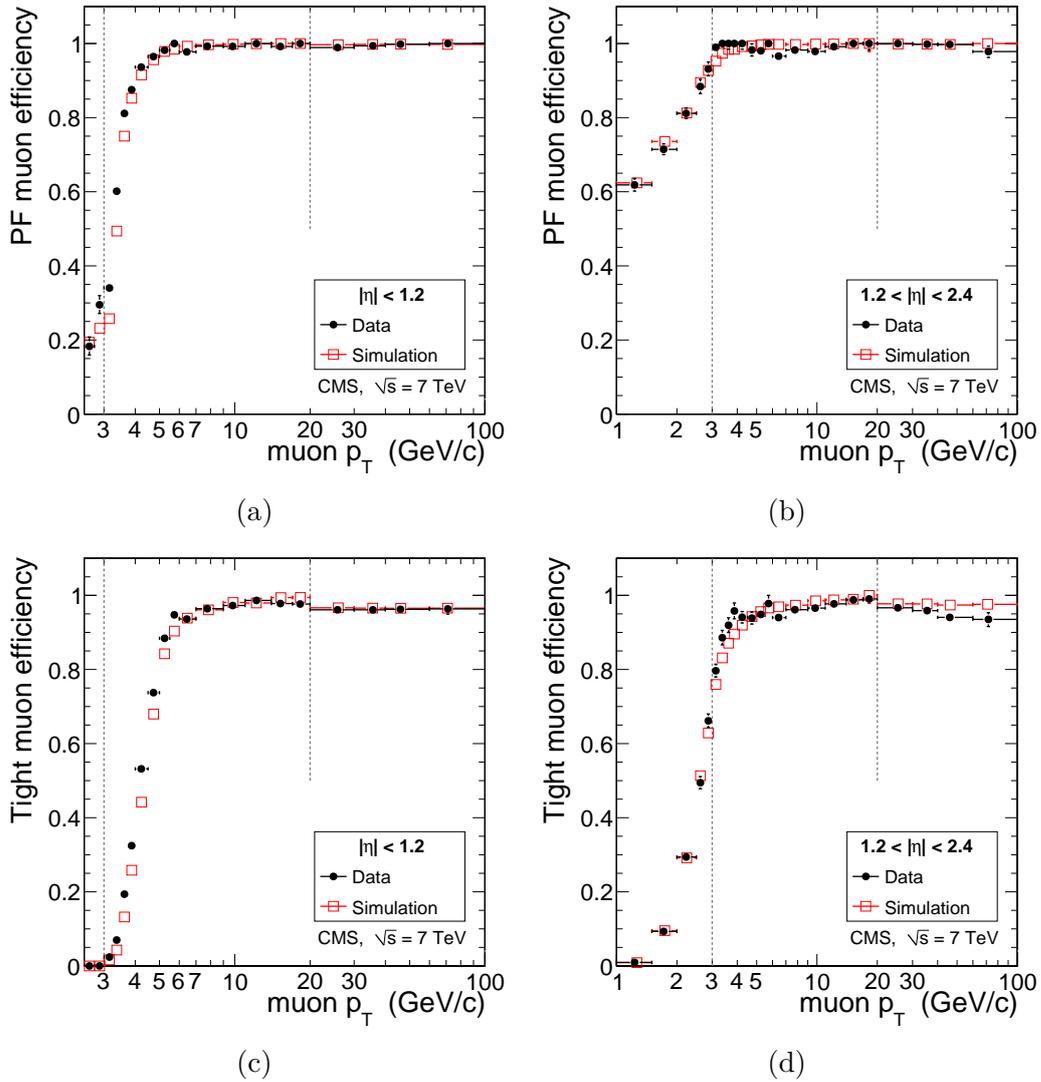


Figure 3.7: Muon reconstruction and identification efficiency measured by tag-and-probe method in $J/\psi \rightarrow \mu^+\mu^-$ events for $p_T < 20$ GeV and in $Z \rightarrow \mu^+\mu^-$ events for $p_T > 20$ GeV. The plots show the efficiency as a function of p_T in the barrel and endcap regions for the PF muon selection (a,b) and the tight muon selection (c,d) [114].

Particle type	veto-cone	P_T
Electrons		
Charged particles	0.01 (EB) / 0.015 (EE)	–
Photons	0.08 (EB, EE)	–
Neutral hadrons	–	–
$\Delta\beta$	–	–
Muons		
Charged particles	0.0001	–
Photons	0.01	> 0.5 GeV
Neutral hadrons	0.01	> 0.5 GeV
$\Delta\beta$	0.01	> 0.5 GeV

Table 3.3: Size of “veto cones” and p_T thresholds used to evaluate the isolation of electrons and muons. For each term in eq. 3.1 size cones are different. The symbol “–” indicates that the requirement on veto cone or p_T is not applied.

This quantity is calculated considering all PF particles, excluding the lepton candidate, within a cone of size $\Delta R = 0.4$ around the direction of the lepton, excluding an innermost region, the “veto cone”. In the case of electrons large veto cones are used, to avoid that tracks coming from photon conversion or photons from bremsstrahlung spoil the isolation. In case of muons, in order to reduce pile-up effects, the veto cone region depends on the transverse momenta of the neutral hadrons and photons considered in the isolation, which are required to exceed a threshold of 0.5 GeV. The size of the veto cones and transverse momentum thresholds (only for muons) used to evaluate I_{rel} are summarized in Tab. 3.3.

In order to remove the pile-up effects, all charged PF particles are required to be originated from the primary hard interaction vertex, requiring $|z_0^{\text{track}} - z^{\text{vertex}}| < 2$ mm.

To further reduce the effects of pile-up the $\Delta\beta$ corrections are applied. These corrections are calculated as in equation 3.2, summing the transverse momenta of the charged particles within the isolation cone and with $|z_0^{\text{track}} - z^{\text{vertex}}| > 2$ mm, not originating from the primary vertex.

$$\Delta\beta = 0.5 \cdot \sum P_T^{\text{charged}}(\Delta z > 2 \text{ mm}) \quad (3.2)$$

To consider the amount of neutral energy with respect to the amount of charged energy in the isolation cone the sum is scaled by a factor 0.5.

In the analysis presented in this thesis requirements to veto electrons or muons will be described in section 4.5.1. Among these I_{rel} is required to be less than 0.3.

3.2.5 Jets

Quarks and gluons, abundantly produced in p-p collisions at LHC, instantaneously fragment and produce a collimated spray of hadrons [115]. The non-stable hadrons created decay to stable particles which, together with the stable particles created

during hadronisation, reach the detector material. The resulting shower of particles appears as a cluster of energy deposited in a localized areas of the calorimeter, called *jet*. This is pictorially represented in Fig. 3.8.

The main challenge of an analysis using jets is to recover the initial energy, momentum and, possibly, the nature of the parton produced in the original interaction conceiving an algorithm that reconstructs jets, allowing a fairly accurate extrapolation of the parton properties.

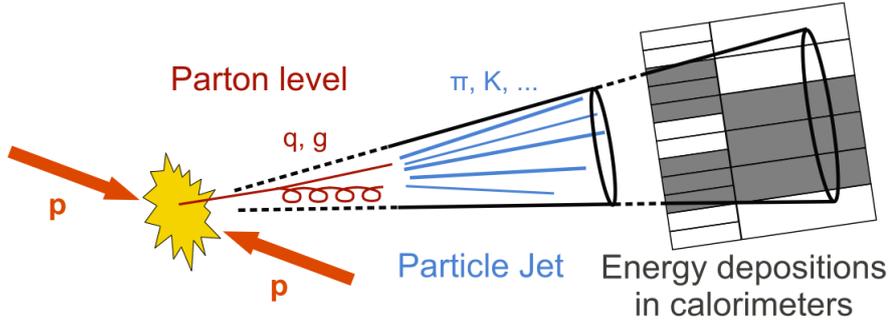


Figure 3.8: A simple representation of a jet and its component.

The algorithm used in CMS analyses is the anti- k_t algorithm [116], as implemented in the FASTJET package [117].

The **Jet Reconstruction** starts from the PF particles (section 3.2.1), which are clustered using the anti- k_t algorithm to form jets. The anti- k_t is an ideal cone algorithm and uses as input the list of the PF particles. In each iteration, it recombines the current set of particles and jet-candidates, which both are referred as entries. To form jets two quantities, the distances between pairs of entries (d_{ij}) and a parameter used to represent distance between an entry and the beam line (d_{iB}), are updated in each step and are defined as [116]:

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.3)$$

$$d_{iB} = p_{Ti}^{-2}, \quad (3.4)$$

where p_{Ti} is the transverse momentum of entry i , ΔR_{ij} is the distance in the $\eta\phi$ -plane between entry i and j , and R is a radius parameter fixed to 0.5 for CMS.

If d_{iB} is smaller than $\min_j d_{ij}$, entry i is promoted to reconstructed jets and excluded from the further iterations. Otherwise, entry i and $j' = \arg \min_j d_{ij}$ are merged into new entry i' and the algorithm proceeds to the next iteration. The algorithm stops when all jet-candidates are promoted to the reconstructed jets or when $\min(d_{ij}, d_{iB})$ is below the threshold d_{cut} .

The interesting jets are the ones related to the hard collision, to reject fake jets and identify jets from pile-up collisions, reconstructed jets are required to pass two levels of identification criteria.

The first requires that reconstructed jets pass a set of *loose* identification criteria,

to remove jets coming from calorimeter noise. The loose requirements are: the presence of at least one PF particle, whose energy comes from both ECAL and HCAL; a fraction of the total jet energy coming from photons or neutral hadrons less than 0.99 of the total jet energy; at least one charged PF object, for jets inside the tracker acceptance; a charged energy fraction greater than zero; finally an electron energy fraction less than 0.99.

The second set of requirements is to identify jets coming from pile-up vertices. The pile-up jets come from low p_T multijets production resulting from pile-up collisions. This contribution is removed by the effective jet area algorithm described in [118] and also using a multivariate approach [119].

The first algorithm relies on the fact that jets coming from pile-up are generally wider.

The multivariate approach, a BDT, uses as discriminating variables: the precise tracking information and the knowledge of the primary vertex, to define the proximity of the charged PF particles in a jet to the primary vertex; other discriminating variables associated to the jet shape, since jets from pile-up are generally wider.

The most discriminating shape variable used by BDT is a p_T -weighed average $\eta\phi$ distance of PF candidates within the jet with respect to the jet direction:

$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 p_{Ti}^2}{\sum_i p_{Ti}^2} \quad (3.5)$$

In this thesis jets are required to pass the *loose* working point (WP) of the discriminator. This WP corresponds to the 95% efficiency for 20 GeV jets within the tracker volume and 85% efficiency for 20 GeV jets, outside the tracker volume.

Jet Energy Corrections Due to the non-linearities in the detector responses, imperfect detector modelling, noise and pile-up effects, the raw jet 4-momentum obtained by the anti- k_t algorithm (p_{raw}^μ) does not represent an optimal estimation of the true jet 4-momentum (p_{true}^μ) at the hadron level. A correction is determined using simulated samples, and applied to a reconstructed jet 4-momentum for all the jets, such that the four-momentum is equal, on average, to that at the hadron level. This correction is applied on samples used in this analysis.

The correction can be factorised into a set of sequential corrections [120]:

$$p_{corr}^\mu = C_{\text{offset}}(p_T^{raw}, \eta) \cdot C_{\text{MC}}(p_T', \eta) \cdot C_{\text{abs}}(p_T'', \eta) \cdot C_{\text{rel}}(p_T''', \eta) \cdot p_{raw}^\mu \quad (3.6)$$

The first term applied C_{offset} , subtracts the contribution from pile-up and detector noise. The second term is the MC correction (C_{MC}) that aims to remove non-linearity on p_T and non-uniformity on η of the detector response, is applied to the offset corrected transverse momentum p_T' . The third term, the absolute correction, C_{abs} , is applied to the MC corrected transverse p_T'' and is designed to give a uniform response in p_T . The last term C_{rel} is applied to the corrected transverse momentum p_T''' and corrects the response to be flat as a function of η .

The correction factors were estimated using the methods and the samples (data and MC) of [120]. For the PF jets with $p_T > 20$ GeV the combined correction is below 20% depending on η as shown in figure 3.9. Further details can be found in [121].

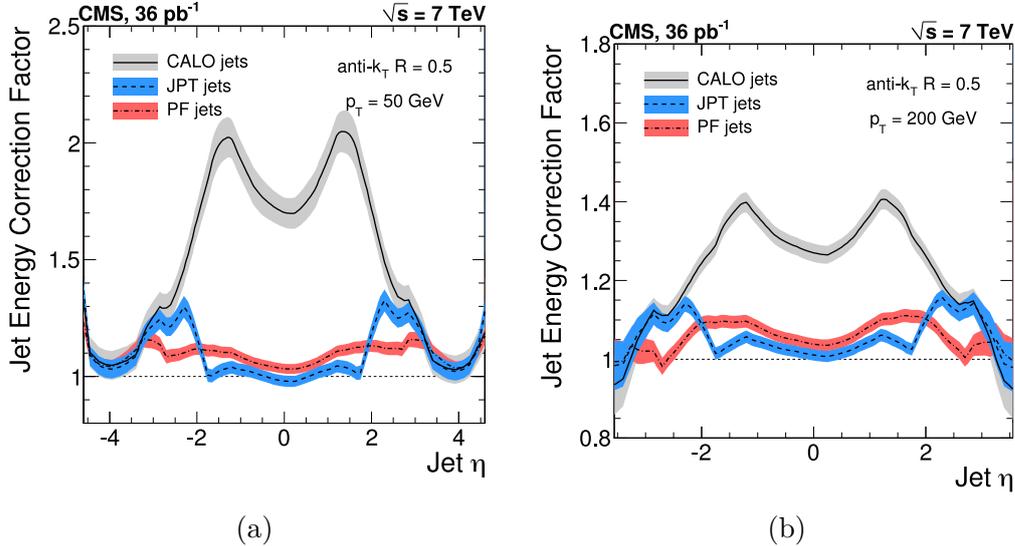


Figure 3.9: The combined jet energy correction factor C as a function of jet η for jets with $p_T = 50$ GeV (a) and $p_T = 200$ GeV (b). The correction factor is shown for three different jet reconstruction algorithms: PF jets, JPT jets and CALO jets. Only PF jets are used in the analysis presented in this thesis. For PF jets the total uncertainty on jet energy correction factor varies between 3 and 5%, depending on p_T [120].

3.2.6 b jets

The identification of jets resulting from the hadronization of b quarks, is the key for many physics analyses at HEP collider experiments. The identification of b jets, known as *b tagging*, is used in the SM Higgs or in top quark analyses since $t \rightarrow Wb$. Searches for new physics also typically involve b tagging, especially in theories where the third generation has a special role. Last but not least, b tagging is an important tool used in the search presented in this thesis, where one of the two SM-like Higgs bosons, originated by the resonance, decays into a pair of b jets.

b jets reconstruction is quite challenging because they can be easily faked by jets originated by gluons and light-flavour quarks. The discrimination of b jets against jets originated from light quarks or gluons, mainly exploits the properties b-hadrons arising from the hadronization of the b quark: their long lifetime (~ 1.5 ps) and their relatively large mass (~ 5 GeV). A simplified representation of a b jet can be found in Fig. 3.10.

The long lifetime means that tracks from the b-hadron decay typically (at the LHC) have a large impact parameter and are originated from a secondary vertex

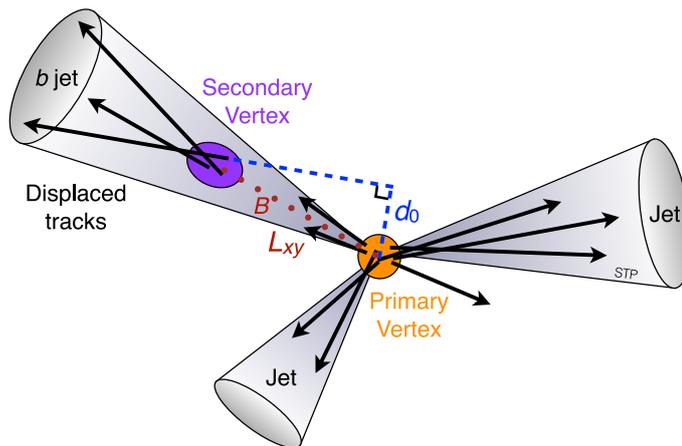


Figure 3.10: A simplified representation of a b jet (credits to [122]).

separated several millimeters from the primary vertex. The relatively large mass disfavors their creation during fragmentation and hence b-hadrons are more likely from the higher momentum hard process than a typical light hadron. Furthermore, the number of particles produced during the fragmentation depends on the mass of the parton and hence a b-hadron will carry a larger fraction of the momentum of the parton than a light hadron. This implies that decay products within a b jet are more collimated within the jet cone and form a larger invariant mass. Additionally, the larger mass of the b-hadron means it will typically decay into more particles than light hadrons. All these properties can be exploited by a tagging algorithm for b jets.

Various b tagging algorithms have been developed at CMS. The one used in thesis is the Combined Secondary Vertex (CSV) which uses a combination of secondary vertices information and tracker-based variables in order to separate b jets from other jets [123]. The use of the tracker-based variables allows to recover b tagging efficiency, when the secondary vertex (SV) cannot be properly reconstructed.

To reconstruct SV, the same algorithm (the Adaptive Vertex Fitter [105], section 3.1.2) used for the primary vertex (PV) reconstruction, is applied to the tracks within a given jet. The ΔR between the track direction and the jet axis is required to be less than 0.3 for the tracks used for the SV reconstruction. Once reconstructed the SV candidates, additional requirements are applied in order to improve b tagging discriminating power: the fraction of common tracks used in the reconstruction of both, the primary and secondary vertexes, should be less than 65%; the significance of the distance between the PV and SV in the transverse plane should be greater than 3σ ; the reconstructed flight direction of each candidate, defined as the unit vector pointing from the PV to SV, has to be within a cone of $\Delta R < 0.5$ around the jet direction. Furthermore, to reduce the contamination due to interactions with the detector and decays of long-living mesons, secondary vertices with a distance in the transverse plane from the PV greater than 2.5 cm and with an invariant mass greater than 6.5 GeV or compatible with K_0 mass are rejected [123].

The information described above and the transverse impact parameter measured of the charged particles⁴ in each jet are used to build two likelihood-based discriminators, in order to distinguish between b jets against charm jets (D_c), and b jets against light quarks or gluons ($D_{light-g}$). The D_c and $D_{light-g}$ are then combined into a single CSV discriminator.

In Fig. 3.11a is reported the distribution of the CSV discriminator in simulated samples and in real data of QCD multijets events collected during 2011 [123]. The stacked, coloured histograms are the contributions of the different components from simulated multijets samples, showing a good agreement between the CSV variable evaluated in the simulated and real data samples. The performance of the CSV discriminator has been evaluated in term of b tagging efficiency as a function of mistag probability. The latter is the probability of selecting a non-b jet using the discriminator. Depending on the expected probability of b tagging misidentification, CMS uses three Working Point (WPs). Each point corresponds to a value of the CSV discriminator above a threshold. The three WPs correspond to a misidentification probability of 10% (Loose b jet) , 1% (Medium b jet) , and 0.1% (Tight b jet) for $p_T \geq 80$ GeV and allow a flexible b jets selection depending on the achievable compromise between purity and efficiency required to tag a b jet at the analysis level. The related threshold values of the CSV discriminator are 0.244 (CSVL), 0.679 (CSVM), 0.898 (CSVT). The three WPs, their threshold values and their misidentification probabilities are summarized in Table 3.4. This study considers b-jets, the jets tagged using the medium WP.

Working Point	CSV threshold	Misidentification probability
Loose	0.244	0.1
Medium	0.679	0.01
Tight	0.898	0.001

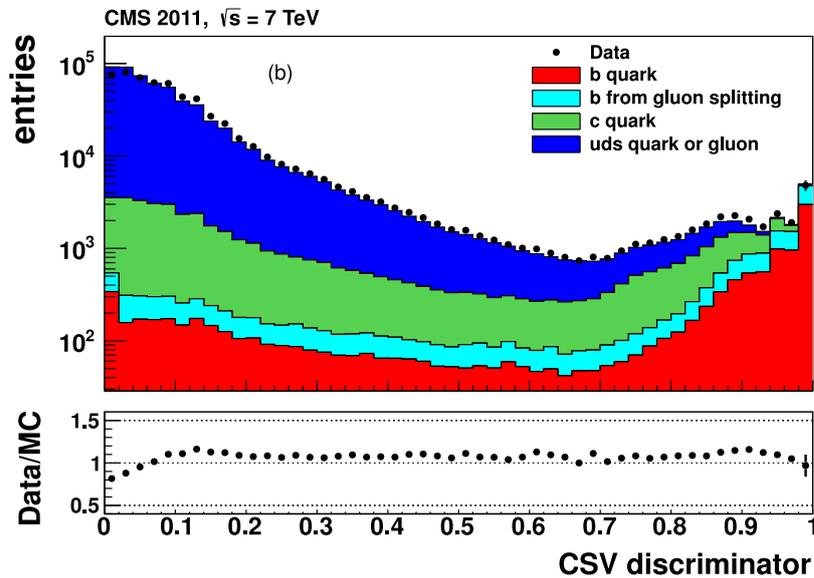
Table 3.4: Definition of b tag working points with the corresponding CSV value thresholds and the expected misidentification probabilities. To pass the WP identification criterion, jet should have CSV tag value greater than the threshold value.

In Fig. 3.11b, the b tagging efficiency as a function of CSV discriminator measured in data and predicted from simulation, using the same samples mentioned above, is shown in the upper panel. The three arrows on the bottom panel of Fig 3.11b indicate the three WPs. In the same panel the Scaling Factors are reported, defined as the ratio of b tagging efficiency in MC and data sample, that will be used to correct the simulation predictions in the analyses [123]. Based on these scale factors, in this work the b tag efficiency and misidentification probability were independently scaled within uncertainties to take into account corresponding systematics (section 5.1.1).

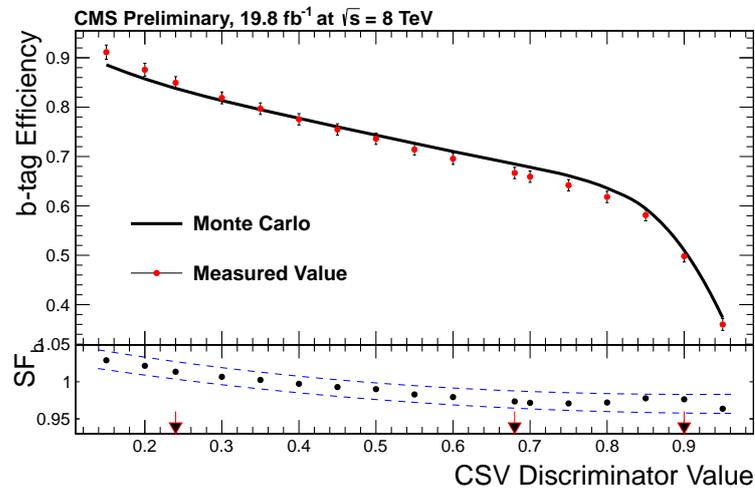
CMS has achieved a b jet tagging efficiency of 85% for a light-parton misidentification probability of 10% in multijets events. For analyses requiring higher purity,

⁴The IP is given the same sign as the scalar product of the jet axis direction with the vector pointing from the primary vertex to this point of closest approach.

a misidentification probability of only 1.5% has been achieved, for a 70% b jet tagging efficiency [123].



(a)



(b)

Figure 3.11: (a) Distribution of the CSV discriminator, where the filled circles correspond to data. The stacked, coloured histograms indicate the contributions of the different components in simulated multijets (“QCD”) samples [123]. (b) Efficiency of b tagging as a function of the CSV discriminator thresholds. The efficiency measured in data and predicted from simulation are shown in the upper panel. The scale factor is shown in the lower panel, where the blue dashed lines represent the combined statistical and systematic uncertainty. The arrows indicate the standard operating points.

3.2.7 Missing Transverse Energy

Neutrinos or neutral weakly interacting particles not yet discovered don't leave any signature in the detector. The presence of these particles can be inferred by the momentum imbalance in the transverse detector plane. The transverse momentum of the colliding proton pair is negligible, and therefore, from the conservation of 4-momentum, the sum of the transverse momenta of all the particles produced in a p-p interaction must be null. As a consequence of the imbalance, the missing transverse energy (\vec{E}_T^{miss}), is measured by the sum of all the observed particles and is defined as

$$\vec{E}_T^{\text{miss}} = - \sum_{\text{visible}} \vec{p}_T \quad (3.7)$$

The missing transverse energy plays an important role at LHC in several analyses, among these the ones aiming to the discovery of new physics or to the precision measurements in the SM processes in particular in the Higgs sector.

The source of real \cancel{E}_T in this analysis is from neutrinos created in electroweak interactions. There are also several sources of false \cancel{E}_T which are often difficult to control. These sources include the mismeasurement of jets energy, the detector inefficiencies, the reconstruction inefficiency and the pile-up.

Several methods were developed by CMS for a precise reconstruction of the missing transverse energy. The best performance in presence of a high number of pile-up interactions is obtained using the Multivariate Particle-Flow Missing Energy (MVA PF \vec{E}_T^{miss}) algorithm [124].

The MVA PF \vec{E}_T^{miss} is computed as a correction to the *hadronic recoil* \vec{u}_T reconstructed from PF particles. To estimate the \vec{E}_T^{miss} the algorithm assumes the presence of a boson (a vector boson Z or the SM Higgs boson) decaying into a pair of leptons and that the main uncertainties of \vec{E}_T^{miss} coming from the other hard scattering products, excluding the boson, can be neglected. This is compatible with our process $H \rightarrow hh \rightarrow bb\tau_h\tau_h$, because the invisible momentum comes mostly from $h \rightarrow \tau_h\tau_h$. Therefore assuming

- \vec{q}_T , the boson momentum in the transverse plane, obtained as the p_T sum over all visible decay products of the boson
- \vec{u}_T , the *hadronic recoil*, the vectorial sum of the transverse momenta of all particles, except those coming from the boson

Momentum conservation in the transverse plane gives

$$\vec{u}_T = -\vec{q}_T - \vec{E}_T^{\text{miss}} \quad (3.8)$$

In order to avoid the dependency from the global coordinate system, the hadronic recoil can be projected parallel (u_{\parallel}) and perpendicular (u_{\perp}) to the \vec{q}_T direction,

considered as the event axis.

Five discriminating uncorrelated variables are exploited by the algorithm to estimate the \vec{E}_T^{miss} . These variables are constructed using linear combinations of \vec{E}_T^{miss} calculated from different types of PF particles, to take into account the different components of an event. The PF types used are the charged PF particles from (or not) the hard interaction, the neutral PF particles in jets passing or failing the MVA pile-up identification (section 3.2.5) and the unclustered neutral PF particles [124]. For each $\vec{E}_T^{\text{miss}}(i)$ variable, the vector $\vec{u}_T(i)$ is computed. A BDT regression then takes as input: the magnitude and the azimuthal angle ϕ of all type of \vec{u}_T , the scalar p_T sum of all PF particles for each \vec{E}_T^{miss} variable, the momentum vectors of the two leading p_T jets and the number of the primary vertices. The BDT computes a correction to both the angle and magnitude of the recoil to match the true recoil. The corrected \vec{u}_T is then added to \vec{q}_T to obtain the negative MVA PF \vec{E}_T^{miss} .

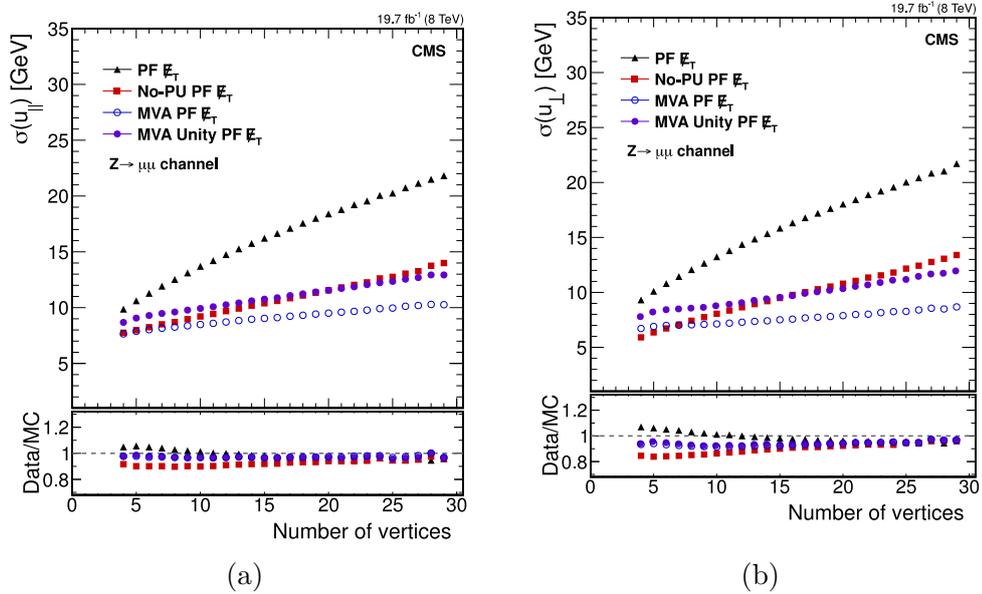


Figure 3.12: Parallel (a) and perpendicular (b) recoil component (with respect to the event axis, \vec{q}_T direction) resolution as a function of the number of reconstructed vertices. The obtained resolution is also shown for other algorithms: PF \vec{E}_T^{miss} (black triangles), No-PU \vec{E}_T^{miss} (red squares), MVA PF \vec{E}_T^{miss} (blue open circles), used in this thesis, and MVA Unity \vec{E}_T^{miss} (violet full circles) in $Z \rightarrow \mu\mu$. The upper frame of each figure shows the resolution in data; the lower frame shows the ratio of data to simulation [124].

The BDT has been trained on the $Z \rightarrow \mu\mu$ process for data and MC samples, because of the absence of neutrinos in the hard scattering process. The performance in the \vec{E}_T^{miss} reconstruction of several algorithms [124] studied by CMS is shown in Fig. 3.12, where the resolutions of u_{\parallel} and u_{\perp} are reported as a function of the number of vertices using data collected at 8 TeV. The best performance is obtained using MVA PF \vec{E}_T^{miss} .

The MVA PF \vec{E}_T^{miss} shows a smaller dependence on pile-up interactions increasing the sensitivity and so the role of the missing transverse energy in the new physics

searches. Indeed the sensitivity of the search of the Higgs boson into a tau pair using the MVA PF \vec{E}_T^{miss} was improved of 20% with respect to the \vec{E}_T^{miss} estimated using the only the PF particles [71].

A parametric data-to-simulation correction is derived and used to rescale the response and smear the resolution of the recoil in simulated events. The corrections are parameterized as a function of the p_T of the boson at generator level and affect only the MC samples that contain h, Z or W. The parameterization was obtained from the discrepancies in the BDT output between $Z \rightarrow \mu\mu$ data and MC samples. For each generated boson momentum range, the MVA \vec{u}_T distribution from simulations is then corrected to match the distribution in data.

3.2.8 τ leptons

The τ is the heaviest among the leptons with its mass of (1.77686 ± 0.00012) GeV [35]. It has a short lifetime of $(2.903 \pm 0.005) \times 10^{-13}$ s, so it decays by weak interactions before being detected and is reconstructed through its decay products. The main τ decay modes are resumed in Table 3.5.

τ lepton decays through the so called *leptonic* decay into electron or muon, a tau neutrino and a neutrino associated to electron or muon (which doesn't enter in the reconstruction because it is undetected). The total branching fraction of the leptonic decays is around 35%. The remaining 65% of the times the tau decays into hadrons plus a tau neutrino. This decay mode, indicated with τ_h , is the so called hadronic tau decay.

Decay mode	Branching Fraction (%)
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$	17.41 ± 0.04
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$	17.83 ± 0.04
$\tau^- \rightarrow l^- \bar{\nu}_l \nu_\tau$	35.24 ± 0.08
$\tau^- \rightarrow h^- \nu_\tau$	11.53 ± 0.06
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	25.95 ± 0.09
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	9.53 ± 0.11
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	9.80 ± 0.06
Other hadronic decays	≈ 7.95
All hadronic modes	≈ 64.76

Table 3.5: All possible τ decay modes with the corresponding branching fractions; h^\pm indicates a π or a K ; l^\pm stays for e^- or μ^- [35].

Since it is impossible to separate the τ signature from the electron and muon, in CMS the τ decay in electron and neutrinos, indicated with τ_e , is reconstructed as an electron (sections 3.2.2 and 3.2.4), while the τ decay in muon and neutrinos, indicated with τ_μ , is reconstructed as a muon (sections 3.2.3 and 3.2.4).

Like in the case of the b jet, the reconstruction and identification of the τ_h decays play a very important role in this thesis, since the final state investigated has two hadronic taus, coming from the decay of one of the two SM-like h bosons.

The Hadron plus strips algorithm is used to identify τ_h as a τ_{jets} , exploiting the properties of such decays, characterized by particles produced in a narrow cone (like a jet) but with a low particle multiplicity content with respect to a jet coming from quark or gluon. This algorithm and the variable used to discriminate τ_{jets} from real jets, the τ isolation, will be described in the following. A simplified representation of a τ jet is reported in Fig. 3.13a.

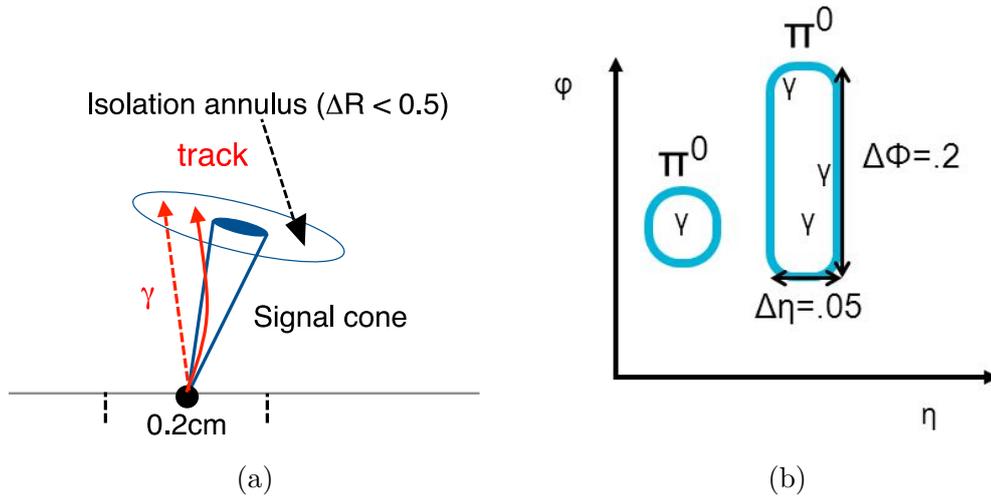


Figure 3.13: (a) A simplified representation of a τ jet (credits to Yuta Takahashi). (b) A schematic representation of the reconstruction of a *strip* in the HPS algorithm.

Hadron plus strips algorithm

CMS has developed two algorithms to reconstruct the τ_h : the hadron plus strips [125, 126] (HPS) and tau neural classifier⁵ (TNC). The HPS algorithm is used in this work and usually in CMS τ analyses. The algorithm uses the PF particles information from each subdetector (section 3.2.1) and exploits the topology of τ_h decays, searching for the hadronic decay modes reported in Table. 3.5, in one or three charged pions and up to two neutral pions.

The τ_h reconstruction starts using a PF jet candidate with $p_T > 14$ GeV and $|\eta| < 2.5$ as seed (section 3.2.5). The first step is the identification of the π^0 in the τ jet candidate, which requires to recover photons from the π^0 decay converted in electron pairs in the tracker material. The bending of electron/positron tracks in the magnetic field brings a broad signature of π^0 in the calorimeter along azimuthal direction. The neutral pions are indeed reconstructed as $\eta - \phi$ “strips” of electromagnetic particles, photons and electrons, reconstructed by the PF algorithm. A strip is open in a $\eta\phi$ -window of size 0.05×0.20 around the most energetic electromagnetic PF particle (Fig. 3.13b). A clustering of all electrons and photons in this window is done, adding the PF momenta to obtain the strip momentum.

⁵The TNC algorithm is not described in this work because it is not exploited for the τ_h reconstruction. Further details can be found in [125]

All reconstructed strips with a $p_T > 2.5$ GeV are considered as π^0 candidates if the reconstructed invariant mass is in the range of [50,200] MeV. After clustering, strips are combined with charged PF particles (excluding electromagnetic particles) within the jet, requiring a p_T of the selected particles greater than 0.5 GeV. In order to reduce the pile-up contribution, only PF charged particles compatible with the estimated τ_h production vertex, are considered, requiring a longitudinal, $|\Delta z|$, and a transverse distances, $|d_{xy}|$, between the particle track and the production vertex position less than 0.4 and 0.03 cm, respectively.

A τ jet candidate is reconstructed depending on the number of charged particles (pions) and strips in the τ_h candidate. Each candidate must be compatible with a combination of charged hadrons and π^0 candidates of a hadronic tau decay, listed in Table. 3.5:

- **Single hadron:** A single charged hadron reconstructed by the PF algorithm without any strips. This decay mode corresponds to $\tau^\pm \rightarrow \pi^\pm \nu_\tau$, with some contribution from $\tau^\pm \rightarrow \pi^\pm \pi^0 \nu_\tau$, where the π^0 is not energetic enough to be detected.
- **Hadron plus one Strip:** this selects the decays to the intermediate resonance $\rho(770)$, $\tau^\pm \rightarrow \rho^\pm \nu_\tau \rightarrow \pi^\pm \pi^0 \nu_\tau$. In this case the combined system must have the invariant mass compatible with the $\rho(770)$.
- **Hadron plus two Strips:** corresponds to decays of $\tau^\pm \rightarrow \pi^\pm \pi^0 \pi^0 \nu_\tau$. Here the consistency of the reconstructed with $a_1(1260)$ resonance is required.
- **Three Hadrons:** this corresponds to the decay mode $\tau^\pm \rightarrow h^\pm h^\mp h^\pm \nu_\tau$ and $\tau^\pm \rightarrow h^\pm h^\mp h^\pm \pi^0 \nu_\tau$. In addition the tracks are required to come from the same vertex, with a maximum impact parameter of 2 mm in the z direction and sum to unit charge.

All charged hadrons and strips considered above are required to be within a narrow cone around the jet axis, of size:

$$\Delta R = \begin{cases} 0.05 & p_T^{\tau_h} > 56 \text{ GeV} \\ 2.8/p_T^{\tau_h} [GeV] & 28 < p_T^{\tau_h} < 56 \text{ GeV}, \\ 0.10 & p_T^{\tau_h} < 28 \text{ GeV} \end{cases}$$

where $p_T^{\tau_h}$ is the transverse momentum of the reconstructed τ_h . The ΔR between the directions of the PF Jet and of the reconstructed τ jet is required to be less 0.1. In case a τ_h decay is compatible with more hypotheses, the hypothesis that gives the highest $p_T^{\tau_h}$ is chosen.

τ_h Isolation

After the HPS reconstruction the τ_h is required to be isolated to reduce the presence of jets originated from quarks and gluons as well as the charged and neutral

particles coming from pile-up interactions. The variable used to define the τ_h isolation, I_τ , is calculated using PF charged particles and photons within a cone of $\Delta R < 0.5$ around the τ_h candidate direction, with $p_T > 0.5$ GeV and not used to form the h candidate. To reduce the pile-up contamination which spoils the isolation, charged particles tracks used in the evaluation are required to be compatible with the τ_h production vertex ($|\Delta z| < 0.2$ cm and $|d_{xy}| < 0.03$ cm) assumed to be the vertex with the highest probability associated to the “leading” (highest P_T) track of the τ_{had} .

The Isolation can be expressed as:

$$I_\tau = \sum P_T^{charged} + \max(P_T^\gamma - \Delta\beta, 0), \quad (3.9)$$

The $\Delta\beta$ term takes into account the contribution of pile-up to the τ_h isolation; it is defined as a scaled (0.4576) sum of the charged particles p_T within a $\Delta R < 0.8$ cone around the τ_h direction. The sum extends to the charged particles which are compatible to originate from the τ_h production vertex [125, 126]. The scale factor (0.4576) makes the τ_{had} identification efficiency insensitive to pile-up. The $\Delta\beta$ term is then expressed as

$$\Delta\beta = 0.4576 \cdot \sum P_T^{charged}$$

Changing the threshold of the isolation variable it is possible to define three working points: “loose”, “medium” and “tight” of the *HPS combined isolation 3-hit discriminator*⁶, defined by requiring the isolation p_T sum not to exceed 2.0, 1.0 and 0.8 GeV respectively. The working points are determined using a MC sample of QCD dijet events, obtaining that the “loose” working point corresponds to a probability of approximately 1% for jets to be misidentified as τ_h . Successive working points reduce the misidentification rate by a factor of two with respect to the previous one. The identification efficiency has been measured in $Z \rightarrow \tau\tau$ data and simulations and is above 40% for all working points, as shown in figure 3.14.

Discriminating τ_h against Electron and Muon

In order to reject muons or electrons misidentified as τ_h , different vetoes are applied. Muons or electrons misidentified as τ_h are suppressed by requiring loose, medium and tight working points (WP) for their identification, depending on the compromise between the efficiency and purity required for the τ_h identification. The probability of a muon to produce a fake hadronic tau is much lower than in the case of electrons.

To suppress a muon faking a τ_h the track of the leading charged hadron of the τ_h candidate it is required to be not reconstructed as a tracker muon.

⁶This name refers to the fact that charged particles associated to tracks with ≥ 3 hits in pixel plus silicon strip detectors are considered in the isolation P_T sum.

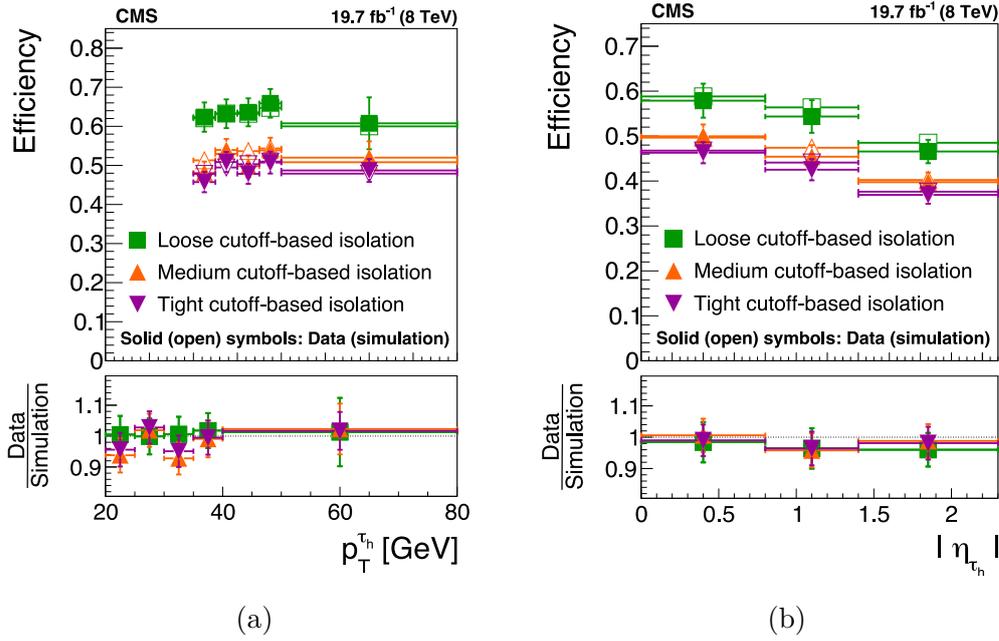


Figure 3.14: τ_h identification efficiency measured in $Z \rightarrow \tau_\mu \tau_h$ events as a function of (a) p_T and (b) η for different isolation working points [126].

The three WPs are:

- **muon-loose:** τ_h candidates are vetoed in case a track segment in the muon system is found within a cone of size $\Delta R = 0.5$ centered on the τ_h direction.
- **muon-medium:** τ_h candidates are vetoed in case of a muon-loose or in case hits within a cone of size $\Delta R = 0.5$ around the τ_h direction are present in CSC, DT or RPC modules located in the two outermost muon stations.
- **muon-tight:** τ_h candidates are vetoed in case of a muon-medium or the total deposited energy by the leading track of τ_h in the calorimeters is less than the 20% of track momentum

To reject electrons faking hadronic taus, a BDT discriminator, utilizing the set of variables defined for electron identification of section 3.2.2, is used. The discriminator has been trained on simulated samples with electrons or hadronic taus in the final state. Different working points were defined (Loose, Medium, Tight and Very-tight) by applying requirements on the BDT output and are chosen in order to obtain the lowest $e \rightarrow \tau_{had}$ fake-rate for a given τ_h identification efficiency.

Chapter 4

Event Selection and modelling

This chapter presents the strategy followed to reconstruct and select the heavy Higgs H candidate in the decay mode $H \rightarrow h[bb]h[\tau_h\tau_h]$. To increase the sensitivity of this search, the signal must be selected in the most efficient way by reducing the presence of the backgrounds with a similar signature. Therefore an accurate modelling of both signal and backgrounds is needed. The analysis uses several techniques employed by previous searches of the SM and MSSM Higgs boson decaying into a tau pair [68, 71], fully exploiting the advantage of solid techniques to extract the signal and to model the data. The physical objects described in the previous chapter are used to reconstruct $H \rightarrow hh \rightarrow bb\tau_h\tau_h$ through several steps. The resulting distributions of the reconstructed H candidate masses (for simulated signal plus backgrounds hypothesis and for data) are used as input to a fit in order to extract the signal, as described in the next chapter.

4.1 Signal and Background Definition

The process studied in this thesis is $pp \rightarrow H \rightarrow hh \rightarrow bb\tau_h\tau_h$ (the *fully-hadronic* channel). The expected signal events are characterized by the presence in the final state of two b jets and two τ jets, which represent the b quarks and the τ leptons signatures. An example of a simulated signal event is reported in Fig. 4.1, where the two b jets and the two τ_h from the hh decay are highlighted.

Unfortunately the signal events can be reproduced by several SM processes when the same particles are in the final state or when quark and gluon jets can reproduce the same characteristics of the b jets or τ_h . The fully hadronic channel is rather challenging in hadron colliders due to the large background produced in association with jets.

The main contributing backgrounds, in order of relevance are:

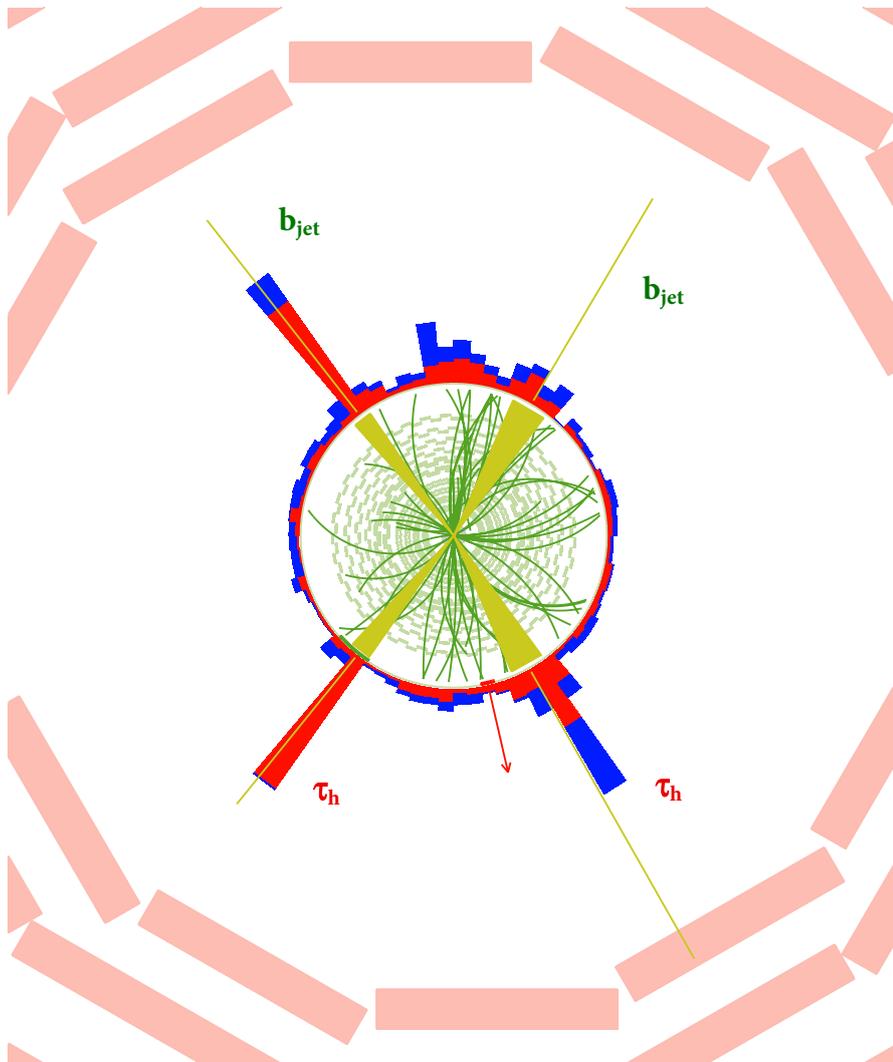


Figure 4.1: Event display of a signal event $pp \rightarrow H \rightarrow hh \rightarrow bb\tau_h\tau_h$ with $M_H = 300$ GeV, where the two b jets and the two τ_h are highlighted. The red and blue towers are respectively the energy deposits collected by the ECAL and HCAL detectors. The yellow “cones” represent the reconstructed jets. The green lines represent the tracks, while the red arrow the missing transverse energy of the event. The red external segments are the muon chambers.

- **QCD multijets:** this background (Fig. 4.2a, 4.2b) is contributing because of the possible misidentification of quark and gluon jets as τ_h and the presence of b jets from pile-up or underlying events. The cross-section ($O(200)$ nb [127]) makes this background one of the most important.
- **$Z/\gamma^* \rightarrow \ell^+\ell^- + \mathbf{jets}$,** where $\ell = e, \mu, \tau$: in this process (Fig. 4.2c), the decay $Z \rightarrow \tau\tau$ behaves like $h \rightarrow \tau\tau$ and the major contribution to the background arises from $Z \rightarrow \tau_h\tau_h$ plus two jets, when the latter are misidentified as b jets. The invariant mass reconstruction of the $\tau_h\tau_h$ candidate can be exploited in discriminating Z from h decays. Like QCD plus multijets, the relative high cross-section ($O(1)$ nb, Fig. 1.12) and the high probability of misidentification make this background one of the dominating after QCD multijets.
- **$t\bar{t} + \mathbf{jets}$:** can perfectly reproduce the signal when the two τ_h come from W decay (Fig. 4.2d). It becomes a dominant and irreducible background when the two b jets are identified and reconstructed. Other contributions arise from misidentification of jets from the hard interaction or from pile-up as τ_h . The production cross-section of this process is $O(250)$ pb (Fig. 1.12).
- **$W \rightarrow \ell\nu_\ell + \mathbf{jets}$,** where $\ell = e, \mu, \tau$: despite its high production cross-section ($O(30)$ nb, Fig. 1.12), this process (Fig. 4.2e) is negligible since, at the most, only one real τ_h comes from W decay and in this case, other three jets in the event must be misidentified as a τ_h plus two b jets, and this happens with a low probability.
- **WW/WZ/ZZ + jets:** the searched final state can be mimicked by real or fake objects, coming from the bosons decays or from multiple interactions. A low contribution is expected (the production cross-sections are $\sigma_{WW} \approx 70$ pb; $\sigma_{WZ} \approx 20$ pb; $\sigma_{ZZ} \approx 7$ pb - Fig. 1.12).
- **Single top, tW channel:** this process has one top, decaying in bW , and one W boson in the final state. Then, one signal b jet is reproduced by the b quark coming from the top decay, the other signal objects can arise from the possible W decays or from pile-up or underlying events jets. A low contribution is expected ($\sigma \approx 20$ pb - Fig. 1.12).

4.2 Data samples

This analysis uses 19.7 fb^{-1} of the data collected by CMS during 2012 at a centre of mass energy of $\sqrt{s} = 8 \text{ TeV}$. Collected data are separated into *Primary Dataset*, according to the specific fired triggers (section 2.2.8). This analysis utilizes the Tau dataset. Since during the data taking period the trigger configuration changed and the Tau dataset has been divided in four secondary datasets depending on the run range (Table 4.1), corresponding to different data periods during which the detector was operated under a stable configuration (since trigger settings, average

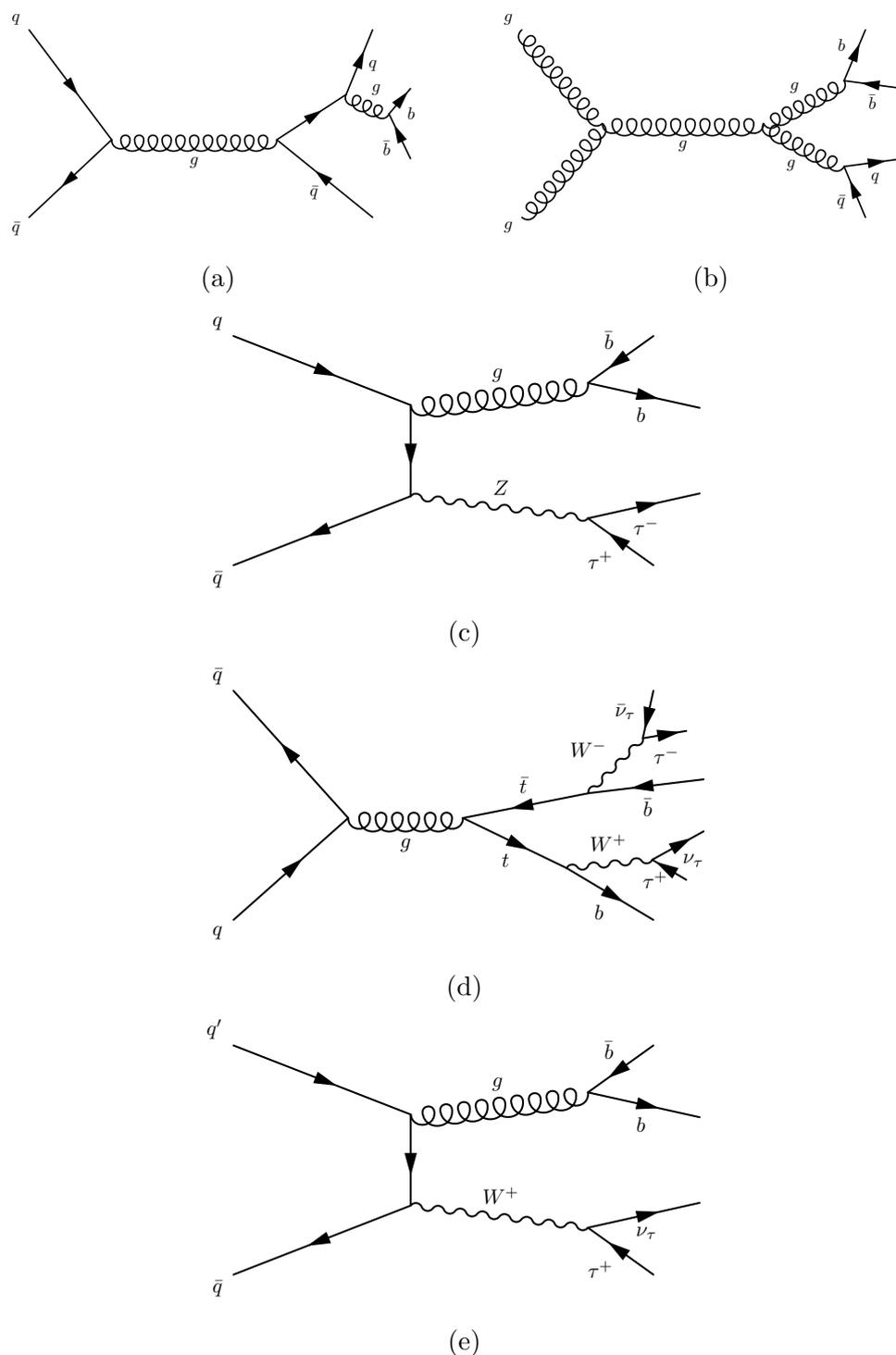


Figure 4.2: Examples of Feynman diagrams of some contributing backgrounds in this analysis, final states containing taus are shown, except for QCD multijets: (a,b) QCD multijets events, (c) $Z \rightarrow \tau\tau$ plus jets process, (d) $t\bar{t}$ production and (e) W plus jets.

instantaneous luminosity, system calibrations, etc. can change over time). The dataset uses only runs certified as good, where all components of the detector and accelerator are reliably working. The good runs are collected in a text file, called JSON file (`Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`) reducing the total delivered luminosity from 21.79 fb^{-1} to 19.7 fb^{-1} analysed in this thesis.

Dataset Name	Run-range	Luminosity
/Tau/Run2012A-22Jan2013-v1/AOD	190456–193621	0.887
/TauParked/Run2012B-22Jan2013-v1/AOD	193833–196531	4.446
/TauParked/Run2012C-22Jan2013-v1/AOD	198022–203742	7.153
/TauParked/Run2012D-22Jan2013-v1/AOD	203777–208686	7.318

Table 4.1: Tau data samples collected by CMS detector and used for this search. They are divided according to the data taking period.

The Tau dataset has been selected by two different triggers designed to select events with two taus (Double Tau Trigger) and a single tau (Single Tau Trigger) (see following section and Table 4.2). The *Double Tau Trigger* requires two *Medium* isolated τ_h candidates (section 3.2.8), the presence of a track within $|\eta| < 2.1$ and with p_T above a threshold, plus eventually an additional jet.

The *Single Tau Trigger* requires the presence of one *loose* isolated τ_h candidate, a track with p_T above a threshold and eventually missing transverse energy.

This analysis is focused on *Double Tau Trigger*, because there are two hadronic τ decays in the signal events. The specific trigger paths used for this thesis will be described in the next section.

4.3 Trigger selection

To be collected data are required to pass the L1 and the HLT paths listed in Table 4.2. These paths have the following requirements at L1 and at HLT (section 2.2.7), depending on the data taking period:

Trigger paths		
	HLT path	L1 seeds
Di-tau triggers	DoubleMediumIsoPFTau35_Trk5_eta2p1 DoubleMediumIsoPFTau35_Trk1_eta2p1	L1_DoubleTau44 OR L1_DoubleJetC64
Di-tau plus jet triggers	DoubleMediumIsoPFTau25_Trk5_eta2p1_Jet30 DoubleMediumIsoPFTau30_Trk5_eta2p1_Jet30 DoubleMediumIsoPFTau30_Trk1_eta2p1_Jet30	L1_DoubleTau44 OR L1_DoubleJetC64

Table 4.2: All the di-tau trigger paths used in $\tau_{had}\tau_{had}$ channel during the data-taking of 2012.

The **L1** trigger (section 2.2.7) [71] requires two L1 τ objects in the region $|\eta| <$

2.17 or, in order to increase the acceptance and the efficiency, two L1 jets in the region $|\eta| < 3$ [128].

- Each τ object is reconstructed as a L1 jet, with a transverse energy $E_T > 44$ GeV, with an energy deposit not spread over more than 2×2 calorimeter trigger towers in a calorimeter trigger region $\Delta\eta \times \Delta\phi \approx 0.35 \times 0.35$ wide, corresponding to 4×4 trigger towers.
- Each of the two L1 jets must have a transverse energy $E_T > 64$ GeV.

After a L1 accept, at the **HLT** a regional PF algorithm is run around the L1 seeds and search in a cone of $\Delta R < 0.5$ for isolated high p_T PF hadronic taus or jets. Then the event has to fire the di-tau trigger selection at HLT requiring: two isolated high- p_T ($p_T > 35$ GeV) PF τ_h (di-tau trigger) and in case of failure, two isolated high- p_T (both with $p_T > 25$ GeV or $p_T > 30$ GeV) PF τ_h plus a L1 seed jet with $p_T > 30$ GeV (di-tau plus jet trigger). Data and simulated samples used by this analysis were selected by these trigger paths. No prescaling was applied.

4.4 Sample composition modelling

The signature of the signal process, can be reproduced by a large variety of background processes (section 4.1). The sensitivity of the search of $H \rightarrow hh \rightarrow bb\tau_h\tau_h$ and its significance, rely on an accurate modelling of each process contributing to the data collected by the trigger mentioned above.

The data modelling used Monte Carlo simulations (appendix B) and data-driven techniques.

The simulation-based techniques are used in the modelling of some particular processes when the level of theoretical (cross-section) and experimental (detector response) precision is sufficient and a huge number of events is not required to precisely model the process (prohibitive time machine). The Monte Carlo (MC) simulations provide an interpretation of the data in HEP and allow to extract measurements on fundamental physical parameters relying on the modelling of the physical processes. The MC simulations play a fundamental role in the physics data analysis because they provide a complete picture of the large multiplicity of particles as the outcome of a hard interaction. When simulated events can be used to model a specific process, MC simulations allow to optimize criteria in order to separate the signal from the other processes or to apply a specific selection to reduce the contribution of a background process. Unfortunately MC simulation techniques are significantly affected by the systematic uncertainties due to a not precise modelling of the detector response, by the limited accuracy in the pile-up and QCD generator-level description, by uncertainties in the luminosity measurements and by uncertainties in the theoretical cross-sections. Moreover in some cases huge samples are required to accurately model some physical processes.

For some processes ($Z \rightarrow \tau\tau + \text{jets}$ and QCD multijets events), where the modelling uncertainties are big, and/or the statistics necessary is prohibitive, data-driven techniques were used. In the data-driven techniques the processes are estimated in a region where this process is dominant, after some specific requirements, and where the signal is expected to be negligible. This region is called *sideband* region. After that the background contribution is extrapolated in a region enriched of signal processes, called *signal* region, where all results are estimated. The parameter space in the sideband region should be as close as possible to the one of the signal region in order to avoid bias and big systematic uncertainties.

In the next section the modelling based on MC samples will be described. While the data-driven methods will be described in section 4.6.

4.4.1 Monte Carlo samples

The MC samples used in this analysis, the generator used and their cross-sections are reported in Table 4.3. After the generation of the events and the simulation of the detector response (for more details see appendix B), events are selected by the triggers (section 4.3), and the high level objects are reconstructed with the same algorithms used for real data (chapter 3).

The Parton distribution functions (PDFs) for the proton used in this analysis are CT10 [129] or CTEQ6L1 [130], together with MSTW2008 [131] following the PDF4LHC prescriptions [132].

The $H \rightarrow hh \rightarrow bb\tau\tau$ signal samples have been modelled with PYTHIA 6.4.26 [96], in a range of the heavy Higgs boson mass, M_H , between 260 and 350 GeV, in step of 10 GeV, using the cross-sections from “low- $\tan\beta$ -high” scenario (section 1.3).

The background processes, like $Z/\gamma^* \rightarrow ee$, $Z/\gamma^* \rightarrow \mu\mu$, $W \rightarrow \ell\nu + \text{jets}$, $t\bar{t}$ and dibosons (WW , WZ , ZZ) plus jets, have been modelled with the MADGRAPH 5.1 [98]. POWHEG 1.0 [97, 133–135] has been used for single top quark production. In addition to the inclusive $Z/\gamma^* \rightarrow \ell\ell$ plus jets and W plus jets backgrounds, samples with different jets multiplicity (up to four) have been simulated, in order to increase the statistics for these backgrounds. These are combined with the inclusive samples using the correct cross-sections ratio, taking into account the jets multiplicity at parton level. To model $Z \rightarrow \tau\tau + \text{jets}$ process, which represents one of the main backgrounds, a mixed modelling based on simulation and on data-driven techniques is used (see section 4.6.1).

The $t\bar{t}$ background can be accurately modelled using MC thanks to the precise measurement of the differential $t\bar{t}$ production cross-section [136]. In this analysis $t\bar{t}$ shape¹ and normalization are taken from MC. The most recent and precise CMS and ATLAS cross-section measurements are used [137].

$W + \text{jets}$ is a minor background. Both yield and shape of the $W + \text{jets}$ background are evaluated from MC samples. The most recent CMS cross-section mea-

¹A shape is defined as the template normalized to the unit area.

measurements at 8 TeV are used [138].

In the $\mathbf{Z} \rightarrow \ell\ell$ plus jets sample, when Z decays in two electrons or muons, MC $Z \rightarrow \ell\ell$ events are normalised to the NNLO cross-section [139]. Also the shape is taken from MC.

In the **Diboson** plus jets events are normalized to NLO cross-sections [140, 141] and the shape is taken from MC.

The **Single Top** process is in the tW channel. The events are normalized to NLO cross-sections [142] and shape is taken from MC.

For $Z \rightarrow \ell\ell$, Diboson, Single Top the low event statistics is faced relaxing the requirements on b tagging (section 4.5.2) only for the shape, without introducing any bias.

Monte Carlo samples		
Signal		
Process	MC generator	σ [pb]
$gg \rightarrow H \rightarrow hh \rightarrow bb\tau\tau$	PYTHIA	–
Boson plus jets		
$Z \rightarrow \ell\ell + \text{jets}$	MADGRAPH	3503.7
$Z \rightarrow \ell\ell + 1\text{jet}$	MADGRAPH	666.3
$Z \rightarrow \ell\ell + 2\text{jets}$	MADGRAPH	215.0
$Z \rightarrow \ell\ell + 3\text{jets}$	MADGRAPH	60.7
$Z \rightarrow \ell\ell + 4\text{jets}$	MADGRAPH	27.3
$W \rightarrow \ell\nu + \text{jets}$	MADGRAPH	36257.2
$W \rightarrow \ell\nu + 1\text{jet}$	MADGRAPH	6381.2
$W \rightarrow \ell\nu + 2\text{jets}$	MADGRAPH	2039.8
$W \rightarrow \ell\nu + 3\text{jets}$	MADGRAPH	612.5
$W \rightarrow \ell\nu + 4\text{jets}$	MADGRAPH	251.0
$t\bar{t}$ plus jets		
$t\bar{t} + \text{jets}$	MADGRAPH	245.3
Single top		
tW channel	POWHEG	22.2
Diboson plus jets		
$WW \rightarrow (\ell\nu)(\ell\nu) + \text{jets}$	MADGRAPH	5.8
$WZ \rightarrow (qq')(\ell\ell) + \text{jets}$	MADGRAPH	2.2
$WZ \rightarrow (\ell\nu)(\ell\ell) + \text{jets}$	MADGRAPH	1.1
$ZZ \rightarrow (\ell\ell)(q\bar{q}) + \text{jets}$	MADGRAPH	2.5
$ZZ \rightarrow (\ell\ell)(\nu\nu) + \text{jets}$	MADGRAPH	0.7
$ZZ \rightarrow (\ell\ell)(\ell\ell) + \text{jets}$	MADGRAPH	0.2

Table 4.3: List of the processes simulated using MC samples, specifying the used MC generator and the corresponding production cross-sections for 8 TeV p-p collisions. ℓ stands for e , μ or τ . Inclusive W decays are considered in $t\bar{t}$ and single top.

MC weights For MC simulation the estimate of the predicted events starts from the determination of the fraction of events passing all the event selection requirements. This number needs to be corrected, in order to take into account the differences between the MC simulation and the real data. A weight, w , is

given to each MC sample and is applied on each event:

$$w = \int \mathcal{L} dt \cdot \frac{\sigma_{pred}}{N_{evt}} \cdot SF_{Data/MC} \cdot w^{PU} \quad (4.1)$$

- $\int \mathcal{L} dt$ is the total integrated luminosity $\approx 19.7 \text{ fb}^{-1}$
- σ_{pred} is the cross-section used as an input in the MC generation (Table 4.3)
- N_{evt} is the total number of the generated MC events
- $SF_{Data/MC}$ resumes the corrections applied to MC to improve the agreement with the real data. This scale factor is a product of trigger efficiencies and τ identification scale factors. A brief description of how these are calculated is given in the Appendix C. The scale factors used in this analysis were calculated for the search of SM and MSSM Higgs boson in $\tau\tau$ final state [68, 71].
- w^{PU} is the pile-up weight. This weight is used in order to better match the MC distribution of the average number of pile-up interactions with data. To model pile-up in simulated samples, for each event of the MC samples a random number of minimum bias events are generated by PYTHIA and injected to the original event. The number of pile-up events follows a Poissonian distribution defined by CMS in [143]. The weight, w^{PU} , is calculated as the ratio between the number of pile-up events simulated in [143] coming from the same bunch-crossing of the hard scattering event and the distribution of pile-up events in data. The dependence of w^{PU} on number of pile-up interactions is represented in Fig. 4.3.

4.5 Event Selection

The reconstruction of $H \rightarrow hh \rightarrow bb\tau\tau$ signal events is one of the main purpose of this analysis and is realised in different steps.

Data and simulated samples collected by the triggers (section 4.3), are processed offline. After the reconstruction of the primary vertex (section 3.1.2) of the hard interaction, an offline selection, called *baseline selection*, is applied to reconstruct the $h \rightarrow \tau_h\tau_h$ candidate. The selection of the di-tau candidate pair follows the inclusive selection already published for the search of the SM $h \rightarrow \tau\tau$ candidate [71] and of the MSSM $\phi \rightarrow \tau\tau$ analyses [68]. Then, additional requirements are applied to the jets for the reconstruction of the $h \rightarrow bb$ candidate in the same event.

As final step, the $H \rightarrow hh$ candidate is reconstructed taking into account the presence of two SM-like h bosons.

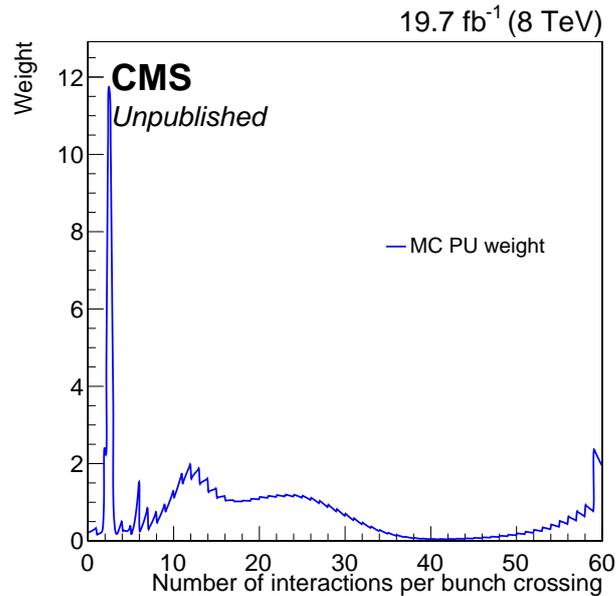


Figure 4.3: Event pile-up weight as a function of the number of pile-up interactions per bunch crossing in MC.

4.5.1 $h \rightarrow \tau_h \tau_h$ selection

After the events have passed the trigger selection, to form a $h \rightarrow \tau_h \tau_h$ candidate a pair of opposite charge of medium isolated hadronic taus ($I_\tau < 1$, section 3.2.8), compatible with the primary hard interaction vertex ($|\Delta z_v| < 0.2$ cm) and with a separation $\Delta R > 0.5$, has to be reconstructed in the event satisfying the following requirements:

- both τ_h must have a $p_T > 45$ GeV and $|\eta| < 2.1$. The high p_T requirement is due to the trigger efficiency turn on curve and has been chosen to keep the highest signal efficiency (appendix C.1). The η region and the p_T threshold chosen ensure good reconstruction performance and stable HLT trigger operation.
- each τ_h must be matched to one of the two HLT tau objects; the match requires that the direction of the HLT tau object is in a cone of radius $|\Delta R| = 0.5$ around the reconstructed τ_h .

This selection was optimized for the SM $h \rightarrow \tau_h \tau_h$ analysis [71].

To discard events in which a muon or an electron can be misidentified as a τ_h , it has been required that the both τ_h have passed the loose anti-muon and loose anti-electron discriminators (section 3.2.8).

Extra leptons veto Events with additional leptons, muons or electrons, are rejected by vetoes to further reduce specific backgrounds, like for instance Diboson.

- Events are vetoed if they contain any muon with $p_T > 10$ GeV, $|\eta| < 2.4$, compatible with the primary vertex ($|\Delta z_V| < 0.2$ cm, $|d_{xy}| < 0.045$ cm), which passes the *Tight* muon identification (section 3.2.3 - Table 3.2) and with the isolation $I_{\text{rel}} < 0.3$ (eq. 3.1)
- Events are also vetoed if they contain any electron with $p_T > 10$ GeV, $|\eta| < 2.5$, compatible with the primary vertex ($|\Delta z_V| < 0.2$ cm, $|d_{xy}| < 0.045$ cm), which passes the *loose* BDT-based identification (section 3.2.2 - Table 3.1) and with the isolation $I_{\text{rel}} < 0.3$ (eq. 3.1)

In case of multiple candidates in an event, the pair with the most isolated τ_h is chosen. If there is still more than one couple, the $h \rightarrow \tau_h \tau_h$ with the highest scalar sum of p_T is taken.

In Fig. 4.4 and 4.5 the p_T and η distributions of the *leading* (highest p_T) and *subleading* (second leading p_T) τ_h candidates are reported, resulting after the baseline selection. These distributions show a good agreement between data and MC.

The “Electroweak” in the legends (red) includes $W + \text{jets}$, diboson and single top processes; the $Z \rightarrow \ell\ell$ (blue) includes $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ plus jets processes. This convention for the legend (if not differently specified) will be used in all the plots of this thesis.

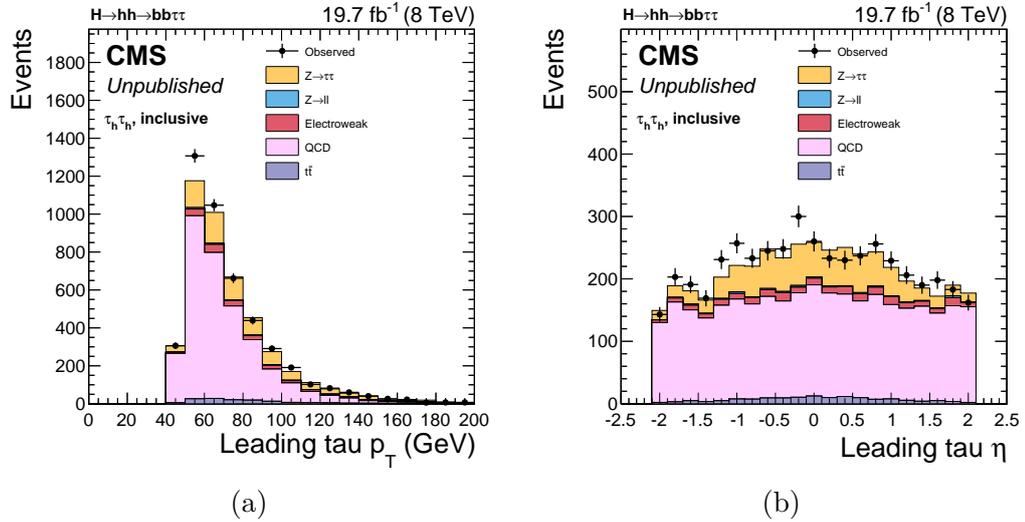


Figure 4.4: Distributions of the *leading* τ_h : p_T (a) and η (b).

4.5.2 $h \rightarrow bb$ selection

After the reconstruction of the $h \rightarrow \tau_h \tau_h$, almost two additional jets must be present in the event to reconstruct a $h \rightarrow bb$ candidate. $h \rightarrow bb$ candidates are reconstructed using pairs of PF jets satisfying the following requirements:

- jets must have a $p_T > 20$ GeV, $|\eta| < 2.4$ to ensure a good performance of the CSV discriminator used for b tagging (section 3.2.6)

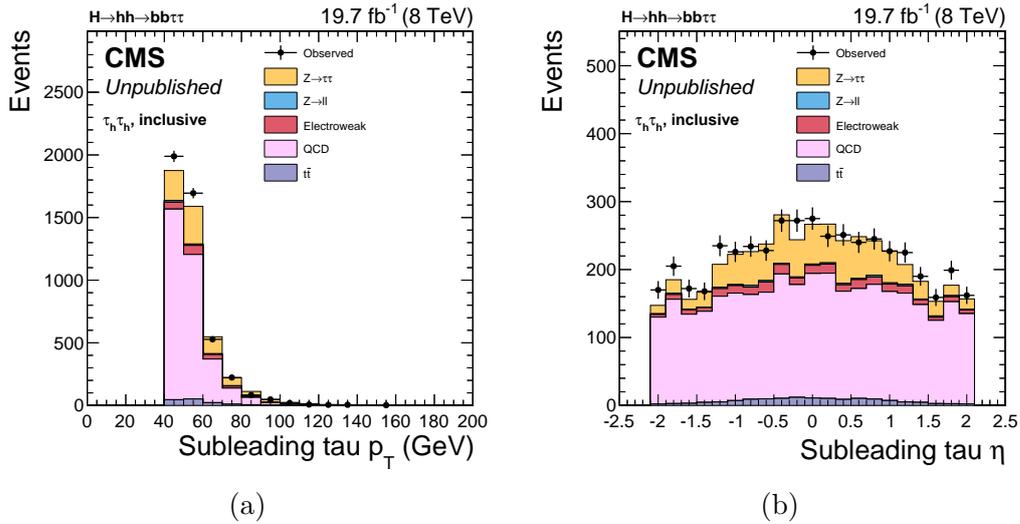


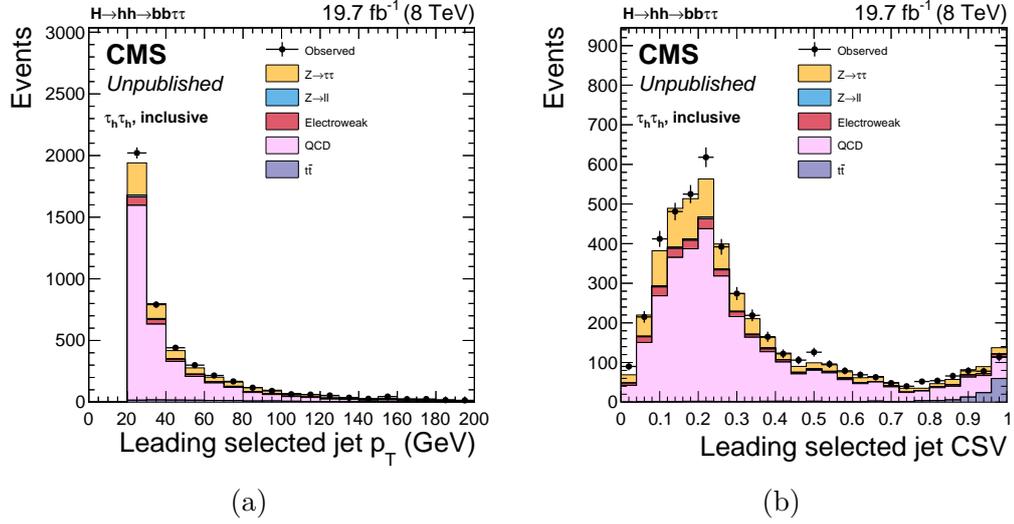
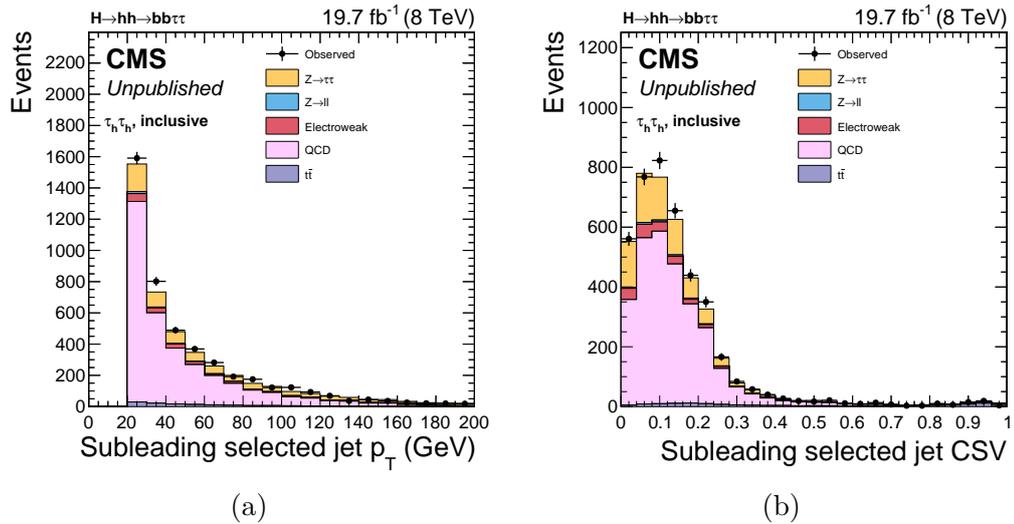
Figure 4.5: Distributions of the *subleading* τ_h : p_T (a) and η (b).

- jets must pass the loose working point of the discriminator against pile-up jets (section 3.2.5)
- jets must be separated from the two τ_h of the h candidate decay, requiring that the direction of the jet is out of a cone of radius $|R| = 0.5$ around each of the reconstructed τ_h . This requirement allows to avoid any overlap with the signal objects.

In order to select the correct b jets coming from h boson, a study at generator level on a signal MC sample has been performed, comparing different methods to select b jets. The study (see following subsection) has revealed that the majority of signal events will have at least one jet passing the *medium* WP of the CSV discriminator (CSVM). Moreover it revealed that the best criterion to reconstruct the candidate $h \rightarrow bb$ is to order jets according to their CSV discriminator value. Thus the *leading* and *subleading* b jets correspond to the jets with the highest and second highest CSV values respectively. These jets are used to reconstruct the invariant mass of the $h \rightarrow bb$ candidate, M_{bb} .

In figures 4.6 and 4.7 the distributions of the p_T and CSV discriminator of the *leading* and *subleading* b jets respectively are reported, satisfying all the requirements mentioned above except the b tag requirement (CSVM). Also for these distributions a good agreement between data and MC has been achieved.

b jets event categorization Events selected are separated into categories whether zero, one or two of the selected jets pass the medium CSV b tagging WP. Since medium CSV b tagging WP has an expected misidentification probability rate of the 1% (section 3.2.6), these categories have different background composition, are orthogonal among them and have a very different signal sensitivity. They can be defined in this way:

Figure 4.6: Distributions of the *leading* b jet: p_T (a) and CSV discriminator (b).Figure 4.7: Distributions of the *subleading* b jet: p_T (a) and CSV discriminator (b).

- **2jets–0tag** Events belong to this category if neither the leading nor the subleading jet passes the medium CSV working point. This category is dominated by backgrounds and is not very sensitive to the signal events.
- **2jets–1tag** Events are in this category if the leading but not the subleading jet passes the medium CSV working point. This category contains half of the signal and it is still dominated by the background.
- **2jets–2tags** Events belong to this category, if both the leading and subleading jets pass the medium CSV working point. This is the most signal-sensitive category.

In addition in this work other three categories are used and defined as follows:

- **inclusive** In this kind of events only the baseline selection for $h \rightarrow \tau_h \tau_h$ is applied without any requirement on jets.
- **2jets–inclusive** Events are collected in this category if there are at least two jets, without any b tagging requirement; then it combines the three categories previously defined.
- **2jets-1tag loose, 2jets-2tags loose** These two categories are added because are useful for background estimation with data–driven methods; the same definition of **2jets–1tag/2tags** is used, but a *loose* WP for the CSV discriminator (CSVL - Table 3.4) is required (section 3.2.6).

In Fig. 4.8 the distribution of the number of b jets passing the CSV medium working point for the inclusive category is reported, showing a good agreement between data and MC.

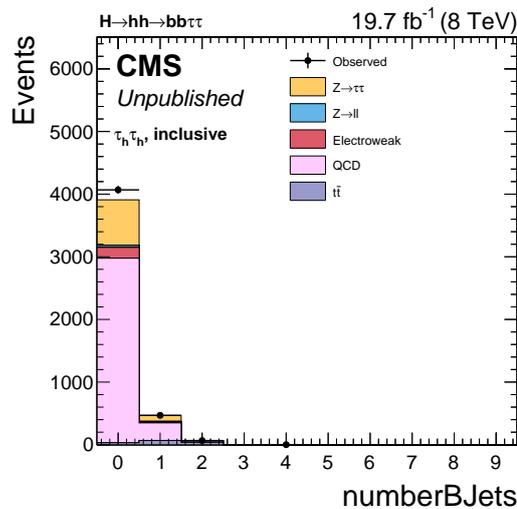


Figure 4.8: Distribution of the number of b jets passing the CSV medium working point for the inclusive category.

Study at generator level on bjets

This study has been performed using the $H \rightarrow hh \rightarrow bb\tau\tau$ MC sample with $M_H = 300$ GeV to determine the criterion to identify the best b jet pair on signal events. As first step, the match between real and reco b jets has been studied, evaluating how many events have two “preselected” jets matched with MC truth b jets. The “preselected” jets are defined as jets to which the selection described in 4.5.2 has been applied, in order to select a pair of jets forming a h candidate and a $\Delta R < 0.3$ has been required between objects. The resulting percentage of events with matched jets relative to the total number of events is around 68% in 2jets-inclusive category.

For this study only events with 2 “preselected” jets matched with b jets at generator level are used. Three methods have been investigated to determine the best criterion to order jets, in a way that the choice of the leading and subleading jets give the best performance:

- 2 jets with highest CSV discriminator values
- 2 jets with the highest p_T values
- The jet pair which provides the minimum χ^2 of the kinematic fit used to reconstruct the four body invariant mass, M_H , for the H candidate (section 4.5.4)

Performance of the three methods is evaluated in terms of purity, defined as the ratio between the number of events with both selected b jets matched with real b jets and the number of events with two “preselected” jets that match two MC truth b jets.

The results of this study are reported in Table 4.4. The best performance is obtained in all categories using the CSV method.

Event category	Selection “purity” (%)		
	CSV-based	p_T -based	χ^2 -based
2jets-inclusive	84.3	53.1	70.2
2jets-0tag	64.9	28.9	45.5
2jets-1tag	76.5	43.6	62.5
2jets-2tags	98	96.5	99.5

Table 4.4: Purity of b jets selected for each category for different methods.

4.5.3 $H \rightarrow hh \rightarrow bb\tau\tau$ selection

The presence of two SM-like Higgs bosons h in the signal events makes consistent the requirement that the invariant mass of the tau pair and the b jet pair are both around 125 GeV.

In the case of the $h \rightarrow \tau_h \tau_h$, due to the presence of two neutrinos from the tau decays in the final state, the invariant visible mass, $M_{\tau\tau}^{vis}$, calculated using the visible tau products, does not peak in the right position and does not allow to discriminate the $h \rightarrow \tau_h \tau_h$ signal from the background, in particular in the case of $Z \rightarrow \tau_h \tau_h$. In order to improve the resolution on mass, the CMS algorithm SVFit [144], exploiting the information of \vec{E}_T^{miss} and of visible products of the τ decays, is used. SVFit is a likelihood-based algorithm, that reconstructs the invariant mass of the h tau pair candidate, $M_{\tau\tau}^{svFit}$, recovering the information of the missing transverse energy on a event by event basis. In figure 4.9 are reported the candidate pair $\tau_h \tau_h$ distributions of the invariant visible mass, $M_{\tau\tau}^{vis}$ (a) and of the invariant mass reconstructed by SVFit, $M_{\tau\tau}^{svFit}$ (b), for signal $H \rightarrow hh \rightarrow bb\tau\tau$ (in blue) and $Z \rightarrow \tau\tau$ plus jets (in red) events. The SVFit improvement on discrimination power is clearly visible. Further details can be found in appendix D.

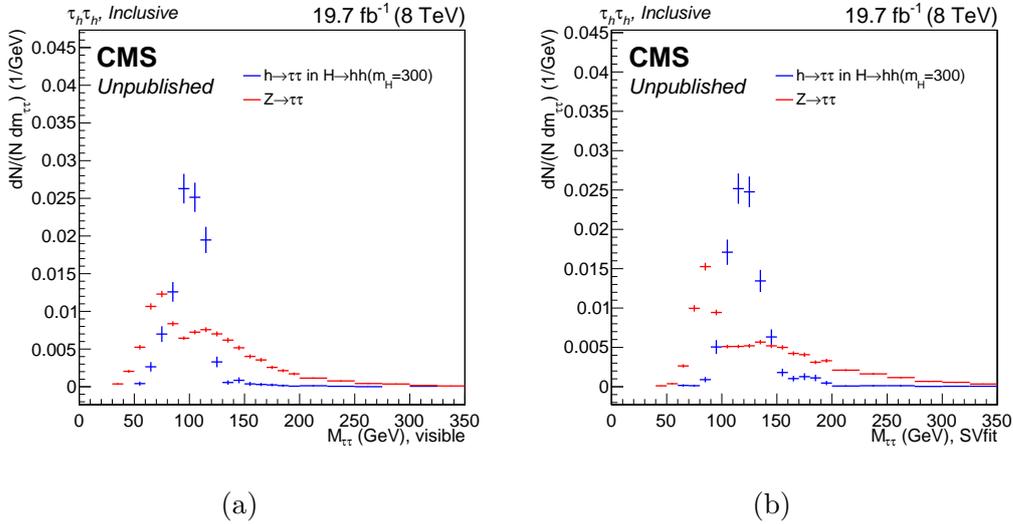


Figure 4.9: Distributions of (a) visible invariant mass of the τ pair candidates and of (b) SVFit invariant mass of the tau pair candidate, for signal $h \rightarrow \tau\tau$ from $H \rightarrow hh$ (in blue) and $Z \rightarrow \tau\tau$ (in red). Distributions are renormalized to the unit area.

In the case of the $h \rightarrow bb$, the M_{bb} is the invariant mass of the two b jets candidates.

In order to keep most of the signal events and reject most of the backgrounds, especially the SM Higgs processes (section 5.4), in all categories it is also required that signal events have to be within the mass windows:

$$90 \text{ GeV} < M_{\tau\tau}^{svFit} < 150 \text{ GeV} \quad (4.2)$$

$$70 \text{ GeV} < M_{bb} < 150 \text{ GeV} \quad (4.3)$$

In Fig. 4.10 the distributions of $M_{\tau\tau}^{svFit}$ (left) and M_{bb} (right), before applying the cut and for each category are reported, for modelled contributions and data events.

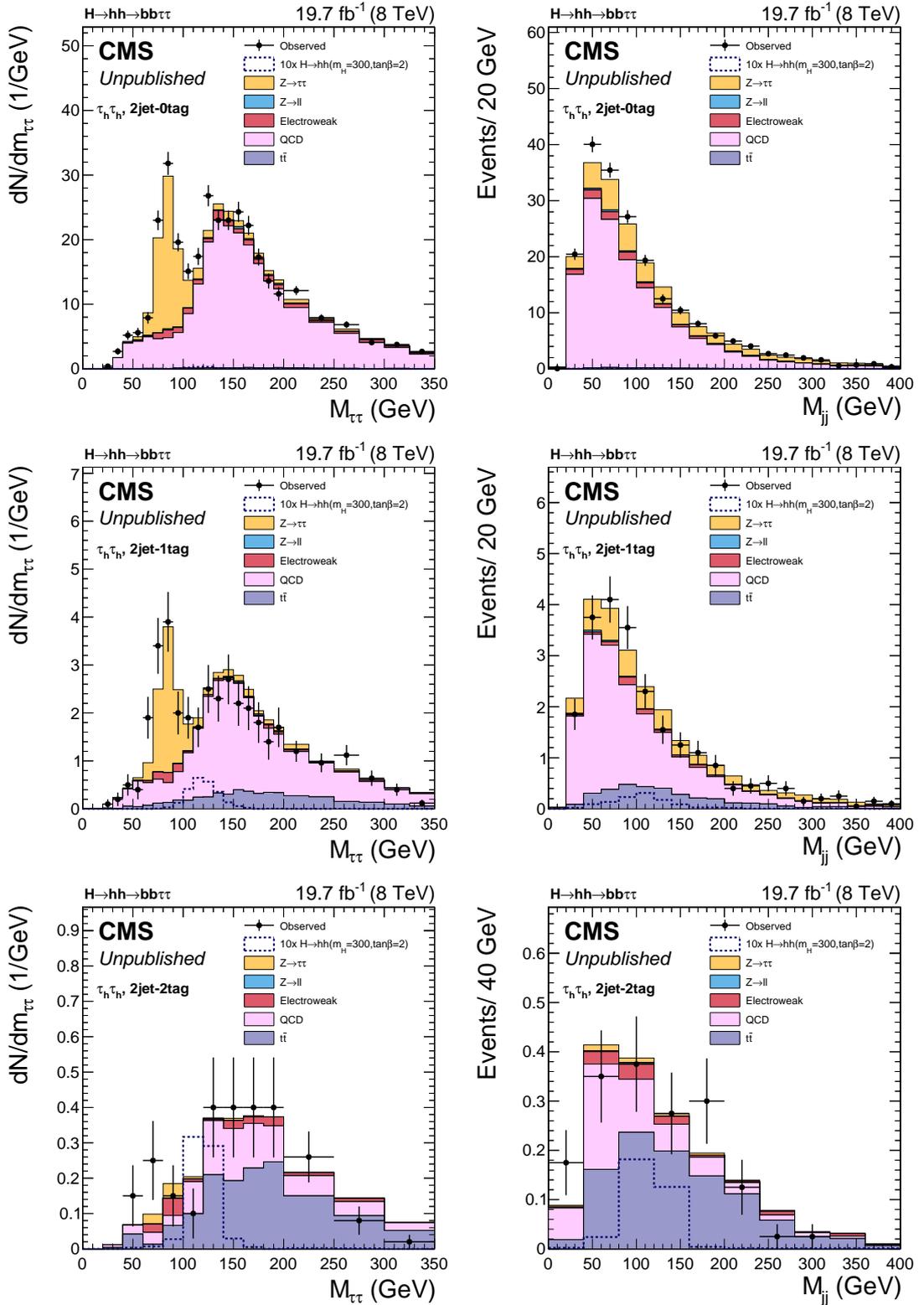


Figure 4.10: Distributions of $M_{\tau\tau}^{svFit}$ (left) and M_{bb} (right) in the $\tau_h\tau_h$ channel. The plots are shown for events in the 2jets—0tag (top), 2jets—1tag (centre) and 2jets—2tags (bottom) categories before applying the mass window requirement. The expected signal multiplied by a factor 10 is shown superimposed as an open dashed histogram for $\tan\beta = 2$ and $M_H = 300$ GeV in the “low- $\tan\beta$ -high” scenario of the MSSM. These distributions are divided by the bin width.

4.5.4 H \rightarrow hh reconstruction

The final step of the signal selection is the reconstruction of the heavy Higgs boson H. The presence of two SM-like Higgs bosons h, consistent with a mass around 125 GeV can be exploited not only applying the “rectangular” cut on the masses of the tau and b jet pairs.

Two approaches are considered to reconstruct the mass of the H candidate (M_H) based on a different estimate of the H 4-momentum.

The first approach reconstructs the H 4-momentum as the sum of the 4 visible objects (two τ_h and two b jets) 4-momenta plus the missing transverse energy ($\vec{E}_T^{\text{miss}} = (\cancel{E}_x, \cancel{E}_y, 0, 0)$) 4-momentum. This approach delivers a robust estimator of the signal, the H candidate invariant mass, $M_{\tau\tau+jj}$, calculated using the H 4-momentum defined above. However, mostly due to the \vec{E}_T^{miss} , the mass resolution is relatively low, as is shown on the left of Fig. 4.11, where $M_{\tau\tau+jj}$, is reported.

In the second approach, a *kinematic fit* procedure [145], which exploits the known event topology, is used. The kinematic fit tool puts the constraint that in signal events $M_{bb} = M_{\tau\tau} = 125$ GeV in order to shift the energy of each final object within its measurement uncertainties to bring the mass of each individual pair as close as possible to 125 GeV. Thanks to the fact that the h boson natural width is of the order of few MeV, the difference in mass are due only to the detector resolution effects.

In the fit procedure, the varied observables are the energies of the taus and of the b jets. The deviations between the measured and the varied observables are normalized to their uncertainties, giving the possibility to construct a χ^2 function. The χ^2 is then minimised in the fitting procedure, giving the possibility to understand how compatible the analysed event is compared to the hypothesis made.

For the b jets it is varied only the energy because their directions, η and ϕ , can be considered very well reconstructed. While the energy of the second b jet can be obtained from the mass constraint and varying the energy of the first b jet:

$$M_h^2 = p_{b_1}^2 + p_{b_2}^{2,\text{new}} + 2p_{b_1}p_{b_2}^{\text{new}} = m_{b_1}^2 + \frac{E_{b_2}^{2,\text{new}}}{\gamma_{b_2}^2} + 2E_{b_1}E_{b_2}^{\text{new}} \left(\underbrace{1 - \vec{\beta}_{b_1}\vec{\beta}_{b_2}}_c \right). \quad (4.4)$$

c is assumed to be constant (at first approximation $\vec{\beta} = \vec{p}/E$ is considered constant), and it can be derived from the kinematics before applying the fit:

$$c = \frac{M_h^2 - m_{b_1}^2 - m_{b_2}^2}{2E_{b_1}E_{b_2}}, \quad (4.5)$$

From the eq. 4.4 under these assumptions the energy of the second b jet is equal

to:

$$E_{b_2}^{\text{new}} = E_{b_1} c\gamma_{b_2}^2 \left(-1 + \sqrt{1 + \frac{M_h^2 - m_{b_1}^2}{(E_{b_1} c\gamma_{b_2})^2}} \right). \quad (4.6)$$

During the fit procedure the mass of the i^{th} bjet can be scaled according to its energy:

$$m_{b_i}^{\text{new}} = m_{b_i} \frac{E_{b_i}^{\text{new}}}{E_{b_i}} \quad (4.7)$$

For each b jet a χ^2 term can be built to quantify how much the measured energy of the b jet i (with $i = 1$ or 2) is modified by the fit:

$$\chi_{b_i}^2 = \left(\frac{E_{b_i}^{\text{fit}} - E_{b_i}^{\text{meas}}}{\sigma_{b_i}} \right)^2 \quad (4.8)$$

$E_{b_i}^{\text{meas}}$ is the measured b jet energy and σ_{b_i} its energy resolution. The uncertainties σ_{b_1} and σ_{b_2} are evaluated from a MC sample as a function of η and p_T as reported in the following analysis [146].

Due to the presence of a big contribution of \vec{E}_T^{miss} coming from the neutrinos present in the τ decays, the energy of the $h \rightarrow \tau_h \tau_h$ is underestimated. For this reason the treatment of the τ pair follows a slightly different procedure. The compatibility of ditau mass with 125 GeV SM-like Higgs boson allows the possibility to use the collinear approximation and consider the directions of the visible decay products to be the same of the original τ leptons, thanks to the high boosted energy of the taus. Moreover the measured direction is very accurate and reliable, so it is considered constant as for b jets. Using the mass constraint only the energy of one tau is varied in the fit procedure and also m_τ is considered constant.

Due to the presence of initial state radiation and underlying events, the heavy Higgs boson H may have a non-zero transverse momentum, $\vec{p}_{T,H}^{\text{meas}}$. The reconstructed recoil ($\vec{p}_{T,\text{recoil}}^{\text{meas}}$) defined as a momentum in the transverse plane that is opposite to the H candidate momentum, satisfies the following equation:

$$\vec{p}_{T,\text{recoil}}^{\text{meas}} = -\vec{p}_{T,\text{miss}}^{\text{meas}} - \vec{p}_{T,b_1}^{\text{meas}} - \vec{p}_{T,b_2}^{\text{meas}} - \vec{p}_{T,\tau_1}^{\text{meas}} - \vec{p}_{T,\tau_2}^{\text{meas}} = -\vec{p}_{T,H}^{\text{meas}} \quad (4.9)$$

where the $\vec{p}_{T,\text{miss}}^{\text{meas}}$ is the missing transverse energy, reconstructed as described in section 3.2.7; the subscript *vis* indicates that for the τ only the visible decay products are used. In order to obtain the correct τ energies, the fitted, $\vec{p}_{T,H}^{\text{fit}}$, and the reconstructed heavy Higgs transverse momenta, $\vec{p}_{T,H}^{\text{meas}}$, are required to be quite close.

The recoil is calculated by the fit as:

$$\vec{p}_{\text{T,recoil}}^{\text{fit}} = -\vec{p}_{\text{T},b_1}^{\text{fit}} - \vec{p}_{\text{T},b_2}^{\text{fit}} - \vec{p}_{\text{T},\tau_1}^{\text{fit}} - \vec{p}_{\text{T},\tau_2}^{\text{fit}} = -\vec{p}_{\text{T},H}^{\text{fit}} \quad (4.10)$$

Any nonzero residual vector $\vec{p}_{\text{T,recoil}}^{\text{res}} = \vec{p}_{\text{T},H}^{\text{fit}} + \vec{p}_{\text{T,recoil}}^{\text{meas}}$ contributes to a χ^2 term as follows

$$\chi_{\text{recoil}}^2 = \vec{p}_{\text{T,recoil}}^{\text{res},T} \cdot V_{\text{recoil}}^{-1} \cdot \vec{p}_{\text{T,recoil}}^{\text{res}} \quad (4.11)$$

V_{recoil} denotes the 2×2 covariance matrix of the reconstructed recoil vector:

$$V_{\text{recoil}} = V_{\vec{p}_{\text{T},\text{miss}}} - V_{b_1} - V_{b_2} \quad (4.12)$$

The total χ^2 function of the kinematic fit is the sum all the individual χ^2 contributions:

$$\chi_{\text{tot}}^2 = \chi_{b_1}^2 + \chi_{b_2}^2 + \chi_{\text{recoil}}^2 \quad (4.13)$$

χ_{tot}^2 is a function of the variables E_{b_1} and E_{τ_1} .

After the iterative minimisation of the χ_{tot}^2 (HHKinFit package [145]), the kinematic fit provides the corrected 4-momentum of the taus and b jets which can be summed to determine a more precise measurement of the heavy Higgs boson mass. The resulting heavy Higgs boson mass is called M_H^{kinfit} . This notation will be used throughout this thesis. Sometime the fit may not converge, which means that the signal object combination does not provide a mass M_h around 125 GeV. This happens rarely (less than 1%) after the mass window requirement for the tau pair and the b jet pair (section 4.5.3).

The distributions of the reconstructed 4 body mass M_H without (left, $M_{\tau\tau+jj}$) and with (right, M_H^{kinfit}) the kinematic fit applied for each category are reported in Fig. 4.11. This figure shows, as expected, that the mass resolution for signal events is quite improved using the kinematic fit, while for the background the M_H distribution is still wide and quite unchanged. The sharp cut at $M_H = 250$ GeV is due to the threshold imposed by the kinematic fit $M_{\tau\tau} = M_{bb} = 125$ GeV and corresponding to $2 \times M_h = 250$ GeV.

Thanks to these properties the kinematic fit mass is chosen as the final discriminant for the signal extraction (section 5.2), since it provides a sharp and narrow peak of the M_H distribution for the signal events and a broader M_H distribution for the backgrounds events.

A summary of the complete event selection is reported in Table 4.5.

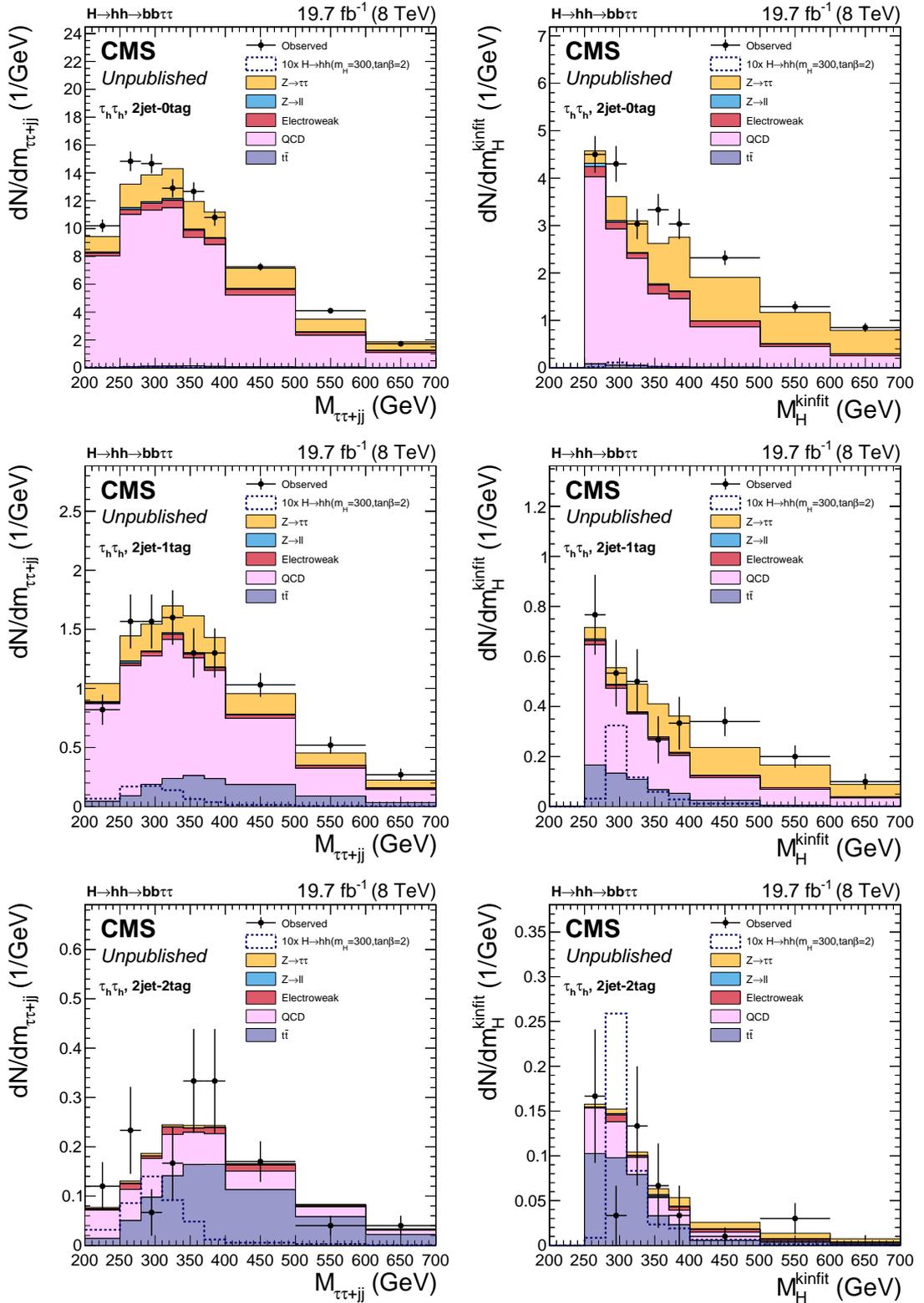


Figure 4.11: Distributions of the reconstructed 4 body mass without (left, $M_{\tau\tau+jj}$) and with (right, M_H^{kinfit}) the kinematic fit applied in the $\tau_h\tau_h$ channel. The plots are shown for events in the 2jets—0tag (top), 2jets—1tag (centre) and 2jets—2tags (bottom) categories. The expected signal multiplied by a factor 10 is shown superimposed as an open dashed histogram for $\tan\beta = 2$ and $M_H = 300$ GeV in the “low- $\tan\beta$ -high” scenario of the MSSM. These distributions are divided by the bin width.

H → hh → bbτ_hτ_h	
τ_h	$p_T > 45 \text{ GeV}, \eta < 2.1$
	$I_\tau < 1 \text{ GeV}$
	Loose anti- <i>e</i> Loose anti- <i>μ</i> No extra electrons or muons
h → τ_hτ_h	Opposite charge
	$90 < M_{\tau\tau}^{svFit} < 150 \text{ GeV}$
Jets	PF jet, $p_T > 20 \text{ GeV}, \eta < 2.4$ pile-up jet ID
b jets	CSV Medium WP
h → bb	$70 < M_{bb} < 150 \text{ GeV}$
H → hh	Kinematic Fit convergence

Table 4.5: H → hh → bbτ_hτ_h, *fully hadronic* channel, requirements.

4.6 Data driven Background modelling

In this section the data-driven methods used to model the $Z \rightarrow \tau\tau$ plus jets² and the QCD multijets³ will be described.

Due to the presence of at least two jets in the event, the composition of these backgrounds, is expected to change in each category. In the 2jets–0tag the major contribution is expected from QCD and $Z \rightarrow \tau\tau$. Meanwhile, the contribution is expected to be drastically reduced in the 2jets–2tags category, where $t\bar{t}$ (modelled by MC, section 4.4.1) background becomes dominant for the presence of two genuine b jets.

These data-driven techniques are applied in all the steps of the selection, starting from the *baseline* selection (section 4.5). Depending on the step, a specific “reference” distribution is used to estimate the shape and the normalization of each data-driven background in the signal region.

In particular, after the *baseline* selection (section 4.5), the $M_{\tau\tau}^{svFit}$ distribution is used for modelling the data-driven backgrounds. While at the end of the selection, the distribution used for the estimation is M_H^{kinfit} , since this distribution is the most sensitive for the signal extraction.

In the following sections the procedures to model these backgrounds, their shapes and normalization, and their reliability are presented.

²This background will be indicated as $Z \rightarrow \tau\tau$ omitting “plus jets”.

³This background will be indicated as QCD omitting “multijets”.

4.6.1 $Z \rightarrow \tau\tau$

The technique utilized to model this background has been optimized for the SM [71] and MSSM [68] $h \rightarrow \tau\tau$ searches. The modelling relies on the use of an *embedded* sample, constructed using a data sample of $Z \rightarrow \mu\mu$ and MC sample of $Z \rightarrow \tau\tau$. The two opposite charge loose PF isolated muons (section 3.2.3) from the Z decay, are replaced by two MC generator level taus. The signals left in the detector by muons are removed and the muons 4-momenta are assigned to the taus. Tau decays are then simulated by TAUOLA [147]. The advantage to use the embedded sample comes from the fact that the majority of the objects are taken from data, as a consequence the behaviour of the underlying events is better described and the systematic uncertainties on jets and \cancel{E}_T are reduced.

In order to avoid any possible bias or contamination from $Z \rightarrow \mu\mu$ selection, an additional weight is applied to each embedded event, called *embedded weight*. It is defined using $Z \rightarrow \mu\mu$ MC sample as the ratio of events that pass the selection applied for $Z \rightarrow \mu\mu$ to the total number of events in the sample depending on p_T and η of each muon.

The total yield of the $Z \rightarrow \tau\tau$ background is estimated in the inclusive category counting the total number of events of the MC $Z \rightarrow \tau\tau$ sample entering in the distribution of M_H^{kinfit} ($N_{\text{inclusive}}(Z \rightarrow \tau\tau)^{MC}$). To determine the yield in each category, the extrapolation factor, $\epsilon_{\text{category}}$, to pass the requirements of the category (respectively 2jets-0tag, 2jets-1tag, 2jets-2tags) is evaluated:

$$\epsilon_{\text{category}} = \frac{N_{\text{category}}(Z \rightarrow \tau\tau)^{\text{embedded}}}{N_{\text{inclusive}}(Z \rightarrow \tau\tau)^{\text{embedded}}} \quad (4.14)$$

where $N_{\text{inclusive}}(Z \rightarrow \tau\tau)^{\text{embedded}}$ and $N_{\text{category}}(Z \rightarrow \tau\tau)^{\text{embedded}}$ are obtained counting the respective number of events of the embedded sample entering in the distribution of M_H^{kinfit} .

Then the yield in each category is obtained using the following equation:

$$N_{\text{category}}(Z \rightarrow \tau\tau) = N_{\text{inclusive}}(Z \rightarrow \tau\tau)^{MC} \cdot \epsilon_{\text{category}} \quad (4.15)$$

For the shape of this background, the distribution of M_H^{kinfit} in the embedded sample is taken for each category. The presence of at least 2 jets in the event and the b tagging requirements reduce a lot the $Z \rightarrow \tau\tau$ contribution. Due to the low event statistics, the shape of the 2jets-1tag and 2jets-2tags categories are evaluated using the 2jets-1tag/2tags-loose categories (section 4.5.2), where b tagging requires a looser WP (Table 3.4). The shape, obtained using the loose b tagging for the distributions of M_H^{kinfit} and shown in Fig. 4.12, is compatible with the one obtained with the medium b tagging WP, providing a Kolmogorov-Smirnov test [148] greater than 95%. This ensures that no bias is introduced by relaxing the b tag requirements.

Since the embedded samples of $Z \rightarrow \tau\tau$ was originated from di-muon data, a

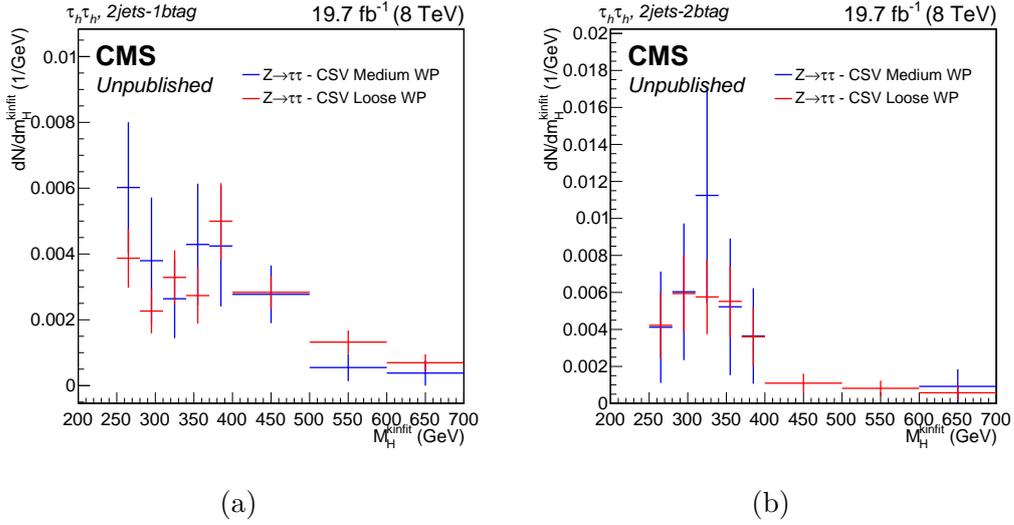


Figure 4.12: Comparison of $Z \rightarrow \tau\tau$ embedded shapes of the M_H^{kinfit} distribution, requiring CSV loose working point (red) and CSV medium (blue) in the 2jets–1tag (a) and 2jets–2tags (b) categories.

considerable contamination of $t\bar{t} \rightarrow \mu\mu + X$ events is present. This contribution cannot be neglected because the $t\bar{t}$ background is one of the most relevant background, especially when the presence of two b jets is required. In order to estimate the contamination in the $Z \rightarrow \tau\tau$ embedded sample, a $t\bar{t} \rightarrow \mu\mu + X$ MC sample is produced with the embedding procedure applied this time on a $t\bar{t}$ MC sample. The $t\bar{t}$ embedded contamination, like for instance in the case of $M_{\tau\tau}^{\text{svFit}}$ distribution, amounts to $\sim 36\%$ in 2jets–1tag and to $\sim 97\%$ in 2jets–2tags categories, after the baseline selection, where the $h \rightarrow \tau_h\tau_h$ is reconstructed. In Fig. 4.13 the $t\bar{t}$ contamination in the $Z \rightarrow \tau\tau$ embedded sample is shown in the 2jets–1tag (4.13a) and 2jets–2tags (4.13b) categories using the $M_{\tau\tau}^{\text{svFit}}$ distribution.

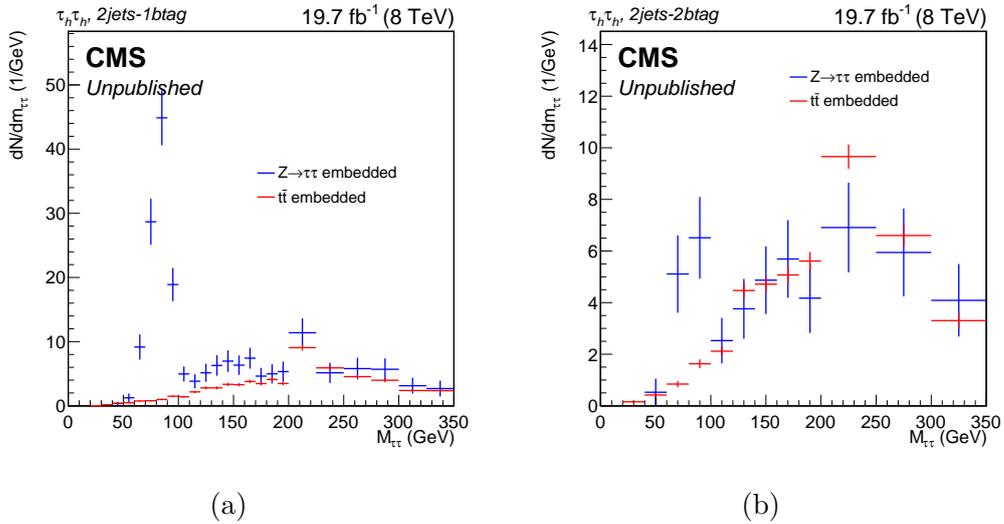


Figure 4.13: $t\bar{t}$ contamination (red) in $Z \rightarrow \tau\tau$ embedded sample (blue) by the $t\bar{t}$ process (red) is shown in the 2jets–1tag (a) and 2jets–2tags (b) categories for the $M_{\tau\tau}^{\text{svFit}}$ distribution.

The $t\bar{t}$ embedded sample uses as normalization the cross-section of the $t\bar{t}$ process (the same used for the normalization of the $t\bar{t}$ MC sample of section 4.4.1). Summarizing, in the modelling of the $Z \rightarrow \tau\tau$, the $t\bar{t}$ embedded background is subtracted from the $Z \rightarrow \tau\tau$ embedded sample.

4.6.2 QCD

This background requires a very huge sample and as a consequence, a prohibitive time machine to be properly modelled by MC simulation. In this case the use of a data modelling technique is mandatory.

The technique assumes a similar behaviour for QCD processes when di-tau candidates are reconstructed with opposite-charge (opposit-sign, OS) or same-charge (same-sign, SS) or when one of the tau candidates is not isolated. Under this hypothesis, in order to properly estimate the shape and yield of the QCD background in the signal region, three QCD enriched control regions are defined, applying some changes in the requirements of the baseline selection.

The regions are then defined:

- **OS isolated:** this is the signal region, where the two τ_h candidates are required to have opposite charge and to be isolated. The shape and yield of the QCD model must be estimated in this region.
- **SS isolated:** the opposite charge requirement for the τ_h pair is inverted to require the same charge.
- **SS anti-isolated:** the τ_h are required to have same sign (SS). One tau passes the isolation cut ($I_\tau < 1$, Medium WP), the other is not isolated with I_τ between 1 and 4 GeV, to avoid any overlap with SS isolated control region.
- **OS anti-isolated:** Opposite charge di- τ_h events are required (OS). One tau passes the isolation cut ($I_\tau < 1$, Medium WP), the other is not isolated with I_τ between 1 and 4 GeV, to avoid the overlap with **OS isolated**, corresponding to the signal region.

A schematic picture of these regions is reported in Fig. 4.14

The choice of the isolation range for the anti-isolated regions is determined by the fact that biases affecting the shape, when isolation is too far from the signal distribution, or a too low statistics sample, affecting the uncertainties on the background (yield and shape), when the range is too narrow, must be avoided.

The figure 4.15a shows that in the inclusive category the QCD yield does not depend on the τ isolation, for both OS and SS regions, confirming the hypothesis assumed to grant the absence of any bias due to a dependence of the QCD yield on the τ isolation.

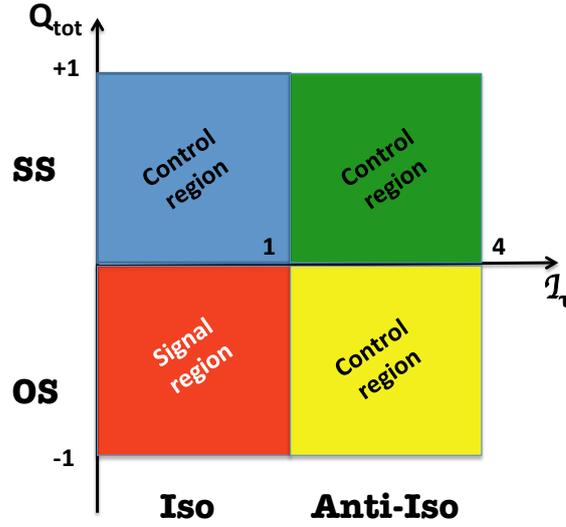


Figure 4.14: A schematic representation of the QCD signal and control regions.

However, due to the significant amount of the fake di- τ_h candidates, originated from di-jets in QCD processes, the number of events in the OS and in the SS regions are not equal. This difference may increase with b tag requirement due to the fact that in the $b\bar{b}$ events there is a bias towards opposite sign combination.

Another assumption made, is that the ratios of the yields between isolated and anti-isolated regions for SS ($N_{iso}^{SS}/N_{anti-iso}^{SS}$) and OS ($N_{iso}^{OS}/N_{anti-iso}^{OS}$) regions are the same. This assumption is verified in the inclusive category, where the contribution of the signal is expected to be negligible, after having applied mass window cuts (section 4.5.3) and kinematic fit convergence (section 4.5.4). The ratio $N_{iso}^{OS}/N_{anti-iso}^{OS}$ is equal on average to 0.25 ± 0.02 (stat.); the ratio $N_{iso}^{SS}/N_{anti-iso}^{SS}$ is equal on average to 0.23 ± 0.02 (stat.). To exclude possible b tag dependency effect, these ratios have been evaluated for each b tag category and compatible values with the previous have been found.

The shape and yield used to model the QCD background in the signal region, are determined utilizing the QCD contribution measured in the OS anti-isolated control region ($N_{anti-iso}^{OS}$), because the main background contribution in this region is given by the QCD. This can be observed in Fig. 4.15b, where the distribution of $M_{\tau\tau}^{svFit}$ is shown with all contributing backgrounds but QCD.

The QCD model yield in the signal region (N_{iso}^{OS}) is obtained using the extrapolation factor, calculated in each b tag category, to properly model the QCD background in the different category, $N_{SS-iso}^{cat}/N_{SS-anti-iso}^{cat}$, through the relation:

$$N_{OS-iso}^{cat} = N_{OS-anti-iso}^{cat} \cdot \frac{N_{SS-iso}^{cat}}{N_{SS-anti-iso}^{cat}}, \quad (4.16)$$

The QCD contribution in each control region, is evaluated subtracting from the

data the contribution of all non-QCD backgrounds, estimated using MC samples (including MC $Z \rightarrow \tau\tau$) in the 2jets–0tag category, while in the 2jets–1tag and 2jets–2tags categories, due to its low contribution the $W + \text{jets}$ is not subtracted. This can be applied because at the most only one real τ_h may come from W decay and the other three jets are misidentified as a τ_h plus two b jets.

Furthermore, due to the low event statistics in the 2jets–1tag and 2jets–2tags categories, the shapes in OS anti-isolated region and the $N_{SS-iso}^{cat}/N_{SS-anti-Iso}^{cat}$ extrapolation factor are calculated in 2jets–1tag–loose and 2jets–2tags–loose. The looser b tagging requirement does not introduce any bias neither in the shape nor in the SS extrapolation factor but it allows to increase the statistics, a smoother shape and more precise QCD background modelling, as shown in Fig. 4.15c and Fig. 4.15d, where the comparison of loose and tighter QCD shapes is reported.

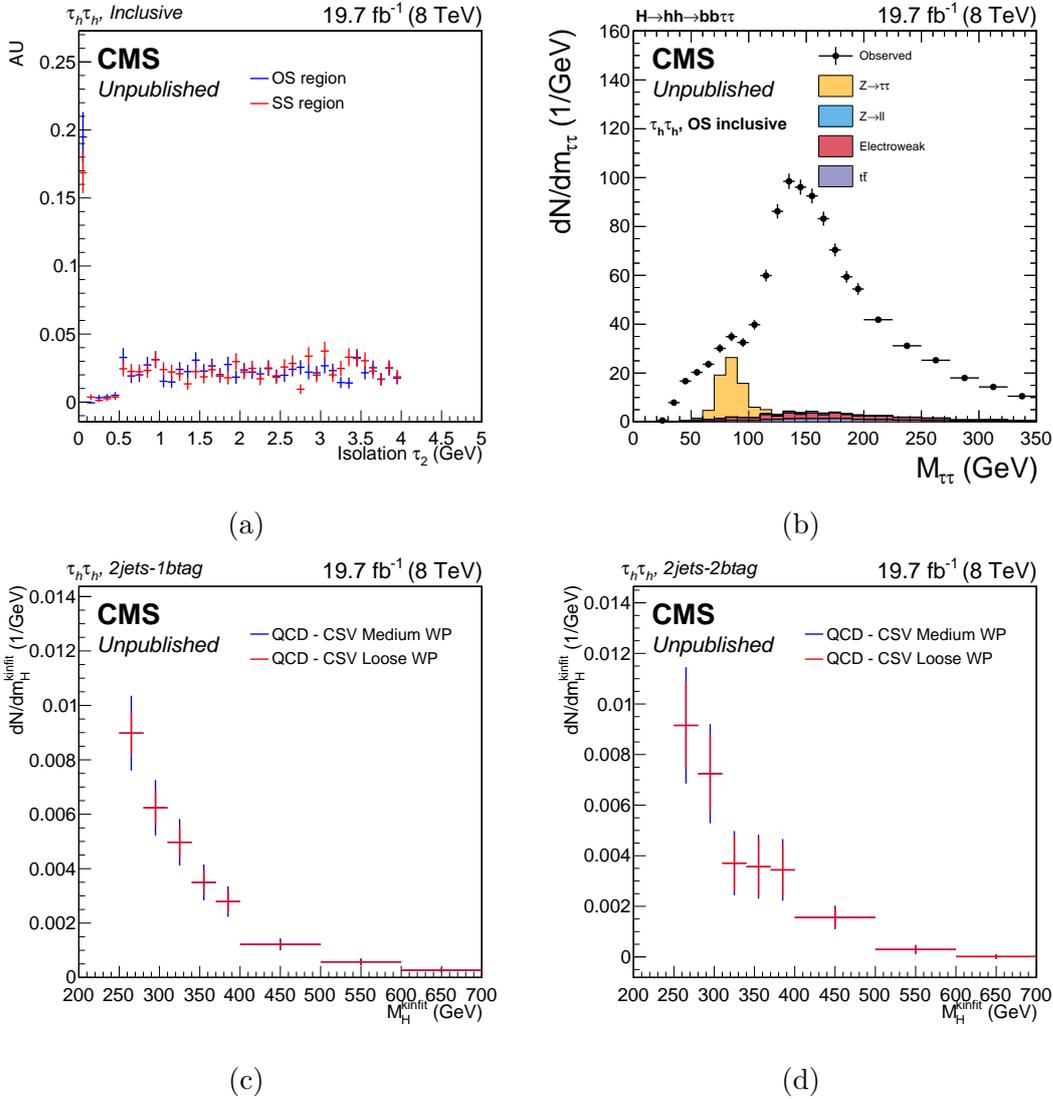


Figure 4.15: (a) Distribution of isolation variable of the subleading τ_h for QCD multijets events in OS (blue) and SS (red) regions in the inclusive category. The mass window cut and kinematic fit are also applied. (b) Distribution of $M_{\tau\tau}^{svFit}$ in the OS anti-isolated control region for the 2jets-inclusive category. The missing background contribution is considered coming from the QCD multijets process. Comparison of QCD multijets shapes of the M_H^{kinfit} distribution, requiring CSV loose working point (red) and CSV medium (blue) in the 2jets-1tag (c) and 2jets-2tags (d) categories.

Chapter 5

Results

The reconstructed H candidate masses for the modelled signal, backgrounds and data are used as input to a fit to extract the signal contribution. The statistical procedure employed was developed by the LHC Higgs Combination Group [149] and adopted by both ATLAS and CMS experiments. The systematic uncertainties considered for this analysis are discussed. The fit takes into account the statistical and the systematic uncertainties. The model independent 95% CL exclusion limits on the production cross-section of a heavy scalar neutral H boson times the branching fraction, $\sigma(gg \rightarrow H \rightarrow) \times B(hh \rightarrow bb\tau\tau)$ are set as a function of M_H . These exclusion limits are also interpreted in the two benchmark scenarios, respectively the “low- $\tan\beta$ -high” MSSM and the 2HDM Type II.

5.1 Systematic uncertainties

The modelling of signal and backgrounds in the distribution of M_H^{kinfit} , presented in the previous chapter, is affected by the systematic uncertainties. These can be divided in two different types:

- **Normalisation uncertainties**, affecting the yield of a particular background or signal
- **Shape uncertainties**, affecting also the shape of M_H^{KinFit} distributions for signal and background, used as input in the fit procedure (section 5.2). In this case the number of events in a particular bin (the shape) is also influenced.

A detailed description of all the systematic uncertainties considered in the analysis is described in the following. The summary of all the systematic sources and the corresponding uncertainties, considered for this analysis, is reported in the Table 5.1.

Affected Quantity	Source	Systematic Uncertainties (%)		
		2jets-0tag	2jets-1tag	2jets-2tags
Yield	Luminosity	2.6	2.6	2.6
	τ_h identification and trigger efficiencies	19	19	19
	b tagging efficiency	4-5	2	4-5
	b mistag rate	–	2.5	–
	Z cross-section	3.3	3.3	3.3
	$Z \rightarrow \tau\tau$ signal extraction	5.7	25.2	175
	$Z \rightarrow \tau\tau$ - $t\bar{t}$ contamination	–	5	49
	$t\bar{t}$ cross-section	10	10	10
	$Z \rightarrow ll$: jet and l misidentification	30	52	67
	Diboson plus single top cross-sections	15	15	15
W plus jets cross-section	20	–	–	
QCD Multijets	10	20	40	
Shape	Tau energy scale	M_H^{KinFit} distribution		
	Jet energy scale	M_H^{KinFit} distribution		

Table 5.1: Summary of the systematic uncertainties affecting the yields and the shape for signal and backgrounds of the M_H^{KinFit} distribution in the different categories as described in sections 5.1.1 and 5.1.2. Some uncertainties may vary with the samples, so in this case ranges are given. W plus jets uncertainties are evaluated only in the 2jets-0tag category, the other two categories being included in the QCD multijets systematic (section 5.1.1). The uncertainties less than 1% are not reported and are replaced by a “–”.

5.1.1 Normalization uncertainties

The sources of systematic which affect the yields for background or signal processes are the following:

- **Luminosity**

The relative uncertainty on the luminosity amounts to 2.6% for 2012 data ([150], Appendix A). This uncertainty is applied to the signal and to all the backgrounds modelled by MC. The relative normalization uncertainty is equal to the relative uncertainty on the luminosity measurements.

- **τ_h identification and trigger efficiencies**

The scale factors for identification and trigger efficiencies (described in [68] and resumed in Appendix C) were used in order to obtain a good agreement between data and MC. The relative uncertainties of these scale factors are 8% for the τ_h identification and 4.5% for each τ_h candidate. These independent relative uncertainties have been summed in quadrature. They have been applied to signal samples and to all the MC backgrounds, when a genuine tau is present. The final relative uncertainty amount is 19% for all MC samples, except for the $W + \text{jets}$ background in which only one real τ_h is present and it amounts to 9.5%.

- **$\ell \rightarrow \tau_h$ and $jet \rightarrow \tau_h$ fake-rates**

The relative uncertainties on the $\ell \rightarrow \tau_h$ and $jet \rightarrow \tau_h$ fake-rates are respectively $\simeq 30\%$ and $\simeq 20\%$ [71]. These affect mostly $Z/\gamma^* \rightarrow \ell\ell$ background, as it will be described in the following paragraph.

- **b tagging efficiency and mistag rate**

The uncertainties on the b tagging efficiencies and mistag rates scale factors as function of the jet p_T and η for the CSVM discriminator (Table 3.4) are evaluated in [123, 151]. These scale factors have been estimated to correct disagreements in b tagging performance between data and MC. In this analysis the effect of their uncertainties on the yields has been evaluated by varying the scale factors applied within their recommended uncertainties ($\pm 1\sigma$ values). The corresponding overall change in the yields of signal and each MC background, for each category, has been evaluated after applying the promotion and demotion procedure. This procedure (described in Appendix C) is used to reclassify a fraction of b tagged jets as non b tagged jets, or vice versa, in order to obtain a better agreement data-MC. For the backgrounds with a variation greater than 1%, the yield change is used as the normalisation uncertainty. Since the b tagging and mistagging uncertainties are treated as independent sources of systematic in the signal extraction, the procedure is repeated four times.

- **Background Normalisation**

The simulated processes are affected by uncertainties on the cross-sections arising from their measurements or their theoretical predictions. Data driven

modelled background processes are in addition affected by statistical uncertainties in the extrapolation factor from the sideband region to the signal region (section 4.6). In this case this uncertainty is combined with the statistical uncertainty on the yield in the sideband region. The sources of systematic affecting the backgrounds normalization considered in this analysis are:

1. The modelling of the $Z/\gamma^* \rightarrow \tau\tau$ background uses a combination of embedded and MC samples (section 4.6.1). The relative uncertainty of 3% (**Z cross-section** in Table 5.1), arising from the measured cross-section of this process [139], is combined with the relative uncertainty on the cross-section of the simulated $t\bar{t}$ process, which contaminates the embedded sample (**Z $\rightarrow \tau\tau$ - $t\bar{t}$ contamination** in Table 5.1), of the order of 10% (**$t\bar{t}$ cross-section** in Table 5.1), and (**Z $\rightarrow \tau\tau$ signal extraction** in Table 5.1) with the relative uncertainty on the extrapolation factor (section 4.6.1).
2. The modelling of the $Z/\gamma^* \rightarrow \ell\ell$ ($\ell = e, \mu$) plus jets background fully relies on the correct simulation of the jet and lepton faking τ_h in this process. In this case the uncertainties in the τ_h misidentification probabilities are considered as source of systematic. There are three different contributions due to the probability that the selected τ_h is a misidentified electron, or muon, or jet. The information of the way how the fake τ_h was misidentified is accessible through the MC truth of a $Z \rightarrow \ell\ell$ ($\ell = e, \mu, \tau$) plus jets sample. Three classes of events are determined depending on such information: ZTT, ZL and ZJ. The ZL class includes the $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events, where the τ_h are faked by real electrons or muons. The ZJ class includes the $Z \rightarrow \ell\ell$ events where the τ_h are faked by quark or gluon jets. The ZTT includes the remaining $Z \rightarrow \tau\tau$ events, which are not included in the ZJ. Since the contributions of ZL and ZJ are strongly reduced by the requirement of at least two jets in the events, these contributions are added together in a unique sample called $Z \rightarrow \ell\ell$ with $\ell = e$ or μ . For each of these samples, the uncertainties in the τ_h misidentification probabilities are taken. The dominant contribution to the systematic uncertainty comes from the statistical uncertainty on the yield estimate of ZJ and ZL. To these statistical uncertainties, the fake-rate uncertainties, described above (see **$\ell \rightarrow \tau_h$ and $jet \rightarrow \tau_h$ fake-rates**), have been added in quadrature. The resulting systematic uncertainties (**Z $\rightarrow \ell\ell$: jet and ℓ misidentification** in Table 5.1) are 30% in 2jets-0tag, 52% in 2jets-1tag and 67% in 2jets-2tags categories.
3. The uncertainties on the measurements of the production cross-sections for the **di-boson** and the **single top** processes contribute to a relative systematic uncertainty which amounts to 15% (**diboson plus single top cross-sections** in Table 5.1).
4. The modelling of the **W plus jets** in 2jet-0tag category. The relative systematic uncertainty on the yield in this category (**W plus jet cross**

section, in Table 5.1), arising from the cross-section measurement of this process, is of 20%. The modelling of this process in the other categories was accounted by the data driven QCD which includes the contribution of this process (section 4.6.2).

5. The **QCD multijets** modelling: in this case the systematic uncertainty (**QCD multijets**, in Table 5.1) arises from the statistical uncertainty on the extrapolation factor and from the uncertainties on the cross-sections of the simulated background processes, subtracted in the QCD modelling procedure (section 4.6.2). In order to investigate if a shape uncertainty has to be taken into account and if the method used to estimate the QCD multijets background is biased by the choice of the control regions, alternative sideband regions enriched of QCD events are used. The QCD total yield is determined with these new sideband regions using the following equation (to be compared with the equation 4.16 of section 4.6.2 used to model the QCD background):

$$N_{OS-iso}^{cat-QCD-alt} = N_{SS-iso}^{cat} \cdot \frac{N_{OS-anti-iso}^{cat}}{N_{SS-anti-iso}^{cat}}, \quad (5.1)$$

The same procedure adopted in section 4.6.2 is then used to obtain the alternative QCD template (histogram). The shapes for the two methods, the standard (blue points) and the alternative (red points), are shown in Fig. 5.1. The shapes of the two methods are quite compatible in each category as provided by the Kolmogorov–Smirnov test [148] greater than 95%. This result confirms that a shape uncertainty can be neglected and also that the yield is not biased by the choice of the control region. Furthermore the choice of the control regions, is not a source of additional systematic, since the difference of the yields between the two methods is compatible with zero, as shown in the table 5.2.

Category	standard QCD	alternative QCD	Difference
2jets–0tag	261±22 (stat.)±4 (syst.)	261±22 (stat.)±4 (syst.)	0±29 (stat.)±0 (syst.)
2jets–1tag	29±4 (stat.)±1 (syst.)	37±8 (stat.)±1 (syst.)	-8.7±8.6 (stat.)±4 (syst.)
2jets–2tags	3.8±1.5 (stat.)±0.2 (syst.)	2.2±2.0 (stat.)±0.1 (syst.)	1.6±2.4 (stat.)±0.1 (syst.)

Table 5.2: QCD yields for the two methods with their statistical and systematic uncertainties and the difference of the yields for each category.

- **Theoretical Uncertainties**

The uncertainties for the signal processes, related to the theoretical calculations, are due to uncertainties in the parton distribution functions (PDF), variations of the renormalization and factorization scales, and uncertainties in the modelling of the underlying event and parton showers [68, 71].

The theoretical uncertainties also depend on the choice of signal hypothesis. For the model independent results (section 5.5) no choice of cross-section is made and hence no theoretical uncertainties are considered. For the MSSM

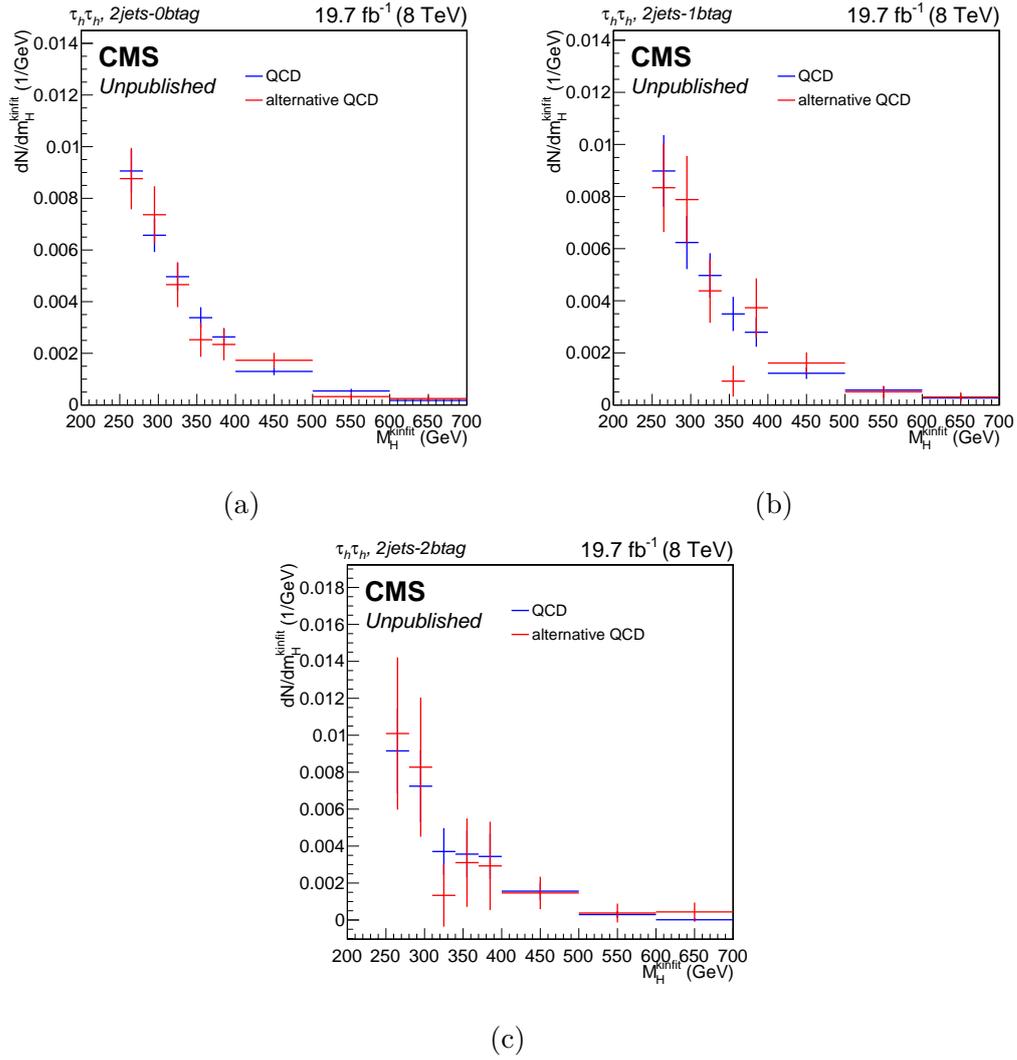


Figure 5.1: Comparison of the QCD shapes in the 2jets-0tag (a), 2jets-1tag (b) and 2jets-2tags (c) categories, using the standard (blue points) and the alternative (red points) methods to model the QCD multijets background.

interpretation the uncertainties depend on M_A and $\tan\beta$ and amount to 2-3% for PDF uncertainties and 5-9% for scale uncertainties, evaluated as described in [70] and using the PDF4LHC recommendations [132]. No theoretical uncertainties are considered in the 2HDM interpretation.

These uncertainties are not reported in the Table 5.1 but are considered in the signal extraction fit.

5.1.2 Shape uncertainties

- **τ_{had} and jet energy scales**

The $m_{\tau\tau}$ and the M_H^{kinfit} shapes are sensitive to the hadronic tau energy scale, which has been evaluated from a fit of the τ_h mass distribution [71]. The recommended value for this uncertainty is $\pm 3\%$. It affects both the shape and the normalisation of the signal and $Z \rightarrow \tau\tau$ processes.

In this analysis, the 4 body mass is reconstructed considering the τ_h candidates and the chosen leading and subleading b jets. Hence the shape of this distribution is influenced by the energy scale of both objects. Jets energy scale uncertainties have been provided by the CMS collaboration as function of jet P_T and η in [120]. The shapes have been generated by shifting up and down these uncertainties that affect all MC backgrounds. The jets energy scale uncertainties are not propagated to the \cancel{E}_T .

The impact of the τ_h and of the jets energy scales on M_H^{KinFit} distribution is evaluated in each category, independently, varying the τ_h and the jets energy scales within their uncertainties, then all the analysis selection is repeated. In Fig. 5.2 two examples of the shape uncertainties effect on M_H^{KinFit} distribution are shown: in Fig. 5.2a the effect of the τ_h energy scale for the signal $H \rightarrow hh \rightarrow bb\tau\tau$ events and in Fig. 5.2b the effect of the jets energy scale on $t\bar{t}$ events, both in the 2jets-2tags category.

- **Bin-by-bin statistical uncertainties**

As described in the section 4.6 for the background estimation, the shape of some backgrounds have been relaxed to obtain a smoother distribution, due to the low statistics in some signal regions. This problem is not completely solved in this way but it has been taken into account in the fit procedure. For the bins with lower statistics, the statistical uncertainty in these bins has been considered as a single uncorrelated source of uncertainty, as explained in [152].

5.2 Signal extraction procedure

The signal extraction is performed using a binned maximum likelihood fit of the final four-body mass distribution, M_H^{KinFit} , for the *signal* and *signal-plus-background* hypotheses. The procedure used to statistically interpret the data was

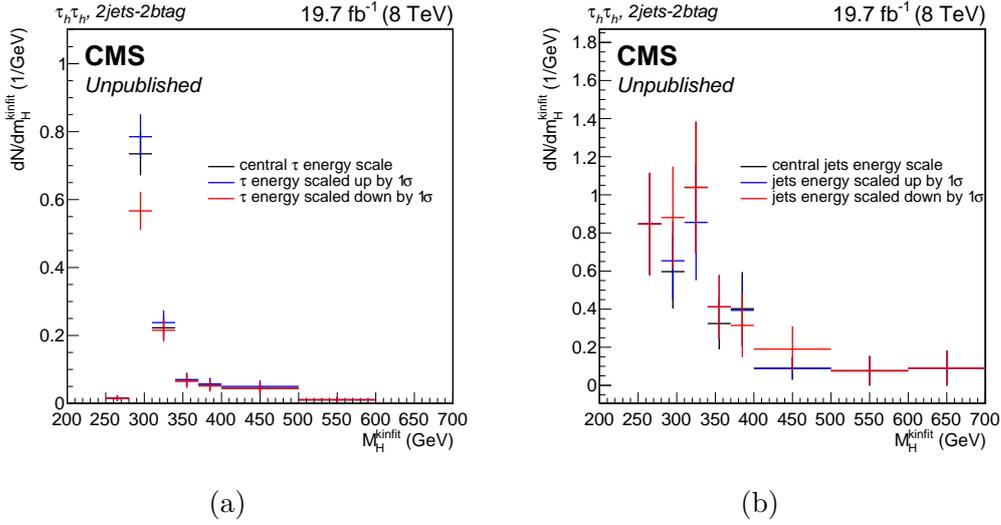


Figure 5.2: Examples of the shape uncertainties effects on the M_H^{KinFit} distribution in the 2jets-2tags category. (a) τ_h energy scale variation ($\pm 1\sigma$ around the central value) for signal $H \rightarrow hh \rightarrow bb\tau\tau$. (b) Jets energy scale variation ($\pm 1\sigma$ around the central value) for $t\bar{t}$ process.

developed by the LHC Higgs Combination Group [149] and is adopted by both ATLAS and CMS experiments.

The distributions of M_H^{KinFit} are obtained for signal and background events, as well as for data, in each b tag category, providing the corresponding number of events and shapes. Then the so called *shape* analysis is performed, where the bins content of the mass distributions are considered in the fit procedure.

The number of events predicted by the modelling in the j -th bin for the i -th event category, can be written as:

$$\nu_{ij}(\mu, \boldsymbol{\theta}_i) = \mu \cdot s_{ij}(\boldsymbol{\theta}_i) + b_{ij}(\boldsymbol{\theta}_i), \quad (5.2)$$

where $s_{ij}(\boldsymbol{\theta}_i)$ and $b_{ij}(\boldsymbol{\theta}_i)$ represent the signal and background yields in the j -th bin of the i -th event category. μ is the *signal strength modifier*: $\mu = 0$ corresponds to the background-only hypothesis, when there is no signal contribution. It represents the unknown rate of the signal and can be given in reference to a benchmark cross-section times branching fraction, in this case that of the heavy Higgs into a pair of SM-like Higgses. $\boldsymbol{\theta}_i$ represents the values of all the *nuisance parameters* in the i -th category. The summary of the nuisance parameters considered in each event category, together with their best estimates ($\hat{\boldsymbol{\theta}}_i$), were presented in the Table 5.1).

Then a likelihood \mathcal{L} is built for the observed data distribution, n_{ij} , function of the predictions for signal and background in each bin, ν_{ij} , and taking into account the

systematic uncertainties described in the previous section:

$$\mathcal{L} = \prod_j \text{Poisson} \left(n_{ij} | \nu_{ij}(\mu, \theta_i) \right) \cdot \prod_k \text{Constraint} \left(\theta_{ik}, \tilde{\theta}_{ik} \right) \quad (5.3)$$

where, a poissonian distribution

$$\text{Poisson} \left(n_{ij} | \nu_{ij} \right) = \frac{\nu_{ij}^{n_{ij}}}{n_{ij}!} \exp(-\nu_{ij}) \quad (5.4)$$

is assumed as the probability to observe n_{ij} data events in the bin j of the event category i . The second product of eq. 5.4 represents the knowledge about the nuisance parameters. The term $\text{Constraint} \left(\theta_{ik}, \tilde{\theta}_{ik} \right)$ is the probability for the true value of a nuisance parameter to be equal to θ_{ik} , given its best estimate $\tilde{\theta}_{ik}$. $\tilde{\theta}_{ik}$ is calculated with the external measurements discussed in section 5.1. The form of this term depends on the type of the systematic uncertainty, affecting the $M_{\text{H}}^{\text{KinFit}}$ distribution, *normalization* or *shape* uncertainty (section 5.1).

The normalization uncertainties are multiplicative factors on the signal or background yield and the constraints are represented by log-normal *pdfs*.

Also shape uncertainties affect the four-body mass distribution of the signal as well as of the background processes. These are accounted using a template morphing technique, called *vertical* morphing [153]. For each nuisance parameter θ_k influencing the shape in the i -th category, multiple templates are produced by varying the best estimate of the corresponding nuisance parameter by ± 1 standard deviation, obtaining three yield estimates. Using these three estimates, each $n_{ij}(\theta_k, \dots)$ is smoothly interpolated in the region $\tilde{\theta}_k \pm 1\sigma$ and the new uniformly distributed nuisance parameter θ'_k is introduced to parameterize this yield variation within the $\tilde{\theta}_k$ uncertainty.

In order to compare the compatibility of data with the *background-only* or *signal plus background* hypothesis, a test statistic is used, which is a single value and which can be used to set upper limits on the rate of signal production. The definition of this test statistics is

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data} | \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data} | \hat{\mu}, \hat{\theta})}, \text{ with the constraint } 0 \leq \hat{\mu} \leq \mu. \quad (5.5)$$

Where μ is the signal strength modifier for the signal hypothesis being tested. $\hat{\theta}_\mu$ are the values of the nuisance parameters which maximise the likelihood for that μ and $\hat{\mu}$ and $\hat{\theta}$ are the values which give the global maximum of the likelihood. The constraint $0 \leq \hat{\mu}$ guarantees that μ has a physical interpretation because the signal rate is positive, while the second constraint on $\hat{\mu}$ defines a one-sided confidence interval. Small values of \tilde{q}_μ indicate good compatibility between the data and the μ hypothesis. The term “data” refers to the distribution obtained from real data or from pseudo-data (pseudo-experiments, toys MC).

The following steps are aimed to find the *observed* value of the test statistic \tilde{q}_μ^{obs} , for a given signal strength, and the values of the nuisance parameters that maximise the likelihood function (eq. 5.3) for the *background-only* ($\hat{\theta}_0^{obs}$) and *signal plus background* ($\hat{\theta}_\mu^{obs}$) hypotheses.

First of all, in order to decide which is the hypothesis to support, it is important to define the probability density function (*pdfs*) $f(\tilde{q}_\mu)$ for each hypothesis. In general the *pdfs* are computed using toy MC pseudo-datasets, generated for each hypothesis and each representing the outcome of an experiment. Unfortunately this procedure is CPU time consuming, therefore in the case of a large data sample, the *pdfs* follow a calculable formula known as *asymptotic* limit approximation [154], where observations are equal to the predictions and the nuisance parameters are equal to their nominal values. The searches for the SM and MSSM Higgs boson in tau pair have demonstrated that also with a limited event statistics this approximation can be used [68, 71].

The nuisance parameters are fixed to the values determined by fitting the observed data, $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$, but are allowed to float within their uncertainties in order to evaluate the test statistic.

After having determined the *pdfs* for both the *background-only* and *signal plus background* hypotheses, $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})$, two p-values are defined. The p-value quantifies the compatibility of the \tilde{q}_μ^{obs} of the observed test statistics with both hypotheses. The p-value is defined as the probability of finding a value of q_μ greater than or equal to the observed value \tilde{q}_μ^{obs} :

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | \text{signal plus background}) = \int_{\tilde{q}_\mu^{obs}}^{\text{inf}} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \quad (5.6)$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | \text{background-only}) = \int_{\tilde{q}_\mu^{obs}}^{\text{inf}} f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu \quad (5.7)$$

Therefore the $CL_s(\mu)$ for a given μ can be defined as the ratio of these two probabilities:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \quad (5.8)$$

If $CL_s \leq \alpha$, with α called significance, it can be stated that the signal is excluded with $(1 - \alpha)$ CL_s confidence level (CL), which is a quite conservative approach. As a convention in the ATLAS and CMS collaborations, the 95% confidence level (CL), with $\alpha = 0.05$, is chosen as default value to quote the 95% CL upper exclusion limits on μ , that are used in this analysis and in general in CMS experiment. Therefore the observed upper limit is obtained with $\geq \mu_{up}$, where μ_{up} is adjusted until $CL_s(\mu_{up}) = 0.05$.

It is very useful to compare the observed limits with the expected median upper limits under the background-only hypothesis. This can be realized simulating

a large set of background-only pseudo-datasets. Then the CL_s and the upper limits are calculated. Finally a cumulative probability distribution of the results is built. From this distribution the median expected limit, ± 1 sigma and ± 2 sigma error bands (which correspond to the 68% and 95% confidence intervals) are built. It is expected that the observed limits have to be within the error bands of the expected limits.

In the case that an excess of events is observed, the p-value is the right quantity to quantify the probability that such an excess could be the result of the background-only hypothesis. Then the p-value to observe an excess as large or larger than the observed data can be defined as:

$$p_0 = \int_{\tilde{q}_0^{obs}}^{\text{inf}} f(\tilde{q}_0|0, \hat{\theta}_0^{obs}) d\tilde{q}_0. \quad (5.9)$$

The maximum likelihood fit is performed for all categories simultaneously. The values of nuisance parameters, output of the fit, which maximise the likelihood function, are taken into account. These new values of the nuisance parameters result in a change in the shape and normalization of the background processes, providing the so called *post-fit* distributions.

5.3 Results

The number of expected events and their uncertainties for signal and background processes, resulting from the fit, are summarized in Table 5.3 and are compared with the number of observed events for each sensitive category.

Process	Yields		
	2jets-0tag	2jets-1tag	2jets-2tags
Z \rightarrow $\tau\tau$ plus jets	83 ± 14	8 ± 3	1 ± 2
QCD	273 ± 18	30 ± 4	3 ± 1
W+jets	17 ± 3	–	–
Z+jets (lepton or jet faking τ)	3.0 ± 0.7	0 ± 0.1	0 ± 0.1
$t\bar{t}$	3.0 ± 0.5	6 ± 1	3.0 ± 0.6
Dibosons + single top	4.0 ± 0.8	1.0 ± 0.2	1.0 ± 0.1
Total Background	382 ± 18	45 ± 5	8 ± 2
H \rightarrow hh \rightarrow $bb\tau_h\tau_h$	0 ± 0.1	1.0 ± 0.2	1.0 ± 0.1
Data	390	45	6

Table 5.3: Fit results for the yields for the data and for the contributing background processes in the $\tau_h\tau_h$ channel. The yields are shown for the 2jets-0tag, 2jets-1tag and 2jets-2tag categories. The yields expected in each category in the hypothesis of a MSSM Higgs signal with $M_H = 300$ GeV and $\tan\beta = 2$ are given for comparison.

The post-fit plots of several quantities that have a direct impact on the H candidate

reconstruction and on the $M_{\text{H}}^{\text{KinFit}}$ distribution are reported in the following figures:

- Fig. 5.3: p_{T} and η distributions of the *leading* τ_{h} in the inclusive category
- Fig. 5.4: p_{T} and η distributions of the *subleading* τ_{h} in the inclusive category
- Fig. 5.5: $|\vec{E}_{\text{T}}^{\text{miss}}| = \cancel{E}_{\text{T}}$ and number of b jets reconstructed in the inclusive category
- Fig. 5.6: the p_{T} and b tag CSV discriminator distributions of the *leading* b jet candidate, in the inclusive category
- Fig. 5.7: the p_{T} and b tag CSV discriminator distributions of the *subleading* b jet candidate, in the inclusive category
- Fig. 5.8: distributions of $M_{\tau\tau}$ and M_{bb} in each category, before applying the mass window requirements (section 4.5.3)
- Fig. 5.9: distributions of the reconstructed 4 body mass in each category without and with the kinematic fit applied (section 4.5.4)
- Fig. 5.10: distributions of the H candidate mass, $M_{\text{H}}^{\text{KinFit}}$ (section 4.5.4), using the kinematic fit in each category, after applying mass window requirements on $M_{\tau\tau}$ and M_{bb}

In the post-fit plots stacked contributions from the modelled backgrounds are superimposed to data (black points). The ‘‘Electroweak’’ in the legends (red) includes $W + \text{jets}$, diboson and single top processes; the $Z \rightarrow \ell\ell$ (blue) includes $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ plus jets processes. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the signal plus background hypothesis to the data. The expected signal, multiplied by a factor 10, is shown superimposed as an open dashed histogram in the hypothesis of $\tan\beta = 2$ and $M_{\text{H}} = 300 \text{ GeV}$, in the ‘‘low- $\tan\beta$ -high’’ scenario of the MSSM. These conventions (if not differently specified) will be used in all the post-fit plots of this thesis. In all the plots, the grey bands represent the uncertainties on the modelled processes provided by the fit. All the post-fit distributions show a quite good agreement between real data and the modelled background processes.

5.4 SM Higgs background contribution

The $\text{H} \rightarrow \text{hh} \rightarrow \text{bb}\tau\tau$ final state can be reproduced by the direct production of SM Higgs boson decaying in a τ pair or a b jet pair. In 2jets-0tag category the dominant production for the SM Higgs boson is the gluon fusion, while in the other two categories the associated production (WH and ZH) can contribute, and in particular the irreducible background, the ZH process. On the other side after

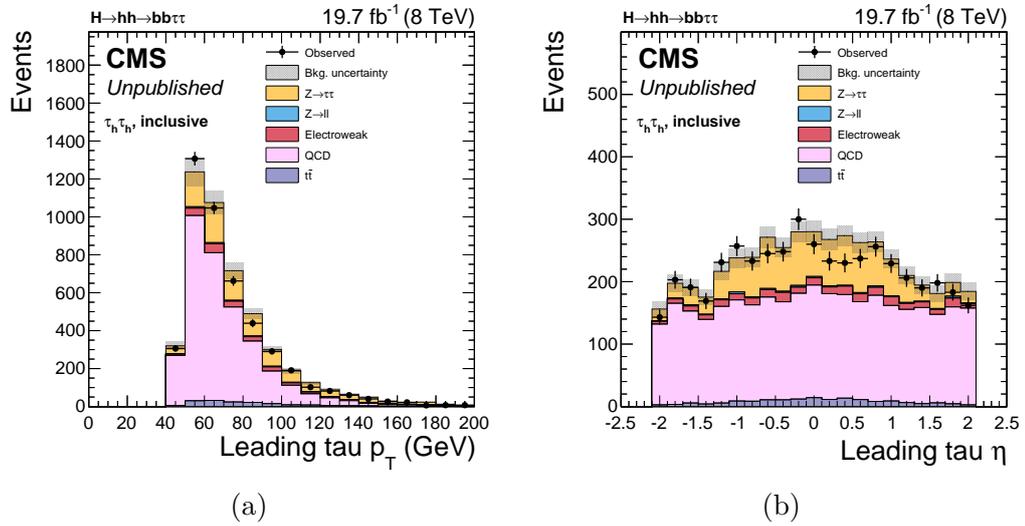


Figure 5.3: Distributions of the *leading* τ_h : p_T (a) and η (b), in the inclusive category.

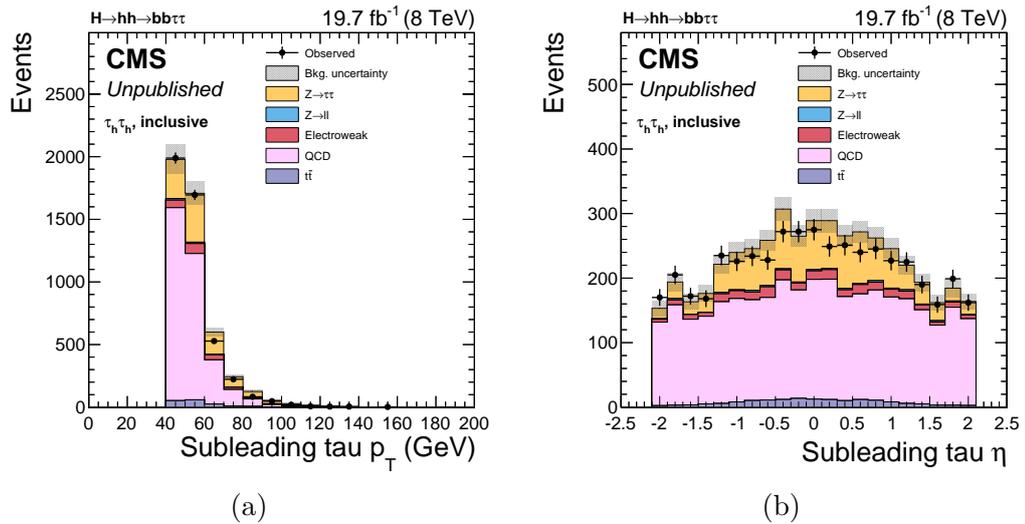


Figure 5.4: Distributions of the *subleading* τ_h : p_T (a) and η (b), in the inclusive category.

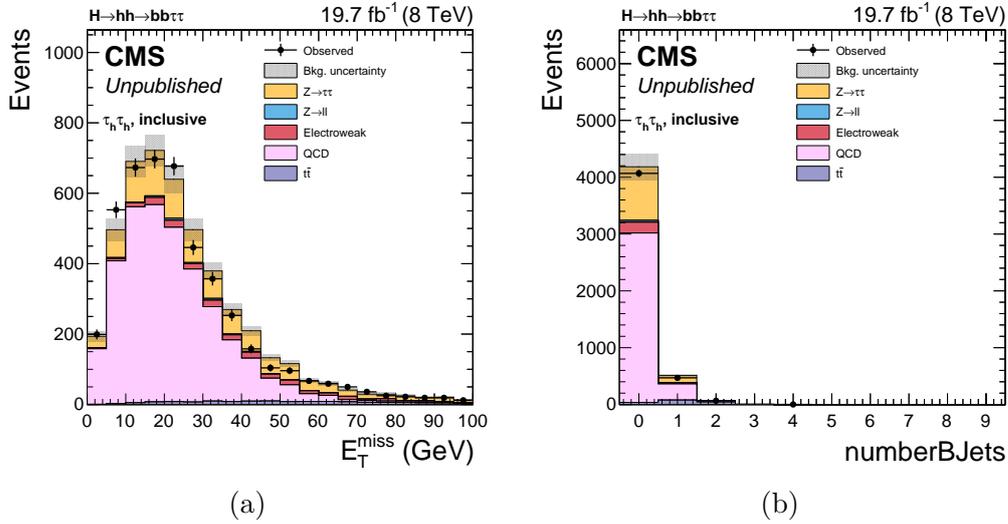


Figure 5.5: Distributions of E_T^{miss} (a) and of the number of b jets (b) in the inclusive category.

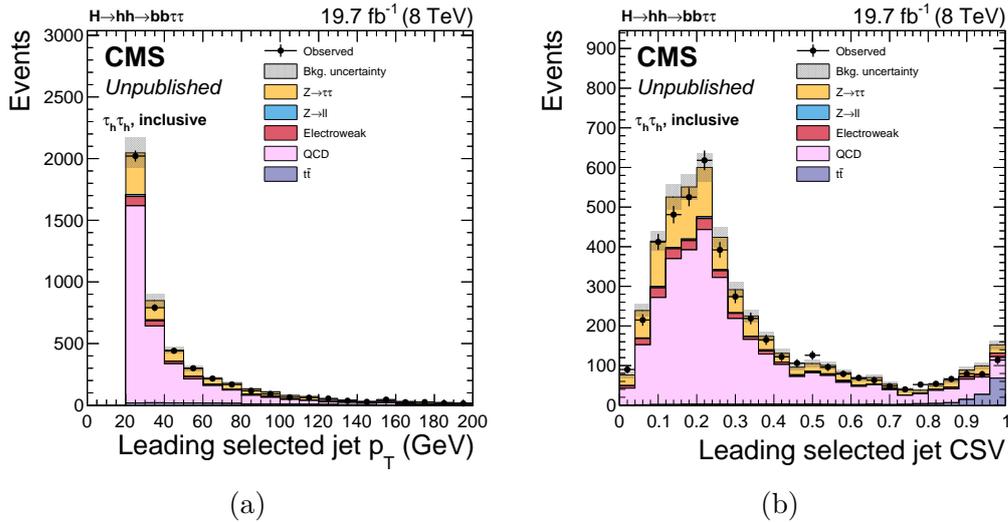


Figure 5.6: Distributions of the *leading* b jet candidate in the inclusive category: p_T (a) and b tag CSV discriminator (b).

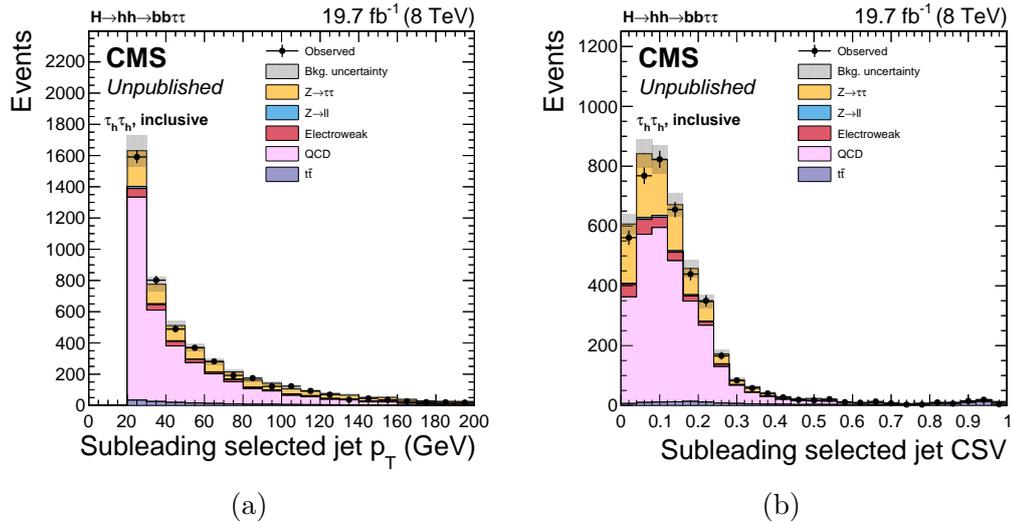


Figure 5.7: Distributions of the *subleading* b jet candidate in the inclusive: p_T (a) and b tag CSV discriminator (b).

the mass window requirements, these SM contributions are reduced to less than 10% compared the other processes. The effect on the expected limits (section 5.2) has been estimated by injecting the SM processes mentioned above and evaluating the new exclusion limits. The change of the limits was smaller than 3%, thus, with the present statistics, these SM backgrounds can be neglected.

5.5 Model independent limits

From Fig. 5.10, where the M_H^{KinFit} distribution is reported for the three categories, it can be clearly observed the absence of an evidence of a new particle. Therefore, the limits on the production cross-section of the heavy Higgs boson, through the gluon fusion process $\sigma(gg \rightarrow H)$, times the branching fraction $H \rightarrow hh \rightarrow bb\tau\tau$ are set. The 95% CL exclusion limits are evaluated independently from the theoretical environment. They are calculated for M_H between 260 and 350 GeV, in step of 10 GeV, using all the simulated signal samples. All categories are combined and in order to cover the full range of M_H , a linear interpolation is used. From Fig. 5.11a, the observed $\sigma \times B$ exclusion limits are between 0.1 and 0.9 pb and are then compatible with the expected exclusion limits within their error bands, in the full range of explored mass under the background-only hypothesis.

The $\tau_h\tau_h$ channel results were combined with the results of $H \rightarrow hh \rightarrow bb\tau_l\tau_h$ CMS analysis [34], where one τ decays hadronically and the other in electron or muon ($\tau_e\tau_h$ or $\tau_\mu\tau_h$ channels), using the same signal extraction technique described in section 5.2. The $\tau_h\tau_h$ channel gives a significant gain in sensitivity especially for M_H greater than 300 GeV. Therefore the observed $\sigma(gg \rightarrow H) \cdot B(H \rightarrow hh \rightarrow bb\tau\tau)$ exclusion limits, obtained from the combination of these three channels, are $\approx 0.13 - 0.3$ pb, depending on the M_H , as shown in Fig. 5.11b. Within their error bands, the observed limits are in agreement with the expected exclusion limits

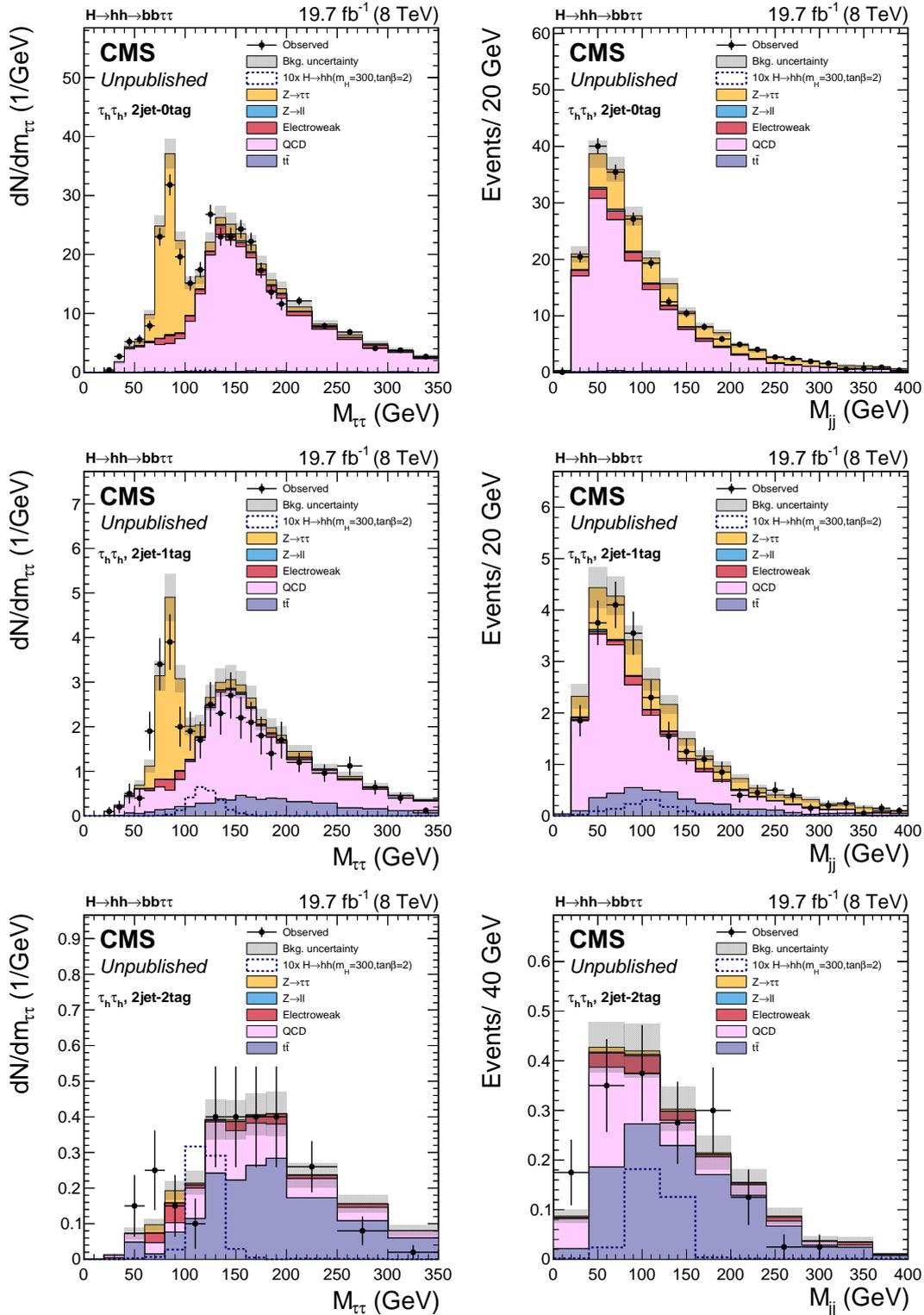


Figure 5.8: Distributions of $M_{\tau\tau}$ (left) and M_{bb} (right) in the $\tau_h\tau_h$ channel. The plots are shown for events in the 2jets-0tag (top), 2jets-1tag (centre) and 2jets-2tags (bottom) categories, before applying the mass window requirements. The content of each bin of these distributions is divided by the bin width.

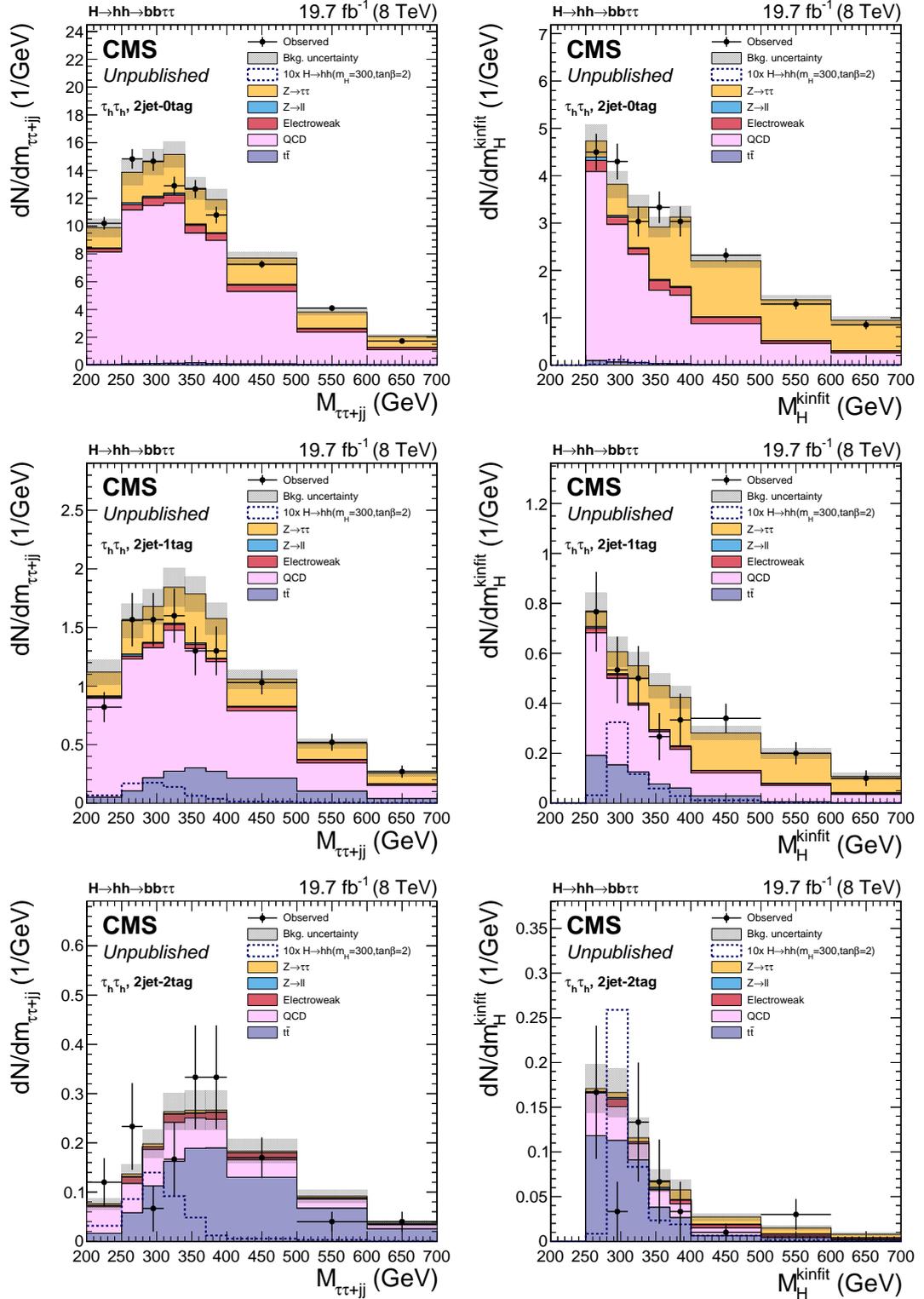


Figure 5.9: Distributions of the reconstructed 4 body mass without ($M_{\tau\tau+jj}$, left) and with (M_H^{KinFit} , right) the kinematic fit applied. The plots are shown for events in the 2jets-0tag (top), 2jets-1tag (centre) and 2jets-2tags (bottom) categories, before applying the mass window requirements. The content of each bin of these distributions is divided by the bin width.

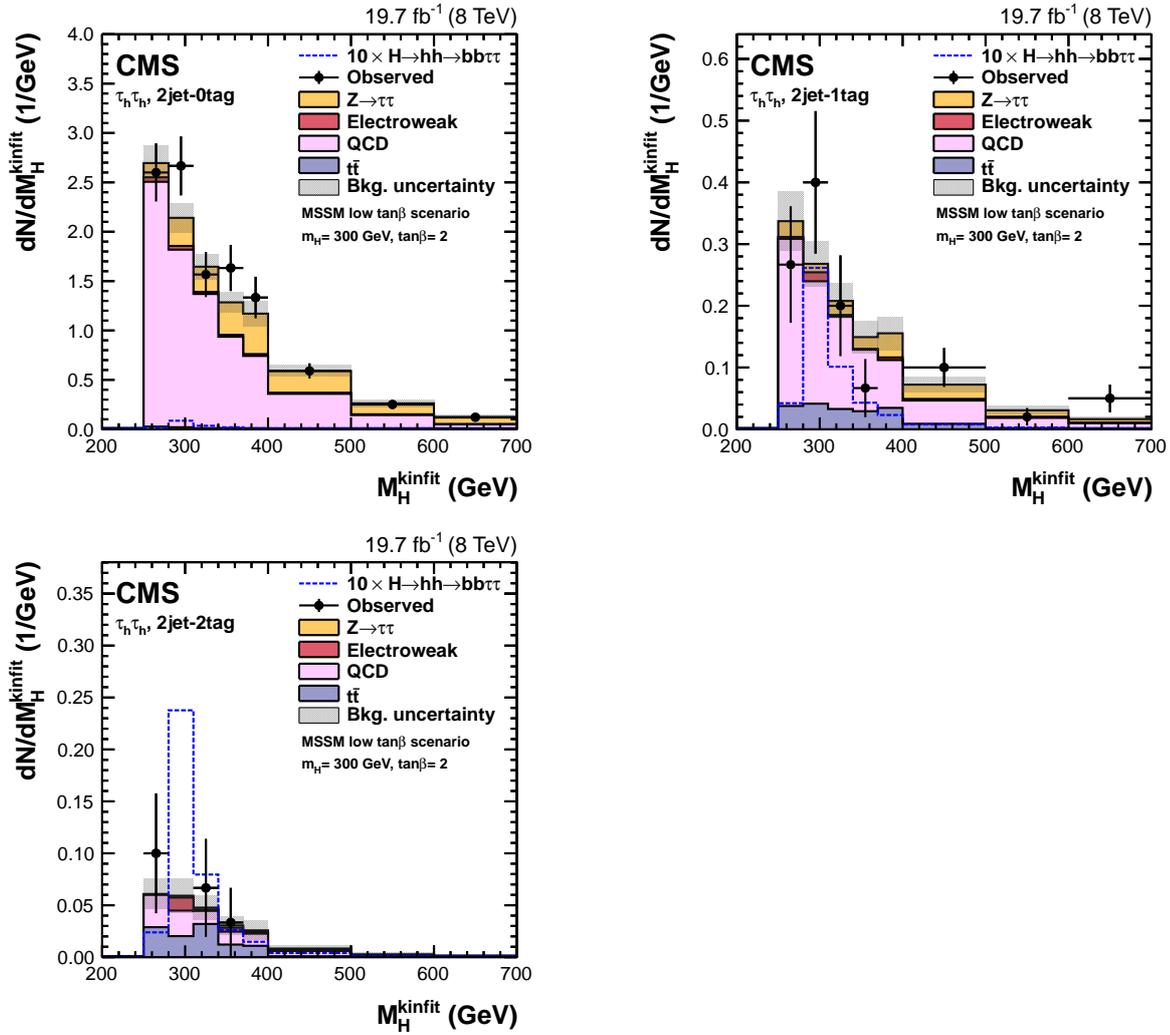
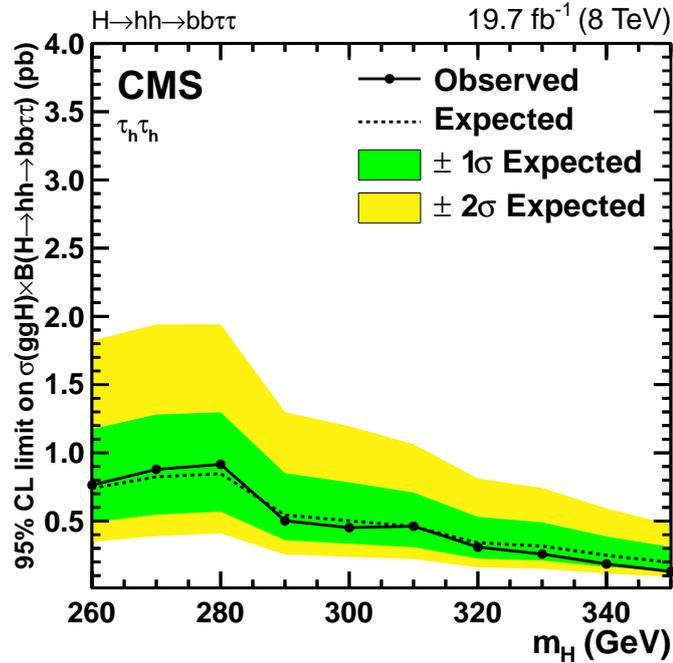
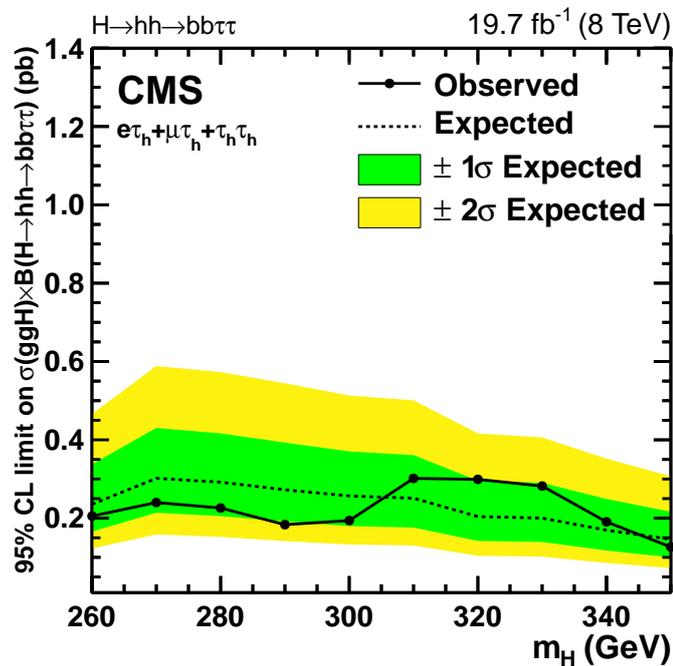


Figure 5.10: Distributions of the H candidate mass in the $\tau_h\tau_h$ channel, M_H^{KinFit} , using the kinematic fit, after applying mass window requirements on $M_{\tau\tau}$ and M_{bb} in the $\tau_h\tau_h$ channel. The plots are shown for events in the 2jets–0tag (top left), 2jets–1tag (top right), and 2jets–2tags (bottom) categories. The content of each bin of these distributions is divided by the bin width.

under the background-only hypothesis.



(a)



(b)

Figure 5.11: Upper limits at 95% CL on the $\text{H} \rightarrow \text{hh} \rightarrow \text{bb}\tau\tau$ cross-section times branching fraction for the $\tau_h\tau_h$ channel (a), and combining three channels: $\tau_h\tau_h$, $e\tau_h$ and $\mu\tau_h$ (b) [34].

5.6 Model dependent limits

The model independent limits on the cross-section times the branching fractions, determined in the previous section, can be subsequently interpreted in the contexts of both: the “low- $\tan\beta$ -high” MSSM scenario and the 2HDM Type II model. These results provide the so called model “dependent” limits in the planes (section 1.3), $[M_A, \tan\beta]$ for the MSSM and $[\cos(\beta - \alpha), \tan\beta]$ for 2HDM, respectively the free parameters space of the two models.

To obtain the final limits, a scan is performed in the two dimensional parameter space. The CL_s is calculated at each point of the two dimensional grid, using the asymptotic approximation, and comparing the signal hypothesis with the background-only hypothesis (section 5.2). The theoretical values of the cross-sections and branching fractions are computed by the LHC Higgs Cross Section Working Group for both models (LHCHSWG [70,155]). Grid points with a CL_s less than 0.05 are excluded at 95% CL. Then, a contour is drawn to connect the excluded points using the linear interpolation between the neighbouring points in the grid. This procedure is called the “horizontal morphing” technique [153,156]. All categories are combined and the shape and the normalization of the M_H^{KinFit} templates are determined for each point of the 2D grid.

The resulting limits for both models are obtained combining the final results of $\tau_h\tau_h$ channel with the semileptonic ones, the $\tau_e\tau_h$ and $\tau_\mu\tau_h$ channels, described in [34]. The exclusion limits at 95% of confidence level for the MSSM and the 2HDM are shown respectively in Fig. 5.12 and 5.13. As can be seen from the plots the absence of an excess of events allows to set upper limits and to exclude a small region of the phase space parameters, which is marginal for the MSSM scenario. In both figures the area highlighted in blue below the black curve marks the observed exclusion, instead the dashed curve and the grey bands show the expected exclusion limits with their relative uncertainties.

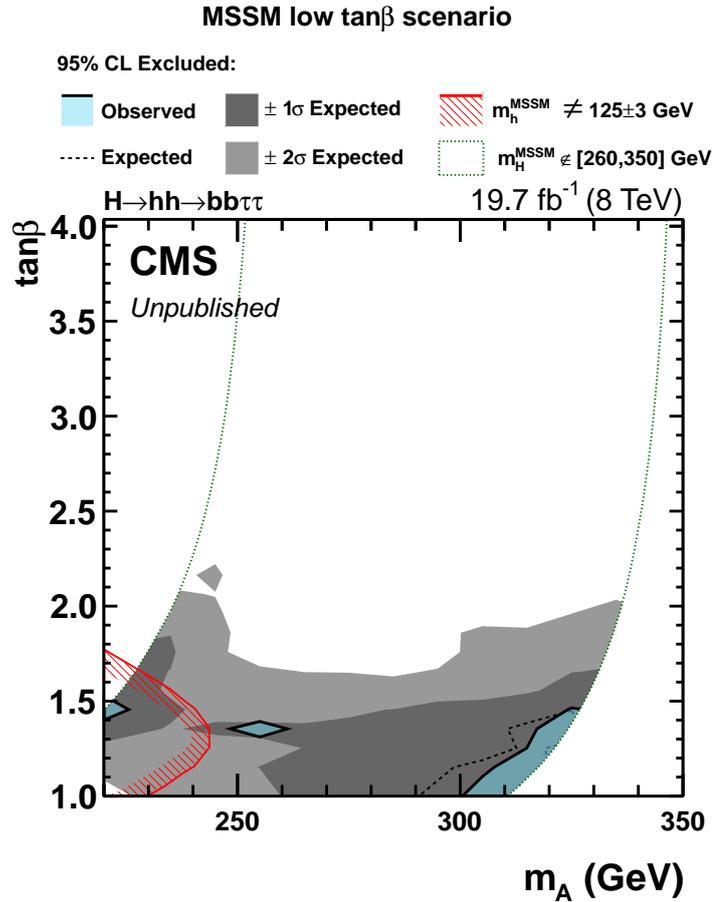


Figure 5.12: The 95% CL exclusion region in the $[M_A, \tan\beta]$ plane for the “low- $\tan\beta$ -high” MSSM scenario, using the results of the $H \rightarrow hh \rightarrow bb\tau\tau$ analysis and combining the three channels: $\tau_h\tau_h$, $\tau_e\tau_h$ and $\tau_\mu\tau_h$. The red area indicates the area already excluded by constraining the SM-like Higgs boson $M_h = (125 \pm 3)$ GeV. The plot also goes up $\tan\beta = 1$ and not lower because the cross-section values were not validated by the LHCHSWG at the time of this thesis. The green dashed lines indicate the region which is outside of the range of the M_H values considered in this analysis.

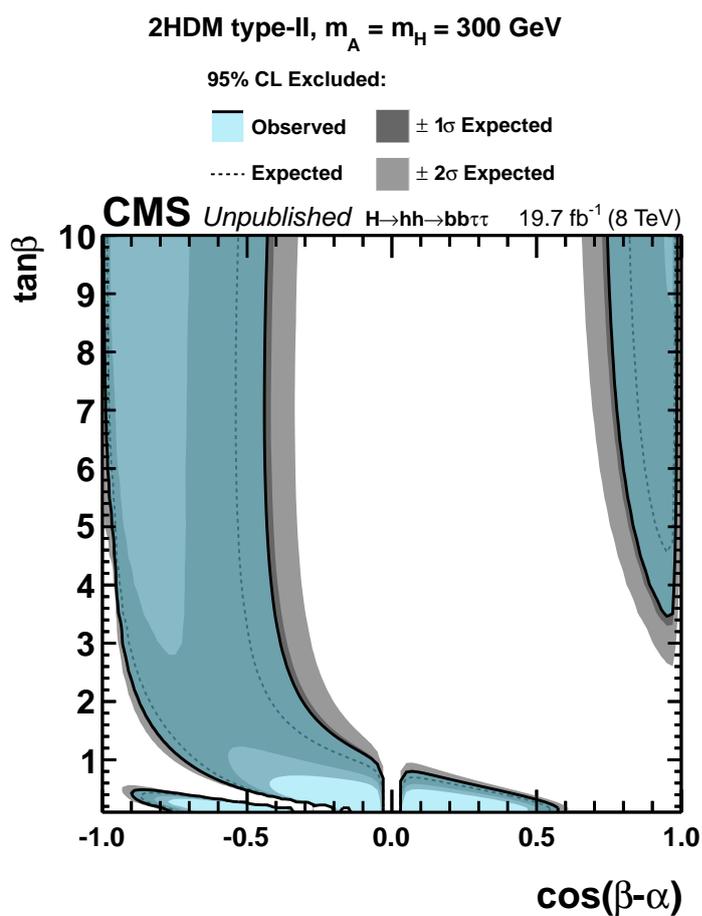


Figure 5.13: The 95% CL exclusion regions in the $[\cos(\beta-\alpha), \tan \beta]$ plane of 2HDM Type II model, for $M_A = M_H = 300$ GeV, using the results of the $H \rightarrow hh \rightarrow bb\tau\tau$ analysis and combining the three channels: $\tau_h\tau_h$, $\tau_e\tau_h$ and $\tau_\mu\tau_h$.

Conclusion and perspectives

In this work of thesis the first search for a heavy scalar boson H decaying in two SM-like Higgs bosons h in the final state $bb\tau_h\tau_h$ is presented. For this search, proton-proton collisions data, collected by CMS experiment during 2012, were analysed.

Since no signal excess was observed in the distribution of the reconstructed invariant mass of the H candidate, model independent exclusion limits were set at 95% CL on the production cross-section of a scalar neutral boson H , $\sigma(gg \rightarrow H)$, times the branching fraction, $B(H \rightarrow hh \rightarrow bb\tau\tau)$, as a function of its mass, in the range $260 \leq M_H \leq 350$ GeV.

These results were combined with the ones obtained by CMS for the $bb\tau_e\tau_h$ and the $bb\tau_\mu\tau_h$ final states [34], giving 95% CL observed exclusion limits in the range ≈ 0.1 and 0.3 pb (Fig. 5.11b), depending on M_H . The observed and expected exclusion limits are in agreement within their error bands.

These limits can be also compared with the results obtained by other CMS searches [28–33] for double Higgs production, using different final states, as shown in Fig. 5.14. For the comparison, the SM branching fractions, $B(h \rightarrow \tau\tau) \approx 0.063$ and $B(h \rightarrow bb) \approx 0.575$, are factored out to obtain the exclusion limits for $\sigma(gg \rightarrow H) \times B(H \rightarrow hh)$. The corresponding values are between 1.4 and 4.1 pb in the M_H range considered in this thesis.

Fig. 5.14 shows that the $bb\tau\tau$ channel provides complementary and competitive results with respect to $bb\gamma\gamma$ and $bbbb$ final states studied by CMS.

The observed exclusion limits for $\sigma(gg \rightarrow H) \times B(H \rightarrow hh)$, provided by ATLAS for the $bb\tau\tau$ final state, are between 1.7 and 4.2 pb for $260 \leq M_H \leq 350$ GeV [25–27], and are comparable with the results presented in this thesis.

The results were also interpreted in the context of the “low- $\tan\beta$ -high” MSSM and 2HDM Type II scenarios. The model dependent exclusion limits were set in the parameter space ($[M_A, \tan\beta]$ for the MSSM, and $[\cos(\beta - \alpha), \tan\beta]$ for 2HDM). The excluded regions, shown in Fig. 5.12 and Fig. 5.13, are marginal for the MSSM scenario.

With the current data taking (Run II), the searches of the double Higgs production will cover a significant region of the MSSM and 2HDM parameter spaces. On the other hand, tools and methods developed for this analysis will be an important legacy for LHC phase II, when the collected luminosity is expected to achieve \approx

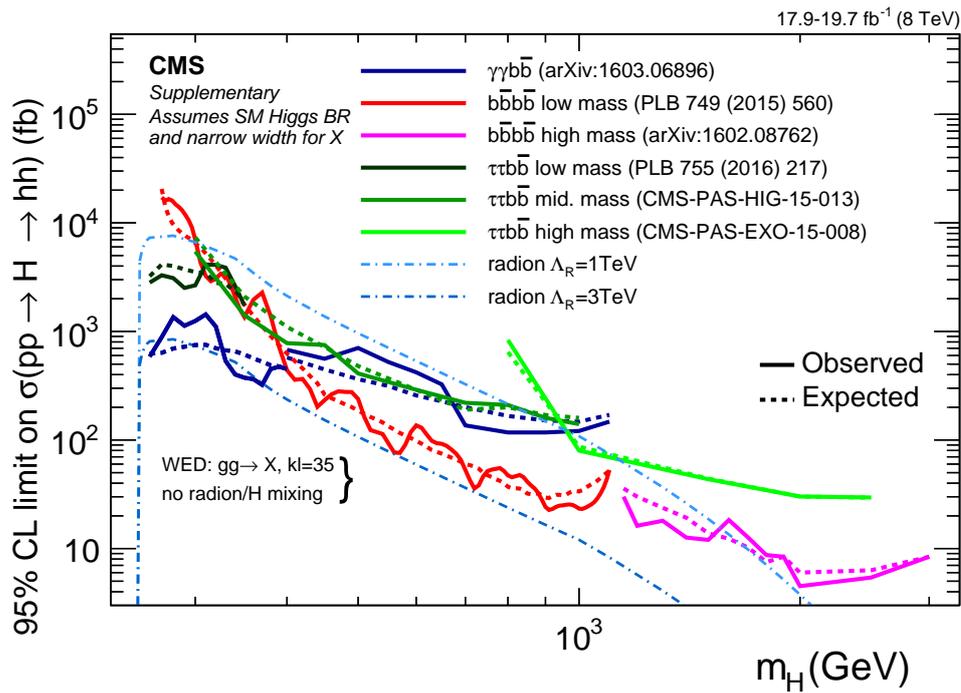


Figure 5.14: Observed and expected 95% CL exclusion limits for the production cross-section of a heavy neutral scalar boson H times $H \rightarrow hh$ branching fraction set by CMS searches. The limits performed in this analysis (dark green lines) are compared with other CMS analyses in different final states, as reported in the legend [157] (credits to hh group).

3000 fb^{-1} , allowing the first measurement of the Higgs (h) self-coupling, probing deeply the SM Higgs nature. Due to its high discovery potential, the study of the double Higgs production is recognized as a key analysis of the LHC physics program.

Appendix A

Luminosity Measurement

In CMS the measurement of the luminosity delivered by LHC [150] is done using two subdetectors: the forward hadron calorimeter (HF) (section 2.2.4) and the silicon pixel detector (section 2.2.2). The HF is capable of estimating the luminosity per bunch-crossing thanks to its high rate acquisition system. The pixel detector is used for its very low occupancy and excellent stability over time.

The estimation of the luminosity relies on the precise measurement of the rate, R , of events:

$$\mathcal{L} = \frac{R}{\sigma_{vis}} \quad (\text{A.1})$$

A.1 relates the measured rates to the luminosity using the visible cross-section, σ_{vis} , which includes the acceptance.

The HF provides an on-line luminosity measurement and uses the info coming from the electronics.

Two methods have been implemented to extract an on-line instantaneous luminosity with the HF. The first method, “zero counting”, uses the average fraction of the HF empty towers to determine the mean number of interactions per bunch crossing. The second method exploits the linear relationship between the average E_T per tower and the mean luminosity [158]. HF provides small statistical uncertainties, but suffers of non-linear dependence with pile-up and of a calibration drift over a long period of time.

For the pixel detector, the counting of the pixel clusters has been chosen for the off-line luminosity measurement, because it operates at stable beam conditions, providing a very small dependence on pile-up and an excellent linear detector response with the increasing of the luminosity. Luminosity is evaluated considering the number of pixel clusters in a zero-bias event.¹ Since the average number of pixel cluster per event, $\langle n \rangle$, is proportional to the number of p-p collisions in the

¹Zero-bias event is an event accepted only requiring that two bunches cross at the CMS interaction point.

2012 LHC run, the instantaneous luminosity is estimated as

$$\mathcal{L} = \frac{\nu \langle n \rangle}{\sigma_{vis}} \quad (\text{A.2})$$

$\nu = 11246$ Hz is the beam revolution frequency. The visible cross-section is calibrated using the Van der Meer (VdM) scan technique [150]. Briefly the VdM scan technique consists in deriving the beam overlap from the shape of the measured rates as a function of the beam separation, obtained by scanning the beams one across the other along the horizontal and vertical planes.

The total integrated luminosity, calculated by the off-line method, recorded by the CMS in 2012 is 21.79 fb^{-1} , as shown in Fig. 2.2b. The overall uncertainty on the integrated luminosity is estimated to be 2.5% (syst.) \pm 0.5% (stat.) [150]. The systematic uncertainties are related to σ_{vis} and to luminosity integration. The details about their estimation are reported in [150].

Appendix B

Monte Carlo simulation

The interpretation of the data from high energy physics particle colliders and their use to extract measurements on fundamental physical parameters often heavily rely on the theoretical modelling of the physical processes and on a detailed simulation of the interactions of particles with detectors. These are called Monte Carlo (MC) simulations. and play a fundamental role in the physics data analysis, providing a complete picture of the large multiplicity of particles as the outcome of a hard interaction.

A complete MC simulation is composed of three processes [35]:

- The generation of the hard process produced by a p–p collision. The hard process is described by the differential cross–section of the partonic process and is weighted by the Parton Distribution Functions (PDFs)¹.
- The simulation of the *parton showering* and of the *hadronisation*. The parton showering is made of partons branched before the hard interaction (*Initial State Radiation* (ISR)) or is made of partons produced after the hard interaction (*Final State Radiation* (FSR)). The hadronisation represents the process where the partons, after the showering, are transformed into a set of unstable hadrons, that can subsequently decay to secondary hadrons.
- The simulation of the geometry and of the material present in the detector structure and the answer of the detector after the passage of an interacting particle.

The most popular event generator is PYTHIA [96], that has been used in this analysis to describe the p–p collisions and their hadronic activity. The temporal sequence of a typical high energy physics event in PYTHIA, is the following:

1. the first process is the collision of the two proton beams; due to the high energies the interaction happens between the partons inside the protons

¹The Parton Distribution Functions (PDFs) represent the momentum distribution of the constituent partons (a general term which includes quarks and gluons) in a proton

2. the ISR starts from a parton in each beam that creates a sequence of branching, like $q \rightarrow qg$
3. a parton coming from each of the two showers realizes the *hard process*, which represents the main event
4. partons resulting from the hard process produce showers in the final state;
5. another important process is the interaction between the partons coming from the hadrons and the following creation of other hadrons;
6. finally most of the produced hadrons are unstable and decay.

PYTHIA simulates these steps and the computation element on which the partonic differential cross-section depends, the matrix element (ME), at first approximation level, the leading order (LO). Moreover, it generates the so called *Underlying event* (UE), which denotes any additional activity beyond the basic process, associated to the ISR and FSR. These different steps can be simulated also at different approximation levels, like the Next to Leading Order (NLO), with other *Matrix Element Calculator*. The widely known generators are MADGRAPH [98] and POWHEG [97, 133–135]. POWHEG is used for the ME calculation at NLO, while MADGRAPH performs either LO or NLO calculations. The POWHEG and MADGRAPH generators are interfaced with PYTHIA for parton showering and fragmentation using the Z2* tune [159] and with TAUOLA [147] for the simulation of the τ decays.

The next step is to model the interaction of particles going through the detector. The tool used is GEANT4 [99]. GEANT4 simulates the detector geometry, the passage of the particles through the detector and their interactions. After that also the electronic response and the digitization of the signals are simulated. In the digitization step the *Minimum-bias* (MB) events are injected, simulating the elastic and inelastic scattering, and the diffractive processes. The MB events are used to model the *pile-up* events in real data.

Appendix C

Data and MC corrections

This section is dedicated to the description of the corrections applied to improve the modelling of the data when using simulated samples.

C.1 Hadronic τ efficiency for $\tau_h\tau_h$ trigger

For the di- τ_h , the trigger efficiency for the single τ_h candidate in the simulated MC samples is corrected taking into account the measured differences in the efficiency of the selection in MC compared with data (section 4.3).

The trigger efficiency curve for a hadronic tau is calculated selecting $Z/\gamma^* \rightarrow \mu\tau_h$ events, using the tag-and-probe technique [103], and is parameterized as function of p_T for data and simulated samples. The details of how the efficiencies have been calculated are described in the reference [160].

In Fig. C.1, examples of the fitted efficiency curves for both the HLT trigger paths used in this thesis (section 4.3) are reported for data and MC simulations.

The trigger efficiency curves, measured in data and MC samples for each hadronic

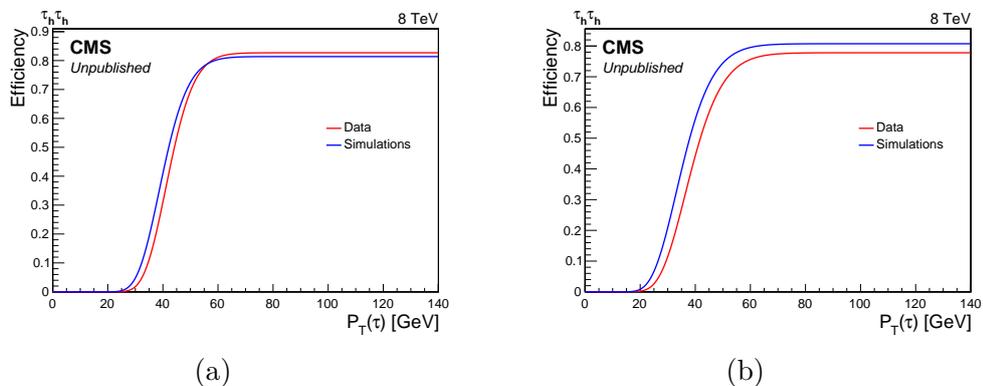


Figure C.1: Examples of fitted trigger efficiency curves for DiTau path (a) and DiTau plus jet path (b). In each plot the distributions of the Data (red) and of the MC simulation (blue) are reported.

tau leg, are then used to calculate the scale factors aiming to improve, on average, the agreement between the efficiency of the simulated trigger and the trigger efficiency as measured in data. The product of the scale factors of the two legs has been applied as an event weight to the MC simulated events, for the di- τ_h trigger:

$$w_i^{trigger} = \frac{F^{data}(p_{Ti})}{F^{MC}(p_{Ti})} \quad (\text{C.1})$$

In the Drell—Yan $Z/\gamma^* \rightarrow \tau\tau$ Embedded sample the $\tau_h\tau_h$ HLT is not simulated and no trigger matching requirements are applied. Therefore, each event passing the offline selection is weighted only taking the product of the trigger efficiencies per tau leg as measured in data.

The same efficiency curves have been already used in the search of the SM $H \rightarrow \tau\tau$ [71] and in the MSSM $\phi \rightarrow \tau\tau$ analyses [68].

C.2 τ_h energy scale corrections

A further Data/MC correction has been applied to the energy of the τ_h for the different hadronic tau decays. These corrections have been calculated by fitting the distribution of the τ_h mass selecting $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\tau_h$ events in data and considering the shape templates for the $Z/\gamma^* \rightarrow \tau\tau$ plus jets and the other contributing background processes. The mass distribution is scaled by a shift parameter, that, together with the normalisation factors are adjusted by a fit, giving the final distribution reported in Fig. C.2. The fits have been performed separately for each τ_h decay mode. The different scale factors and energy scale calibrations obtained according to the τ_h decay mode are reported in Tab. C.1. These corrections have been applied to all MC where a reconstructed τ_h candidate is matched with a MC true hadronic tau and to $Z \rightarrow \tau\tau$ embedded samples.

Decay Mode	Normalization Factor	Energy Scale
<i>One Hadron (h^\pm)</i>	0.88	–
<i>One Hadron plus n Strips ($h^\pm\pi^0$ plus $h^\pm\pi^0\pi^0$)</i>	1.0	1.012
<i>Three Hadrons ($h^\pm h^\pm h^\mp$)</i>	1.0	1.012

Table C.1: Normalization and Energy scale corrections for the decay modes of hadronic taus.

C.3 \cancel{E}_T resolution and response

Another correction has been applied to signal samples and to $Z/\gamma^* \rightarrow \ell\ell$ and W plus jets backgrounds, called *Recoil Correction* (section 3.2.7). This corrects

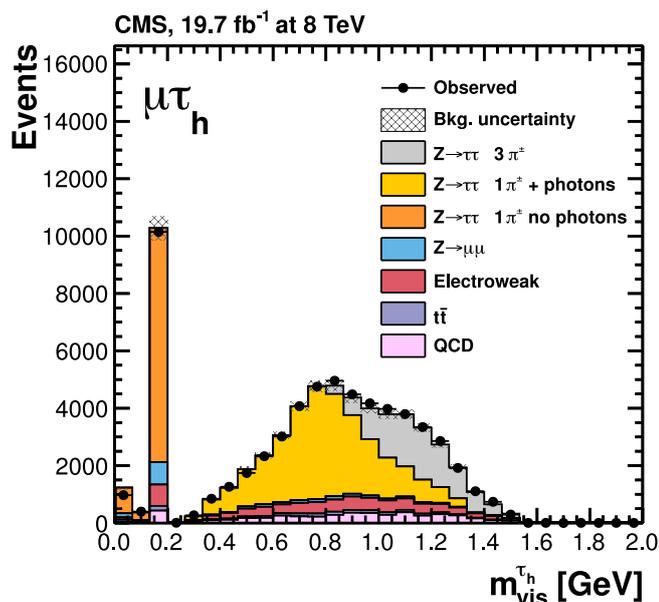


Figure C.2: Observed and predicted distributions for the visible τ_h mass in the $\mu\tau_h$ channel. The $Z \rightarrow \tau\tau$ contribution is split according to the decay mode reconstructed by the HPS [125,126] algorithm as shown in the legend [71].

MC for residual differences in \cancel{E}_T response and resolution between data and MC simulation, as described in [161].

C.4 *b* tagging efficiency and mistag rate

In this analysis the *b* tagging requirement is used to identify the *b* jets in the final state and to categorize the event section 4.5.2. In order to improve the agreement between simulated samples and data, the *b* tagging efficiency and the mistag rate are corrected in the MC samples through the application of efficiency and mistag scale factors. A detailed description of their estimate for the 2012 data is in [151]. The simulated samples are corrected using a *promote/demote* method, where a subset of *b* tagged jets is re-classified as un-tagged, or, viceversa, the un-tagged jets are promoted as *b* tag jets. The promotion or demotion probabilities for each jet are estimated using the scale factors (SF), as a function as p_T , η and jet-flavour and the *b* tagging efficiencies in MC (Eff). These probabilities are defined by:

$$\begin{aligned}
 P(\text{demote}) &= 1 - SF && \text{when } SF < 1 \\
 P(\text{promote}) &= \frac{(SF - 1)}{\frac{SF}{\text{Eff}} - 1} && \text{when } SF > 1
 \end{aligned}$$

Appendix D

Di-tau mass reconstruction – SVFit algorithm

The precise reconstruction of the invariant mass of the tau pair candidate, $M_{\tau\tau}$ is important to discriminate the $h \rightarrow \tau\tau$ process from the irreducible background $Z \rightarrow \tau\tau$. Because of the presence of the two neutrinos from τ decays in the final state, the invariant mass of the visible $\tau_h\tau_h$ pair ($M_{\tau\tau}^{vis}$) has a resolution, which does not allow to well separate the $h \rightarrow \tau_h\tau_h$ signal from the $Z \rightarrow \tau_h\tau_h$ background, since the mass of the Z boson ($M_Z \approx 91.2$ GeV) is close to the h mass. The accuracy in the mass reconstruction can be restored exploiting also the energy taken by the neutrinos, the missing transverse energy. To precisely reconstruct the invariant mass of the $h \rightarrow \tau\tau$ candidate, the algorithm SVFit [144] is used. This algorithm was already used by CMS in the SM and MSSM Higgs searches with a pair of taus in the final state [68, 71].

The SVFit is a likelihood based algorithm combining the information of the visible decay products of the τ and the measured \cancel{E}_T reconstructed event-by-event, to determine the invariant mass of the di-tau pair. Depending on the tau decay, the kinematic of the tau pair decay depends on a different number of parameters. 5 parameters are needed to specify the hadronic tau decays: the momentum, the polar and the azimuthal angles of the tau lepton in the laboratory frame, plus the two decay angles in the rest-frame of the tau lepton. In the decays of the tau leptons into electrons or muons, two neutrinos are produced and the invariant mass of the di-neutrino system ($m_{\nu\nu}$) constitutes a 6-th parameter. This parameter is zero in the case of the hadronic tau decays. The unknown parameters of the h τ pair candidate are then constrained by 3 observables: the momentum, the polar and the azimuthal angles of the visible tau decay products, measured in the laboratory frame, leaving 2 (3) unconstrained parameters for hadronic (leptonic) tau decay. A further constraint is applied on the total momentum of the neutrinos, using the \cancel{E}_x and \cancel{E}_y , the x and y components of the reconstructed missing transverse momentum. Finally the unconstrained parameters are:

- x , the fraction of tau lepton energy (in the laboratory frame) carried by visible decay products

- ϕ , the angle between the tau lepton momentum vector and the momentum vector of the visible decay products
- $m_{\nu\nu}$, the mass of the neutrino system present in leptonic tau decays

Given x , ϕ and $m_{\nu\nu}$, the energy and the momentum of the tau lepton in the laboratory system are fully determined [144].

$M_{\tau\tau}$ values are then reconstructed by combining the measured observables \cancel{E}_x and \cancel{E}_y with a probability model, which includes terms from tau decay kinematics and from the \cancel{E}_T resolution. The model makes a prediction for the probability density $p(\vec{x}|\vec{y}, \vec{a})$ to observe the values $\vec{x} = (\cancel{E}_x, \cancel{E}_y)$ measured in an event, given that the unknown parameters specifying the kinematics of the tau pair decay have values $\vec{a} = (x_1, \phi_1, m_{\nu\nu}^1, x_2, \phi_2, m_{\nu\nu}^2)$, where labels 1 and 2 refer to the two taus.

The likelihood model computes the probability $P(M_{\tau\tau}^i)$ for each value of mass $M_{\tau\tau}^i$, that the values of \vec{x} are observed, given that the momenta of the visible decay products are equal to the observed $\vec{y} = (p_1^{vis}, p_2^{vis})$:

$$P(M_{\tau\tau}^i) = \int \delta(M_{\tau\tau}^i - M_{\tau\tau}(\vec{y}, \vec{a})) p(\vec{x}|\vec{y}, \vec{a}) d\vec{a}.$$

The best estimate $\hat{M}_{\tau\tau}$ for the tau pair mass is taken to be the value $M_{\tau\tau}^i$ that maximises the probability.

The global likelihood function is composed of three likelihood functions. Two describe the decay parameters of the two tau leptons. The last quantifies the compatibility of a tau decay hypothesis with the \cancel{E}_T measured in an event, assuming that the neutrinos produced in tau decays are the only source of \cancel{E}_T . The likelihood functions used to model the tau decay kinematics are different for leptonic and hadronic tau decays. The leptonic function is based on the matrix elements [162]:

$$\mathcal{L}_{\tau}^{lep} = \frac{d\Gamma}{dx dm_{\nu\nu} d\phi} \propto \frac{m_{\nu\nu}}{4m_{\tau}^2} [(m_{\tau}^2 + 2m_{\nu\nu}^2)(m_{\tau}^2 - m_{\nu\nu}^2)]$$

within the physically allowed region $0 \leq x \leq 1$ and $0 \leq m_{\nu\nu} \leq m_{\tau}\sqrt{1-x}$.

For the hadronic decay of the tau, a model based on two-body phase-space is used, assuming all visible decay products of the tau to constitute one ‘‘particle’’:

$$\mathcal{L}_{\tau}^{had} = \frac{d\Gamma}{dx d\phi} = \frac{1}{2\pi} \left(\frac{1}{1 - \frac{m_{vis}^2}{m_{\tau}^2}} \right)$$

within the physically allowed region $\frac{m_{vis}^2}{m_{\tau}^2} \leq x \leq 1$.

The \cancel{E}_T likelihood can be expressed by the following formula:

$$\mathcal{L}_{MET} = \frac{1}{2\pi\sqrt{|V|}} \cdot \exp \left(-\frac{1}{2} \begin{pmatrix} \cancel{E}_x - \sum p_x^{\nu} \\ \cancel{E}_y - \sum p_y^{\nu} \end{pmatrix}^T \cdot V^{-1} \cdot \begin{pmatrix} \cancel{E}_x - \sum p_x^{\nu} \\ \cancel{E}_y - \sum p_y^{\nu} \end{pmatrix} \right) \quad (\text{D.1})$$

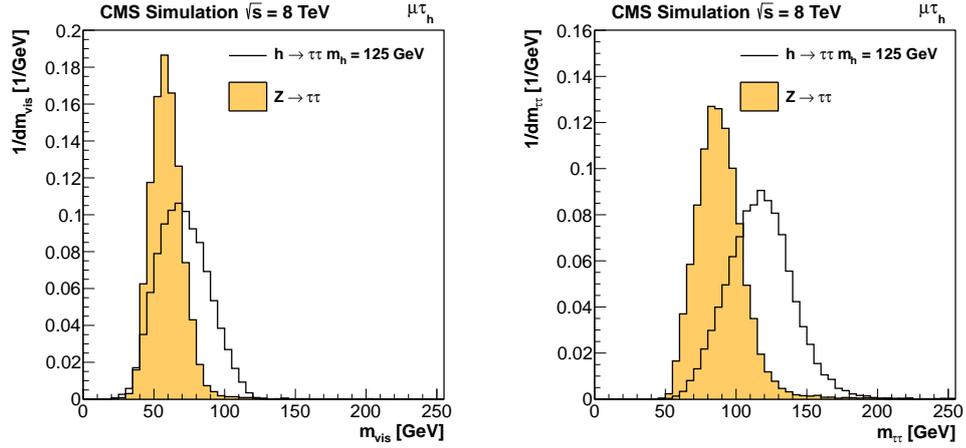


Figure D.1: Examples of the distributions of $M_{\tau\tau}$ candidate reconstructed by the SVfit algorithm (right) and of the mass of all visible tau decay products (left), for $h \rightarrow \tau\tau$ signal events and for the $Z \rightarrow \tau\tau$ background, in the final state $\tau\tau \rightarrow \mu\tau_h$ [144].

V is the covariance matrix that takes into account the expected resolution of the \cancel{E}_T reconstruction. The resolution of \cancel{E}_T is determined for each event by the MVA MET algorithm (section 3.2.7), which is around 10–15 GeV. The $|V|$ denotes the determinant of V . Differences between the sum of neutrino momenta, $\sum p_x^\nu$, and \cancel{E}_x ($\sum p_y^\nu$ and \cancel{E}_y) may arise due to resolution effects and are accounted for in the probability model, assuming a Gaussian resolution.

Fig. D.1 shows an example of the distribution of the visible mass (on the left) and of the SVfit mass (on the right) for the SM $H \rightarrow \tau\tau$ and for $Z \rightarrow \tau\tau$ processes, in the $\mu\tau_h$ final state. It can be seen that the SVfit mass improves the shape separation. In case of the SM $H \rightarrow \tau\tau$ searches, the sensitivity of the CMS analyses was improved of a 30% by using SVfit with respect to the performance obtained by M_{vis} [71, 144].

Bibliography

- [1] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [4] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [5] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [6] M. E. Peskin and D. V. Schroeder, “An Introduction to quantum field theory”. 1995.
- [7] M. D. Schwartz, “Quantum Field Theory and the Standard Model”. Cambridge University Press, 2014.
- [8] ATLAS, CMS Collaboration, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.
- [9] 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”, CMS-PAS-HIG-15-002, 2015.
- [10] LSND Collaboration, “Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam”, *Phys. Rev.* **D64** (2001) 112007, doi:10.1103/PhysRevD.64.112007, arXiv:hep-ex/0104049.

- [11] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.
- [12] M. Trodden, “Electroweak baryogenesis: A Brief review”, in *Proceedings, 33rd Recontres de Moriond 98 electroweak interactions and unified theories*, pp. 471–480. 1998. arXiv:hep-ph/9805252.
- [13] Planck Collaboration, “Planck 2015 results. XIII. Cosmological parameters”, arXiv:1502.01589, 2015.
- [14] A. Boyarsky, O. Ruchayskiy, D. Iakubovskiy, and J. Franse, “Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster”, *Phys. Rev. Lett.* **113** (2014) 251301, doi:10.1103/PhysRevLett.113.251301, arXiv:1402.4119.
- [15] E. Gildener and S. Weinberg, “Symmetry Breaking and Scalar Bosons”, *Phys. Rev.* **D13** (1976) 3333, doi:10.1103/PhysRevD.13.3333.
- [16] S. P. Martin, “A Supersymmetry primer”, doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001, arXiv:hep-ph/9709356, 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [17] I. J. R. Aitchison, “Supersymmetry and the MSSM: An Elementary introduction”, arXiv:hep-ph/0505105, 2005.
- [18] A. Djouadi, “The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model”, *Phys. Rept.* **459** (2008) 1–241, doi:10.1016/j.physrep.2007.10.005, arXiv:hep-ph/0503173.
- [19] G. C. Branco et al., “Theory and phenomenology of two-Higgs-doublet models”, *Phys. Rept.* **516** (2012) 1–102, doi:10.1016/j.physrep.2012.02.002, arXiv:1106.0034.
- [20] A. Djouadi et al., “The post-Higgs MSSM scenario: habemus MSSM?”, *Eur. Phys. J. C* **73** (2013) 2650, doi:10.1140/epjc/s10052-013-2650-0, arXiv:1307.5205.
- [21] A. Djouadi et al., “Fully covering the MSSM Higgs sector at the LHC”, *JHEP* **06** (2015) 168, doi:10.1007/JHEP06(2015)168, arXiv:1502.05653.
- [22] A. Djouadi and J. Quevillon, “The MSSM Higgs sector at a high M_{SUSY} : reopening the low $\tan\beta$ regime and heavy Higgs searches”, *JHEP* **10** (2013) 028, doi:10.1007/JHEP10(2013)028, arXiv:1304.1787.
- [23] N. Craig, J. Galloway, and S. Thomas, “Searching for Signs of the Second Higgs Doublet”, arXiv:1305.2424, RU-NHETC-2013-07, 2013.

- [24] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”, doi:10.5170/CERN-2013-004, arXiv:1307.1347, CERN-2013-004, FERMILAB-CONF-13-667-T, 2013.
- [25] ATLAS Collaboration, “Searches for Higgs boson pair production in the $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector”, arXiv:1509.04670, CERN-PH-EP-2015-225, 2015.
- [26] ATLAS Collaboration, “Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector”, *Phys. Rev. Lett.* **114** (2015) 081802, doi:10.1103/PhysRevLett.114.081802, arXiv:1406.5053.
- [27] ATLAS Collaboration, “Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, arXiv:1506.00285, CERN-PH-EP-2015-099, 2015.
- [28] CMS Collaboration, “Searches for heavy Higgs bosons in two-Higgs-doublet models and for $t \rightarrow ch$ decay using multilepton and diphoton final states in pp collisions at 8 TeV”, *Phys.Rev.* **D90** (2014) 112013, doi:10.1103/PhysRevD.90.112013, arXiv:1410.2751.
- [29] CMS Collaboration, “Search for resonant pair production of Higgs bosons decaying to two bottom quark–antiquark pairs in proton–proton collisions at 8 TeV”, *Phys. Lett.* **B749** (2015) 560–582, doi:10.1016/j.physletb.2015.08.047, arXiv:1503.04114.
- [30] CMS Collaboration, “Search for two Higgs bosons in final states containing two photons and two bottom quarks”, arXiv:1603.06896, CMS-HIG-13-032, CERN-EP-2016-050, 2016.
- [31] CMS Collaboration, “Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks”, *Submitted to: Eur. Phys. J. C* (2016) arXiv:1602.08762.
- [32] CMS Collaboration, “Model independent search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state”, CMS-PAS-HIG-15-013, 2016.
- [33] CMS Collaboration, “Search for resonant pair production of Higgs bosons decaying to $b\bar{b}$ and $\tau^+\tau^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV”, CMS-PAS-EXO-15-008, 2015.
- [34] 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 $h \rightarrow \tau\tau$ ”, *Phys. Lett.* **B755** (2016) 217–244, doi:10.1016/j.physletb.2016.01.056, arXiv:1510.01181.
- [35] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

- [36] C.-N. Yang and R. L. Mills, “Conservation of Isotopic Spin and Isotopic Gauge Invariance”, *Phys. Rev.* **96** (1954) 191–195, doi:10.1103/PhysRev.96.191.
- [37] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.
- [38] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.
- [39] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc.* **C680519** (1968) 367–377.
- [40] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3 Collaboration, “Precision electroweak measurements on the Z resonance”, *Phys. Rept.* **427** (2006) 257–454, doi:10.1016/j.physrep.2005.12.006, arXiv:hep-ex/0509008.
- [41] UA1 Collaboration, “Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c² at the CERN SPS Collider”, *Phys. Lett.* **B126** (1983) 398–410, doi:10.1016/0370-2693(83)90188-0.
- [42] UA2 Collaboration, “Evidence for $Z^0 \rightarrow e^+ e^-$ at the CERN anti-p p Collider”, *Phys. Lett.* **B129** (1983) 130–140, doi:10.1016/0370-2693(83)90744-X.
- [43] UA1 Collaboration, “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $s^{1/2} = 540\text{-GeV}$ ”, *Phys. Lett.* **B122** (1983) 103–116, doi:10.1016/0370-2693(83)91177-2.
- [44] UA2 Collaboration, “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider”, *Phys. Lett.* **B122** (1983) 476–485, doi:10.1016/0370-2693(83)91605-2.
- [45] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.
- [46] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132, doi:10.1016/0031-9163(64)91136-9.
- [47] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.
- [48] A. Djouadi, “The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model”, *Phys. Rept.* **457** (2008) 1–216, doi:10.1016/j.physrep.2007.10.004, arXiv:hep-ph/0503172.

- [49] SLD Electroweak Group, SLD Heavy Flavor Group, DELPHI, LEP, ALEPH, OPAL, LEP Electroweak Working Group, L3 Collaboration, “A Combination of preliminary electroweak measurements and constraints on the standard model”, [arXiv:hep-ex/0312023](#), SLAC-R-701, LEPEWWG-2003-02, ALEPH-2003-017-PHYSICS-2003-005, DELPHI-2003-072-PHYS-937, L3-NOTE-2825, OPAL-PR-392, SLD-PHYSICS-NOTE-78, CERN-EP-2003-091, LEP-EWWG-2003-02, 2003.
- [50] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”, [doi:10.5170/CERN-2011-002](#), [arXiv:1101.0593](#), CERN-2011-002, 2011.
- [51] “LEP design report”. CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.
- [52] OPAL, DELPHI, LEP Working Group for Higgs boson searches, ALEPH, L3 Collaboration, “Search for the standard model Higgs boson at LEP”, *Phys. Lett.* **B565** (2003) 61–75, [doi:10.1016/S0370-2693\(03\)00614-2](#), [arXiv:hep-ex/0306033](#).
- [53] “Design report Tevatron 1 Project”. FERMILAB, Batavia, 1982.
- [54] CDF, D0 Collaboration, “Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron”, *Phys. Rev. Lett.* **109** (2012) 071804, [doi:10.1103/PhysRevLett.109.071804](#), [arXiv:1207.6436](#).
- [55] ATLAS Collaboration, “Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector”, *Eur. Phys. J.* **C75** (2015), no. 10, 476, [doi:10.1140/epjc/s10052-015-3685-1](#), [arXiv:1506.05669](#).
- [56] CMS Collaboration, “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV”, *Phys. Rev. D* **92** (2015), no. 1, 012004, [doi:10.1103/PhysRevD.92.012004](#), [arXiv:1411.3441](#).
- [57] ATLAS Collaboration, “Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector”, *Eur. Phys. J.* **C75** (2015), no. 7, 335, [doi:10.1140/epjc/s10052-015-3542-2](#), [arXiv:1503.01060](#).
- [58] CMS Collaboration, “Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs”, *Phys. Lett.* **B736** (2014) 64, [doi:10.1016/j.physletb.2014.06.077](#), [arXiv:1405.3455](#).
- [59] U. Amaldi, W. de Boer, and H. Furstenau, “Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP”, *Phys. Lett.* **B260** (1991) 447–455, [doi:10.1016/0370-2693\(91\)91641-8](#).

- [60] J. Grigo, J. Hoff, K. Melnikov, and M. Steinhauser, “On the Higgs boson pair production at the LHC”, *Nucl. Phys.* **B875** (2013) 1–17, doi:10.1016/j.nuclphysb.2013.06.024, arXiv:1305.7340.
- [61] S. Dawson et al., “Working Group Report: Higgs Boson”, in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*. 2013. arXiv:1310.8361.
- [62] J. Baglio et al., “The measurement of the Higgs self-coupling at the LHC: theoretical status”, *JHEP* **04** (2013) 151, doi:10.1007/JHEP04(2013)151, arXiv:1212.5581.
- [63] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak Interactions with Lepton-Hadron Symmetry”, *Phys. Rev.* **D2** (1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.
- [64] ATLAS Collaboration, “Search for charged Higgs bosons decaying via $H^\pm \rightarrow \tau^\pm \nu$ in fully hadronic final states using pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *JHEP* **03** (2015) 088, doi:10.1007/JHEP03(2015)088, arXiv:1412.6663.
- [65] CMS Collaboration, “Search for charged Higgs bosons with the H^\pm to tau nu decay channel in the fully hadronic final state at $\sqrt{s} = 8$ TeV”, CMS-PAS-HIG-14-020, 2014.
- [66] ATLAS Collaboration, “Search for charged Higgs bosons in the τ +jets final state with pp collision data recorded at $\sqrt{s} = 8$ TeV with the ATLAS experiment”, ATLAS-CONF-2013-090, 2013.
- [67] ATLAS Collaboration Collaboration, “Search for neutral Higgs bosons of the minimal supersymmetric standard model in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, arXiv:1409.6064, CERN-PH-EP-2014-210, 2014.
- [68] CMS Collaboration, “Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions”, *JHEP* **10** (2014) 160, doi:10.1007/JHEP10(2014)160, arXiv:1408.3316.
- [69] CMS Collaboration, “Search for Neutral MSSM Higgs Bosons Decaying into A Pair of Bottom Quarks”, arXiv:1506.08329, CMS-HIG-14-017, CERN-PH-EP-2015-133, 2015.
- [70] E. Bagnaschi et al., “Benchmark scenarios for low $\tan \beta$ in the MSSM”, Technical Report LHCHXSWG-2015-002, CERN, Geneva, Aug, 2015.
- [71] CMS Collaboration, “Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons”, *JHEP* **05** (2014) 104, doi:10.1007/JHEP05(2014)104, arXiv:1401.5041.

- [72] ATLAS Collaboration, “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”, *JHEP* **04** (2015) 117, doi:10.1007/JHEP04(2015)117, arXiv:1501.04943.
- [73] CMS Collaboration, “Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks”, *Phys. Rev.* **D89** (2014), no. 1, 012003, doi:10.1103/PhysRevD.89.012003, arXiv:1310.3687.
- [74] 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”, *JHEP* **01** (2015) 069, doi:10.1007/JHEP01(2015)069, arXiv:1409.6212.
- [75] CMS Collaboration, “Evidence for the direct decay of the 125 GeV Higgs boson to fermions”, *Nature Phys.* **10** (2014) 557–560, doi:10.1038/nphys3005, arXiv:1401.6527.
- [76] CMS Collaboration, “Summaries of CMS cross section measurements”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>, 2015.
- [77] J. Haffner, “The CERN accelerator complex. Complexe des accélérateurs du CERN”, OPEN-PHO-ACCEL-2013-056, 2013. General Photo.
- [78] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.
- [79] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.
- [80] TOTEM Collaboration, “Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at $\sqrt{s} = 8$ TeV”, *Phys. Rev. Lett.* **111** (2013), no. 1, 012001, doi:10.1103/PhysRevLett.111.012001.
- [81] CMS Collaboration, “Public CMS Luminosity Information”. <http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>, 2015.
- [82] CMS Collaboration, V. Karimäki et al., “The CMS tracker system project: Technical Design Report”. CERN-LHCC-98-6. CERN, Geneva, 1997.
- [83] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [84] CMS Collaboration, “CMS: The electromagnetic calorimeter: Technical design report”. CERN-LHCC-97-33. CERN, Geneva, 1997.

- [85] C. W. Fabjan, “Calorimetry in High-Energy Physics”, *NATO Sci. Ser. B* **128** (1985) 281–336, doi:10.1007/978-1-4684-5018-7_5.
- [86] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”. CERN-LHCC-97-31. CERN, Geneva, 1997.
- [87] USCMS, ECAL/HCAL Collaboration, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”, *Eur. Phys. J.* **C60** (2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5, 10.1140/epjc/s10052-009-1024-0. [Erratum: *Eur. Phys. J.*C61,353(2009)].
- [88] CMS Collaboration, G. Acquistapace et al., “CMS, the magnet project: Technical design report”. CERN-LHCC-97-10. CERN, Geneva, 1997.
- [89] CMS Collaboration, “CMS, the Compact Muon Solenoid. Muon technical design report”. CERN-LHCC-97-32. CERN, Geneva, 1997.
- [90] CMS Collaboration, “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC”, *JINST* **8** (2013) P11002, doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.
- [91] CMS Collaboration, “CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems”. CERN-LHCC-2000-038. CERN, Geneva, 2000.
- [92] CMS Collaboration, S. Cittolin, A. Rácz, and P. Sphicas, “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project”. CERN-LHCC-2002-026. CERN, Geneva, 2002.
- [93] CMS Collaboration, “CMS software”. <https://github.com/cms-sw>, 2015.
- [94] CMS Collaboration G. L. Bayatyan, et al., “CMS computing: Technical Design Report”. CERN-LHCC-2005-023. CERN, Geneva, 2005.
- [95] CMS Collaboration, “CMS physics: Technical design report Volume 1: Detector Performance and Software”, CERN-LHCC-2006-001, CMS-TDR-008-1, 2006.
- [96] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [97] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: s - and t -channel contributions”, *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076.
- [98] J. Alwall et al., “MadGraph 5: going beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

- [99] GEANT4 Collaboration, “GEANT4 — a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [100] W. Adam, B. Mangano, T. Speer, and T. Todorov, “Track reconstruction in the CMS tracker”, CERN-CMS-NOTE-2006-041, CMS-NOTE-2006-041, 2005.
- [101] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.
- [102] J. R. Cash and A. H. Karp, “A Variable Order Runge-Kutta Method for Initial Value Problems with Rapidly Varying Right-hand Sides”, *ACM Trans. Math. Softw.* **16** (September, 1990) 201–222, doi:10.1145/79505.79507.
- [103] CMS Collaboration, “Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **01** (2011) 080, doi:10.1007/JHEP01(2011)080, arXiv:1012.2466.
- [104] Rose, K., “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239, doi:{10.1109/5.726788}.
- [105] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting”, *J. Phys.* **G34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.
- [106] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS-PAS-PFT-09-001, 2009.
- [107] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, CMS-PAS-PFT-10-001, 2010.
- [108] CMS Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, CMS-PAS-PFT-10-002, 2010.
- [109] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC”, *eConf C0303241* (2003) TULT009, doi:10.1088/0954-3899/31/9/N01, arXiv:physics/0306087. [J. Phys.G31,N9(2005)].
- [110] S. Baffioni et al., “Electron reconstruction in CMS”, *Eur. Phys. J.* **C49** (2007) 1099–1116, doi:10.1140/epjc/s10052-006-0175-5.
- [111] CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

- [112] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, “Electron Reconstruction in the CMS Electromagnetic Calorimeter”, CMS-NOTE-2001-034, CERN-CMS-NOTE-2001-034, 2001.
- [113] A. Hocker et al., “TMVA - Toolkit for Multivariate Data Analysis”, *PoS ACAT* (2007) 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [114] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, [doi:10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002), [arXiv:1206.4071](https://arxiv.org/abs/1206.4071).
- [115] G. P. Salam, “Towards Jetography”, *Eur. Phys. J.* **C67** (2010) 637–686, [doi:10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6), [arXiv:0906.1833](https://arxiv.org/abs/0906.1833).
- [116] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm”, *JHEP* **04** (2008) 063, [doi:10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063), [arXiv:0802.1189](https://arxiv.org/abs/0802.1189).
- [117] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J.* **C72** (2012) 1896, [doi:10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2), [arXiv:1111.6097](https://arxiv.org/abs/1111.6097).
- [118] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett.* **B659** (2008) 119–126, [doi:10.1016/j.physletb.2007.09.077](https://doi.org/10.1016/j.physletb.2007.09.077), [arXiv:0707.1378](https://arxiv.org/abs/0707.1378).
- [119] CMS Collaboration, “Pileup Jet Identification”, CMS-PAS-JME-13-005, 2013.
- [120] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002, [doi:10.1088/1748-0221/6/11/P11002](https://doi.org/10.1088/1748-0221/6/11/P11002), [arXiv:1107.4277](https://arxiv.org/abs/1107.4277).
- [121] CMS Collaboration, “Plans for Jet Energy Corrections at CMS”, CMS-PAS-JME-07-002, 2008.
- [122] S. Poprocki, “Search for WZ+ZZ Production with Missing Transverse Energy and b Jets at CDF”. PhD thesis, Cornell U., 2013.
- [123] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, [doi:10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013), [arXiv:1211.4462](https://arxiv.org/abs/1211.4462).
- [124] CMS Collaboration, “Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 02, P02006, [doi:10.1088/1748-0221/10/02/P02006](https://doi.org/10.1088/1748-0221/10/02/P02006), [arXiv:1411.0511](https://arxiv.org/abs/1411.0511).
- [125] CMS Collaboration, “Performance of tau-lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, [doi:10.1088/1748-0221/7/01/P01001](https://doi.org/10.1088/1748-0221/7/01/P01001), [arXiv:1109.6034](https://arxiv.org/abs/1109.6034).

- [126] CMS Collaboration, “Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS”, [arXiv:1510.07488](#), CMS-TAU-14-001, CERN-PH-EP-2015-261, 2015.
- [127] S. Badger, B. Biedermann, P. Uwer, and V. Yundin, “NLO QCD corrections to multi-jet production at the LHC with a centre-of-mass energy of $\sqrt{s} = 8$ TeV”, *Phys. Lett.* **B718** (2013) 965–978, [doi:10.1016/j.physletb.2012.11.029](#), [arXiv:1209.0098](#).
- [128] CMS collaboration, “The CMS Trigger”, *CMS Paper* (2015). *Draft CMS paper*.
- [129] H.-L. Lai et al., “New parton distributions for collider physics”, *Phys. Rev.* **D82** (2010) 074024, [doi:10.1103/PhysRevD.82.074024](#), [arXiv:1007.2241](#).
- [130] J. Pumplin et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **07** (2002) 012, [doi:10.1088/1126-6708/2002/07/012](#), [arXiv:hep-ph/0201195](#).
- [131] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC”, *Eur. Phys. J.* **C63** (2009) 189–285, [doi:10.1140/epjc/s10052-009-1072-5](#), [arXiv:0901.0002](#).
- [132] M. Botje et al., “The PDF4LHC Working Group Interim Recommendations”, [arXiv:1101.0538](#), 2011.
- [133] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, [doi:10.1088/1126-6708/2004/11/040](#), [arXiv:hep-ph/0409146](#).
- [134] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, [doi:10.1088/1126-6708/2007/11/070](#), [arXiv:0709.2092](#).
- [135] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, [doi:10.1007/JHEP06\(2010\)043](#), [arXiv:1002.2581](#).
- [136] CMS Collaboration, “Measurement of differential top-quark pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J.* **C73** (2013), no. 3, 2339, [doi:10.1140/epjc/s10052-013-2339-4](#), [arXiv:1211.2220](#).
- [137] CMS Collaboration, “Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV”, CMS-PAS-TOP-14-016, 2014.

- [138] CMS Collaboration, “Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev. Lett.* **112** (2014) 191802, doi:10.1103/PhysRevLett.112.191802, arXiv:1402.0923.
- [139] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $O(\alpha(s)^2)$ ”, *Phys. Rev.* **D74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.
- [140] CMS Collaboration, “Measurement of W^+W^- and ZZ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett.* **B721** (2013) 190–211, doi:10.1016/j.physletb.2013.03.027, arXiv:1301.4698.
- [141] CMS Collaboration, “Measurement of WZ production rate”, CMS-PAS-SMP-12-006, 2013.
- [142] N. Kidonakis, “NNLL threshold resummation for top-pair and single-top production”, *Phys. Part. Nucl.* **45** (2014), no. 4, 714–722, doi:10.1134/S1063779614040091, arXiv:1210.7813.
- [143] CMS Collaboration, “Pileup Information”.
<https://twiki.cern.ch/twiki/bin/view/CMS/PdmVPileUpDescription>, 2015. *Internal CMS webpage*.
- [144] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, “Reconstruction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood techniques”, *J. Phys. Conf. Ser.* **513** (2014) 022035, doi:10.1088/1742-6596/513/2/022035.
- [145] University of Hamburg Higgs Group, “HHKinFit – a kinematic fitting package to fit heavy Higgs decays”, *CMS Note* **2014/163** (2014). *Internal CMS document*.
- [146] E. Schlieckau, “Measurement of Top-Quark Mass and Inclusive Top-Quark-Pair Production Cross Section in pp Collisions at $\sqrt{s} = 7/8$ TeV with CMS”. PhD thesis, University of Hamburg, 2014.
- [147] N. Davidson et al., “Universal interface of TAUOLA: Technical and physics documentation”, *Comput. Phys. Commun.* **183** (2012) 821, doi:10.1016/j.cpc.2011.12.009, arXiv:1002.0543.
- [148] John Wiley and Sons Collaboration, “Handbook of Methods of Applied Statistics”, 1967.
- [149] ATLAS Collaboration, “Procedure for the LHC Higgs boson search combination in summer 2011”, ATL-PHYS-PUB-2011-011, ATL-COM-PHYS-2011-818, CMS-NOTE-2011-005, 2011.
- [150] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, CMS-PAS-LUM-13-001, 2013.

- [151] CMS Collaboration, “Performance of b tagging at $\sqrt{s}=8$ TeV in multijet, ttbar and boosted topology events”, CMS-PAS-BTV-13-001, 2013.
- [152] R. J. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Comput. Phys. Commun.* **77** (1993) 219–228, doi:10.1016/0010-4655(93)90005-W.
- [153] Higgs TauTau group, “Search for MSSM Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions”, *CMS Note* **2013/171** (2013). *Internal CMS document*.
- [154] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: *Eur. Phys. J.* **C73**,2501(2013)].
- [155] R. Harlander et al., “Interim recommendations for the evaluation of Higgs production cross sections and branching ratios at the LHC in the Two-Higgs-Doublet Model”, arXiv:1312.5571, KA-TP-41-2013, LU-TP-13-44, PSI-PR-13-17, WUB-13-19, LHCHXSWG-2013-001, 2013.
- [156] A. L. Read, “Linear interpolation of histograms”, *Nucl. Instrum. Meth.* **A425** (1999) 357–360, doi:10.1016/S0168-9002(98)01347-3.
- [157] CMS Collaboration, “Public CMS - Higgs PAG Summary Plots”. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryResultsHIG#ttH_combination_for_Moriond_2016, 2016.
- [158] CMS Collaboration, “Measurement of CMS Luminosity”, CMS-PAS-EWK-10-004, 2010.
- [159] CMS Collaboration, “Jet and underlying event properties as a function of charged-particle multiplicity in proton–proton collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J.* **C73** (2013) 2674, doi:10.1140/epjc/s10052-013-2674-5, arXiv:1310.4554.
- [160] R. Manzoni, “Search for a Higgs boson decaying into a $di-\tau$ pair in the double hadronic final state”. PhD thesis, University of Milano Bicocca, 2014.
- [161] CMS Collaboration, “Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector”, CMS-PAS-JME-12-002, 2012.
- [162] B. K. Bullock, K. Hagiwara, and A. D. Martin, “Tau polarization and its correlations as a probe of new physics”, *Nucl. Phys.* **B395** (1993) 499–533, doi:10.1016/0550-3213(93)90045-Q.