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

Search for Higgs bosons decaying to tau leptons with the CMS experiment at the LHC

Doctoral dissertation presented by

Lucia Perrini

in fulfilment of the requirements for the degree of Doctor in Sciences

Jury de thèse

Prof. V. Lemaître (President)	UCL, Belgium
Prof. G. Bruno (Advisor)	UCL, Belgium
Prof. F. Maltoni	UCL, Belgium
Prof. B. Clerbaux	ULB, Belgium
Prof. C. Delaere	UCL, Belgium
Dr. G. Gòmez-Ceballos Retuerto	MIT, USA

October 9^{th}, 2015

"Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less."
Marie Curie

Acknowledgments

The experience of the Ph.D. has gone beyond any possible expectation. I grow up as a physicist, but also as a woman. And it is now time to thank all the people that have contributed to make precious these last four years.

Thanks to my supervisor, Prof. Giacomo Bruno. Thanks Giacomo for having given me the opportunity to start the Ph.D., for your constant support, your precious and practical helps, for what you have taught me, for your guidance in every single step of this amazing experience and for having impressed me with your smartness and incredible knowledge. Thanks Loic, for having jumped in my life when I needed to re-orient my point of view on how much passionate doing research can be. Thanks for having believed in me more than I was believing in myself. Your efficiency, your pragmatism, your curiosity, your knowledge and brutal honest will continue guiding me. Thanks Pavel, my first unforgettable officemate. Thanks for having brought me in the "tau-friends" group and having gifted me with the possibility of creating contacts with other physicists, for having shared our pain during the terrible period of the cross check, for your kindness and your confidentiality.

Thanks to Prof. Vincent Lemaitre, Prof. Fabio Maltoni, Prof. Christophe Delaere, Prof. Barbara Clerbaux, Dr. Guillelmo Gòmez-Ceballos: thanks for having reviewed my thesis and helped me in improving its contents and its shape, for all your comments and suggestions, for our useful discussions. Thanks for having made the private defence a nice memory and an educational experience: I will never forget the atmosphere, the time you gave me to think and concentrate, the privilege of feeling one of you. Thanks.

Thanks to Dr. Anna Colaleo and Prof. Salvatore Nuzzo, because initiating someone to this incredible journey of being an experimental particle physicist is a challenge and you have reached the goal in the best way possible.

Thanks to all my CP3 friends. First of all, Ginette. Yes, you are a dear friend for me, beyond your being the most qualified and efficient secretary I know.

You helped me in breaking those disagreeable walls full of shy, you gave me hugs, and I gave you hugs, whenever the time was right. And if I manage to speak a decent french, it is just for you. Thanks Elisa, Alessio, Joze, Adrien, Roberto, Tristan, David, Miguel, Bob, Laurent, Michele, Olivier, for having shared nice moments inside and outside the lab. Thanks to my coloured dear Ludi and to my "pizza-afecionado" and crazy friend Fede, simply for having always been there (in good times and bad times). Thanks to Nico, for having managed to recreate a bit of our own hometown also in Brussels. Thanks Ale, Paolo, Nicola, Philippe: meeting you has been one of the most important and funny experience of these four years spent in Brussels. After all, it does not happen every day to share something with so cute beavers! Thanks Anna P. and Alice for all the soothing moments and Paula for the pain shared about our theses (e per le tante foto e per le tante figurine). And thanks to Angela, Raffaella, Maria Teresa, Alessia, Resy, for that friendship that transcends time and space.

Grazie alla mia seconda famiglia, Giuliana, Duccio, Giorgia e Malfattone, per la vostra presenza ed il vostro affetto. Grazie mamma, babbo, Mari e Luca. Grazie per ESE's sempre, per sopportare e comprendere i miei silenzi e le mie decisioni, per soffrire e gioire con me. Per essere cosí orgogliosi di me. Grazie Greta, perché sei il dono che ha reso piú belle le nostre vite.

E GRAZIE Anto, per il tuo essere e il tuo fare, perché il viaggio vissuto con te é una benedizione.

Abstract

With the first LHC proton-proton (p-p) collisions at a center-of-mass energy $\sqrt{s} = 7$ TeV occurred in spring 2010, the world witnessed the beginning of a new era in particle physics. From that milestone onward, the Standard Model (SM), the theory that describes matter constituents and their interactions, has been further verified and a large number of searches for new phenomena have been performed. The SM foresees the existence of the Higgs boson, which is the particle that gives mass to fermions and gauge bosons. The discovery of a SM Higgs-like particle in bosonic channels has been announced on July 2012 by the ATLAS and CMS experiments at LHC. Signals of such a particle in the fermionic decay channels were, instead, still not revealed when the work presented in this thesis started. Moreover, it is known that the SM has limitations which might be explained by extended theoretical scenarios.

This thesis has dealt on one side with finding traces, with the CMS experiment, of the SM Higgs boson decay in the fermionic channels and on the other with finding evidence of possible scenarios beyond the SM. In the former context, the search for the SM Higgs boson h decaying into a pair of tau leptons and produced in association with a Z boson decaying into either a dielectron or dimuon pair, has been performed. By combining the result of this search with a few others, evidence for the decay of the newly discovered particle in the ditau channel has been found, demonstrating its Higgs-like nature as predicted by the SM. In the second part of the thesis, a search for a new resonance decaying to a Z boson and a lighter resonance, as predicted by models with two Higgs doublets (2HDM, which are extensions of the SM), has been performed. The decay of the Z boson into dilepton pairs and that of the light resonance into a pair of tau leptons have been considered. With no significant excess observed over expected backgrounds, 95% confidence level (CL) exclusion limits on the signal cross section have been set.

The performance of the algorithm used in the CMS experiment to reconstruct and identify tau leptons, crucial for the analyses presented in this thesis, have also been widely studied and presented in this manuscript.

Contents

Introduction

The harmony and the beauty of the Nature has motivated the humankind over the centuries to wonder about the laws that rule it. In the 5^{th} century B.C., Empedocle romantically introduced his own vision of the matter composition: four ultimate elements, water, earth, air and fire are seen as attracted and separated by each other by the force of love and hate, respectively. The first concept that all the visible matter consists of indivisible unities called *atoms* came few years later with the Greek atomist doctrine (the word *à-tomos*, in fact, means *indivisible* in Greek), conceived by Leucippus and formalized by his disciple Democritus. The first experimental evidence that, in reality, atoms have their own internal substructure, arrived in 1897 with the discovery of the electron by Thomson [1]. That milestone was only the beginning of the search for the fundamental constituents of the Universe, and the interactions amongst them, a challenging adventure that still continues to fascinate the worldwide physics community.

More than one century of reasoning, theoretical computations, tremendous experimental efforts, and crucial discoveries have led, today, to claim that most of the existing Nature can be described by a powerful theory called Standard Model (SM). The SM describes all fundamental interactions, other than gravity, existing amongst a small number (twenty-four) of elementary particles. From an experimental point of view, the SM is impeccable. It has predicted the existence of the weak force carriers, the gauge bosons W and Z, before their observation at the UA1 and UA2 experiments [2, 3, 4, 5] at CERN (*Conseil Européen pour la Recherche Nucláire*) in Geneva, Switzerland. Also the charm and the top quarks have been observed [6, 7, 8] only after having been predicted by the theoretical model. However, the discovery of a Higgs-like particle (considering the combination of the $\gamma\gamma$, ZZ and WW channels), announced by the ATLAS [9] and CMS [10] experiments, at the Large Hadron Collider (LHC) at CERN on July the 4^{th} 2012 [11, 12], has actually marked the beginning of a new era. Such a key milestone has confirmed one of the main SM predictions [13, 14] left experimentally unconfirmed for about fifty years: the mech-

anism through which the matter elementary constituents (leptons and quarks, the so-called *fermions*) and the weak force carriers (the so-called *gauge bosons*) acquire mass. A period of prudence and overwhelming experimental efforts followed, because if, from one side, the announcement generated enthusiasm, on the other side, the fact that the new resonance was exactly "the" Higgs boson predicted by the SM, was not yet confirmed. Studying if and how the newly discovered particle decays to fermions was the main goal to achieve a complete comprehension about this particle and its intrinsic properties. It is in this context that the search for the SM Higgs boson produced in association with a Z boson and decaying into a pair of tau leptons, which is presented in this thesis, found its fundamental motivation. Despite the lower branching ratio (cfr. Section 1.2.3), the decay in $\tau\tau$ is preferred to the one in $b\bar{b}$ given that it can be more efficiently selected out of the enormous multijet background from QCD processes at the LHC. The results obtained have been included in the wider $h \to \tau\tau$ search program performed by the CMS collaboration. In the corresponding paper [15], the first evidence of the 125 GeV Higgs boson decaying to a pair of tau leptons has been documented.

So far, the results obtained at the LHC seem to be aligned along the direction that sanctions the triumph of the SM. The new boson discovered seems more and more compatible with its SM description with increasing precision measurements. Moreover, not only no hints of new physics beyond the SM have been found, but also from the side of precision physics, nothing in disagreement with the SM predictions has been observed (for example, the first evidence of the rare decay $B_s^0 \to \mu^+\mu^-$, whose branching ratio has been measured to be compatible with the SM prediction [16]). However, there is a point of view from which this overwhelming experimental success of the SM is frustrating: it is well known that this theory cannot be the ultimate particle theory. The SM is not able to describe, for instance, the dark matter [17, 18] and the baryogenesis [19]. The theory does not include the gravity interaction and, in addition, it appears that a quantum description of gravity would lead to a non-renormalizable theory [20], differently from what the SM theory is. More precisely, the absence of the gravity is not the main problem in itself, at least because phenomenologically it plays a role only around the Planck mass ($M_P \simeq \mathcal{O}(10^{19})$ GeV). It is however ironic noticing that, even regardless of gravity, it is just at the Planck scale that the SM shows its weakness. In fact, the energy scale of the electroweak interactions is set by the vacuum expectation value of the Higgs boson $v \simeq 246$ GeV. It will be later explained that, differently from fermions and gauge bosons, the mass of the SM Higgs receives quadratic corrections which can only be limited by explicitly introducing a cut-off energy $\Lambda_{\text{cut-off}}$. Under the strong assumption that there are no

beyond SM theories up to the Planck scale, the only natural value that $\Lambda_{\text{cut-off}}$ can assume is the one of the Planck mass. This induces huge, and then *unnatural*, corrections to cure the divergences affecting the (small) Higgs mass. This severe problem is known as the *hierarchy* problem [21]. The most reliable conviction among the physics community is that there is a more general ultimate theory, which is able to explain the phenomena appearing at very high energy scale and that coherently accommodates the SM description at low scale. The Higgs sector is the connection between the SM and the unknown new physics phenomena. Among several models built to introduce new heavy particles able to solve the hierarchy problem, in this thesis the two-Higgs-doublet model (2HDM) is considered. According to the most general 2HDM formulation, two complex Higgs doublets exist and these lead to five physical states: two charged bosons $H^{\pm}$, two neutral scalar bosons h/H and one neutral pseudoscalar boson A. The search for a process in which a high mass resonance, either a H or a A boson, decays to a Z boson and a lighter resonance, either a A or a H boson, respectively, is presented in this thesis. Also in this case, the lighter resonance decay studied is the one in a pair of tau leptons. The results obtained in the context of this analysis, which have been included in a CMS Physics Analysis Summary [22], have been combined with those obtained in the $b\bar{b}$ channel [23].

The thesis is organized in order to guide the reader, at the best of the writer possibilities, towards a proper understanding of the final results. The SM theoretical framework, the phenomenology of the SM Higgs and experimental results obtained at the LHC concerning the discovery of the Higgs boson and its characterization is introduced in Chapter 1. The phenomenology of the Higgs sector in the context of the 2HDM is also introduced in the same chapter. The description of the LHC collider and the CMS experimental apparatus, to which the experimental analyses described in this thesis respond, is the subject of Chapter 2. Chapter 2 also contains the description of the simulation techniques adopted to generate hard process events, given the importance that the modelling of such events has for a proper optimization of the analysis strategies. Chapter 3 is devoted to the description of the techniques adopted, within the CMS collaboration, to reconstruct and identify physical objects. A particular attention is given to the particle-flow (PF) technique, which has been developed behind the idea of reconstructing and identifying a given object by coherently exploiting all the information of each CMS subdetector. Given the crucial role that the tau lepton has in the context of the experimental searches, a separated chapter is devoted to the reconstruction and the identification of this object. The description of the Hadron Plus Strips (HPS) algorithm [24] used in CMS to reconstruct tau leptons is provided in

Chapter 4. The SM and the 2HDM analyses are the subjects of the Chapter 5 and 6, respectively. Conclusions and future outlooks are then presented before approaching to the appendices. In particular, more results obtained in the study of the algorithm performance, which has played an important role in the doctoral work and are going to be published in [25], are shown in Appendix A. In Appendix B the statistical methods used to interpret the final experimental results obtained in the context of both SM and 2HDM analyses are described. Subchannels ditau mass distributions, related to the SM analysis, are documented in Appendix C. The last appendices of the manuscript, Appendix D and E, include additional supporting material for the two searches performed.

As far as concern Chapters 4, 5 and 6, and the Appendices A and C, the label [*A*] is used to indicate those plots originally made by the author of the manuscript. The label [*AC*], instead, is used to indicate those plots which the author have brought contributions. For all those plots, which have been obtained exploiting CMS data or/and simulated samples, the label "CMS" is adopted only in case they are already public.

1

The Standard Model and Beyond

The experimental searches reported in this thesis have been inspired by the Standard Model (SM) theory and the simplest possible extensions of the SM, the two-Higgs-doublet model (2HDM). This chapter has been fully designed to introduce both theoretical contexts and, in particular, to highlight the role of Higgs bosons. The SM will be firstly described as a theory able to describe all the elementary matter constituents and three of the four fundamental interactions these particles experience. The phenomenology of the SM Higgs boson will then be introduced and the particle arising in the context of the Brout-Englert-Higgs (BEH) mechanism will be also described from an experimental point of view. A more extended vision of the Higgs sector will be then provided looking beyond the SM theory. Despite the experimental remarkable success the SM theory has achieved, the incompleteness of the theory will be treated in the chapter and the 2HDM scenario will be motivated and introduced. The first LHC experimental constraints on the Type-II 2HDM scenario will be presented at the end of the chapter.

1.1 The Standard Model of particle physics

The Standard Model (SM) of particle physics [26, 27, 28] is a theory describing the known matter in terms of its elementary constituents and their inter-

actions. The theory describes three out of the four fundamental interactions existing in Nature: the electromagnetic, the strong and the weak interactions. The gravitation interaction is not included in the theory, but this does not pose a limitation on the description of the fundamental interactions at the LHC, because at the TeV scale[1], gravity is approximately forty orders of magnitude smaller than the strong nuclear force.

Despite the complexity of the theory formalism, the SM has an intrinsic simplicity in the way in which it allows to be explained. This peculiarity is going to be exploited in the next section, before providing an overview of the SM as a field theory.

1.1.1 Elementary particles and interactions

The *building blocks* of the matter are the fermions, which are elementary[2] particles having spin 1/2. The interactions amongst fermions are mediated by the exchange of elementary particles with spin 1, the gauge bosons.

Fermions are divided into leptons and quarks. Six leptons and six quarks exist in Nature and those are beautifully organized in three generations, each composed of two leptons and two quarks. The generations are organized (cfr. Table 1.1) in order to accommodate particles from the lightest and most stable to the heaviest and less stable ones.

	Generation I		**Generation II**		**Generation III**		Electric	Color?
Group	Symbol	Mass	Symbol	Mass	Symbol	Mass	charge	
leptons	ν_e	<2 eV	ν_μ	<2 eV	ν_τ	<2 eV	0	NO
	e^-	0.511 MeV[a]	μ^-	105.7 MeV	τ^-	1.777 GeV	-1	
quarks	u	2.3 MeV	c	1.27 GeV	t	173.2 GeV	$+2/3$	YES
	d	4.8 MeV	s	95 MeV	b	4.18 GeV	$-1/3$	

Table 1.1: The three generations of fermions and their properties. Mass values are taken from [29].

[a] In this thesis the system of natural units is adopted: $\hbar = c = 1$. Therefore, energy, momentum and mass are expressed with the same units.

The first generation is composed of the electron e^-, its corresponding neutrino ν_e, and the quarks up u and down d. The second generation accommodates

[1] The unit of energy used in particle physics is the *electronvolt* eV, which correspond to $1.6 \cdot 10^{-19}$ J. By definition, 1 eV corresponds to the amount of energy that an electron gains (or loses) moving across an electric field generated by a potential difference of 1 V.

[2] The term "elementary" means that these particles do not have an internal structure, at least according to the distances that can be probed with the current energies, which correspond to approximately 10^{-19} m.

the muon μ^-, its neutrino ν_μ, the charm c and strange s quarks. The tau τ^- and the tau neutrino ν_τ, together with the quarks top t and bottom b, constitute the third generation. Leptons and quarks have their own "antiparticles" (antileptons and antiquarks). An antiparticle has the same mass, spin and lifetime of its counterpart, but opposite charges.

Among the leptons, electrons, muons and taus are massive and electrically charged, while neutrinos are electrically neutral and assumed to be massless by the theory. All quarks are massive and characterized by a fractional electric charge. Additionally, quarks present the color charge, which exists in three possible states conventionally indicated as "red" (R), "green" (G) and "blue" (B). However, only neutral color charge states (conventionally indicated as "white"), which are bound states of quarks, exist in Nature: the *hadrons*. Hadrons include *mesons* and *baryons*. The former are bosons (spin 1) made of one quark and one antiquark, as the $\pi^+(u\bar{d})$, the latters are fermions composed of three quarks, as the proton $p(uud)$. To extend the view on hadrons, it has to be cited the prospect of having a more complex quark content, beyond the minimal composition characterized by $q\bar{q}$ or qqq, which was proposed by Gell-Mann [30] and Zweig [31] in 1964. These studies have inspired, two years later, the development of a quantitative model for two quarks plus two antiquarks composition [32], upon the introduction of structures composed of four quarks plus one antiquark [33, 34], the so-called *pentaquark* baryons [35]. On July 2015, the LHCb experiment has reported the observation of exotic structures in the charmonium-pentaquark state $J/\psi p$ [36], considering the decay process $\Lambda_b^0 \to J/\psi K^- p$.

The interaction of the twenty-four fermions (twelve particles plus twelve antiparticles) is mediated by the exchange of twelve elementary bosons. The photon γ, massless and electrically neutral, mediates the electromagnetic interaction among all fermions, but neutrinos. Eight gluons g, massless, electrically neutral, mediate the strong interaction amongst quarks and have color charge themselves. The gauge bosons $W^\pm$ and Z are the carriers of the weak interaction amongst all fermions. The photon, the Z boson and the gluons, are identical to their antiparticles; the W^+ antiparticle, instead, is the boson W^- and viceversa. The SM gauge bosons and their properties are summarized in Table 1.2.

1.1.2 The Standard Model as a gauge theory

The SM is a *Quantum Field Theory* (QFT) in the sense that all elementary particles and interactions are represented as a *field* defined in each point of the

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

Table 1.2: The SM force mediators and their properties. Mass values are taken from [29].

space-time. Before moving forward, it is worth to underline that the SM is a *renormalizable* theory. This peculiarity allows to reabsorb unphysical divergences, arising in intermediate steps of theoretical computations, inside the definition of physically measurable quantities. The SM renormalizability makes the theory finite and predictive.

The SM formalism is developed in terms of a Lagrangian field theory, which is governed by the *gauge principle* [37] based on the concept of *symmetry*. A system is considered symmetric if the solutions of a set of equations describing it remain unchanged even if the properties of the system that the equations describe change. Particular important for the following discussions are the definitions of *global* and *local* symmetries. A global symmetry acts identically and simultaneously in all the points of the space-time; a local symmetry, instead, depends on the space-time point in which it is applied and all the local variations can be independent by each other.

What is the idea behind the gauge principle?

By generalizing the global symmetry of a generic Lagrangian describing a free particle ψ to a local symmetry, the existence of a gauge boson field X_μ interacting with the particle ψ arises. This would be the end of the story in case of an *abelian* symmetry. However, in the case of *non-abelian* symmetries, more than one gauge boson appears, which does not only interact with fermions, but also self-interacts.

These concepts might be clarified if, to a given gauge symmetry, a mathematical group is associated. The SM Lagrangian is locally invariant under the gauge group

$$G_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \, . \tag{1.1}$$

The group $SU(3)_C$ responds to the symmetry under local color transformations. $SU(3)_C$ rules the strong interaction in the context of the quantum field theory called *Quantum Chromodynamics* (QCD). Despite the QCD has been the

last gauge theory to be formalized in the context of the SM, it features an elegant example of the full consistency of the gauge principle. For this reason it will be described first. The generator of $SU(3)_C$ are the 8 Gell-Man matrices, to which the local invariance under the color symmetry associate eight physical gauge fields: the gluons. The symmetry group has the peculiarity of being non-abelian, which mathematically translates in the generation of cubic and quartic gluon self-interaction terms. The gluons, in fact, carry themselves the color charge. The gauge principle let also the known quarks, which exists in color triplets, interacting with gluons. Moreover, the gauge boson fields have to be massless to preserve the local symmetry.

The strength of the strong interaction is governed by the strong coupling constant α_s, which is, however, not properly a constant. The coupling α_s, in fact, logarithmically decreases as a function of the energy Q^2 transferred in the process. This behaviour, known as *asymptotic freedom* [38] makes the strong interaction becoming weaker with decreasing distances, or increasing energies. At low energies ($\Lambda_{QCD} \approx 250$ MeV) the interactions among quarks become very strong, which experimentally translates in the observation of *colour confinement*: as already explained, quarks are observed only bounded together in hadrons. The perturbation theory, which represents the possibility to develop the calculus as a power series in α_s (cfr. Section 1.2.1), is not able to describe this process in which α_s diverges. However, theoretical techniques [39] have been developed to predict such a process. Therefore, while on the one hand it is difficult to describe low energy interactions of hadrons, on the other hand, high energy hadron collisions, such as those made possible at the LHC, can be described by means of perturbation theory, depicted via Feynman diagrams.

The symmetry group $SU(2)_L \otimes U(1)_Y$ is the one governing the electroweak interactions. The group describing the electromagnetic interaction, $U(1)_{em} \subset SU(2)_L \otimes U(1)_Y$ responds to the *Quantum Electrodynamics* (QED), the first gauge theory to be formulated. Glashow, Salam and Weinberg [26, 27, 28], at the end of the 60's, unified the weak and the electromagnetic interaction in a single sector: the electroweak sector.

$U(1)_Y$ is an abelian group describing the local symmetry under *weak hypercharge* transformations. The weak hypercharge Y is the generator of $U(1)_Y$ and, the local gauge invariance let the gauge boson field B_μ to be directly associated with the generator Y. $SU(2)_L$ is a non-abelian group describing the *weak isospin* symmetry. The three generators of $SU(2)_L$ are the three components of the weak isospin $\vec{T} \equiv (T_1, T_2, T_3)$, to which three gauge boson fields, $\vec{W}_\mu \equiv (W^1_\mu, W^2_\mu, W^3_\mu)$ are associated. The fermionic fields ψ interact with the gauge boson fields B_μ and $\vec{W}_\mu$, and, additionally, given the non-abelianity

of the $SU(2)_L$ group, the bosons $\vec{W}_\mu$ self-interact (these, in fact, carry weak isospin). Physically, the weak isospin and hypercharge are responsible of the fermion *chirality* transformations[3]. Left-handed fermions form isospin doublets ($T = 1/2$) are

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \qquad \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \qquad \begin{pmatrix} t \\ b \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \tag{1.2}$$

while right-handed fermions form singlets ($T = 0$) are

$$u_R,\ d_R,\ e_R, \qquad c_R,\ s_R,\ \mu_R, \qquad t_R,\ b_R,\ \tau_R\,. \tag{1.3}$$

The right-handed state of neutrinos has not yet been observed directly, although there is indirect evidence of its existence from experimental results on neutrino oscillations [40, 41]. The weak interaction only acts on left-handed particles (and right-handed antiparticles) and, thus, violates parity [42], which is the reversal transformation of the the spatial coordinates. The weak isospin and hypercharge are related to the electric charge Q ($Q = T_3 + Y/2$), which is the generator of the abelian group $U(1)_{em}$. The gauge boson field assigned to the generator Q is the photon field A_μ, with which the gauge local invariance of the system under electric charge transformation let all charged fermions interacting. Moreover, given the abelianity of the symmetry group, A_μ does not interact with itself: the photon, in fact, does not carry any electric charge.

Why, however, the weak gauge bosons are four despite it has been previously explained that the weak force carriers are three?

The answer is that the gauge fields $\vec{W}_\mu$ and B_μ do not correspond to physical particles. The mediators of the weak interaction, $W^\pm$ and Z, and the photon too, are combinations of the electroweak unphysical fields:

$$W^\pm_\mu = \frac{W^1_\mu \mp i W^2_\mu}{\sqrt{2}} \qquad Z_\mu = \frac{g' W^3_\mu - g B_\mu}{\sqrt{g^2 + g'^2}} \qquad A_\mu = \frac{g' W^3_\mu + g B_\mu}{\sqrt{g^2 + g'^2}}\,. \tag{1.4}$$

The symbols g and g', in Equation 1.4, refers to the electroweak couplings. These couplings are such that

$$e = g \sin\theta_W = g' \cos\theta_W\,, \tag{1.5}$$

[3]The chirality is an intrinsic property of a particle, strictly related to the quantum mechanics description of the particle itself. Being ψ a fermionic field, its *left-handed* and *right-handed* chiralities are $\psi_L = \frac{1}{2}(1 - \gamma^5)\psi$ and $\psi_R = \frac{1}{2}(1 + \gamma^5)\psi$, respectively (where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, with γ_i Dirac matrices).

where e is the electric charge and θ_W is known as the Weinberg angle, one of the eighteen SM parameters, which has been experimentally measured [43].

As for the color symmetry, the local electroweak gauge symmetry forbids mass terms for the gauge-fields, whereas in Nature they are all massive but one, the photon. In addition, the $SU(2)$ invariance forbids a mass term for fermions. The needed gauge boson ($W^{\pm}$, Z) and fermion masses have to be generated in a gauge invariant way. The spontaneous breaking of the electroweak symmetry is the elegant solution adopted to provide the mass generation.

1.1.3 The Brout-Englert-Higgs mechanism

As it was formulated at the beginning of the 60's, the structure of the SM gauge theory allowed only massless interaction mediators. Forcibly inserting a mass term in the Lagrangian would break the gauge invariance and making the theory not renormalizable anymore. Therefore, in order to leave the theory still gauge-invariant and renormalizable, the mass terms are elegantly introduced via the Brout-Englert-Higgs (BEH) mechanism [13, 14], theorized in 1964 independently by the physicists R. Brout and F. Englert in Belgium, and P. Higgs in Edinburgh. The backbone of the BEH mechanism is the *Spontaneous Symmetry Breaking* (SSB) of the electroweak sector, which is going to be described in the following.

Let $V(\Phi)$ being a quadratic potential describing the evolution of the field Φ under the symmetry transformation of the $SU(2)_L \otimes U(1)_Y$ group and depending on the parameters μ^2 and λ:

$$V(\Phi) = \frac{1}{2}\mu^2\Phi^2 + \frac{1}{4}\lambda\Phi^4 \,. \tag{1.6}$$

The complex scalar field Φ is a weak isospin doublet, which has hypercharge equal to unity, is obtained as a linear combination of four real and scalar fields:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{1.7}$$

Considering only the case in which $\lambda > 0$, the potential V assumes different configurations and different ground-states according to the value of the parameter μ^2. The potential ground-state is the state in which the energy density is at its minimum. In case of $\mu^2 > 0$, the potential V has one unequivocally

defined ground state, identified with the origin of the reference frame considered. If $\mu^2 < 0$, the quadratic term assumes the wrong sign, to be considered as the mass term related to a physical field (particle), and the potential fundamental state is not unequivocally defined anymore. The infinite minima geometrically lie on a hypersphere identified by the following equation:

$$|\Phi|^2 = -\frac{\mu^2}{2\lambda} \,. \tag{1.8}$$

The two different configurations of the potential V are two-dimensionally depicted in Figure 1.1 for the case corresponding to $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right).

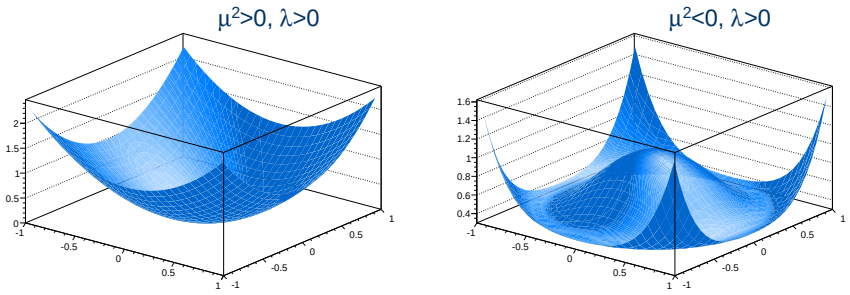

Figure 1.1: Schematic representation of the potential V as a function of the field components, as explicited in Equation 1.6. In both cases the parameter λ is considered positive. On the left the potential corresponding to $\mu^2 > 0$ is reported, the one corresponding to the SM potential ($\mu^2 < 0$) is, instead, shown on the right.

Identifying the *vacuum expectation value* (VEV) of the scalar doublet with $v = \sqrt{-\mu^2/\lambda}$, and choosing one particular minimum

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} , \tag{1.9}$$

it can be observed that the ground state is no longer invariant under an arbitrary transformation of the $SU(2)_L \otimes U(1)_Y$ (the electroweak symmetry breaking has occurred), while, the neutral component of the minimum still remains invariant under the electromagnetic symmetry $U(1)_{em}$. It is crucial to notice, at this point, that the ground state of the Higgs field is no longer invariant under the symmetry $SU(2)_L \otimes SU(2)_R$, but it becomes invariant under the symmetry $SU(2)_{L+R}$, called *custodial symmetry* [44]. It is possible to arbitrarily

choose the configuration in which parametrizing the expansion of the field Φ around the vacuum state $\langle\Phi\rangle$ (cfr. Equation 1.9):

$$\Phi(x) = \frac{e^{i\tau\cdot\vec{\theta}(x)/v}}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix}, \tag{1.10}$$

in which four new scalar fields ($\theta_i(x)$ and $h(x)$ with $i = 1,2,3$) arise. The three fields $\theta_i(x)$ correspond to the massless *Goldstone bosons*, according to the *Goldstone theorem* [45][4]. Given the gauge nature of the theory, these massless unphysical bosons can be re-absorbed into the massive and physical gauge bosons as their longitudinal degrees of freedom, leaving only the scalar field $h(x)$, denoted as the Higgs field, in the expression of $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix}. \tag{1.11}$$

Given the field Φ, expressed as in Equation 1.11, it can be proven that the mass of the physical gauge bosons $W^{\pm}$, Z are:

$$M_W = \frac{1}{2}vg' \qquad M_Z = \frac{1}{2}v\sqrt{g^2+g'^2} \tag{1.12}$$

while the photon remains massless. The $SU(2)_L \otimes U(1)_Y$ symmetry has been broken to $U(1)_{em}$. It is straightforward to deduce from Equations 1.12 and 1.5 that W and Z masses are linked by the relation:

$$\rho \equiv \frac{M_W^2}{M_Z^2\cos^2\theta_W} = 1\,. \tag{1.13}$$

Equation 1.13, which has been verified experimentally [46], proves that even if the electroweak symmetry is broken, the custodial symmetry protects the relation between the masses of the W and Z boson. In particular, ρ must stay close to the unity also when including higher-order corrections in the description of the electroweak interactions.

[4]The Goldstone theorem states that: if a given Lagrangian is globally invariant under symmetry transformations of a n-dimensional group H and, the minimum-energy configurations of the system are also invariant for transformations of a m-dimensional group I, such that $I \subset H$, then H will be spontaneously broken in I and the Lagrangian will describe a system in which $n - m$ scalar massless particles arise. These particles are the *Goldstone bosons*. In this context, $H = SU(2)_L \otimes U(1)_Y$, which has $m = 4$ generators ($\vec{T}$ and Y), while the residual symmetry group $I = U(1)_{em}$, with the $n = 1$ generator Q. Therefore, the number of Goldstone bosons is $m - n = 3$.

Without undermining the elegance of the theory, the fields $\Phi(x)$ can be also used to generate the fermion masses. The gauge invariance allows to introduce additional terms in the SM Lagrangian [47], describing the Yukawa couplings between the field $\Phi(x)$ and the fermion fields. The mass of a given fermion results to be $M_f = g_{hff}v/\sqrt{2}$, where g_{hff} is the Yukawa coupling associated with the fermion considered.

1.1.4 The Higgs boson characterization

By substituting the scalar field Φ in the potential $V(\Phi)$, reported in Equation 1.6, the mass of the Higgs boson h can be obtained:

$$M_h = v\sqrt{2\lambda}\,. \tag{1.14}$$

The SM Higgs boson mass depends, then, on two parameters: v and λ. The value of v, which represents the energy scale at which the electroweak symmetry breaking occurs, is determined starting from the Fermi constant G_F, a parameter on which the vacuum expectation value depends: $v = (\sqrt{2}G_F)^{-1/2} \approx 246$ GeV. The parameter λ, instead, is unknown from the theory, therefore, the mass of the Higgs boson itself remains not predicted. From the potential $V(\Phi)$, also cubic and quartic SM Higgs self-interaction couplings arise, whose analytical expressions are:

$$g_{h^3} = 3M_h^2/v \qquad \text{and} \qquad g_{h^4} = 3M_h^2/v^2\,, \tag{1.15}$$

respectively. Concerning the Higgs boson couplings with the gauge bosons, the SM does not predict a direct coupling with photons, but allows processes in which either one or two Higgs bosons interact with the gauge bosons $V = W^{\pm}/Z$. The couplings for the vertex with one and two Higgs bosons are:

$$g_{hVV} = -2M_V^2/v \qquad \text{and} \qquad g_{hhVV} = -2M_V^2/v^2\,, \tag{1.16}$$

respectively. Similarly for fermions, the Yukawa coupling is:

$$g_{hff} = \frac{M_f}{v}\,. \tag{1.17}$$

The Feynman diagrams corresponding to the interactions just described are summarized in Figure 1.2.

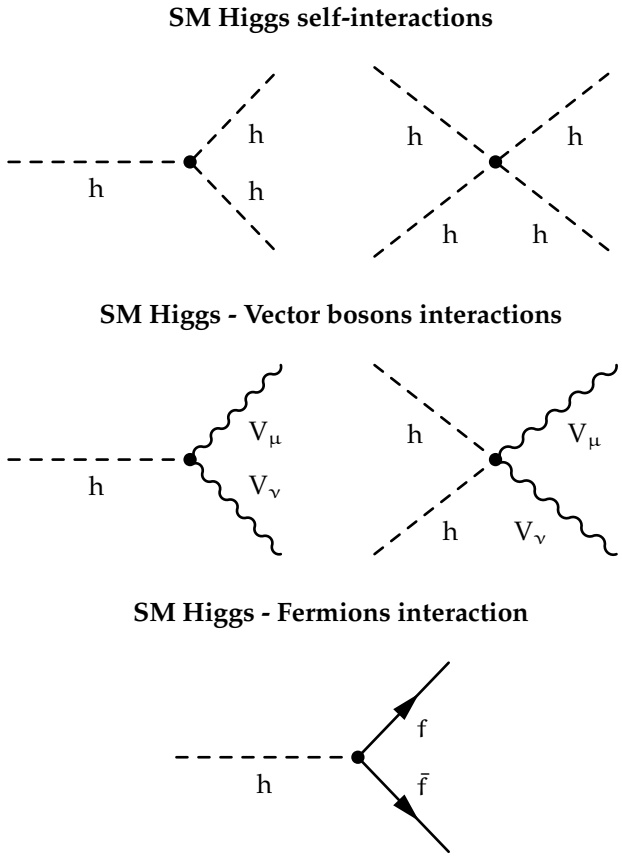

Figure 1.2: Feynman diagrams representing the SM Higgs boson self-couplings (first row) and its coupling to bosons (second row) and fermions (third row).

1.2 Phenomenology of the SM Higgs boson at the LHC

The production mechanisms and the decay channels of the SM Higgs at the LHC are going to be the subject of the following sections. Despite an extensive description of the LHC collider will be provided in Chapter 2, the SM Higgs boson phenomenology at hadron colliders can be understood only if the physics behind proton collisions is explained.

1.2.1 p-p collisions

As previously anticipated (cfr. Section 1.1), the proton has an internal structure consisting of gluons and quarks (*partons*): in particular, gluons bind together two quarks u and one quark d. In literature, these three quarks are called *valence* quarks as almost all the proton characteristics depend on these real particles. All around the valence quarks, a "sea" of virtual quarks and gluons fills the quantum vacuum (these partons are indeed labelled as *sea* partons) and they can be only probed, i.e. become real particles, with a high energy collision. To describe high-energy proton collisions, one works in the approximation that partons are collinear to the proton momentum and carry a fraction x of this momentum (the sum over all partons of each fraction x must return 1). The functions representing the momentum distributions of partons within the proton are called *Parton Distribution Functions* (PDFs) and must be measured experimentally. These functions, which depend on the momentum Q^2 transferred in the scattering, represent the probability densities to find a parton carrying a momentum fraction x at a squared energy scale Q^2. An example of PDFs is shown in Figure 1.3: here the PDFs are represented as a function of x and for two different values of Q^2. It can be observed that the gluon distribution function dominates at low x values, while the parton distribution functions of the valence quarks in the proton are dominant at high x values. The high center-of-mass energy available at the LHC allows low x values to be probed and thus the gluon-fusion Higgs production mechanism to be dominant (cfr. Section 1.2.2). Various sets of PDFs exist: the PDF sets used for the simulation of the signal process used in the 2HDM analysis presented in this thesis (cfr. Chapter 6) are the one provided at LO, NLO and NNLO accuracy by the MSTW2008 collaboration [48]

The cross section of the hadronic process $pp \to X$, according to the QCD factorization theorem [49] can be written as:

$$\sigma_{pp\to X} = \sum_{a,b} \iint dx_a dx_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \hat{\sigma}_{ab\to X}(x_a, x_b, \mu_F, \alpha_s(\mu_R)) \,. \tag{1.18}$$

The total cross section $\sigma(pp \to X)$ depends on the differential cross section of the subprocess $ab \to X$, weighted by the PDF of the partons a and b. The sum takes into account all contributions from the different initial partons (spin, flavor, colour, etc.), while the integration is performed over the momentum fractions x_a and x_b of the partons. The symbol μ_R represents the *renormalization scale*, which is the energy at which the strong coupling constant α_s is measured. The symbol μ_F represents the *factorization scale*, i.e. the mathematical scale at which the decoupling of short and long range physics happens.

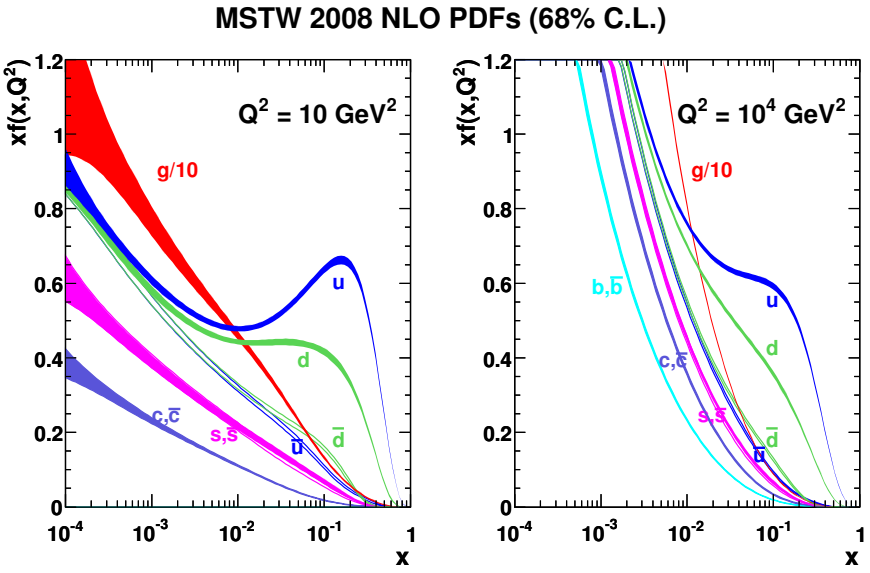

Figure 1.3: Proton Parton Distribution Functions (PDFs) as a function of the four-momentum fraction carried by the partons (x) and for energy scale of the scattering process equal to $Q^2 = 10\,\text{GeV}^2$ (left) and $Q^2 = 10^4\,\text{GeV}^2$ (right) [48].

The most typical choice made regarding the values of the scales is setting $\mu_F = \mu_R = Q$. The QCD asymptotic freedom (cfr. Section 1.1.2) allows a perturbative treatment of the strong interactions at high energies. The partonic differential cross sections depend on the energy, the phase space of the process, and the square of the *matrix element* of the process. According to the perturbation theory, the calculation of the matrix element can be described as a power series in $\alpha_s(Q^2)$: the first non-zero term is leading order (LO), the second term, with one extra power of $\alpha_s(Q^2)$, is next-to-leading order (NLO), the third term is next-to-next-to-leading order (NNLO), and so on. In the following, the SM Higgs boson production cross section predictions at the LHC will be discussed. In particular, at the time of performing the searches described in this thesis (cfr. Chapters 5 and 6) the state of the art contemplated accurate measurements of the cross sections at the NLO and NNLO accuracy. However, recently, the first calculations of the cross sections for the production of a Higgs boson at hadron colliders at the next-to-next-to-next-to-leading order (N^3LO) in perturbative QCD have been published [50].

1.2.2 Production mechanisms

According to the SM predictions, the Higgs boson can be produced in p-p collisions through four main processes: the gluon fusion $gg \to h$, the vector-boson fusion (VBF) $q\bar{q} \to h + 2jets$, the associated production with a W or Z and the associated production with a $t\bar{t}$ pair. The cross sections of these processes depend not only on the Higgs boson mass, but also on the energy at which the p-p collisions occur. The cross sections of the SM Higgs production mechanisms are reported as a function of the Higgs mass in Figure 1.4 for the LHC p-p collisions at a center-of-mass energy $\sqrt{s} = 7$ TeV and 8 TeV. An overview of the main production mechanisms is provided in the following, referring to the CERN Yellow Report [51].

GLUON FUSION The dominant Higgs boson production mechanism is the gluon fusion mediated by loops of quarks t and b. The LO Feynman diagram of the the process is reported in Figure 1.5. As can be observed from Figure 1.4, the gluon-fusion cross section almost decreases with the increase of the Higgs mass, but, at $M_h \approx 350$ GeV, the region in which $M_h \approx 2M_t$ GeV, there is a kink because the amplitude of the process acquire an imaginary component. The cross section reported in Figure 1.4 represents the most updated calculation including NNLO plus NNLL (next-to-next-to-leading logarithmic) order QCD corrections and NLO corrections from electroweak and mixed QCD-electroweak terms. The theoretical uncertainty on the predicted cross section from neglected higher order terms ranges between $8 - 12\%$ [51], depending on the Higgs boson mass. It is of the same order of magnitude the uncertainty related to the gluon PDF.

VECTOR-BOSON FUSION In almost all the mass region, the VBF cross section is an order of magnitude lower than the gluon-fusion one. The LO Feynman diagrams are reported in Figure 1.6 for the channels t, u and s [52].

These three channels are characterized by different topologies, which allow each channel to contribute in different regions of the phase-space. The s-channel, which can be seen as the production of the Higgs boson h in association with a hadronically decaying gauge boson V, contributes in the region in which the invariant mass of the two hadronic jets is close to the mass of the on-shell resonance V, while the interference with the s-channel can be safely neglected outside this mass region. The t and u-channels represent the most important contributions in the region of the phase-space in which the two outgoing jets have a high separation in pseudorapidity[5] ($|\eta| > 4$) and a high invariant mass ($M_{jj} \gtrsim 700$ GeV). Therefore, to totally discriminate the s-channel

[5] The pseudorapidity is defined in Section 2, (cfr. Equation 2.3).

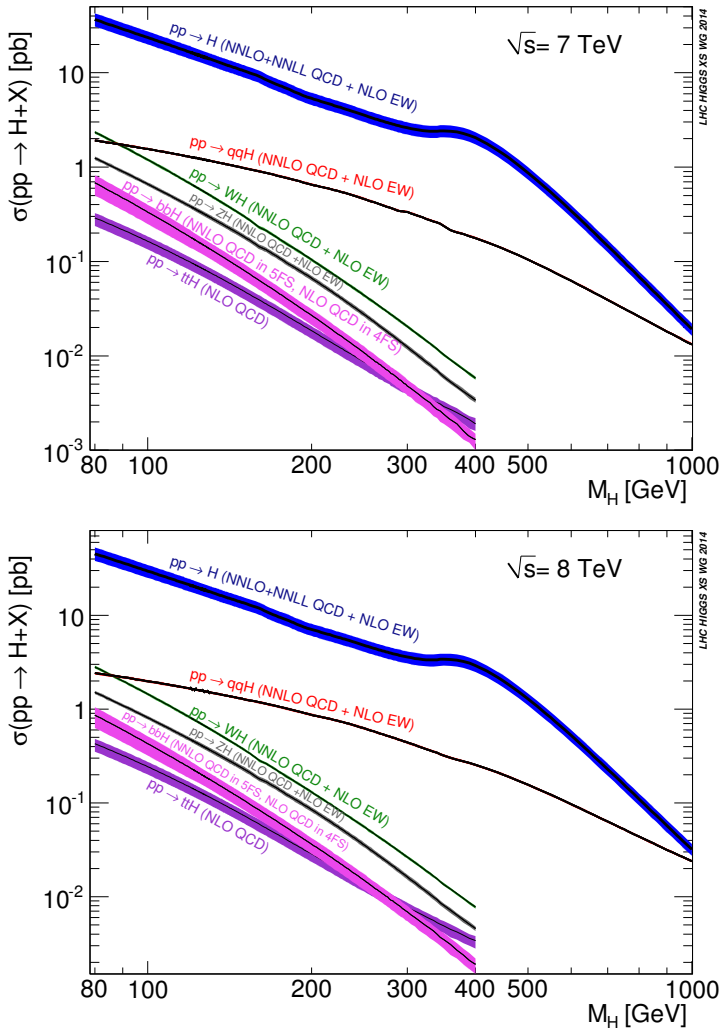

Figure 1.4: Theoretical cross sections of the main SM Higgs production mechanisms reported as a function of the Higgs boson mass. The computation has been performed considering p-p collisions at a center-of-mass energy $\sqrt{s}$ equal to 7 TeV (top) and 8 TeV (bottom) [51].

from the t and u-channels, the kinematical requirements just described need to be imposed. The impact of NLO EWK and QCD corrections is of the order of 5% and negative, with an associated uncertainty of $5-10\%$ [51]. However, the state of the art includes NNLO QCD corrections, which contribute to re-

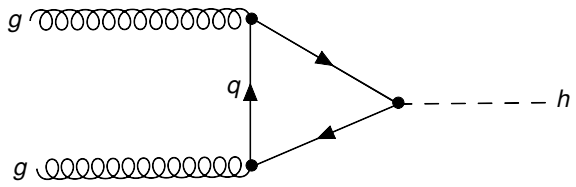

Figure 1.5: Leading order Feynman diagram for the $gg \to h$ production mechanism.

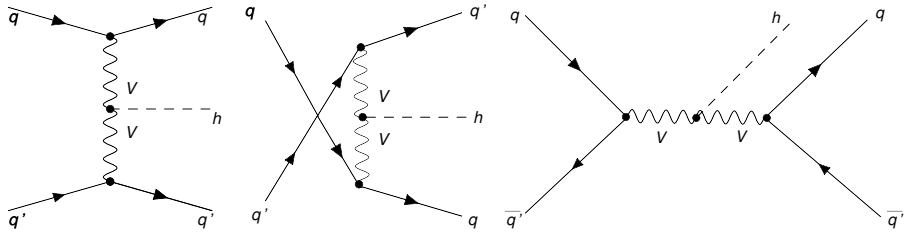

Figure 1.6: Leading order Feynman diagram for the VBF process $q\bar{q} \to h + 2$ jets in the channels t (left), u (middle) and s (right).

duce the theoretical uncertainty to $1 - 2\%$ [53]. The theoretical uncertainties related to the quark PDF range between 2.5% (low M_h) to 7.5% (high M_h).

PRODUCTION IN ASSOCIATION WITH W/Z Despite the lower cross section with respect to the previously described processes, in the low Higgs mass region the Higgsstrahlung production mechanism, in which the gauge boson V decays leptonically, results to have a favorable experimental signature. In this process, two quarks scatter giving rise to a virtual boson V^*, which then decays into a on-shell real boson W/Z and an Higgs boson. The presence of one or two isolated and high momentum charged leptons coming from the gauge boson decay allows to easily trigger the signal events. Similarly to the VBF process, also in this case the cross section has been computed including NNLO and NLO corrections for QCD and EWK, respectively. The overall theoretical uncertainty from higher orders corrections is of about 1% for the Wh process. Concerning the associated production with a Z boson, there is a 8% contribution to the total cross section coming from NNLO diagrams with gluons in the initial state. This component leads to an overall theoretical uncertainty of about $2-6\%$, depending on the Higgs boson mass. The theoretical uncertainty assigned to the PDF knowledge amounts to $3 - 4\%$ for both processes [51]. The LO Feynman diagrams for the processes just described are shown in Figure 1.7.

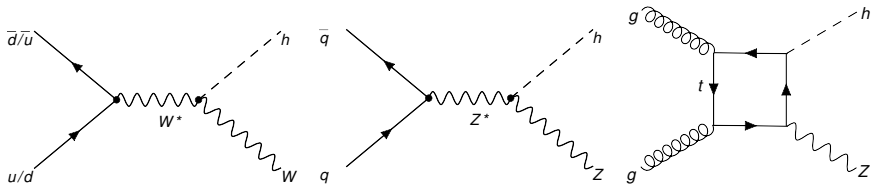

Figure 1.7: Left and middle: leading order Feynman diagrams for Wh and Zh associated production mechanisms. The rightmost diagram represents a higher-order contribution to the Zh process with gluons in the initial state.

The favorable experimental signature provided by the leptons arising from the Z boson decay makes the Zh production mechanism a good candidate for the search of the Higgs boson at the LHC. Such a search has been one of the subject of this thesis and will be described in Chapter 5.

PRODUCTION IN ASSOCIATION WITH HEAVY QUARKS The fourth largest SM Higgs production mechanism cross section is the one related to the production in association with t or b quarks ($pp \to t\bar{t}h/b\bar{b}h$). These processes, which play a role for $M_h < 150$ GeV, are crucial for directly measuring the Higgs Yukawa couplings to t/b quarks, even though they are very challenging to be experimentally reconstructed because of a large number of hadronic jets in the final state. For the $t\bar{t}h$ process, the cross section has been computed at the NLO in QCD only, with an uncertainty from unknown higher-order terms of about 10%. Slightly higher uncertainties are, instead, those related to the $b\bar{b}$ associated production cross section, which has been obtained at the NNLO in QCD considering the four-flavor scheme (4FS) and at the NLO QCD in the five-flavor scheme (5FS) [54]. The LO Feynman diagrams for the $t\bar{t}h$ production mechanism are shown in Figure 1.8. The same diagrams are valid for the $b\bar{b}h$ process in the 4FS.

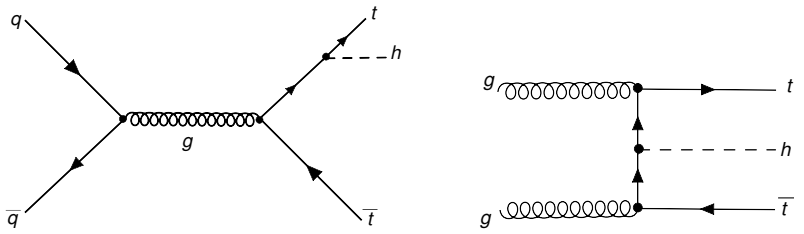

Figure 1.8: Representative leading order Feynman diagrams for the $t\bar{t}h$ production mechanism. The same diagrams correspond to the $b\bar{b}h$ production mode in the 4FS.

1.2.3 Decay processes

According to the SM predictions, the mean lifetime of the SM Higgs boson is $\approx 10^{-22}$ s: this means that it is only possible detecting the SM Higgs particle via its decay products. The theoretical total decay width Γ_h, obtained from the inverse of the lifetime, is shown in Figure 1.9 (left) as a function of the SM Higgs boson mass. Known this and the partial width corresponding to a given decay $h \to X$, the *branching ratio* BR, which is the probability related to that decay, is defined as:

$$BR(h \to X) = \frac{\Gamma(h \to X)}{\Gamma_h} \,. \tag{1.19}$$

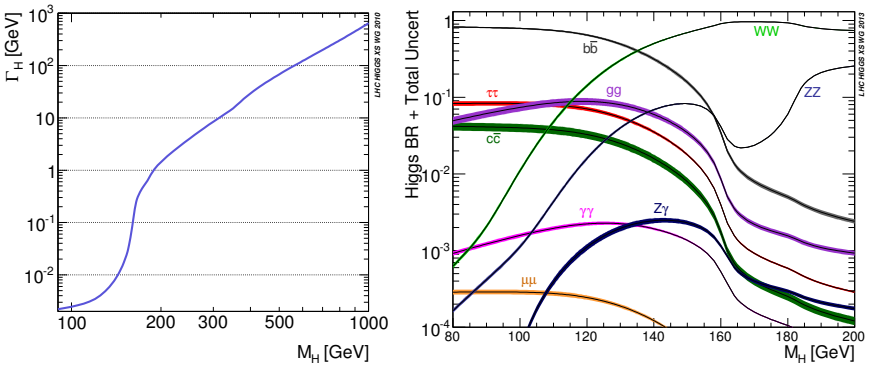

Figure 1.9: Theoretical SM Higgs boson width (left) and branching ratios (right) reported as a function of the Higgs boson mass [51]. Theoretical uncertainties on branching ratios are also shown with coloured bands [55].

The SM Higgs branching ratios are shown, as a function of the Higgs mass, in Figure 1.9 (right). It can be observed that, depending on the Higgs boson mass, some decay channels are favored with respect to others. This is due to the dependency of the partial decay width on the decay kinematics and on the square of the Higgs coupling to the particle involved in the decay. As it has been already discussed in Section 1.1.4, at LO, the Higgs boson can decay to a pair of fermions, through Yukawa interactions, and to a pair of gauge bosons (cfr. Figure 1.2). In the region below the *WW* threshold, the decays into fermions are enhanced because of the dependency of the BR on $(M_f/v)^2$: the most dominant decays are those to $b\bar{b}$ and $\tau\tau$ pairs. In this region, despite not allowed at tree-level via a direct Higgs coupling to gluons, but only through

a fermionic loop, the $h \to gg$ might compete with the $\tau\tau$ and $c\bar{c}$ decays. Despite being experimentally clean, the $h \to \gamma\gamma$ decay is highly theoretically suppressed in this region (such a process can only happen through loops of fermions W bosons). The decays to WW and ZZ pairs start to be dominant for $M_h \geqslant 2M_V$.

Despite theoretically favored by the high BR in the low mass region, the $h \to b\bar{b}$ decay mode is not ideal from an experimental point of view: in an highly dense hadronic environment, as the LHC one, the backgrounds featuring such a process are immense. For this reason, especially in view of the discovery of the 125 GeV Higgs boson in fermionic channels, a topic that will be treated in the next section, the $h \to \tau\tau$ decay channel becomes highly favored and crucial for directly measuring the fermionic Yukawa couplings and for determining the Higgs properties.

1.3 SM Higgs boson experimental searches

The first searches for the SM Higgs boson were performed at the *Large Electron-Positron* collider (LEP) [56] operating at CERN from 1989 to 2000. With the data collected by the four experiments (ALEPH, DELPHI, L3 and OPAL) during the second operational phase of the collider, whose maximum center-of-mass energy reached was $\sqrt{s} = 209$ GeV, the exclusion limit set on the Higgs boson mass at 95% CL was $M_h > 114.4$ GeV [57] (for all the details related to the statistical interpretation of the experimental results, the reader can refer to Appendix B.). Subsequently, the Higgs-hunting program was continued at Tevatron [58], the $p - \bar{p}$ collider, operating at the Fermi National Accelerator Laboratory in the US from 1987 to 2011 at $\sqrt{s} = 1.96$ TeV. The latest combined results, presented by the CDF and D0 collaborations, have excluded the following mass ranges: $100 < M_h < 106$ GeV and $147 < M_h < 179$ GeV. Interestingly, an excess of events, corresponding to a global significance of 3.1σ, was observed in the $h \to b\bar{b}$ channel in the mass range $120 < M_h < 135$ GeV [59]. This result has been published just after the discovery of a new particle with mass 125 GeV, which has been separately announced by the ATLAS [9] and CMS [10] collaborations (referring to the two general purpose experiments built on the LHC circumference) on the 4^{th} of July 2012. The new particle was discovered by exploiting data samples corresponding to integrated luminosities of 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and 5.3 fb^{-1} at $\sqrt{s} = 8$ TeV (for all the technical definition of the LHC and CMS working parameters, the reader can refer to Chapter 2), and combining the results of the $\gamma\gamma$, ZZ and WW decay channels [11, 12]. From that milestone, the tremendous effort put

in the study of additional decay channels and in the newly discovered particle characterization has led to claim at the compatibility with the predicted SM Higgs boson. However, the level of accuracy of the measurements still does not exclude new physics scenarios. An overview of the main Higgs boson related measurements are provided in the following.

HIGGS BOSON MASS The most sensitive channels, $\gamma\gamma$ and ZZ, have been exploited for a precise mass measurement also because of their high invariant mass resolution ($1-2$ GeV). A summary of the results obtained by ATLAS and CMS experiment individual analyses, and from combined analyses, is provided in Figure 1.10. Combining the ZZ and the $\gamma\gamma$ channels, ATLAS and CMS have measured $M_h = 125.36 \pm 0.37\,(\text{stat}) \pm 0.18\,(\text{syst})$ GeV [60] and $M_h = 125.03^{+0.26}_{-0.27}\,(\text{stat})^{+0.13}_{-0.15}\,(\text{syst})$ GeV [61], respectively.

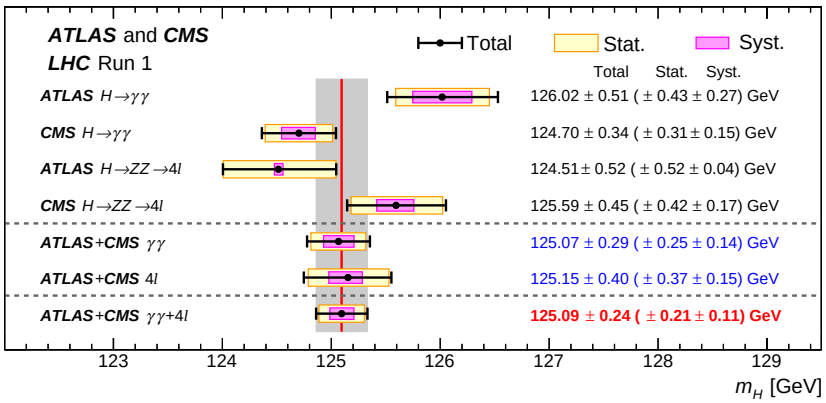

Figure 1.10: Summary of Higgs boson mass measurements performed individually in the context of ATLAS and CMS analyses. The ATLAS and CMS combined results are also provided for the $h \to ZZ \to 4\ell$ and $h \to \gamma\gamma$ decay channels separately. Systematic (magenta bands), statistical (yellow bands), and total (black error bars) uncertainties are reported. The red vertical line and the corresponding gray column indicate the central value and the total uncertainty of the combined measurement, respectively [60].

SPIN PARITY The spin and parity quantum numbers, labelled as J^P, have been studied including also the $h \to WW \to 2\ell 2\nu$ channel. Various spin ($J = 0, 1, 2$) and parity ($P = +1, -1$) hypotheses have been tested. The latest CMS results exclude any concurrent hypothesis of the $J^P = 0^+$ (scalar particle with spin 0) at more than 99% CL, leaving only 0^+ as an option [62]. The same results have been obtained in the equivalent analysis performed by the ATLAS experiment [63]. Moreover, an experimental proof of the custodial symmetry

validity has been obtained in the same channels from both ATLAS [64] and CMS [61] experiments.

HIGGS WIDTH Being the expected SM Higgs boson width highly below the experimental mass resolution ($\Gamma_h \approx 4$ MeV), upper limits have been set at the 95% CL on the ratio $\Gamma_h^{obs}/\Gamma_h^{SM}$. Both CMS and ATLAS have performed such a study considering the off-shell Higgs boson production processes in the following decay channels: $h \to ZZ \to 4\ell$ and $h \to ZZ \to 2\ell 2\nu$. The simultaneous fit of the two contributions leads to an upper limit of $\Gamma_h < 22$ MeV at 95% CL, which correspond to 5.5 (ATLAS) [65] and 5.4 (CMS) [66] times the value predicted by the SM.

SIGNAL STRENGTH The *signal strength* μ quantifies the strength of the observed signal with respect to the expected SM value and it is formally represented by the ratio $\sigma_{95\%}/\sigma_{SM}$. A summary of the ATLAS and CMS signal strength measurements relative to bosonic ($h \to \gamma\gamma/ZZ/WW$) and fermionic ($h \to b\bar{b}/\tau\tau$) channels is provided in Figure 1.11 for ATLAS (left) and CMS (right) experiments. The combined results are $\mu = 1.30^{+0.18}_{-0.17}$ [64] and $\mu = 1.00^{+0.13}_{-0.13}$ [61] for ATLAS and CMS, respectively.

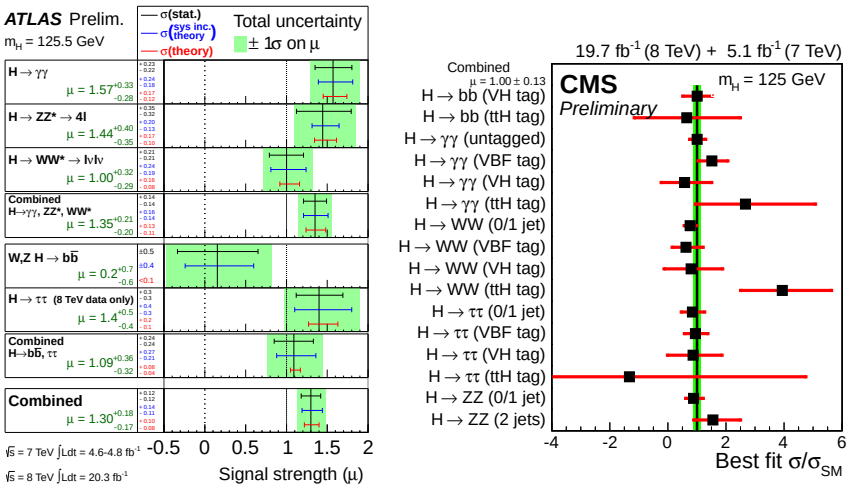

Figure 1.11: Summary of the individual and combined signal strengths computed in the context of different analyses by ATLAS (left) and CMS (right). The combined results are $\mu = 1.30^{+0.18}_{-0.17}$ and $\mu = 1.00^{+0.13}_{-0.13}$ for ATLAS and CMS, respectively [64, 61].

Of particular interest are the excesses over the expected background-only hypothesis that CMS has observed both in the $b\bar{b}$ (corresponding to a significance of 2.1σ [67]) and in the $\tau\tau$ (with a significance of 3.2σ [15]) final states.

ATLAS has observed an excess too in the $\tau\tau$ final state, with a significance of 4.1σ [68], but not in the $b\bar{b}$ one, being less sensitive as far as concern this final state.

YUKAWA COUPLING TO TOP QUARKS The only possible way to measure the Higgs Yukawa coupling to top quarks is probing the $t\bar{t}h$ associated production mechanism. Such a coupling, in fact, cannot be measured through the Higgs to top pair decay, being the top quark heavier than the Higgs boson. CMS has observed a 3.5σ excess with respect to the background-only hypothesis [69] and the signal strength measurement accounts to $\mu = 2.8^{+1.0}_{-0.9}$. The results from ATLAS are consistent with the SM expectations [70].

1.4 Going beyond the SM

Despite the widely experimentally proven success of the SM and the discovery of the Higgs boson, the SM does not yet have all the credentials to be considered as the ultimate theory of Nature. The SM is not able to explain well established empirical facts, such as the gravitational force, the baryogenesis [19], the dark matter and the dark energy [17, 18]. For these reasons, extended theoretical frameworks have been developed over the last thirty years, which incorporate the validity of the SM description beyond the TeV scale, up to the *Grand Unified Theory* (GUT) scale ($M_{GUT} \approx 10^{16}$ GeV, $l_{GUT} \approx 10^{-32}$ m) or even the Planck scale ($M_P \approx 10^{19}$ GeV, $l_P \approx 10^{-35}$ m). The SM, in fact, is *a priori* self-consistent up to energies between $10^8 - 10^{16}$ GeV [71, 72]. However, theoretical limitations connected to the hierarchy problem [21], which precisely involves the Higgs sector, suggests that new physics should enter already at the TeV scale. The hierarchy problem arises due to the unnaturalness of the SM. Conceptually, a theory is considered natural if the description provided at low energies is not affected by the parameter values at higher energy scales.

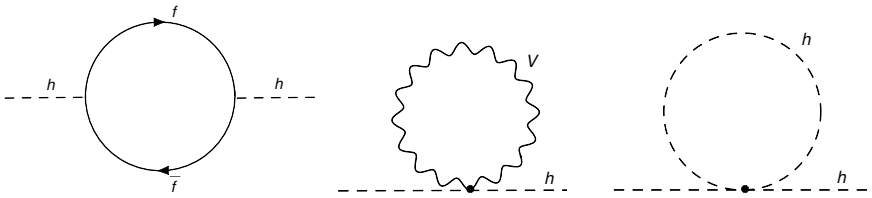

Figure 1.12: One-loop radiative corrections to the Higgs boson mass.

It has already been explained in Section 1.1.4 that the Higgs boson interacts with itself via cubic and quartic interactions, with a strenght proportional to M_h^2. Higher-order corrections to the Higgs mass involve fermions, gauge bosons and Higgs boson loops, as displayed in Figure 1.12. Beyond the tree level, the Higgs mass can be written as:

$$M_h^2 = M_0^2 - \delta M_h^2 \,, \tag{1.20}$$

where M_0^2 represent the unphysical "bare" mass of the Higgs boson, i.e. a UV-divergent parameter of the SM Lagrangian, which one renormalizes with a mass counterterm δM_h^2. Radiative corrections to the Higgs mass are quadratically divergent in the cut-off parameter Λ, which represents the energy scale at which the theory remains still valid and consistent:

$$\delta M_h^2 \propto \Lambda^2 \qquad \text{and then} \qquad M_h^2 \approx M_0^2 - \Lambda^2 \,. \tag{1.21}$$

Assuming the SM valid up to the Planck scale, then the obvious cut-off is the Planck mass: $\Lambda^2 = \mathcal{O}(10^{38})$ GeV^2. This means that also the bare Higgs boson mass squared has to be fine-tuned to the same order of magnitude, $\mathcal{O}(10^{38})$ GeV^2, to cancel out the divergences and make the SM Higgs mass renormalizable at the electroweak scale. The fact that huge effects have to cancel each other with high precision in order to cure the divergences related to a "small" quantity, reflects the naturalness problem of the SM. Considering the cut-off parameter Λ equal to the Planck energy is an arbitrary choice, but plausible in case no phenomena are present in between the electroweak scale and the Planck scale. In other words, the SM naturalness can be restored if new *Beyond* SM (BSM) physics phenomena appear at $\Lambda \approx 1$ TeV. In that case, no large fine-tuning is necessary to achieve divergence cancellation. Moreover, another serious theoretical problem is due to the assumption that the SM is valid up to the Planck scale. It concerns the stability of the electroweak vacuum [73], in light of the present values of the Higgs and the top quark mass [74]. The most reliable trend within the community of physicists is believing that there is a more fundamental theory of particles and interactions at the TeV scale, which reduces to the SM at low energies in order to explain its experimental success. An extended Higgs sector may play a key role in connecting the SM to unknown new physics phenomena. A paradigmatic example is the two-Higgs-doublet Model (2HDM), which is described in Section 1.5.1 and used as the theoretical framework for the analysis described in Chapter 6.

1.5 The two-Higgs-doublet Model

Explicited by its name, the two-Higgs-doublet Model (2HDM) [75] is a theoretical framework consisting of two Higgs doublets, having the same quantum numbers. Aside from being a simple extension of the SM, the 2HDM structure aims to solve the hierarchy problem and is strongly supported by several theoretical motivations. The 2HDM provides a way to explain the asymmetry between matter and antimatter observed in the Universe [19]. Another important motivation is *Supersymmetry* (SUSY) [76], which is a theory based on an extra symmetry between fermions and bosons, implying that every boson must have a corresponding supersymmetric fermion counterpart and viceversa. Supersymmetric embeddings of the SM require at least two Higgs doublets, exactly like axion models [77], that would explain how the strong interaction does not violate the CP symmetry[6] with an effective low-energy theory with two Higgs doublets. The 2HDM can also be viewed as the simplest Higgs sector extension which can describe a Higgs coupling fit with flexible coupling variations [78].

Finally, it has also been recently noted [79] that certain realizations of the 2HDM can accommodate the muon anomalous magnetic moment [80] without violating the present theoretical and experimental constraints. Similarly to the SM doublet Φ (cfr. Equation 1.7), in the 2HDM one considers two complex doublets of $SU(2)_L$, having $Y = +1$ each of them:

$$\Phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \qquad \Phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_5 + i\phi_6 \\ \phi_7 + i\phi_8 \end{pmatrix}. \tag{1.22}$$

Imposing the gauge invariance, the corresponding most general renormalizable potential $V(\Phi_1, \Phi_2)$ has the following expression:

[6] The CP symmetry is the combination of the parity transformation P, already introduced in Chapter 1.1.2, and the charge conjugation transformation C.

$$
\begin{aligned}
V(\Phi_1,\Phi_2) = {} & M_{11}^2\Phi_1^\dagger\Phi_1 + M_{22}^2\Phi_2^\dagger\Phi_2 - \left(M_{12}^2\Phi_1^\dagger\Phi_2 + \text{ h.c.}\right) \\
& + \frac{\lambda_1}{2}\left(\Phi_1^\dagger\Phi_1\right)^2 + \frac{\lambda_2}{2}\left(\Phi_2^\dagger\Phi_2\right)^2 \\
& + \frac{\lambda_3}{2}\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_2\right) + \frac{\lambda_4}{2}\left(\Phi_1^\dagger\Phi_2\right)\left(\Phi_2^\dagger\Phi_1\right) \\
& + \left[\frac{\lambda_5}{2}\left(\Phi_1^\dagger\Phi_2\right)^2 + \lambda_6\left(\phi_1^\dagger\phi_1\right)\left(\phi_1^\dagger\phi_2\right) + \lambda_7\left(\phi_2^\dagger\phi_2\right)\left(\phi_1^\dagger\phi_2\right) + \text{ h.c.}\right],
\end{aligned}
\tag{1.23}
$$

where the free parameters M_{11}, M_{22} and λ_{1-4} are real, while M_{12} and λ_{5-7}, are complex. The most general formulation of the 2HDM potential, then, has 14 degrees of freedom, 6 real plus 8 from the complex parameters. However, under the assumption of certain symmetries and requiring the absence of CP violation, this number can be decreased. In order to preserve the phenomenological viability of the 2HDM, the presence of tree-level flavour-changing neutral currents (FCNC)[7], which are allowed due to the exchange of (one or more) Higgs bosons, has to be suppressed. A classical way to keep tree-level FCNCs suppressed is imposing the Z_2 asymmetry (thus assuming natural flavor conservation) and, as a consequence, avoiding any CP-violation in the scalar sector. In terms of parameters, this means $M_{12}^2 = \lambda_{6,7} = 0$. However, the most common strategy is to consider non-null value for M_{12}^2, accepting a softly broken Z_2 symmetry and making the CP-violation in the scalar sector possible. The assumption of the CP-symmetry invariance, which is valid for the experimental searches documented in this thesis, constraints the remaining non-null complex parameters (M_1^2 and λ_5) to be real.

The two minima chosen in correspondence of the lowest (and simplified) energy configuration are:

$$
\langle\Phi_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_1\end{pmatrix} \qquad \langle\Phi_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_2\end{pmatrix}. \tag{1.24}
$$

The BEH mechanism applies in the context of the 2HDM as in the SM, with the only difference that, on top of the 3 Goldstone bosons reabsorbed by the longitudinal polarization of the, now massive, $W^\pm$ and Z, there are 5 physical Higgs bosons. The five physical states are: two charged bosons $H^\pm$, two

[7] With Flavor Changing Neutral Currents (FCNCs) are indicated those transitions that change the flavor of a fermion, but its charge. These transitions are forbidden at tree-level in the SM and strongly suppressed at higher levels by the Glashow-Iliopoulos-Maiani (GIM) mechanism [81].

neutral scalar bosons h/H (CP-even) and one neutral pseudoscalar boson A (CP-odd). Aside from the physical Higgs masses, two important parameters are:

$$\tan\beta = \frac{v_2}{v_1} \qquad \text{and} \qquad \alpha\,, \tag{1.25}$$

where, $\tan\beta$ is the ratio between the two vacuum expectation values, and α represents the mixing angle for CP-even Higgs bosons. The identification of the SM Higgs boson h_{SM}, having $M_h = 125$ GeV, with the scalar h, constraints the phenomenologically reliable parameter space regions to not depart from the SM-like condition: $\cos(\beta-\alpha) \ll 1$ (and $\sin(\beta-\alpha) \approx 1$). This situation can either correspond to the *decoupling* [82] or to the *alignment limit* [82]. Additionally, the v_1 and v_2 are constrained by the condition

$$v = \sqrt{v_1^2 + v_2^2} = 246\ \text{GeV}\,. \tag{1.26}$$

It can be shown [75] that the masses of the four remaining physical states are correlated with the residual degrees of freedom of the potential (cfr. Equation 1.23). Therefore, under the validity of all the discussed assumptions, one can identify six free parameters that fully characterize the model:

$$M_A\,,\, M_H\,,\, M_{H^\pm}\,,\, M_{12}^2\,,\, \tan\beta\,,\, \cos(\beta-\alpha)\,. \tag{1.27}$$

Four canonical types of 2HDM arise [78], depending on the way in which the Higgs doublets couple to fermions. These four possibilities are summarized in Table 1.3.

	Type-I		**Type-II**		**Type-III**		**Type-IV**	
	Φ_1	Φ_2	Φ_1	Φ_2	Φ_1	Φ_2	Φ_1	Φ_2
$Q = 2/3$ quarks	×	√	×	√	×	√	×	√
$Q = -1/3$ quarks	×	√	√	×	√	×	×	√
Charged leptons	×	√	√	×	×	√	√	×

Table 1.3: Overview of the four 2HDM types depending on the way in which fermions couple to the Higgs doublets.

It is interesting to notice that the only parameters that determine the interactions of the various Higgs fields with fermions (cfr. Table 1.4) and gauge bosons (cfr. Table 1.5) are α and β. So, it is sufficient exploiting just these two parameters to discuss the phenomenology of the different declinations of the 2HDM. This is what is going to be done in the next section, with a particular focus on the Type-II scenario.

1.5.1 Phenomenology of the Type-II Higgs sector

The Type-II Higss sector couplings to fermions and gauge bosons are reported in Tables 1.4 and 1.5, respectively, not considering couplings involved charged Higgses. The expressions summarized in these tables have to be considered normalized to the SM-like Higgs couplings. It can be observe that the tree-level couplings of the neutral scalar Higgs bosons h and H to gauge bosons and fermions have the same structure as the corresponding couplings of the SM Higgs boson. The pseudoscalar A only couples to fermions and the Feynman rule contains an additional factor $i\gamma^5$. The ratios of the 2HDM coupling constants to the SM ones only depend on β and the mixing angle α. Moreover, it is important to notice that the couplings of the light scalar Higgs h approach the corresponding SM values for $\beta - \alpha \to \pi/2$, irrespective of the value of β.

	up-type quarks $u\bar{u}, c\bar{c}, t\bar{t}$	down-type quarks and charged leptons $d\bar{d}, s\bar{s}, b\bar{b}, e^+e^-, \mu^+\mu^-, \tau^+\tau^-$
h	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$
H	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$
A	$i\gamma^5\cot\beta$	$i\gamma^5\tan\beta$

Table 1.4: Tree-level couplings of the neutral 2HDM Higgs bosons to fermions, normalized to the SM ones.

$\cos(\beta-\alpha)$	$\sin(\beta-\alpha)$
HW^+W^-	hW^+W^-
HZZ	hZZ
ZAh	ZAH

Table 1.5: Tree-level couplings of the neutral 2HDM Higgs bosons to gauge bosons, normalized to the SM ones [83].

Focusing on the comparison between the H and A decay modes to bottom quarks and tau leptons, one can observe that they only depend on the mass ratio $m_\tau^2/3m_b^2$, which is approximately 10% [75]. It is worth noticing that the same argument remains valid for the Type-I scenario, but not for the Type-III and Type-IV, in which the Higgs bosons couplings to down-type quarks differ from those to charged leptons (cfr. Table 1.3). In the Type-IV scenario, for instance, the ratio of the A branching fraction into $\tau\tau$ to the one into $b\bar{b}$ is proportional to $\tan\beta$: even for relatively small $\tan\beta$, the $\tau\tau$ mode will dominate (in fact, for $\tan\beta > 3$, the branching ratio exceeds 90%). What happens in the Type-III scenario is exactly the opposite. In this case, one can see that for $\tan\beta \geqslant 1$ the ratio of the A branching fraction into $\tau\tau$ to the one into $b\bar{b}$ is

about 10% or less [75]. This latter scenario offers new possibilities for interpreting some of the final results shown in Chapter 6.

Coming back to the Type-II scenario, for the purpose of this thesis, interesting processes are those in which either an intermediate heavy H [84, 85] or A are produced in the gluon-fusion process, leading, in turn, to a sequential decay into $Z + A/H$ [86]. The Feynman diagram corresponding to the just cited processes is reported in Figure 1.13 for the processes $H/A \to ZA/H$.

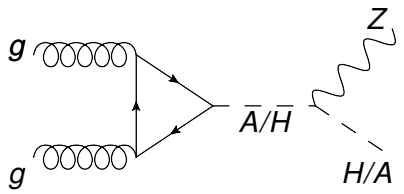

Figure 1.13: Feynman diagram for the $pp \to H/A \to ZA/H$ processes in which the intermediate boson is produced via gluon fusion in the 2HDM.

Given that the coupling AHZ is proportional to $-\sin(\beta - \alpha)$ (cfr. Table 1.5, it is not suppressed in the decoupling limit and for low value of $\tan\beta$ [87]. As it has been recently studied [88], the theoretical projections at $\sqrt{s} = 14$ TeV show that the cross sections of both processes $A \to ZH$, for $M_A > M_Z + M_H$, and $H \to ZA$ for $M_H > M_Z + M_A$, can exceed 1 pb, being promising signatures for the discovery of new physics at the LHC. As further motivation, it has been recently shown [87] that 2HDM scenarios in which the process $A \to ZH$ would be within the reach of the LHC would in addition favor an electroweak phase transition that would be at the origin of the baryogenesis process in the early Universe, thus explaining the currently observed matter-antimatter asymmetry.

From an experimental point of view, the study of both the Type-II 2HDM processes $H \to ZA$ and $A \to ZH$ in the context of the LHC, has been carried out for the first time during the doctoral work documented in this thesis. The analysis, which is described in Chapter 6, has been performed exploiting data collected at the LHC during 2012 and considering the lightest (pseudo)scalar, involved in the process, decaying into a pair of tau leptons. The decay into a pair of tau leptons has lower branching ratio than the $b\bar{b}$ one, given the dependency of the Yukawa coupling on the fermion masses. However, as in the SM case, the $\tau\tau$ final state has the advantage of being more efficiently selected out of the enormous hadronic background.

The phenomenology of the charged Higgses $H^{\pm}$, being out of the scope of this thesis, will be left to the curiosity of the reader [75].

Given the Higgs couplings to fermions and gauge bosons (cfr. Table 1.4), some general statements can be made on how the production mechanisms change in the 2HDM with respect to the SM ones (cfr. Section 1.2.2).

GLUON FUSION In case the top-quark loop is the only one contribution to the Feynman diagram (cfr. Figure 1.5), the difference in the SM cross section, which will be indicated as σ^{SM}_{ggF} is only due to the rescaled coupling of the Higgs to top quark. For the processes $gg \to h$, $gg \to H$ and $gg \to A$, the 2HDM cross section will be $\sigma^{SM}_{ggF} \times (\cos\alpha/\sin\beta)^2$, $\sigma^{SM}_{ggF} \times (\sin\alpha/\sin\beta)^2$ and $\sigma^{SM}_{ggF} \times \cot\beta$, respectively. On the one hand, the suppression of the coupling to the top quark implies that the gluon-fusion production rate decreases with respect to the SM one. On the other hand, the enhanced coupling to down-type fermions entails a non-negligible bottom-mediated contribution, especially for large $\tan\beta$ (even though not favoured in the Type-II 2HDM). The b-loop contribution to the amplitude is obtained by multiplying by $-\tan\alpha\tan\beta$ the one relative to the t-loop contribution.

$V, b\bar{b}, t\bar{t}$ ASSOCIATED PRODUCTIONS AND VBF The arguments that have been just described remain valid also in case scalar or pseudoscalar Higgses are produced in association with heavy-quark pairs ($b\bar{b}$ or $t\bar{t}$). Concerning the associated production with vector bosons and the VBF production mechanisms, it is straightforward to notice that the pseudoscalar A cannot be produced via these modes, since there are no W^+W^-A and Z^+Z^-A vertices. For the scalar Higgses h and H, instead, the SM cross sections are multiplied by the factors $(\sin(\beta-\alpha))^2$ and $(\cos(\beta-\alpha))^2$, respectively.

1.5.2 LHC experimental constraints Type-II 2HDM

Experimental constraints on Type-II 2HDM have been set in the context of searches performed exploiting the data collected by CMS during 2012. Such searches refer to the $A \to Zh \to \ell\ell b\bar{b}$ [89] process and to the combination of $H \to hh \to b\bar{b}\tau\tau$ and $A \to Zh \to \ell\ell\tau\tau$ [90] ones. In all these three analyses, the lightest scalar h is assumed to be the SM Higgs boson. Moreover, these results have been obtained assuming also $m_H = m_{H^{\pm}} = m_A$, $\lambda_{6,7} = 0$ and $m^2_{12} = m^2_A \tan\beta/(1+\tan\beta)$, thus considering a softly broken Z_2 symmetry and avoiding CP-violation at tree-level (cfr. Section 1.5.1). Type-II model dependent upper limits have been computed for $m_A = 300$ GeV in the plane $\tan\beta$-$\cos(\beta-\alpha)$. The results are shown in Figure 1.14 for the $A \to Zh \to \ell\ell b\bar{b}$

analysis (left) and the combination of $H \to hh \to b\bar{b}\tau\tau$ and $A \to Zh \to \ell\ell\tau\tau$ (right).

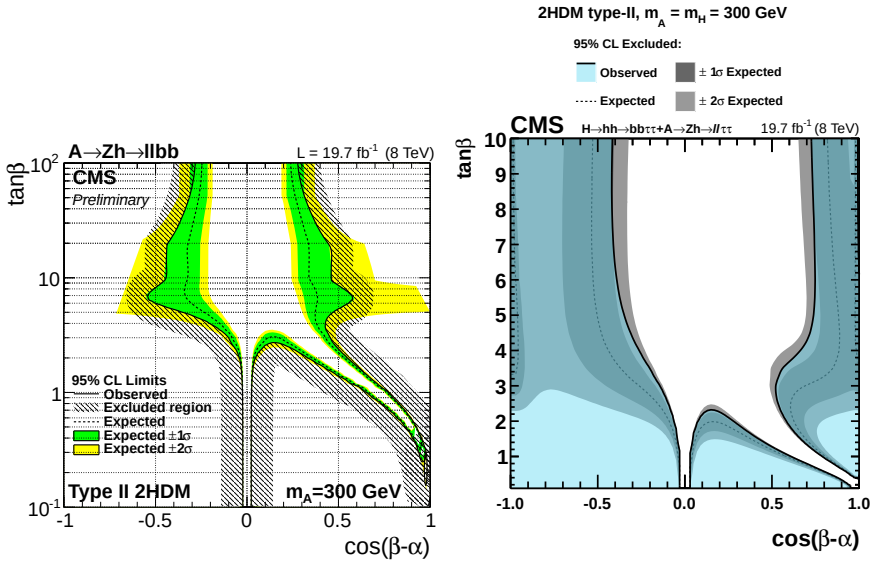

Figure 1.14: Observed and expected 95% CL exclusion limits on the signal strength modifier $\mu = \sigma_{95\%}/\sigma_{TH}$ for Type-II (right) models in the context of the $A \to Zh \to \ell\ell b\bar{b}$ analysis (left) and the combination of $H \to hh \to b\bar{b}\tau\tau$ and $A \to Zh \to \ell\ell\tau\tau$ (right). The results are shown as a function of $\tan\beta$ and $\cos(\beta-\alpha)$, considering $m_A = 300\,\text{GeV}$. Expected $\pm1\sigma$ and $\pm2\sigma$ exclusion bands are also shown [89, 90].

In the alignment limit $\cos(\beta - \alpha) \to 0$, the scalar h behaves as the SM-like Higgs boson and the $A \to Zh$ and $H \to hh$ branching ratios vanish at Born level (cfr. Table 1.5. Moreover, the narrow region with no exclusion power is caused by the vanishing branching ratio of the $h \to b\bar{b}$ and $h \to \tau\tau$ decay processes (cfr. Table 1.4). The CMS collaboration has also recently published compatible results obtained in the context of the search for the processes $H \to hh$ and $A \to Zh$ in diphoton and multilepton final states [91]. The ATLAS experiment has also obtained results in the search for the process $A \to Zh \to \ell\ell b\bar{b}$ [92], which are consistent with the CMS ones.

The results in Figure 1.14 show that the low $\tan\beta$ region, for $\cos(\beta - \alpha) \to 0$, is not yet excluded. Such a region of the parameter space will precisely be the target of the $H/A \to ZA/H \to \ell\ell\tau\tau$ analysis presented in this thesis (cfr. Chapter 6). The results of this latter analysis could be combined with the results of the searches mentioned above. The reasons for the complementarity

of these analyses is that they are sensitive to different regions of the $\tan\beta$-$\cos(\beta-\alpha)$ parameter space. The motivations can be understood looking at Tables 1.4 and 1.5: the AZh and AZH couplings are proportional to $\cos(\beta-\alpha)$ and $\sin(\beta-\alpha)$, respectively. From a technical point of view, one would have to take into account all the possible correlations arising from the sources of systematic uncertainties common to the analyses to be combined.

2

Experimental setup: the LHC and the CMS experiment

The whole work presented in this thesis is based on data collected by the Compact Muon Solenoid (CMS) experiment [10]*, one of the four main experiments installed along the Large Hadron collider (LHC) circumference* [93]*. In this chapter, the main features of the LHC collider are presented, succeeded by the description of the CMS subdetectors used for particle identification and reconstruction. An overview of the CMS computing system, essential for a successful operation of the experiments at the LHC, is also provided. Given the importance of the event generation and detector simulation for a proper optimization of the analysis strategies, the chapter is concluded describing how event generations and detector simulations are performed within the CMS collaboration.*

2.1 The Large Hadron Collider

The LHC [93] technical features have to be framed thinking about the motivation behind the building of this machine, the world's largest and powerful accelerator of protons. The LHC has been designed to operate at a center-of-mass energy of $\sqrt{s} = 14$ TeV, with an instantaneous luminosity $\mathcal{L}(t)$ of the

order of 10^{33} cm^{-2} s^{-1} in the first few years and 10^{34} cm^{-2} s^{-1} afterwards[1]. The concept of luminosity will be explained later in the text (cfr. Equation 2.1), but the reader might already realize that those designed conditions are unique and crucial to address all the fundamental questions which the LHC has been built for: proving the existence of the SM Higgs boson, discovering extended symmetries or extra dimensions requires energy of tens of TeV and a huge number of collisions. Without p-p collision it would have been impossible achieving the same physics goals, to which LHC aims. Accelerating electrons (and positrons) at the same conditions, for instance, is prohibitive because of the loss of energy by synchrotron radiation[2]. Also proton-antiproton collisions, used at Tevatron [58], do not fulfill the need of achieving the luminosity requested, due to the limited production efficiency and collimation of antiprotons and also to the different internal nucleons structure.

In the context of this thesis, with the term "event", a p-p interaction at the LHC will be denoted. The production rates for different physical processes depend strongly on the energy scale of the process, as shown in Figure 2.1 where production cross sections for different physical processes at the LHC, and corresponding event rates at an instantaneous luminosity 10^{33} cm^{-2} s^{-1} are reported as a function of the center-of-mass energy $\sqrt{s}$. The cross section for inelastic p-p collisions at $\sqrt{s} = 8$ TeV is approximately 75 mb [94], corresponding to a rate of 10^8 Hz at $\mathcal{L}(t) = 10^{33}$ cm^{-2} s^{-1}. It can be further observed that, in order to reach sensitivity to rare processes like the production of a SM Higgs boson, an instantaneous luminosity of the order of 10^{33} cm^{-2} s^{-1} is needed to obtain a production rate of 10^{-2} Hz at 8 TeV.

The LHC is hosted inside the 27 km-long tunnel dug for the Large Electron-Positron (LEP) [56] collider. This tunnel is located under both French and Swiss territories. The collider is characterized by superconducting radio-frequency (RF) accelerating cavities, focusing quadrupole magnets and 1232 superconducting dipole magnets for bending the protons and producing a magnetic field of 8.3 T. In order to maintain superconductive properties, magnetic coils and superconductors are cooled at 1.9 K with superfluid Helium. Two beam pipes host protons circulating in opposite directions. In order to be accelerated by RF cavities, proton beams have to be *bunched*: the LHC has been designed

[1] The LHC has been also design to provide heavy ion (Pb) collisions with $\sqrt{s} = 1150$ TeV [93] and a luminosity exceeding 10^{27} cm^{-2} s^{-1}. However, given the scope of this thesis, heavy ion collisions will not be treated further.

[2] The synchrotron radiation is the electromagnetic radiation emitted when charged particles are accelerated radially. It can be shown that the amount of energy lost in each turn in the circular accelerator is $\Delta E = (4\pi e^2/3R)\beta^3(E/m)^4$, where R is the radius of the accelerator. The dependency on the mass implies that an electron looses a quantity of energy 10^{13} times bigger than the one lost by a proton.

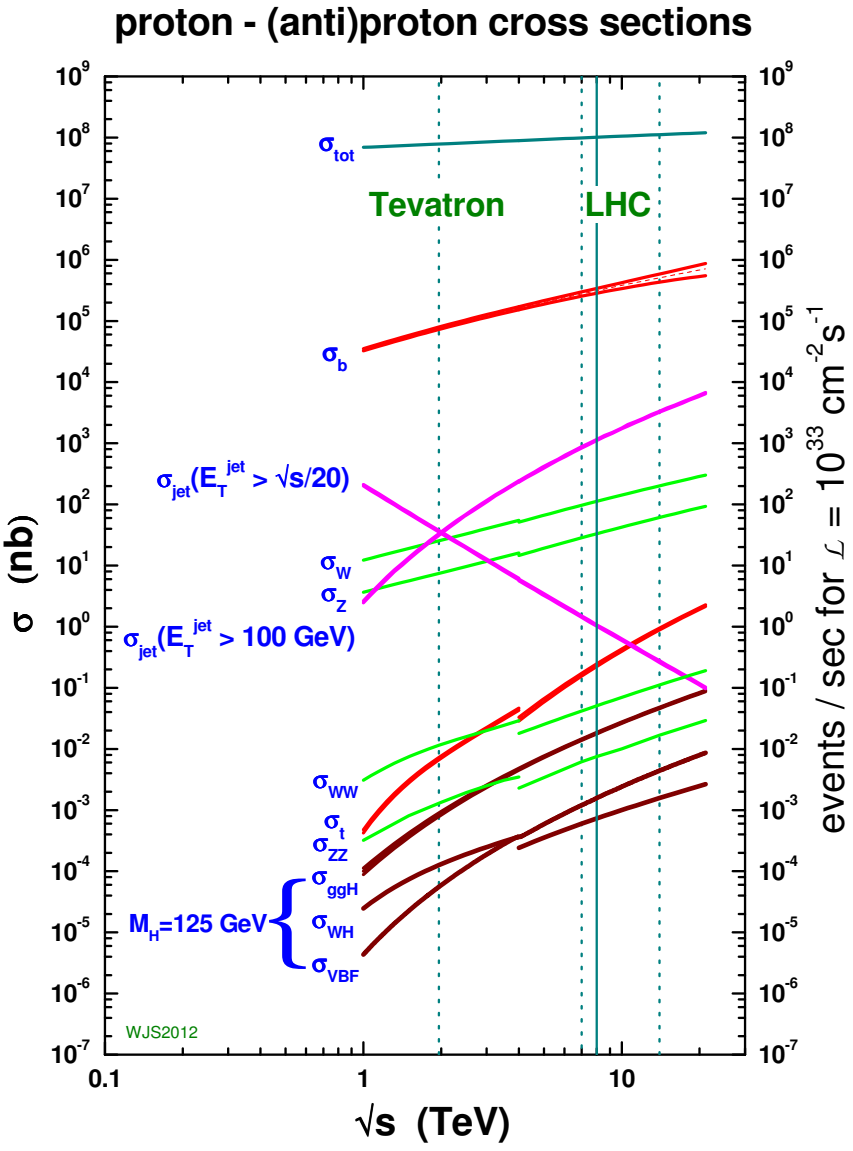

Figure 2.1: Production cross sections and corresponding event rates for different processes at LHC reported as a function of the center-of-mass energy $\sqrt{s}$ and in correspondence of a instantaneous luminosity $\mathcal{L}(t) = 10^{33}$ cm^{-2} s^{-1} [95].

to contain 2808 bunches of 1.15×10^{11} protons each, with a separation in time of 25 ns. The instantaneous luminosity, which is the number of collided particles per unit area (cm^2) per unit time (s) and is indicated as $\mathcal{L}(t)$ and it is given by [96]:

$$\mathcal{L}(t) = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n \beta^*} F \,. \tag{2.1}$$

The quantity $\gamma = E_{beam}/M_p$ is the Lorentz factor of protons in the beam, f is the bunch frequency, k_B is the number of bunches per beam, N_p is the number of protons per bunches, F is a reduction factor due to a non-π intersecting angle of the beams. The symbol ϵ_n indicates the transverse emittance, which quantifies how much particles are spread in the beam in terms of their positions and momenta. The more all particles are confined in a small region of the phase space (small emittance), compatible with the collider design momentum and the beam-pipe size, the more the probability that an interaction occurs increases (higher luminosity). The betatron function β^* quantifies the beam focusing by the magnetic optics at the interaction point. Overall, a high luminosity can be achieved through a high populated bunches of low emittance that collide at high frequency in dedicated interaction points such that the beam optics provide low values of the betatron functions. In correspondence of the design values of the parameters entering in Equation 2.1, which are summarized in Table 2.1, the instantaneous luminosity reached is $\mathcal{L}(t) = 10^{34}\ cm^{-2}s^{-1}$.

E_{beam} [TeV]	k_b	N_p	f [Hz]	F [m]	ϵ_n [mm·mrad]	β^* [m]
7	2808	1.15×10^{11}	11245	0.76	3.75	0.55

Table 2.1: Design values of the most relevant LHC parameters [96].

In a certain time-interval, the average number of events N for a process with cross section σ is given by

$$N = \sigma \int \mathcal{L}(t) dt \,. \tag{2.2}$$

Before being injected in the LHC tunnel, protons are pre-accelerated through the complex accelerator chain hosted at CERN. A schematic view of the CERN accelerators, and their connections, is shown in Figure 2.2. Protons are produced from hydrogen ionization and firstly accelerated up to 50 MeV in a

linear accelerator (LINAC2), to be subsequently injected into the Proton Synchrotron Booster (PSB), which produces the first bunches and brings the protons to an energy of 1.4 GeV. After that, these bunches of protons are injected into the Proton Synchrotron (PS), which accelerates the protons up to 26 GeV and delivers bunches, with 25 ns of time-separation, to the Super Proton Synchrotron (SPS). The SPS accelerates the protons to an energy of 450 GeV, making them ready to be injected into the LHC.

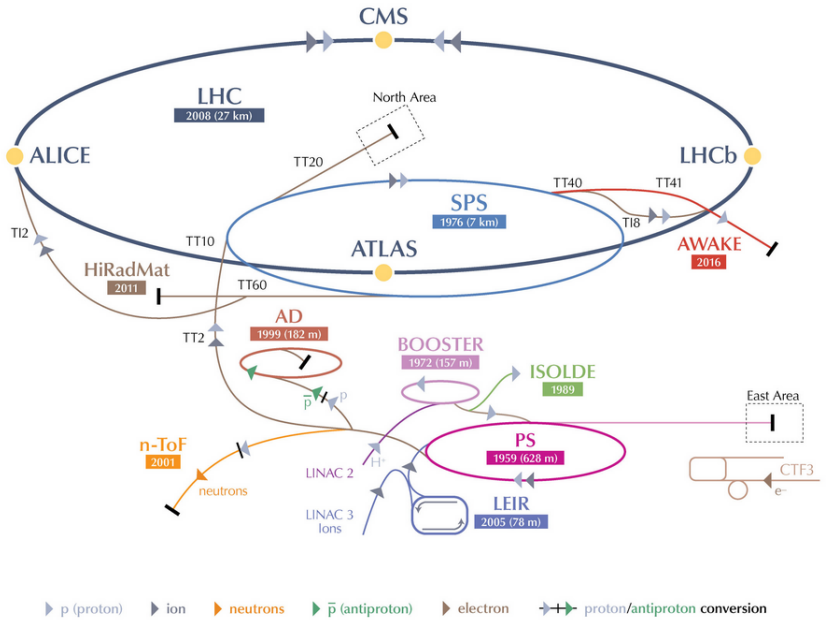

Figure 2.2: Schematic view of the CERN accelerator complex [97].

Four main different experiments are located along the accelerator ring, corresponding to the collision points of the two proton beams: ATLAS (A Toroidal LHC Apparatus) [9], CMS [10], LHCb [98] and ALICE (A Large Ion Collider Experiment) [99]. ATLAS and CMS are general-purpose experiments and they are devoted to investigate a wide range of physics, LHCb is designed to study the physics of b quarks and the CP-violation, ALICE is specialized in studying heavy ions and the quark-gluon plasma.

The LHC started its operation on September 2008, to be stopped nine days later because of a severe magnet quenching that caused an extensive damage

to fifty superconducting magnets, their mountings and the vacuum pipe. In 2009 the machine became operational again, with a reduced beam energy. The 2010, after a ramp-up of the beam energies, saw the start of the LHC research program with collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV. The data taking continued in 2011 at the same center-of-mass energy. During 2010 and 2011 the machine commissioning continued along with the data taking. During this time the instantaneous luminosity increased continuously and, the total integrated luminosity delivered to CMS was 44.2 pb^{-1} and 6.1 fb^{-1} in 2010 and 2011, respectively. The center-of-mass energy was increased to 8 TeV in 2012, while the integrated luminosity $\mathcal{L}$ delivered to the CMS by the end of the first period of data taking (Run I), was 23.3 fb^{-1}. In this period the LHC instantaneous luminosity reached $7.7 \cdot 10^{33}$ cm^{-2} s^{-1}. The integrated luminosity delivered by the LHC during Run I is shown in Figure 2.3. Despite the high luminosity is crucial to reach sensitivity to rare processes, it has the drawback of having in parallel to the hard interaction of interest additional interactions per bunch crossing, which superimpose each other in the detector: the so-called *pileup*. In 2012, about twenty p-p interactions per bunch crossing were expected for an instantaneous luminosity corresponding to $5{-}6{\cdot}10^{33}$ cm^{-2} s^{-1}.

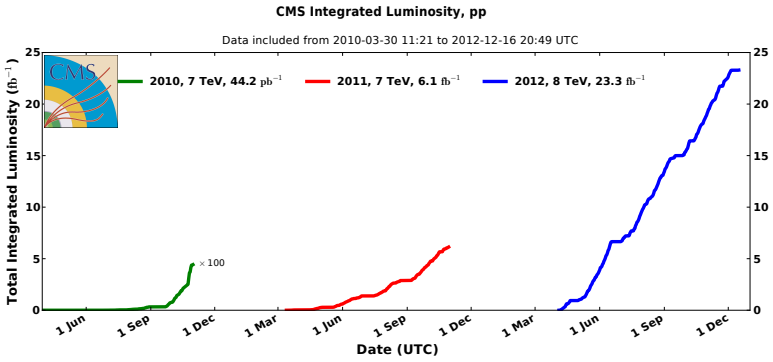

Figure 2.3: Integrated luminosity delivered to the CMS experiment during the Run I (2010, 2011 and 2012) of data taking [100].

The 2013 saw the beginning of the first period of long shutdown (LS1) for the LHC. In April 2015, the LHC has started the second period of data taking (Run II), with a center-of-mass energy of $\sqrt{s} = 13$ TeV. However, the results presented in this thesis have exploited the data collected during the LHC Run I only.

2.2 The CMS detector

The major characteristics of the CMS experiment [10, 96] are summarized in its acronym. CMS is much less extended that the ATLAS detector, hence it is **Compact** with a cylindrical shape of 15 m-long diameter and a length of 21.5 m, for a total weight of about 14000 t. **Muon** stands for the detector ability to measure very energetic muons, while **Solenoid** refers to the shape of the magnet hosted within the detector.

CMS is installed 100 m underground at the LHC interaction point 5 (P5) near the Cessy village in France. As the ATLAS detector, CMS has been designed in the early 1990s with the aim of achieving several goals, based on a wide range of physics phenomena at the TeV scale. The discovery of the Standard Model Higgs boson constituted the benchmark for which the experiment has been built. Additionally, the precise measurements of additional standard model processes (QCD, electroweak, flavour physics, etc.), the search for supersymmetric particles and new massive vector bosons, complete the challenge physics program the CMS detector has been built to fulfill. The CMS experiment has been designed in order to achieve such experimental goals in a very challenging environment in which the interesting events for the physics drown in the pileup sea. The presence of about 20 inelastic events every 25 ns causes a severe increase of the electronic occupancy, requiring the building of high-granularity subdetectors. Precise detectors are also required in order to properly reconstruct those objects belonging to the event of interest and measure their properties. Additionally, the 25 ns bunch spacing constraints the trigger system, the time-response of each subdetector and the readout to cope with a collision rate of 40 MHz. The high radiation levels, due to copious flux of particles, requires radiation-hard detectors and front-end electronics.

2.2.1 The CMS geometry

The CMS apparatus has an approximate cylindrical symmetry around the beam axis and it is segmented into a central part, called *barrel*, and two lateral components, called *endcaps*. The schematic overall layout of the CMS detector is reported in Figure 2.4.

The origin of the coordinate system adopted by CMS coincides with the nominal collision point, which is located inside the experiment. The y-axis points vertically upward, the x-axis radially inward toward the center of the LHC, and the z-axis along the beam direction toward the Jura mountains from LHC

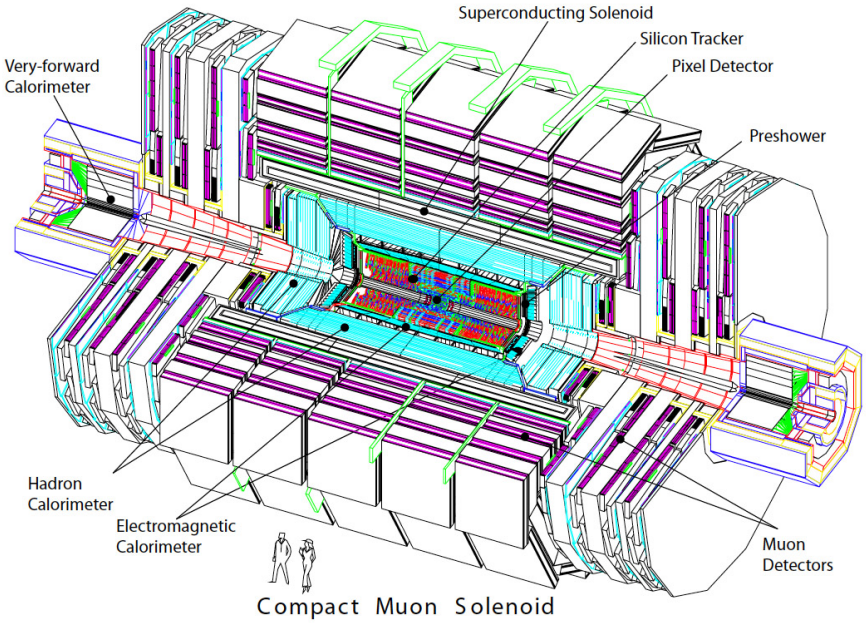

Figure 2.4: Tridimensional representation of the CMS apparatus [10].

P5. The azimuthal angle $\phi \in [-\pi, +\pi]$ is measured from the x-axis in the transverse x-y plane. The polar angle θ is measured with respect to the z-axis. However, instead of the polar angle θ, the pseudorapidity η, defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] , \tag{2.3}$$

is used. In principle, the pseudorapidity is not invariant under Lorentz boosts along the z-axis, but, in the high-relativistic approximation, as the one valid for the physics at the LHC, it is equivalent to the Lorentz-invariant rapidity y:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) . \tag{2.4}$$

As the reader can observe, by comparing Equation 2.3 with Equation 2.4, the rapidity has the disadvantage of being hardly measurable for highly relativistic particles as fully dependent on their kinematics. This is not the case, instead, for the pseudorapidity, given its purely geometrical nature.

Other invariants under Lorentz boosts along the z-axis are the transverse momentum p_T defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \, , \tag{2.5}$$

where p_x and p_y are the projections of the four momentum p^μ of a given particle on the x and the y axes, respectively and the transverse energy $E_T = E \sin\theta$. The distance ΔR between two particles with difference in pseudorapidity and azimuthal angle equal to $\Delta\eta$ and $\Delta\phi$, respectively, defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \, , \tag{2.6}$$

is also a Lorentz invariant. The description of the entire CMS apparatus is provided in the following sections.

2.2.2 Magnet

In order to achieve a high momentum resolution in a very compact detector, a solenoidal superconducting magnet, generating a magnetic field of 3.8 T, is used in association with an iron return yoke in the muon detector system (cfr. Section 2.2.6). Such an intense magnetic field is achieved by letting a current of 18 kA circulating in the cables. The total energy stored by the magnet system amounts to 2.5 GJ. The CMS magnet is 12.9 m long and it is constituted of an internal bore with a radius of 5.9 m, in which the tracker system (cfr. Section 2.2.3) and the calorimeters (cfr. Sections 2.2.4 and 2.2.5) are hosted. The magnet is composed of 2168 turns and it is cooled at the temperature of -268.5 °C. The return field is guided by an external iron-yoke, which is 1.55 m and 1.45 m thick in the barrel and in the endcap, respectively.

The single solenoid configuration, associated with an iron return yoke in the muon system, allows to produce a magnetic field that has always a direction parallel to the beams, thus curving the muon tracks in the transverse plane. The small size of the beams in this plane allows the transverse position of the vertex to be measured with an accuracy of $\leqslant$ 20 µm. Such an accurate measurement plays a less important role, for instance, for muons in the ATLAS magnet configuration, where the magnetic field created by a small inner solenoid plus toroids in air after the calorimeters, bends the trajectory of more energetic muons in the longitudinal plane. Moreover, the size of the CMS magnetic volume is larger with respect to the ATLAS solution, implying a better resolution for charged tracks that do not emerge in the muon system. The ATLAS configuration, instead, is optimized for momentum resolution on high energy muons because it avoids the use of a dense material (like iron) and maximizes the distance over which position measurements are made.

2.2.3 Inner tracking system

The inner tracking system is the innermost CMS subdetector. The main purpose of the tracker is the identification and the characterization of charged particle tracks, the primary and the secondary interaction vertices in the acceptance region of $|\eta| < 2.5$. It has been designed to reconstruct tracks with a transverse momentum resolution of $\delta p_T/p_T \approx 0.15 \cdot p_T$ (TeV) $\bigoplus 0.5\%$ in the central region ($|\eta| < 1.6$), gradually degrading to $\delta p_T/p_T \approx 0.6 \cdot p_T$ (GeV) $\bigoplus 0.5\%$ as $|\eta|$ approaches to 2.5.

In order to cope with the high flux of charged particles coming out from each bunch crossing, whose measurement and reconstruction require high granularity and fast time-response, the CMS tracker system has been fully built in silicon. The technology used is totally driven by the decrease of the charged-track density with the increase of the distance with respect to the interaction point.

Silicon pixel detectors are used in the region closest to the interaction point ($r < 20$ cm), where a high granularity is needed. Silicon microstrip detectors, instead, are used elsewhere. A schematic view of the inner tracker is depicted in Figure 2.5.

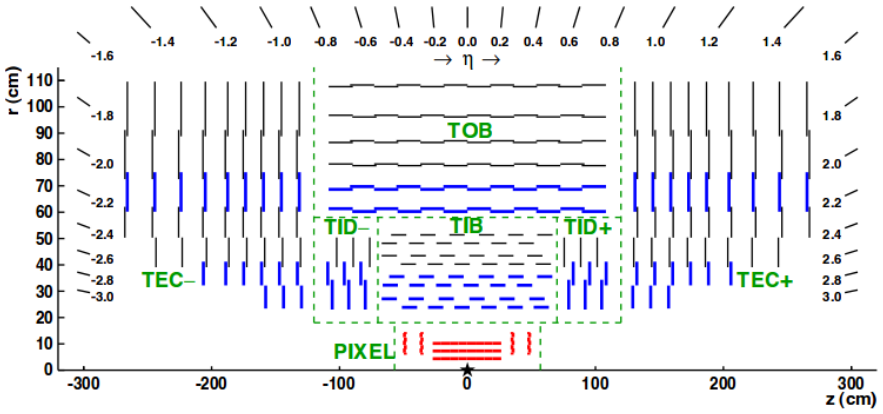

Figure 2.5: Longitudinal view of the layout of the CMS inner tracking system [101].

The pixel detector consists of matrices of 100×150 μm^2 silicon pixels arranged on three concentric barrel layers and four endcap disks, two per each extremity. The barrel layers are positioned at $r = 4.4$, 7.3 and 10.2 cm from the beam line, while the endcap disks are located at $|z| = 34.5$ and 46.5 cm. The total active area of the pixel detector is of about 1 m^2. Given the small size of the

silicon pixel, the measured spatial resolution is about 10 μm in the barrel and about 20 μm in the endcaps.

The outermost layers of the tracker, for which the microstrip technology has been used, are further arranged into separated substructures: an inner section in the barrel (TIB) completed with endcap disks (TID) and an outer barrel section (TOB) with endcap components (TEC) at the extremities. The TIB consists of four barrel layers and covers up to $|z| < 65$ cm. It is accompanied by the TID, which is composed of three disks at each end. The entire TIB/TID structure extends up to $r = 55$ cm. TIB/TID silicon microstrip sensors have a thickness of 320 μm and a strip pitch that varies from 80 μm in the innermost barrel layer to 140 μm in the outermost disks. The TIB/TID system is surrounded by the TOB, which extends up to $r = 116$ cm. It is composed of six layers of 500 μm thick microstrip sensors with a strip pitch of 183 μm for the first four layers and 122 μm for the last two ones. In the endcaps, the TEC, which consists of nine disks, extends in the regions 124 cm $< |z| <$ 282 cm and 22.5 cm $< |z| <$ 113.5 cm. Each disk consists of seven rings of silicon microstrip detectors with a thickness varying from 320 μm for the first four rings, to 500 μm for the last two rings and with an average pitch varying from 97 μm to 184 μm depending on the distance from the beam line.

The sensors strips are arranged parallely to the beam axis in the barrel and radially in the disks. Therefore, unlike the pixel modules, the strips ones do not allow to measure the point of passage of the particle along the strip, but provide only two coordinates of it: the r-ϕ and the z-ϕ coordinates. Moreover, in order to provide a measurement of the third coordinate, the first two TIB and TOB layers, the first two TID rings, and the first, the second and the fifth rings of the TEC, are characterized by "stereo" modules. These components have a second microstrip detector module which is mounted back-to-back and tilted at an angle of 100 mrad for a more precise measurement of the z-coordinate. For the TIB (TOB), this configuration leads to a single-point resolution of about 23-34 μm (35-52 μm) in r-ϕ and 230 μm (530 μm) in z.

2.2.4 Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) [102] has the main function of measuring the energy of particles that generate electromagnetic showers (i.e. photons and electrons). It is placed outside the tracker and inside the coil of the magnet.

The ECAL is an homogeneous calorimeter whose active material is a crystal scintillator made of lead tungstate ($PbWO_4$). The calorimeter is divided into

two parts: the ECAL barrel (EB), which covers the pseudorapidity range $|\eta| < 1.479$ and the ECAL endcaps (EE), which extend the detector pseudorapidity range to $1.479 < |\eta| < 3.0$. The most distinguishing feature of the CMS ECAL is the lead tungstate scintillating crystals, which have a short radiation length $X_0 = 0.89$ cm (the radiation length is the distance needed to attenuate, on average, the energy incident radiation by a factor $1/e$) and a small Molière radius $R_M = 2.2$ cm (the Molière radius represents the size needed to fully contain the electromagnetic showers in the transverse direction with respect to that of the shower itself). These features guarantee an excellent containment of the showers in a compact volume with high granularity. In particular, crystal-lengths of 23 cm in the barrel and 22 cm in the endcaps are long enough to collect all the particle energy in the large majority of cases. The EB consists of 61200 crystals with a section of 22×22 mm^2 and a granularity of 0.0174×0.0174 in $\Delta\eta \times \Delta\phi$. Each EE, located at $|z| = 314$ cm, is made of 7324 crystals with a section of 30×30 mm^2. Both barrel and endcap crystals are displaced off-point from the nominal-vertex position to mitigate the detection inefficiency that would result from particles going along the edges of the crystals. The crystals are arranged according a η-ϕ grid in barrel, while the disposition is organized along x-y in endcaps.

Together with tightness, compactness and high granularity, the crystals have a fast response given that the 80% of the scintillation light is emitted in 25 ns. The crystals are also radiation resistant. One of the defects of this kind of crystal is that a large amount of the stored energy is dissipated through thermal emission, then, in order to cope with the relatively low light yield (30γ/MeV), photodetectors with intrinsic high gain, and able to operate in a high magnetic field, are used. Silicon avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) are used as photodetectors in the barrel and in the endcaps, respectively. Moreover, given that the sensitivity of the crystals and the APDs changes with the temperature, a cooling circuit is used to keep the operating temperature constantly at (18 ± 0.05) °C.

An additional component, the ECAL preshower (ES), is placed in front of the EE, covering the pseudorapidity range $1.653 < |\eta| < 2.6$. It is a sampling calorimeter composed of two layers of lead radiator, interspersed with silicon strip sensors with a 63×63 mm^3-extended surface. The preshower has been mostly designed to discriminate genuine photons from those coming from the electromagnetic decay of neutral pions. The reason why only the endcap regions are equipped with preshowers is twofold: the worst spacial resolution in the EE and the fact that, in this region, the angular separation between the two photons coming from a π^0 decay is expected to be small enough to make

challenging the π^0-γ separation. The longitudinal view of one quarter of the CMS ECAL is shown in Figure 2.6.

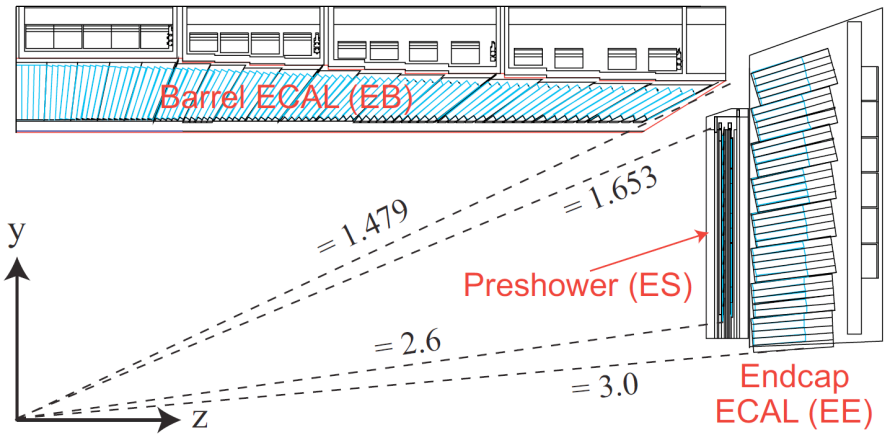

Figure 2.6: Longitudinal view of one quarter of the CMS ECAL detector in which the positions of the different components (ECAL barrel, endcap and preshower) are visible. The dashed lines correspond to fixed η values [10].

The ECAL energy relative resolution, expressed as a function of the energy of the incoming particle, can be parametrized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_N}{E}\right)^2 + c^2 \text{ (with E in GeV).} \tag{2.7}$$

The stochastic term a takes into account the intrinsic statistical fluctuations of the shower development and light collection. The term with σ_N takes into account the noise associated with the electronics, which receives contributions also from the leakage current induced by neutron irradiation especially in the barrel, and the pileup. The constant term c is related to the calorimeter features (i.e. inter-calibration errors, geometrical effects, etc.). The energy resolution parameter values determined in test-beam using electrons with energy in the range $[20, 250]$ GeV, and considering a 3×3 crystal configuration, are: $a = 0.028$ GeV$^{1/2}$, $\sigma_N = 0.12$ GeV and $c = 0.003$ [103].

The ECAL energy resolution has been measured in 7 TeV data (corresponding to an integrated luminosity of about 5 fb^{-1}): for $E_T \approx 45$ GeV electrons from Z boson decays, the resolution is better than 2% in the central region ($|\eta| < 0.8$) and between 2% and 5% elsewhere. For electrons emitting photons for bremsstrahlung, and having more than 94% of the clustered energy contained

within a 3×3 array of crystals, the energy resolution improves to 1.5% for $|\eta| < 0.8$ [104].

2.2.5 Hadronic Calorimeter

The hadron calorimeter (HCAL) [105] has the main function of measuring the charged and neutral hadron energy and also providing hermeticity for the measurement of the missing transverse energy $\not{E}_T$. It is divided into the HCAL barrel (HB), which covers the pseudorapidity range $|\eta| < 1.4$ and the HCAL endcap (HE), which is extended in the range $|\eta| < 1.4$ and $1.3 < |\eta| < 3.0$, as can be seen from Figure 2.7.

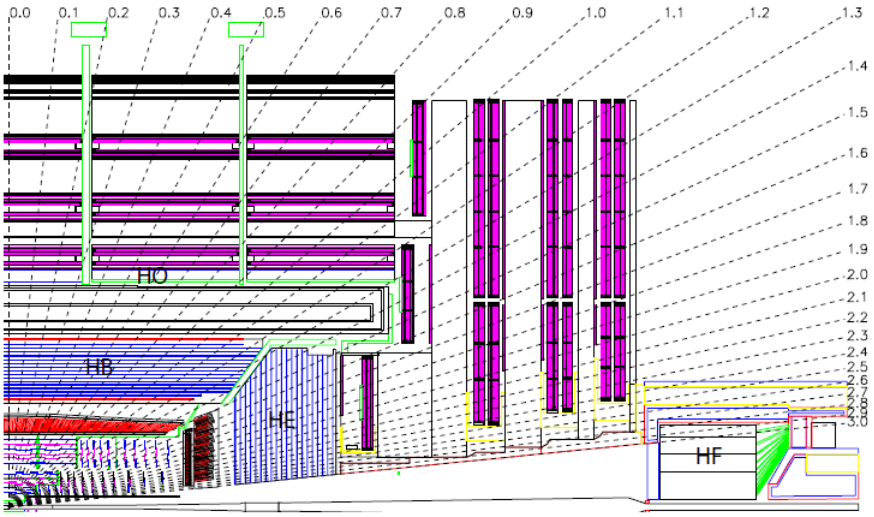

Figure 2.7: Longitudinal and schematic view of the CMS HCAL. The dashed lines correspond to fixed η values [10].

The HCAL is a sampling calorimeter, meaning that, differently from the ECAL, it is constituted of layers of a dense absorber alternated with an active medium. Both the HB and the HE are placed outside ECAL and inside the coil of the solenoid. For this reason, and given that the intense magnetic field requires the usage of non-ferromagnetic materials, stainless steel and copper alloys (brass) have been chosen as absorbers, while tiles of plastic scintillator are used as active medium. Each scintillator is read out with embedded wavelength-shifting fibers, connected to high-attenuation-length clear fibers, which con-

vey the light signal to the readout system, where it is amplified by a hybrid photodiode.

The high granularity in η-ϕ is obtained by segmenting the scintillators from $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in the barrel up to $\Delta\eta \times \Delta\phi = 0.350 \times 0.174$ in the endcaps. The HCAL thickness, in terms of nuclear absorption length λ_I (which is the equivalent of the radiation length X_0 for hadronic calorimeters[3]), goes from approximately 5.82 λ_I in the barrel up to 10.6 λ_I in the endcaps. However, the barrel-depth is not enough to contain the full shower, then an additional module, the HCAL Outer (HO), is placed outside the magnet. The HO detector, which covers the region $|\eta| < 1.26$ (cfr. Figure 2.7), contains scintillators serving as "tail-catcher", given that they sample the energy from penetrating hadron showers leaking through the rear of the calorimeters. The forward HCAL (HF) extends the acceptance of the detector to the region $3.0 < |\eta| < 5.2$ and it is located 11 m away from the interaction point (cfr. Figure 2.7). This location causes the HF to work in an high-radiation dose environment: for this reason, the material chosen as absorber is steel embedded with quartz fibers, which act as the active material. The fibers, once crossed by the the hadronic shower, produce light due to the Cherenkov effect. The inclusion of the HO layers extends the total depth of the calorimeter system to a minimum of 11 λ_I.

The energy resolution is $\sigma/E = 65\%/\sqrt{E} \oplus 5\%$ in the barrel, $\sigma/E = 83\%/\sqrt{E} \oplus 5\%$ in the endcaps and $\sigma/E = 100\%/\sqrt{E} \oplus 5\%$ in the HF [105].

2.2.6 Muon system

Muons play a major role in many appealing SM and exotic scenarios made available by the LHC physics program. Muon detection was crucial for the discovery of the SM Higgs boson in the golden-channel $h \to ZZ$ (in the four leptons final state), or also for the observation of the rare decay $B_S^0 \to \mu^+\mu^-$ [106]. For these reasons, a sophisticated and complex system has been built to reconstruct muons and measure their kinematic properties, but also to fulfill the function of event-trigger. Given the need to cover a large area affected by radiations and a variable-intensity magnetic field, three different kind of gas-ionization detectors have been used in the muon system [107], as can be seen from Figure 2.8: drift tubes (DT), catode strip chambers (CSC), and resistive plate chambers (RPC).

[3] In general, for a given type of material, $\lambda_I > X_0$.

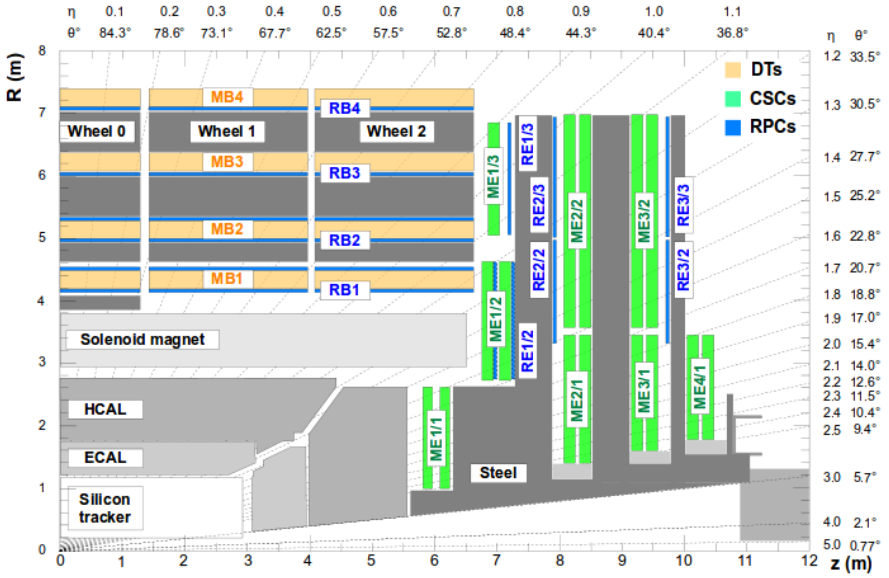

Figure 2.8: Longitudinal view of the CMS detector in which the different components of the muon system are detailed schematized. The dashed lines correspond to fixed η values [108].

DRIFT TUBES The DT chambers are made up with a 50 mm-thick gold-plated stainless-steel anode wire and an aluminum strip cathode, placed on the shortest side of the cell. The cell is filled with a gas mixture of CO_2 (85%) and Ar (15%). Because of the long drift time ($\approx$ 400 ns [108]), the DT chambers suit as tracking detectors in the barrel region ($|\eta| < 1.2$), where the muon flux is weak (< 10 Hz/cm^2) and the residual magnetic field is low. The barrel region is segmented into five wheels, each 2.5 m-long. Each wheel is divided into 12 sectors and one single sector consists of 250 drift tubes chambers arranged in four stations (MB1, MB2, MB3, MB4), each interspersed by an iron-yoke layer. Each chamber consists of twelve layers of drift tubes organized into three units of four layers each called *superlayer*. The two superlayers at the extremities consist of wires arranged parallely with respect to the beam line, while the one in the middle have wires oriented perpendicularly to the beam. Therefore, the former provide eight measurements in the r-ϕ projection, the latter four measurements along the z-axis. Then, given the higher number of hits available, the radial resolution (100 µm) is better than the one related to the z-direction (150 µm). The angular resolution in ϕ is 1 mrad. Except for the

MB4 station, which hosts the extreme superlayers only, all the other stations are fully equipped with all three superlayers.

CATHODE STRIP CHAMBERS The CSC subdetectors are placed in the endcaps only ($0.9 < |\eta| < 2.4$), where there are high radiation doses and a large, and non-uniform magnetic field. The entire CSC system consists of four stations of CSC (ME1, ME2, ME3, ME4), each interleaved with one layer of the magnet iron-yoke. Each chamber has a trapezoidal shape and is filled with a gas mixture of Ar (40%), CO_2 (50%) and CF_4 (10%). Additionally, it consists of six gaps, each made of a multi-wire anode layer interspersed by cathod-strips planes. Wires are ϕ-oriented and perpendicularly arranged with respect to the strips, which are segmented radially. Therefore, each gap provides six measurements of the ϕ-coordinate (strips) and six measurement of the r-coordinate (wires). In particular, the ϕ-coordinate is measured by calculating the center of gravity of the ionic charge induced by avalanche on the strips once the gas has been ionized, while r is obtained by the electronic charge collected on the anode. The single point resolution is of about 75-150 μm in r-ϕ.

RESISTIVE PLATE CHAMBERS As can be seen from Figure 2.8, a set of RPC is placed both in the barrel ($0 < |\eta| < 1.2$) and in the endcaps ($0.9 < |\eta| < 1.6$). The double-gap RPC detectors consist of two single-gap chambers operating in avalanche mode. The two units are assembled back-to-back in order to share a segmented pick-up read-out strips in between, which allows the usage of lower high voltages (HV) in each single-gap with an effective gain in efficiency. Each chamber consists of two electrodes of bakelite externally covered by a layer of graphite, which is used as HV electrode. The whole unit is covered by mylar sheets, serving as HV insulators. The gap between the electrodes is 2 mm-thick and is filled with a gas mixture of $C_2H_2F_4$ (96%), C_4H_{10} (3.5%) and SF_4 (0.5%). Not ideal for space measurements given the coarse spacial resolution, the RPC are, instead, crucial for the event-triggering (excellent timing resolution of about 1 ns [10]).

2.2.7 Trigger and Data acquisition

At the design luminosity ($\mathcal{L}(t) = 10^{34}$ cm^{-2} s^{-1}), the frequency at which protons collide (40 MHz for 25 ns spaced beams) translates into approximately 20 inelastic interactions per bunch crossing and about 1 MB/event of zero-suppressed data in the CMS read-out systems. Performing a permanent storage of such amount of data is technically awkward, therefore a drastic rate reduction must be achieved. To accomplish this goal, a trigger and data ac-

quisition (TriDAQ) system is used. CMS adopts a two-level system of rate reduction, which consists of a Level-1 (L1) Trigger [109], followed by a High-Level Trigger (HLT) [110].

The L1 trigger is a real on-line and hardware-based event filter. It reduces the data-rate of a factor $\mathcal{O}(10^3)$, providing an output rate limit of 100 kHz. During the finite L1 trigger latency-time of 3.2 µm, in which the trigger decides whether accepting an event or not, the information of 128 bunch crossings are stored in pipelined FIFO memories. Due to the short decision time, only the information provided by the calorimeters and the muon system are exploited. In particular, for trigger purposes, calorimeters are segmented into *trigger towers* with a size equal to $0.087 \times 5^\circ$ in $\Delta\eta \times \Delta\phi$ up to $|\eta| < 1.74$ and larger at higher η values. ECAL, HCAL, HF, DT, CSC and RPC contribute to build the L1 trigger components at different levels: local, regional and global. The local components are extracted from energy deposits measured in the calorimeter trigger towers and track segments, or hit patterns in the muon chambers. These primitives are then combined to build trigger objects which give rise to the regional components. At this level, DT and CSC systems join the segments to complete the muon tracks, while the RPC perform unambiguously the tracks/bunch crossing matching. In the calorimeters[4], the electron and photon identification is performed and additional information about the muon isolation and the compatibility with minimally ionizing particles (MIP) are also collected. Once measured, all these regional components are ranked according to whether they meet or not some thresholds and quality requirements. The highest-ranked components are then built independently by the Global Calorimeter Trigger (GCT) and the Global Muon Trigger (GMT) and transferred to the final Global Trigger (GT), the top entity of the Level-1 hierarchy. At this point, based on the best-ranked objects selected, the GT decides whether rejecting the event or not through the Timing, Trigger and Control system (TTC).

Once the L1 trigger processing is finished, the accepted data are read out from the detector through the Data Acquisition System (DAQ). Thus the complete information is passed to the event builder network, and then to the fully-software based HLT system. The HLT reduces the data-rate to about 100 Hz in about 10 µs for most of the event processed. The main pillar on which the HLT algorithms are based is the regional reconstruction: the idea is processing only those events found already interesting by the L1 trigger and fast discard

[4]The first upgrade (Stage-1) of the Calorimeter Trigger has been performed for the LHC Run II to ensure that physics data collections are not compromised by the change of the running conditions. Some electronic components and links (from copper to optical fibers) have been improved [111].

all the rest in order to minimize the latency. This has led to the development of two "virtual" trigger levels: the first level (called Level-2) accesses only the muon and calorimetric data, the second level (referred as Level-3) includes the information provided by the tracking system.

Given the strong algorithm flexibility, the HLT "menu" has been enriched, in the last few years, with many HLT sets of *trigger paths*, each customized to be tuned with the collider configuration.

2.3 The CMS computing system

Events accepted by the HLT are finally written on disks and made available for data analysis. About 30 PB of recorded data are produced annually and physicists all around the world are called to determine which amount of collision events is interesting to be analyzed. In order to facilitate the storage of such a big amount of data and guarantee a flexible and an efficient connection within the experimental physics community, in 2002 a hierarchic computing system has been introduced [112]. Such a system, called *Worldwide LHC Computing Grid* (WLCG) [113], is organized according to the "Grid" paradigm: three hierarchical levels of tiers of computing facilities are in place to allow all computing resources (worldwide distributed) to be connected.

TIER-0 (T0) This center is located at CERN and it allocates data from the CMS TriDAQ system. The data organization in Primary Dataset (PD), depending on the associated HLT paths, is also performed in these sites.

TIER-1 (T1) The datasets are distributed from the T0 to the T1 centers. The latter are systems of permanent storage, which also provide the necessary CPU for skimming and calibration processes. Currently, seven computer centers are located in Europe, Asia and US, which are connected to CERN by optical-fibre links.

TIER-2 (T2) More numerous (about seventy) and smaller are the Tier-2 (T2) centers, characterized by conspicuous CPU resources able to provide capacity for analysis, calibration activities and Monte Carlo (MC) simulation.

TIER-3 (T3) Local computing resources.

2.4 Event simulation

A fundamental interplay exists between real and simulated collision events: just consider how the discovery of new phenomena would allow the knowledge of the phenomenological modeling for simulation to be extended and improved and how this latter achievement might, in turn, lead to a better set-up of the future analysis strategies. A schematic view of how a generic $pp \rightarrow X$ event is framed in the context of the event simulation is shown in Figure 2.9. The simulation of an event produced in a p-p collision can be factorized in different steps, that are going to be described in the following.

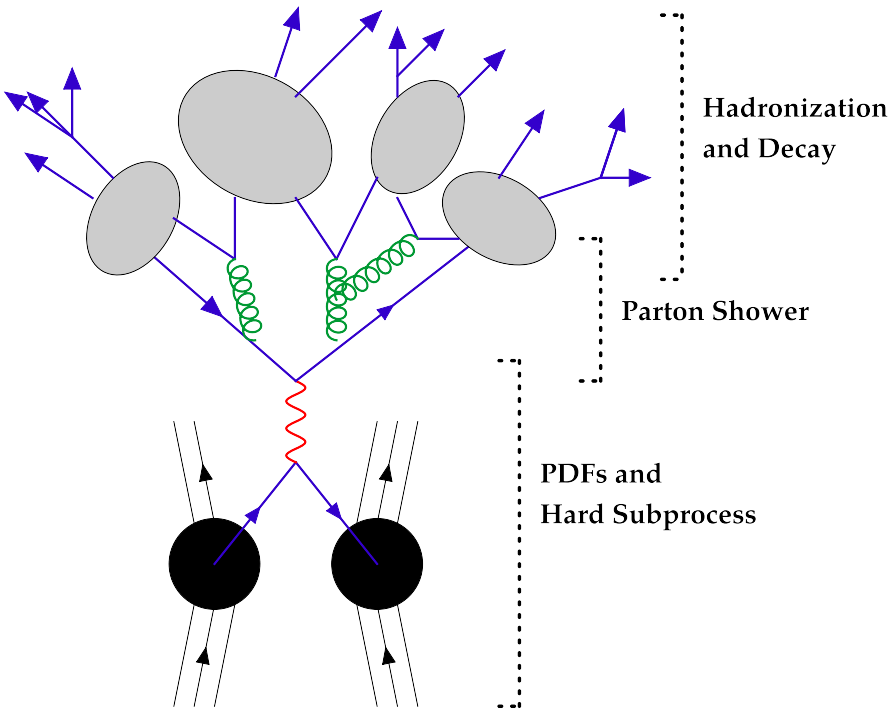

Figure 2.9: Schematic representation of the generic structure of a p-p collision with the main phases of the collision development highlighted: the hard scattering, the parton showering, the hadronization process and the underlying events.

PDFs AND HARD SUBPROCESSES As already explained in Section 1.2.1, the cross section of a generic hadronic process $pp \rightarrow X$ can be written as a function of the differential cross section of the partonic process weighted by the partonic PDFs. The simulation of the hard scattering process, then, consists

in computing the key term, which the partonic differential cross section depends on: the matrix element. The PDFs description and the computation of the matrix element can be performed by general purpose generators like PYTHIA [114]. More than being only a matrix element computation tool, this generator is capable of modelling the entire generator chain, taking care also of the parton showering, hadronization and underlying event descriptions (such steps will be described in the following paragraphs). In PYTHIA, the computation of the matrix element is performed at the first order in perturbation theory (LO). MADGRAPH/MADEVENT [115, 116] is a dedicated generator devoted to an accurate calculation of the matrix element. MADGRAPH allows to calculate cross sections at the LO including up to five extra parton in the matrix element and real corrections. The most recent version of MADGRAPH, MADGRAPH5_AMC@NLO [117], allows the virtual loop corrections to be taken into account and either LO or NLO calculations to be performed. POWHEG [118] is also a widely used generator able to calculate the matrix elements of a given process at NLO.

PARTON SHOWER Before the hard interaction, partons can branch into other partons. Also partons produced in the hard interaction can branch themselves, leading to a shower of secondary partons, the so-called *parton shower*. In particular, in the former case the radiation is called *Initial State Radiation* (ISR), in the latter case it is called *Final State Radiation* (FSR). The parton shower approach is a treatment of a QCD parton splitting, which is a perturbative approximation valid up to the order of Λ_{QCD} (cfr. Section 1.1.2). It has been already introduced that PYTHIA is able to perform the parton shower themselves on top of the matrix element computation. In general, parton shower tools are able to accurately describe collinear/soft jet emission, but they can easily break down above a given scale where a purely matrix element approach should be used. Therefore, depending on the process to be studied, it might be often preferable using matrix element tools like MADGRAPH and POWHEG interfaced with a parton shower program. However, the drawback of using matrix element and parton shower tools at the same time is the possibility of counting twice the same event. Generally, in fact, there are two ways to describe an event having $(N + 1)$ jets in the final state:

a) final state with $(N + 1)$ partons in which soft or collinear radiations give rise to $(N + 1)$ jets (typical description well performed by a parton shower tool),

b) final state with N partons, where an hard emission leads to $(N + 1)$ jets (typical description of matrix element plus parton shower tools).

In order to get rid of possible double-counting, different "matching" techniques are used. The main idea is dividing the phase space in two different regions: the region with soft and collinear emissions will be described by parton showering, while the region with harder emissions will be taken care of by the ME generator. The MC@NLO package is properly devoted to this. It first computes the NLO matrix element including $(N + 1)$ partons and then it analytically evaluates how, in the parton shower, N partons would migrate to the $(N + 1)$ parton phase space. This estimation is used as a correction term (which can also be negative in about 10-15% of the total events) to the $(N + 1)$ parton matrix element calculation. The matching scheme integrated in POWHEG allows to overcome negative weights [118].

HADRONIZATION AND DECAYS After the parton showering, the partons enter in a physical regime, which is not possible to be described with the perturbative QCD theory. As already explained in Section 1.1.2, at low energies ($\Lambda_{QCD} \approx 250$ MeV), the coupling α_s acquires extremely high values causing the hadronization, i.e. the process in which partons confine in color-singlet hadrons. Despite the breaking down of the perturbation theory not able to describe such processes, phenomenological models have been developed to simulate the hadronization process, such as the Lund string model [39] which is the one implemented in PYTHIA. All those particles with a decay length $c\tau < 10$ mm, are then let decaying. The generator used to simulate the decay of the tau lepton is TAUOLA [119, 120], which includes also the description of the spin polarization.

UNDERLYING EVENTS Additional contributions are those brought by underlying events. These events have to be taken into account to give a complete realistic description of the p-p collisions. Two main processes contribute to populate the underlying events class: colored remnants of the protons undergoing hard interaction themselves (beam remnants) and possible secondary hard (or soft) interactions between those partons that have not contributed in the main interaction (multiple particle interaction).

The final step of a p-p collision simulation is modelling the interaction of the final particles produced in the event with the CMS detector. In order to achieve such a goal, the GEANT4 toolkit [121] is used to simulate the detector structure (geometry, active and inactive components, material, magnetic-field map, etc.) and also the interaction of a given particle with the detector material. Reproduced on the real data taking archetype, the modeled interactions create simulated electronic signals which are digitized before going through the reconstruction steps.

3

Particle reconstruction and identification

While traversing the detector, almost all particles interact with the matter releasing energy deposits. Neutrinos are an exception as their probability to interact with the matter is extremely low and their presence is inferred by the overall unbalance in the event transverse momentum. In CMS, particle reconstruction and identification is performed starting from the reconstruction of very basic objects: tracks, which are reconstructed in both the inner detector and in the muon system, and calorimeter clusters. Moreover, an efficient reconstruction of the event primary vertex is crucial to reject background events due to pileup. Secondary vertices are also important for identifying long-lived particles as heavy flavor hadrons and tau leptons are. Therefore, given the importance of understanding how tracks, vertices and energy deposits are reconstructed and measured in CMS, these topics will be firstly described in this chapter. Subsequently, an overview of the "standard CMS reconstruction" will be provided for electrons and muons only, given that standard-reconstructed leptons have been used in the physics analysis described in Chapter 5. The description of the particle-flow reconstruction [122] *will be provided afterwards. This method, recommended and widely used in CMS, allows the information coming from each subdetector to be globally combined and all particles in the event to be coherently identified and reconstructed. A general but not exhaustive focus will be adopted on electrons, muons, jets and missing transverse energy, being these objects important for the anal-*

yses described in this thesis. Additionally, a brief mention to the hadrons and photons reconstructions will be also provided, given the important role they play in the hadronically-decaying tau reconstruction, which will not be treated in this chapter, but in Chapter 4.

3.1 Detector-level reconstruction

3.1.1 Track reconstruction in the inner detector

Tracks in the inner detector are reconstructed using the Combinatorial Track Finder (CTF) algorithm, which is an extensive adaption of the Kalman Filter (KF) [123]. The CTF method is applied in an iterative procedure, called *iterative tracking*, in which the tracks of the high-p_T charged particles emerging from the primary vertex are reconstructed first so that their position measurements can be excluded when attempting to reconstruct tracks with more complex topologies (typically lower-p_T particles or secondary particles emerging from highly displaced vertices).

The track reconstruction process requires the availability of particle position measurements, each obtained from a cluster of strips or pixels with signals above predefined thresholds. Such position measurements, which are also called *hits*, result from the local reconstruction in the inner tracker. Hits are grouped to form the *track seeds*, which constitute the starting point of the Kalman algorithm. The CTF collects additional hits by extrapolating the track parameters to the next outer detection layer. Every new hit is used to improve the measurement of the track parameters. When the last detection layer is reached, the procedure is continued in the opposite direction (inward) in order to obtain estimates of the track parameters on the innermost detector layers and at the vertex.Once the trajectory is completed, it is refitted with the Kalman Filter and smoother methods to accurately estimate the parameters of the helix. Moreover, additional quality criteria (the number of detection layers not contributing with hits to the track, the track χ^2, and the number of hits shared with other track candidates) are required in order to reduce the number of fakes. Spurious hits (*outliers*), identified as those that contribute to a large increase in the overall track χ^2, are removed from the set of track hits and the tracks parameters are recomputed in order to improve the accuracy of the measurements. Certain quality requirements are then further imposed to select good tracks.

For most of the offline analyses, *high purity* tracks are used and these are selected after imposing stringent requirements on the number of layers having hits, the χ^2 normalized with respect to the degrees of freedom (χ^2/dof), the distance from the centre of the beam spot to the beam line in the transverse plane and the distance along the beam line from the closest pixel-only vertex[1]. Once the track selection process is completed, the tracks found in each iteration are merged into a single collection.

Measured in simulated $t\bar{t}$ events, considering also 2011 LHC pileup conditions, the averaged tracking efficiency for prompt and charged particles ($p_T > 0.9$ GeV) is about 94% for $|\eta| < 0.9$. In the higher pseudorapidity region, up to $|\eta| < 2.5$, the efficiency decreases to 85%. The main cause of inefficiency is the hadron-nuclear interactions in the tracker material. The fake rate is at the few percent level for $1 < p_T < 20$ GeV, while it quickly increases in the complementary low and high p_T regions. At low p_T, the presence of multiple scattering processes increases the probability to assign wrong hits to a track. At higher p_T, instead, secondary particles are produced in the nuclear-hadron interaction.

Figure 3.1 shows the transverse, longitudinal impact parameter[2] and p_T resolutions as a function of the track-p_T in $t\bar{t}$ events at $\sqrt{s} = 7$ TeV. At low p_T, both the impact parameters and p_T resolutions are progressively deteriorated by multiple scattering effects. At higher p_T, the impact parameter resolutions are mostly dominated by the position resolution of the innermost hit in the pixel detector. The tracker material accounts for 20-30% of the p_T resolution in the high p_T region. It has been checked that the impact that pileup has on track resolution is negligible [124].

3.1.2 Vertex reconstruction

The primary vertex reconstruction is crucial for determining the location and the associated uncertainty of all p-p interaction vertices in the event. The reconstruction method, which uses all the available tracks, consists in three steps. First of all, tracks are selected according to their compatibility of being produced promptly in the primary interaction region (selection criteria are imposed on the transverse impact parameter significance, the number of

[1] CMS has an independent reconstruction of tracks and primary vertices based purely on pixel hits. Pixel vertices are reconstructed using tracks built from hits in the pixel detector only. The information from pixel-only reconstruction can be useful to reduce the combinatorics of hit pairs.

[2] The transverse impact parameter d_0 is defined as the distance of closest approach to the beam line. The longitudinal impact parameter is the z coordinate of the point on the track that determines d_0.

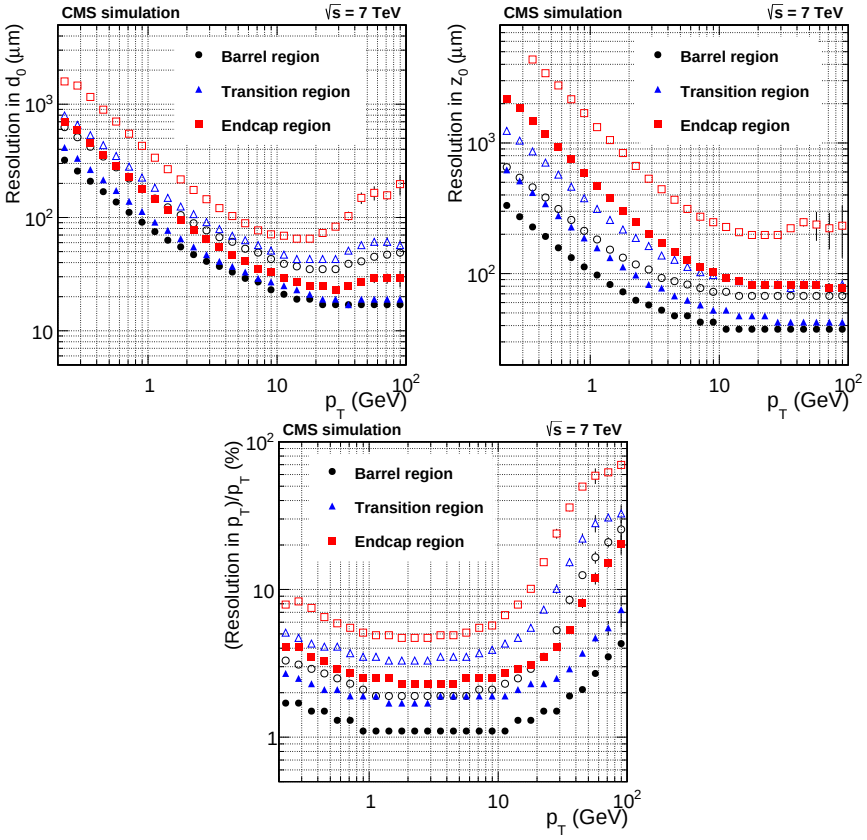

Figure 3.1: Transverse (top left), longitudinal (top right) impact parameter and p_T (bottom) resolutions evaluated as a function of p_T and for different η regions, for tracks in $t\bar{t}$ events. The 2011 LHC pileup conditions has been considered. Solid (open) symbols correspond to the half-width of the 68% (90%) intervals centered on the mode of the distribution in residuals [124].

strips and pixel hits associated with the track, and the χ^2/dof). Secondly, these tracks are clustered on the basis of the z-coordinate of their closest approach distance with respect to the beam line. As third step, a fit is carried out on all the reconstructed tracks.

The track clustering is performed by the Deterministic Annealing (DA) algorithm [125], which allows the determination of the primary vertex on the basis of the process through which a thermodynamic system reaches its minimum energy. Each cluster of tracks assigned to a vertex is then fitted with an iterative weighted Kalman filter (the Adaptive Vertex Fitter [126]), which precisely

decrees the final primary-vertex position. Both the primary-vertex efficiency and resolution strongly depend on the number of tracks used in the fitting procedure. The primary-vertex reconstruction efficiency is close to 100% when more than two tracks are used to reconstruct the vertex and decrease to about 98% when only two tracks are used. For vertices with at least 50 associated tracks, the position resolution is estimated to be 10-12 µm in both the transverse and longitudinal plane. The resolution worsens to 100-150 µm in case of clusters with less than 10 tracks [124].

3.1.3 Energy reconstruction in the calorimeters

As already anticipated in Sections 2.2.4 and 2.2.5, both ECAL and HCAL are used to measure energies deposited by particles. In particular, the ECAL is crucial for electrons and photons reconstruction, but also to unambiguously detect and measure the energy and the direction of stable neutral and charged hadrons (given that hadronic showers can start in the ECAL volume). The HCAL, in conjunction with the ECAL, is crucial for the jet energy and the missing transverse energy measurements. In the ECAL, the energy measurement is performed through algorithms of clustering, which are different in barrel and in endcap. The goal of these algorithms is to collect all the energy deposits of a given electromagnetic shower and recover also the radiated energy. The general idea behind the different clustering algorithms is identifying the energy clusters, i.e. groups of adjacent ECAL cells around a cell with a local energy-deposit maximum (the seed). The group of energy deposits must be compatible with an electromagnetic shower caused by an electron or a photon. Subsequently, nearby clusters are collected into a *supercluster* (SC), which extends in the ϕ direction in order to recover the energy irradiated by bremsshtrahlung by electrons. More details about the different clustering algorithms are provided in [96]. In the HCAL, signals collected by the readout cells, arranged according to the same η-ϕ pattern of the crystals (cfr. Section 2.2.5), are added to the ECAL SC in order to build the so-called calorimeter *towers*. The energy associated with a single tower is calculated as the sum of all contributing readout cells which has satisfied a dedicated selection. The towers, representing the starting point for the purely calorimeter-based jets (calo-jets) reconstruction [127], are treated as massless particles with the energy corresponding to the tower energy and the direction given by the interaction point and the center of the tower itself.

3.1.4 Track reconstruction in the muon system

The reconstruction of the track in the muons system starts, as that in the tracker, with the computation of position measurements resulting from local reconstruction. The muon hits are identified in the DT, CSC and RPC systems using different techniques depending on the subdetector structures which have been previously described in Section 2.2.6. In the DT chambers, 1D-hits in individual tubes are then grouped into 2D-segments on either the r-ϕ or the r-z depending on the superlayer. 3D-segments are then built combining the 2D-segments reconstructed in a given full chamber. In the CSC subdetector, muon hits are computed exploiting the charge induced on the cathode strips and anode wires in each of the six layers of a chamber and then combined to provide the track segments. The muon hits in the RPC, despite being way less precise than those in the DT and CSC, are used in the track fit. They are especially useful in the overlap region between the barrel and the endcap.

3.2 Standard CMS reconstruction

The expression "standard CMS reconstruction" is used to indicate the reconstruction of more complex physics objects performed without the particle-flow algorithm [122], which will be described in Section 3.5. The standard reconstruction of electrons and muons only will be described here, as these collections have been used for the SM Zh analysis (cfr. Chapter 5). What distinguishes standard objects from particle-flow objects is the level of purity of the collections and the precision on the measurement of the object parameters. According to the standard reconstruction strategy, it may happen that some low-level detector information, such as energy deposits in the calorimeter, are used to reconstruct several candidates. The standard reconstruction supports a preliminary cross-cleaning between collections based on the ΔR distance (cfr. Equation 2.6), but if overlapping objects are found, none of them is removed and just an extra information regarding the overlap is added to the collection. Thus, the removal of these ambiguities is not tackled in depth in the reconstruction process, as it happens, instead, in the particle-flow algorithm, but it is left to the final user. Nothing prevents the analyzer, of course, from carefully cleaning all standard collections in the offline analysis process and, additionally, requiring that those object satisfy the particle-flow identification and isolation criteria (cfr. Section 3.6), as it has been done in the SM Zh analysis.

3.3 Electrons

Electrons trajectories are bent by the magnetic field, leaving hits in the inner tracker and depositing their energy in clusters of ECAL crystals. Therefore, the measurement of the energy deposits is crucial to reconstruct an electron. Approximately the 97% of the incident energy of a single electron is contained in a 5×5 cluster of crystals. However, the presence of material in front of the ECAL (corresponding to 1-2 X_0 depending on the η region) causes electron *bremsstrahlung* (which is enhanced with respect to muons by a factor m_e^2/m_μ^2) and, possibly subsequent photon conversion, leading to a spread of the radiated energy along ϕ due to the strong magnetic field. In order to recover this energy, matrices of crystals more extended along the ϕ direction (the superclusters) are built from clustering algorithms (cfr. Section 3.1.3).

The electron track reconstruction, mainly based on the standard steps followed for the track reconstruction in CMS (cfr. Section 3.1.1), is characterized by specific seeding algorithms and a dedicated track building and fitting procedure. The seeding strategy, optimized for high-p_T and isolated electrons is based on the ECAL-driven seeding. A relatively narrow ECAL supercluster (the extensions in ϕ and η are ± 0.3 and 0.09, respectively) is used to extrapolate the electron trajectory backward to the interaction point. Only those superclusters having a corresponding HCAL cluster with less that 15% of the total SC energy are used as seeds. If, in the extrapolation procedure, pixel hits are found matching the supercluster, the primary electron tracks reconstruction can start. A different and optimized seeding strategy has been developed for low-p_T and non-isolated electrons in the context of the particle-flow algorithm and it will be explained in Section 3.5.1.

The electron trajectories are reconstructed using a dedicated modeling of the electron energy loss, which includes the losses caused by bremsshtrahlung. This technique is called the Gaussian Sum Filter (GSF) method [128], and is a non-linear extension of the KF to cases characterized by non-gaussian energy loss distributions, as those related to bremsstrahlung emission and further conversion of secondary photons.

3.4 Muons

Muons are grouped in three categories: *standalone*, *global* and *tracker*.

The standalone muons are reconstructed using the segments built in the muon system only, as previously described in Section 3.1.4. The segments resulting from the local reconstruction are used to generate seeds, which are used by the Kalman algorithm for the track finding and fitting. The global muons are reconstructed combining tracks reconstructed in the inner tracking system (Section 3.1.1) with those reconstructed in the muon system (Section 3.1.4). A standalone muon track is extrapolated from the innermost layer of the muon system toward the outermost tracker surface in order to find a set of tracks compatible with the momentum and the position of the standalone-muon track. Once a subset of tracks compatible with the standalone-muon track is selected, a global refit algorithm acts combining iteratively the collection of tracker hits corresponding to each good track candidate with the collection of muon hits corresponding to the standalone-muon track. Given the lack of hits in at least two stations of the muon spectrometer, or the limited hit availability in just one DT/CSC station with an additional hit in the RPC, low momentum muons are likely reconstructed as tracker muons. In this case, the information of the tracker system is simply checked against the compatibility with segments in the DT or in the CSC subdetectors. Moreover, given that no fit of the track is performed, the momentum vector of a tracker muon is exactly the same as that of the track built in the tracker only.

All muon types are reconstructed with efficiencies between 97-99% for $5 < p_T < 500$ GeV. The large majority of muons (about 99%) used in the offline data analyses consists of global and tracker muons.

3.5 The particle-flow reconstruction

The particle-flow reconstruction algorithm [122] provides a global description of the event exploiting the redundancy of the subdetectors in order to achieve an optimal determination of the type, energy and direction of all stable particles in the event. The elements described in Section 3.1 are the main ingredients that the particle-flow algorithm uses to reconstruct and identify higher-level objects (particles). Practically, through the usage of a *linking algorithm*, tracks, clusters and muon tracks are linked into *blocks*, according to requirements on the η-ϕ distance between them. The particle-flow algorithm scans the content of a block to assign its elements univocally to a particle. Once an element is assigned to a particle, that element is removed from the block. The assignment of an element to a particle runs iteratively until no elements are left within a box. The granularity of the CMS detectors preserves from having boxes with a high multiplicity of elements, reducing the particle

reconstruction and identification dependency on the event complexity. Moreover, in terms of CMS subdetectors used in the particle-flow reconstruction process, the resolution power of the preshower (cfr. Section 2.2.4) is not used for the following motivation: when a charged hadron and a photon produce a single energy deposit in the ECAL endcap, the preshower detects the photon and measure its position, but it does not allow the fraction of the ECAL energy deposit to be assigned to the photon to be determined.

3.5.1 Electrons

As already anticipated in Section 3.3, the standard electron reconstruction, which exploits the ECAL cluster to seed the electron track finding, is optimized for energetic and isolated electrons, but not for low-p_T and/or non-isolated electrons. In the context of the particle-flow algorithm, the seeding efficiency for the latter class of electrons is increased through a pre-selection process at the basis of the tracker-driven seeding strategy. In case of no-bremsstrahlung emission, the momentum p of the electron track (precisely reconstructed with the KF algorithm, cfr. Section 3.1.1) is compared with the corresponding cluster energy E, by computing the ratio E/p. If E/p is approximately equal to unity, the track is selected to be a seed. In case of electrons radiating energy for bremsstrahlung, the KF procedure does not suit for properly reconstructing the electron track. The algorithm can stop following the electron trajectory, leading to a track with a small number of hits, or with a large χ^2, in case all the hits are collected. Therefore, the GSF algorithm [128] is used to build the track seeds taking into account these track peculiarities. The track-based seeding efficiency for non-isolated electrons with $p_T > 2$ GeV is improved by a factor two with respect to the ECAL-based seeding, while for isolated electrons the gain is between 15-50% for $2 < p_T < 5$ GeV and 1-2% for $p_T > 10$ GeV [129]. ECAL-based and tracker-based seeds are merged into a single collection from which the track reconstruction by the GSF method starts. The track obtained is then linked to a given calorimeter cluster if the extrapolated position from the outermost tracker measurement in the calorimeter is within the cluster boundaries. In the ECAL, the track is extrapolated to a depth corresponding to the expected maximum of the electron longitudinal shower extensions, while in the HCAL, up to one interaction length. In case of bremsstrahlung, the energy radiated is recovered by extrapolating tangents to the tracks towards ECAL. If, as just explained, the extrapolation falls within the cluster boundaries, that cluster is linked to the track as a potential bremsstrahlung photon. A cartoon of an electron radiating a photon is shown in Figure 3.2.

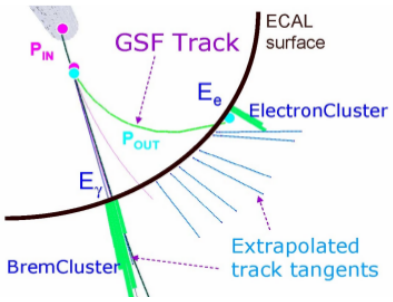

Figure 3.2: Visualization of an electron emitting a single Bremsstrahlung photon.

All the subdetector information, collected during the electron reconstruction, are exploited to build the final observables used for the particle-flow electron identification. These variables mainly refer to the amount of bremsstrahlung energy loss along the electron trajectory, the geometrical and momentum matching between the electron trajectory and associated clusters, and to the shape of the electromagnetic shower. All these observables are combined into a single discriminator ξ using a multivariate Boosted Decision Trees (BDT) method [130]. The distribution of the ξ discriminator, obtained using data and simulated $J/\psi \to e^+e^-$ events, is shown in Figure 3.3. Depending on the requirement on the discriminator ξ, different working points, with corresponding different efficiencies and fake rates, are defined. The working point $\xi > -0.1$ allows the 65% of electrons with $20 < p_T < 120$ GeV to be properly identified in simulated $t\bar{t}$ events with a negligible corresponding amount (1%) of misidentified pions [129].

As summarized in Table 3.1, depending on the ξ discriminator value required and on the p_T and η of the electron, three different working points are defined: very loose, loose and tight.

p_T (GeV) and η selection		Very Loose	Loose	Tight
$p_T < 20$	$\lvert\eta\rvert < 0.8$	$\xi > 0.50$	$\xi > 0.925$	-
	$0.8 < \lvert\eta\rvert < 1.479$	$\xi > 0.12$	$\xi > 0.915$	-
	$\lvert\eta\rvert > 1.479$	$\xi > 0.60$	$\xi > 0.965$	-
$p_T > 20$	$\lvert\eta\rvert < 0.8$	$\xi > 0.50$	$\xi > 0.905$	$\xi > 0.925$
	$0.8 < \lvert\eta\rvert < 1.479$	$\xi > 0.12$	$\xi > 0.955$	$\xi > 0.975$
	$\lvert\eta\rvert > 1.479$	$\xi > 0.60$	$\xi > 0.975$	$\xi > 0.985$

Table 3.1: BDT discriminator thresholds corresponding to the three working points used for particle-flow electron identification.

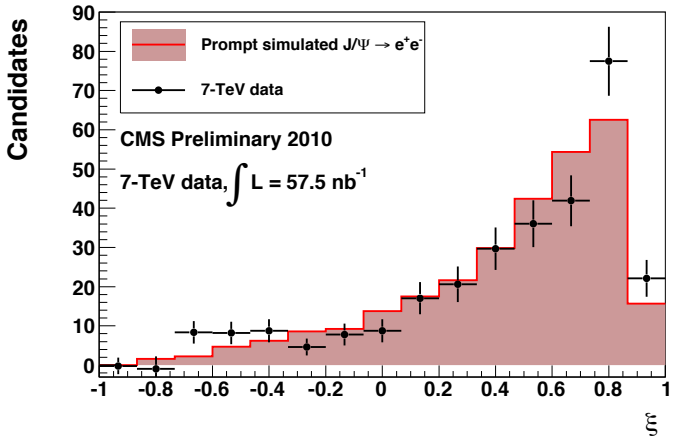

Figure 3.3: Distribution of the particle-flow electron identification discriminator ξ, obtained with a multivariate BDT method. The histogram obtained in simulated $J/\psi \to e^+e^-$ events (filled area) is compared with the one obtained using 2010 data (black points). Additional details can be found in [131].

3.5.2 Muons

Particle-flow muons are reconstructed starting from the method described in Section 3.4. More dedicated selection criteria are then applied to enhance the reconstruction efficiency of muons in jets. Additionally, given that a very energetic charged pion, passing through the hadronic calorimeter, might reach the innermost muon station and being identified as a muon, also the information of the calorimeters is exploited to keep the probability to identify charged hadrons as genuine muons small.

By definition, a particle-flow muon is a global muon if:

- it has at least one valid hit in the muon system
- the sum of the transverse energy of all its neighboring tracks and calorimeter cells within a cone of radius $R = 0.3$ around the muon direction is less than 10% of the muon transverse momentum.

The last requirement allows the contribution from charged hadrons to be highly suppressed, given the incompatibility between the measured track momentum and the energy deposits.

In order to identify genuine muons with high purity, two working points are defined for the particle-flow muon identification: the loose and the tight. According to the particle-flow tight identification criteria muons have to be strictly global and reconstructed as a particle-flow muon (see above). Additional requirements are imposed on the quality of the track reconstruction (in particular on the number of valid hits in the pixel and in the muon system, the number of tracker layers, the matched muon station and the value of the χ^2/dof related to the track fit) and on the compatibility with the primary vertex (requirements are imposed on d_{xy} and d_z, the transverse and the longitudinal impact parameters, respectively). The tight selection takes place before any other possible categorization, therefore, once performed, all the other reconstructed muons are recovered as loose muons. More specifically, a loose muon has to be reconstructed by the particle-flow algorithm and being labelled as global or tracker: standalone muons are rejected because they are typically measured with a poor momentum resolution. The detailed requirements corresponding to loose and tight particle-flow muon identification working points are listed in Table 3.2.

Requirement	Loose	Tight
Global	or Tracker	true
Particle-flow muon	true	true
χ^2/dof	-	<10
Valid pixel hits	-	>0
Tracker layers with measurement	-	>5
Valid muon hits	-	>0
Matched muon stations	-	>1
d_{xy} (cm)	-	0.2
d_z (cm)	-	0.5

Table 3.2: Muon identification criteria. The symbols d_{xy} and d_z indicate the transverse and the longitudinal impact parameter, respectively, of the track assigned to the muon with respect to the primary vertex.

3.5.3 Hadrons and photons

Once assigned to a particle-flow candidate, reconstructed tracks, calorimeter clusters, hits and segments in the muon chambers, are removed from the block. The remaining block elements are used to reconstruct possible hadrons (neutral or charged) and photons. The unambiguous reconstruction of neutral, charged hadrons and photons is fully based on the comparison between the momentum of the remaining tracks in the block with the calibrated en-

ergy[3] detected in the calorimeter clusters. If several tracks are linked to a single HCAL cluster, the cluster energy is compared with the sum of the momenta of all linked tracks. If, conversely, more than one HCAL cluster is linked to a single track, the comparison is performed between the track momentum and the energy deposited in the closest cluster to the track. The same strategy is applied for ECAL clusters. However, an additional control is needed in case several ECAL clusters are linked to one single track. Some clusters, in fact, might come from hadronic shower fluctuations or from photons overlapping with each other. In the former case the closest cluster-track link is preserved, in order to not double-count the hadron energy already considered in the HCAL cluster; in the latter one, the closest cluster-track link is ignored to allow the photon detection. The algorithm distinguishes one case from the other by ordering ECAL clusters according to the distance to the track: the farthest cluster is associated with shower fluctuations if the sum of the linked, possible existing, HCAL clusters and all the other ECAL clusters is smaller than the total track momentum. In case this condition is not satisfied, a particle-flow photon is created. As a result of the comparisons just explained, between the track(s) momentum and the cluster(s) energy, two situations can arise:

- **the total calibrated ECAL and HCAL energy is smaller than the total linked-track(s) momentum by at least a factor three the uncertainty related to the energy measurement:** the algorithm looks for additional particle-flow muons and fake tracks, by exploiting the redundancy of the measurements in the tracker and the calorimeters. Each track remaining after having removed those tracks identified as muons or fakes gives rise to a particle-flow charged hadron. The momentum of the charged hadron can be either equal to the total track(s) momentum, if the calorimetric energy is not compatible with the momentum value within the uncertainty, or, otherwise, it is obtained from a fit including track(s) and cluster(s).

- **the total calibrated ECAL and HCAL energy is larger than the total linked-track(s) momentum by more than one standard deviation of the cluster energy:** additional particle-flow neutral particles are created in this case. Photons are reconstructed with higher priority with respect to neutral hadrons, given that the number of these massless particles is expected to be a factor two higher than the number of neutral hadrons in jets. In particular, if the total ECAL cluster energy is smaller than the

[3] Some corrections are implemented for taking into account non-linear effects in the calorimeters [122].

difference between the total calorimetric energy and the track momentum (excess), the ECAL cluster energy is used to build a photon and the remaining part of the excess gives rise to a neutral hadron. In the opposite case (ECAL cluster energy greater than the excess), the energy excess gives rise to a photon and the remaining ECAL energy is interpreted as an early shower of the charged hadron. The remaining clusters not linked to any tracks give rise either to neutral hadrons or photons, depending on whether they are HCAL or ECAL clusters, respectively.

3.5.4 Jets

According to the most general definition, a jet is reconstructed starting from a set of objects and using an algorithm to cluster them. In the last few years, the most common algorithm chosen to perform the jet clustering, within the CMS collaboration, is the Anti-k_T algorithm [132]. This clustering algorithm is based on the following definition of the distance between two particles i and j:

$$d_{ij} = min\left(k_{ti}^{2p}, k_{tj}^{2p}\right)\frac{\Delta_{ij}^2}{R^2} \tag{3.1}$$

and the distance of each particle i from the so-called "beam"

$$d_{iB} = k_{ti}^{2p}, \tag{3.2}$$

where $\Delta_{ij}^2 = \Delta y_{ij}^2 + \Delta\phi_{ij}^2$ and k_{ti}, y_i and ϕ_i are the transverse momentum, the rapidity and azimuthal angle of the particle i, respectively. The parameter R is set to 0.5, while p, the parameter governing the relative power of the energy with respect to the geometrical scale Δ_{ij}, is equal to -1. The algorithm general behaviour tends to cluster soft particles with hard ones rather than clustering together soft particles (the distance d_{ij} between a hard and a soft particle is in general much smaller than the distance between two soft particles). If a hard particle has no hard neighbours within a distance $2R$, then the algorithm approximately accumulates all the soft particles within a circle of radius R, giving rise to a conical jet. If two hard particles are close to eachother, such that $R < \Delta_{ij} < 2R$, then two hard jets are created. In the case in which $\Delta_{ij} < R$, the two particles are clustered in one single jet. The shape of the jet is totally driven by the quantitative relationship between the transverse momenta of the particles: the more the particles are hard, the more they affect the jet shape. This feature makes this algorithm quite stable also in high-pileup environments. More details about the Anti-k_T algorithm can be found in [132].

As far as concern particle-flow jets, the input for the clustering is represented by all the particles reconstructed by the particle-flow algorithm (charged hadrons, photons, neutral hadrons, electrons and muons). By definition, then, the "raw" jet momentum is obtained by summing the momenta of all particles in the jet. The raw jet momentum is usually not calibrated in the sense that it does not match, on average, the momentum of the initiating parton. Such a behavior is due detector non-linearities, electronic noise, spurious particles and pileup contamination. Dedicated corrections are applied sequentially to calibrate the jet momentum. First of all, contamination from underlying and pileup events are subtracted using the FastJet algorithm [133]: starting from the median energy density obtained from all particle-flow jets reconstructed in the event with $p_T > 5$ GeV, the transverse energy density ρ is determined and an amount equal to ρ times the jet area is subtracted from the jet transverse momentum. As a second step, some variations in the jet energy response as a function of the pseudorapidity and the transverse momentum are flattened. Finally, residual differences between data and simulation are further removed by applying specific calibration factors to the simulated events [127].

Corrected particle-flow jets resolutions measured in dijet data are reported in Figure 3.4 in two different bins of pseudorapidity. Two expected resolutions are also shown on the plot: one computed without any data-to-simulation correction (red-dashed line) and one estimated including corrections (red-solid line). The latter represents the best estimate of the jet p_T resolution in data. Overall, data are found to be in good agreement with the corrected-expected results, within the statistical and systematic uncertainties. The jet energy resolution amounts typically to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV [127].

Concerning the identification, in order to separate real jets from detector noise, jets are required to satisfy a set of quality requirements based on the particle content of the jet. These criteria are summarized in Table 3.3.

Moreover, to cope with the large number of pileup interactions, a dedicated pileup-jet identification algorithm [134] has been deployed to separate hadronic jets and pileup jets. The algorithm exploits both the "vertex" and the "shape" information related to a given jet. In particular, in the tracker acceptance the algorithm exploits the possibility to whether associate charged hadrons with the primary vertex or not (vertex information). Outside the tracker acceptance, where the association charged hadrons-primary vertex cannot be used, the algorithm exploits the difference in the jet shapes as signal jets result to be more collimated than those from pileup (shape information). Vertex and shape information are combined through a multivariate BDT providing the pileup jet identification available for all jets used in CMS. The loose

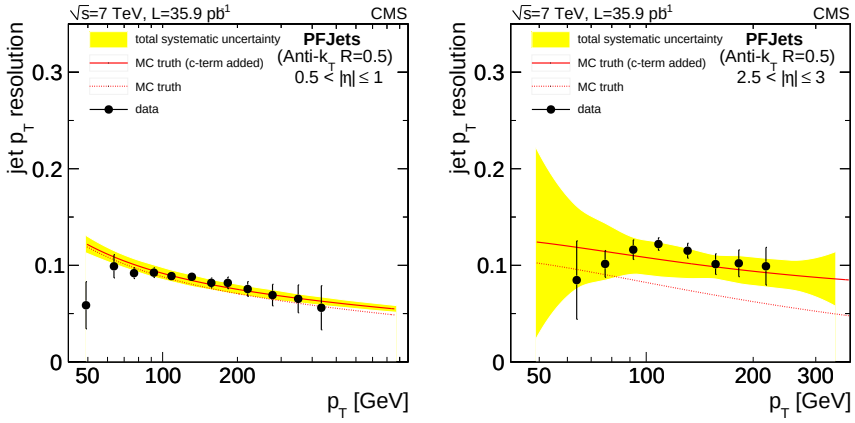

Figure 3.4: Energy resolutions measured in dijet data events (black dots) for Anti-k_T clustered jets obtained from particle-flow objects. The pseudorapidity regions considered, for sake of illustration, are: $0.5 < |\eta| < 1$ (left) and $2.5 < |\eta| < 3$ (right). The expected resolutions before (red-dashed line) and after (red-solid line) data-to-simulation corrections are also shown. The associated systematic-uncertainty bands are drawn in yellow [127].

working point has been used for both analyses presented in this thesis. As estimated in Z+jets events, this working point corresponds to a signal efficiency of about 99% reached for a background rejection of $90-95\%$ (85%) for $30 < p_T < 50$ GeV ($20 < p_T < 30$ GeV) in the barrel. In the endcaps, the probability to reject a pileup jet is of about 60% (40%) for $20 < p_T < 30$ GeV ($30 < p_T < 50$ GeV) at signal efficiencies of 90%. In the same p_T ranges, the background rejection remains the same also in the forward region, while the signal efficiency slightly decreases to 80% [134].

Jets resulting from heavy flavor quark production are critical features for many high energy processes and, therefore, those are further identified in CMS using the Combined Secondary Vertex (CSV) algorithm [135]. It is a sophisticated algorithm that combines information about the impact parameter significance, the secondary vertex and the jet kinematics to identify b-quark jets. These variables are used not only to identify jets originating from heavy quarks from those originating from light quarks and gluons, but also to discriminate between b jets and c jets. Several working points for the CSV output discriminant are available: the loose (CSVL), medium (CSVM), and tight (CSVT) operating points are defined as the CSV values such that the misidentification probability for light-parton jets is close to 10%, 1%, and 0.1%, respectively, at jet p_T of about 80 GeV, as measured independently in QCD-multijet and

Requirements	
$\|\eta\| < 4.7$	
Number of constituents	> 1
Neutral hadron energy fraction	< 0.99
γ energy fraction	< 0.99
Muon energy fraction	< 0.8 (for 2012 analysis only)
Additional cuts applied for $\|\eta\| < 2.4$	
Number of charged constituents	> 0
Charged hadron energy fraction	> 0
Electron energy fraction	< 0.99

Table 3.3: Jet quality requirements used to discriminate real jets from detector noise.

$t\bar{t}$ events. In correspondence of the CSVM (CSV$>$ 0.697), the working point used in the analyses presented in this thesis, the efficiencies to tag b quarks, c quarks, and light quarks/gluons are approximately 65% (as can be observed from Figure 3.5 in which the b-tag efficiency, measured in $t\bar{t}$ events in data and compared with simulation, as a function of the CSV discriminator value CSV is shown), 15%, and 1%, respectively [135].

3.5.5 Missing transverse energy

As mentioned earlier in this thesis, neutrinos are detectable only indirectly, through the balance of the total transverse energy. In the context of the particle-flow algorithm, the raw missing transverse energy, defined with Equation 3.3, is the negative of the vector sum of the transverse momenta of all particles reconstructed in the event:

$$\overrightarrow{\not{E}_T} = -\sum_{i=1}^{N_{PF}} \vec{p}_{T,i} \, . \tag{3.3}$$

As seen for jets, also for the particle-flow missing transverse energy $\not{E}_T$ some corrections are needed in order to get rid of non-linear detector response effects, inefficiencies, thresholds in the calorimeter clustering. The absolute measurement bias is reduced by including all the jet energy corrections (JEC) already described in Section 3.5.4. Relative to the missing transverse energy, this translates into replacing the vector sum of the transverse momenta of the particles clustered as a jet with the vector sum of the transverse momenta of the corrected jets:

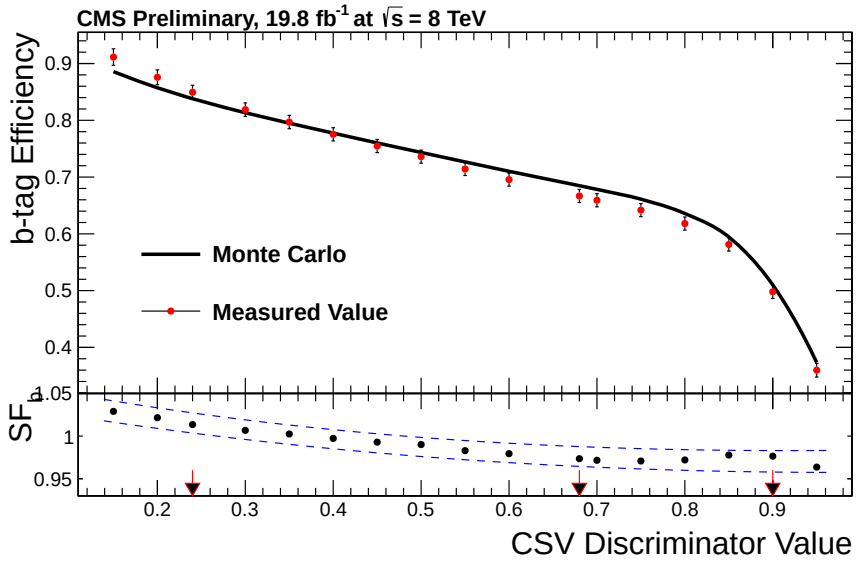

Figure 3.5: b-tagging efficiency measured in $t\bar{t}$ events in data (red dots) compared to the expectations from simulations (solid black line), as function of the CSV b-tagging discriminator. The efficiency-ratio SF is shown in the lower panel with the corresponding uncertainty band (dashed blue lines). The three arrows indicate the loose, medium and tight working points, corresponding to approximately to mistagging rates of 10%, 1% and 0.1% [135].

$$\vec{\not{E}}_{\mathrm{T}}^{\,\mathrm{Type-I}} = \vec{\not{E}}_{\mathrm{T}} + \sum_{\mathrm{jets}} \left(\vec{p}_{\,\mathrm{T,jet}} - \vec{p}^{\,\mathrm{JEC}}_{\mathrm{T,jet}} \right) . \tag{3.4}$$

In the ideal case, the $\vec{\not{E}}_{\mathrm{T}}$ would not degrade with the increase of the pileup interactions, as these events are characterized by a small amount of invisible particles. However this is not the case in the reality and, therefore, an additional correction is implemented to take into account the pileup degradation of the missing energy. This correction removes all the charged hadrons not originating from the primary vertex but also the neutral pileup particles. However, estimating the pileup neutral component is not easy. For this reason, under the assumption of no missing energy characterizing pileup events at true level, the neutral and charged pileup contributions are assumed to be exactly equal, and opposite in direction. Additionally, to take into account

some ϕ asymmetries, which cause a shift in the $\vec{\not\!\!E}_T$ components along the x and y detector axes, further corrections are applied [136].

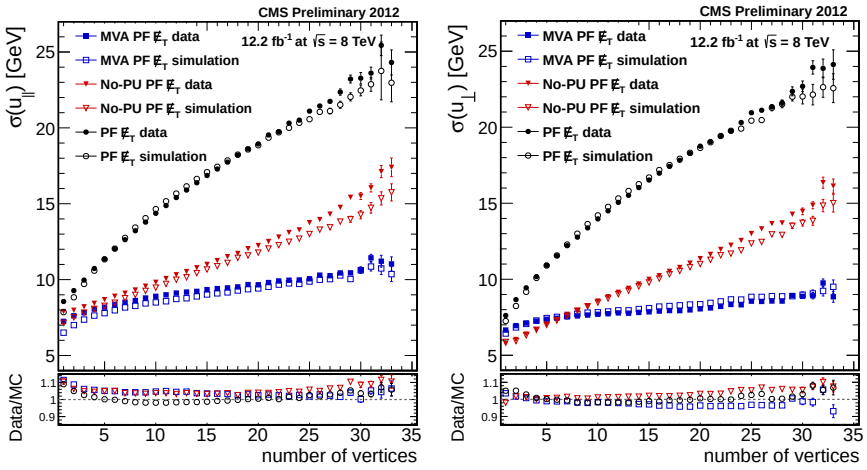

Figure 3.6: Resolution of the reconstructed recoil projected on the axis parallel (left) and perpendicular (right) to the Z boson direction, computed as a function of the number of vertices, in $Z \to \mu\mu$ simulated events. Results obtained from validation in 2012 data ($\sqrt{s} = 8$ TeV and $\mathcal{L} = 12.2$ fb^{-1}) are also shown [137].

Two algorithms have been specifically developed in 2012 to cope with a larger number of pileup interactions: the no pileup and the multivariate particle-flow missing transverse energy algorithms [136], indicated as No-PU PF $\not\!\!E_T$ and MVA PF $\not\!\!E_T$, respectively. The former is based on the computation of the transverse momentum imbalance by separately weighting contributions from the main interaction and pileup particles. The latter, instead, is based on a set of multivariate regressions that provides an improved measurement of the missing energy in the presence of a high number of pileup interactions. Computed in $Z \to \mu\mu$ simulated events and validated in 2012 data, the resolution of the reconstructed recoil projected on the axis parallel and perpendicular to the Z boson direction is reported as a function of the number of vertices for the PF, No-PU PF and MVA PF $\not\!\!E_T$ in Figure 3.6. It can be observed the smallest resolution of the MVA PF $\not\!\!E_T$, for a number of vertices up to thirty-five and the less dependency to pileup. In correspondence of twenty vertices, which is the average number of pileup vertices in 2012 p-p collisions, the purely particle-flow based resolutions are approximately a factor two higher than the MVA ones. Further details about the particle-flow missing energy reconstruction algorithms and their performance can be found in [138, 136].

3.6 Lepton isolation

Before concluding the chapter, it is important to introduce and describe a crucial variable for lepton identification according to the particle-flow paradigm: the lepton isolation. The particle-flow lepton isolation is computed from the flux of particle-flow candidates found within a cone of radius $\Delta R = 0.4$ ($\Delta R = 0.5$ in case of tau leptons, cfr. Section 4.2.2), built around the lepton direction.When dealing with electron candidates, the neutral flux is corrected by using the average energy density due to pileup and underlying events in the central region of the detector (ρ) and an effective area ($\mathcal{A}_{\text{eff}}$) correction which normalizes this estimator in such a way that the isolation is independent of the number of pileup interactions.

The absolute lepton isolation is defined as:

$$I^{PF}_{abs} = \sum_{CH} p_T + \max\left(\sum_{NH} p_T + \sum_{\gamma} p_T - 0.5 \cdot \sum_{CH,PU} p_T, 0\right) \tag{3.5}$$

In this formula, $\sum_{CH} p_T$ is the scalar sum of the transverse momenta of the charged hadrons, electrons and muons originating from the primary vertex and located in a cone of size $\Delta R = 0.4$ ($\Delta R = 0.5$ in case of tau leptons) centered on the lepton direction. The sums $\sum_{NH} p_T$ and $\sum_{\gamma} p_T$ represent the corresponding quantities for neutral hadrons and photons, respectively. The term $\sum_{CH,PU} p_T$ is the contribution of charged hadrons from pileup vertices in the cone and it appears multiplied by a factor 0.5, which is roughly equal to the ratio between the amount of neutral to charge hadrons. The entire term $0.5 \cdot \sum_{CH,PU} p_T$, then, represents an estimate of the contributions from neutral particles arising from pileup events. The subtraction of this contribution is called "$\Delta\beta$ correction". This correction is built starting from the p_T sum of charged particles within a maximal distance of $\Delta R = 0.4$ ($\Delta R = 0.8$ for taus) from the lepton direction, associated with a vertex other than the primary one. Therefore, by construction, the $\Delta\beta$ correction relates the energy deposit measured from charged pileup particles (e.g. $\pi^{\pm}$, $K^{\pm}$) to the expected flow of energy arising from neutral pileup particles (e.g. π^0, K^0_L). The relative isolation is defined as $I^{PF}_{rel} = I^{PF}_{abs}/p_T$, where p_T is the transverse momentum of the considered lepton.

4

Tau reconstruction in CMS

Hadronically decaying tau leptons are fundamental for the analyses presented in this thesis and for many other physics analyses at the LHC, both in the context of the SM and beyond. This chapter is totally devoted to the description of the reconstruction and identification of hadronic taus in CMS. Particular emphasis will be put on the performance of the algorithm used, which have been obtained as part of the doctoral work presented in this thesis. An overview on the algorithm validation performed using 2012 data, and the measurement of the tau energy scale, will be provided at the end of the chapter. All the methods and results reported in this chapter are documented in the paper [25]*, whose publication in the Journal of Instrumentation (JINST) is expected by the end of the year.*

4.1 Tau physics

Having a mass $m_\tau = 1.777$ GeV [29], the tau is the heaviest lepton among the three lepton families and it is the only lepton decaying hadronically via weak interactions. The tau lifetime, multiplied by c, is $c\tau \approx 87$ μm, which is approximately a factor m_μ^5/m_τ^5 shorter than the muon one. A tau lepton can decay fully-leptonically to a lighter lepton plus two neutrinos with a $\mathrm{BR} \approx 35\%$ and to one or more hadrons and neutrinos with a $\mathrm{BR} \approx 65\%$. In the former case,

electrons and muons originating from tau decays are reconstructed and identified according to the algorithms already described in Chapter 3, given the difficulty of distinguishing them from electrons and muons coming directly from the p-p interaction vertex. In the latter case, instead, a dedicated algorithm, which will be described in the next section, is used to reconstruct individual hadronic tau decay modes. The main decay channels of the tau lepton τ^- are summarized in Table 4.1.

Decay channel	BR [%]	Resonance	
		Type	Mass (MeV)
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$	17.8		
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$	17.4		
$\tau^- \to h^- \nu_\tau$	11.5		
$\tau^- \to h^- \pi^0 \nu_\tau$	26.0	ρ^-	770
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	10.8	a_1^-	1260
$\tau^- \to h^- h^+ h^- \nu_\tau$	9.8	a_1^-	1260
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$	4.8		
Other hadronic modes	1.8		
All hadronic modes	64.8		

Table 4.1: Main decay modes for the τ^-. The generic symbol $h^\pm$ indicates a charged meson (either a pion or a kaon). The decays $\tau^- \to h^- \pi^0 \nu_\tau$, $\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$ and $\tau^- \to h^- h^+ h^- \nu_\tau$ proceed via intermediate vector meson resonances [29]. All decay modes are still valid under charge conjugation.

It has to be specified that, despite kaons are allowed in the hadronic tau decays, the decays to pions are enhanced with respect to those to kaons, which are, instead, Cabibbo-suppressed [139]: $BR(\tau^- \to \tau^- \nu_\tau) \approx 14 \cdot BR(\tau^- \to K^- \nu_\tau)$ [29].

4.2 The Hadron Plus Strips algorithm

The energy with which taus are produced in the laboratory frame at hadron colliders far exceeds the value of the tau mass itself: for this reason, in tau hadronic decays, the final-state hadrons assume the topology of a collimated bunch of particles, resembling jets initiated by gluons or quarks. As a consequence, the main challenge in identifying hadronic tau decays is the separation from quark/gluon jets, whose cross section production, additionally, overwhelms by almost eight orders of magnitude the tau production at the

LHC. The topological difference between quark/gluon jets and the hadronic taus (for which the symbol τ_h will be adopted in the text) is used to reduce such a background. Typically, quarks and gluons jets are characterized by a high multiplicity of soft hadrons. Therefore, a fake tau arising from such a jet will be likely surrounded by a large number of hadrons. Quark/gluon jets fragmenting into a few hard hadrons, are challenging, or rather impossible, to be distinguished from hadronically decaying taus. Moreover, reducing the probability of misidentifying electrons or muons as hadronic tau candidates is also crucial for lots of analyses at the LHC. The Hadron Plus Strips (HPS) [24] is the CMS algorithm devoted to achieve these goals. The HPS algorithm is seeded by particle-flow jets (cfr. Section 3.5.4). Subsequently, for each jet, the particle content is analyzed: charged particles are combined with π^0s (if any), which are in turn reconstructed from photons (strips, cfr. Section 4.2.1). Once tau candidates are built, the final tau identification is performed using also dedicated discriminators which allow the jet/$e/\mu \to \tau_h$ misidentification rates to be reduced.

4.2.1 Decay mode finding

Considering the main hadronic tau decay modes summarized in Table 4.1, it can be observed that, on top of the charged-particle reconstruction, the reconstruction of the neutral pions through the $\pi^0 \to \gamma\gamma$ decay is crucial. The probability that at least a photon converts in e^+e^- pairs, within the volume of the tracking detector (1.5 X_0), only depends on the traversing material for photons with energy exceeding 0.5 GeV [29]. The footprints of converted photons in the ECAL lead to strips of size 0.05 in η and 0.20 in ϕ. The elongation is broader in the ϕ direction to accommodate soft electrons, from photon conversion, whose tracks are strongly deviated by the magnetic field before reaching the ECAL surface. Electrons and photons of $p_T > 0.5$ GeV and $E_T > 0.5$ GeV, respectively, are used to build strips. Strips containing one or more electron or photon constituents, whose sum of transverse momenta exceed 2.5 GeV, are kept as π^0 candidates. The charged particles included in the tau reconstruction are required to have $p_T > 0.5$ GeV. Selection criteria are also applied on the number of hits released by these charged particles in the tracker ($N_{hit} \geqslant 3$) and on the quality of the track fit ($\chi^2 < 100$). Moreover, in order to significantly reduce the effect of the pileup and the contribution from spurious tracks, the longitudinal and impact parameters of the charged-particle track with respect

to the hypothetical production vertex of the τ_h[1], are required to be less than 0.2 cm and 0.03 cm, respectively.

Strips and charged particles inside the jet are combined to build the hadronic tau candidate according to four possible categories:

1. one charged particle;
2. one charged particle plus one strip;
3. one charged particle plus two strips;
4. three charged particles.

These categories are designed to account for about 75% of the tau width into hadrons, targeting the decay modes $h^{\pm}$, $h^{\pm}\pi^0$ and $h^{\pm}h^{\mp}h^{\pm}$. Despite not explicitly searched for, the $h^{\pm}\pi^0\pi^0$ might still accidentally be reconstructed in one of the previous categories: for this reason, in the following, the general notation $h^{\pm}\pi^0$s will be used.

A given tau candidate is reconstructed according to a combinatorial approach: for each jet, multiple different τ_h decay mode hypotheses are tested before choosing the final one. The four-momentum (p_T, η, ϕ and mass) of each τ_h candidate hypothesis is given by the vectorial sum of the momenta of charged particles and strips that are included in the respective decay mode combination. Given that no neutrinos are included in the reconstructed-tau four momentum, the kinematics of the tau candidate account only for the visible final-state particles of the hadronic decay. In case any one of the charged particles included in the τ_h candidate hypothesis are identified as electrons or muons by the particle-flow algorithm, the mass of the particle is set equal to the mass of the charged pion [29].

Dedicated selections are imposed on the reconstructed tau candidate mass in order to check the compatibility of each hypothesis with the corresponding expected signature (cfr. Table 4.1). For the first decay mode no selections are imposed, since the tau candidate consists of one single visible particle. For the second, pre-requirements are applied to the strip mass first: it has to be compatible with the π^0 mass (i.e. to be in the range $[50, 200]$ GeV) if the strip consists of more than one electromagnetic particle, otherwise the nominal π^0 mass value [29] is assigned to the strip four-momentum. Subsequently, the mass of the charged particle-strip pair is required to be in the range $[0.3, 1.3 \cdot$

[1]The hypothetical production vertex of the τ_h is taken to be the vertex closest to the highest p_T charged particle within the jet.

$\sqrt{p_T[\,\mathrm{GeV}]/100}]$ GeV, thus compatible with the resonance $\rho(770)$. For taus with $p_T < 100$ GeV ($p_T > 1044$ GeV) the upper limit considered on the mass window is 1.3 GeV (4.2 GeV). In the third decay mode, the mass of the two strips is required to be compatible with the π^0 one and the mass of the final system (charged particle plus two strips) to be in the range $[0.4, 1.2 \cdot \sqrt{p_T[\,\mathrm{GeV}]/100}]$ GeV (compatibility with the $a_1(1260)$ signature). In particular, the upper limit on the mass window is set to 1.2 GeV (4.0 GeV) for taus with $p_T < 100$ GeV ($p_T > 1111$ GeV). The mass of the three charged particles has to be in the range $[0.8, 1.5]$ GeV, for the fourth category. It can be noticed that, despite being characterized by the same intermediate resonance, the mass window is enlarged in the third category with respect to the fourth one: this to take into account resolution effects that affect the reconstruction of high-p_T taus.

Decay mode hypotheses that fail the requirements imposed on the mass window are discarded. Strips and charged particles are also required to fall within the so-called "signal cone" centered on the τ_h direction defined by the leading track of the τ_h signal constituents. The signal cone has a dynamical radius depending on the tau candidate momentum p_T:

$$\Delta R = \min\left[\max\left[\frac{3.0(\mathrm{GeV})}{p_T}, 0.05\right], 0.10\right] . \tag{4.1}$$

The multiplicity of possible hadrons plus strips combinations satisfying all the requirements is solved by retaining only one single tau candidate, the one with the highest p_T. Any other electromagnetic or charged particle found outside the signal cone and within a cone of size $\Delta R = 0.5$ (the "isolation cone"), centered on the τ_h direction, is treated as an isolation particle. The way in which isolation particles are used to discriminate hadronic taus from QCD jets is the subject of the following section.

4.2.2 Tau isolation discriminators

As anticipated in Section 4.2, a tau reconstructed from a quark/gluon jet has a high probability to be surrounded by additional soft hadrons. Therefore, the isolation becomes a crucial variable to distinguish these kind of jets from hadronically decaying taus. Two types of tau isolation discriminators have been developed, based on the "cut-based" and the MVA approach.

CUT-BASED The cut-based isolation is computed by summing the transverse momenta of charged particles of $p_T > 0.5$ GeV and photons of $E_T > 0.5$ GeV

reconstructed with the particle-flow algorithm, within an isolation cone of size $\Delta R = 0.5$, centered on the τ_h direction. The tau absolute isolation I^{τ}_{abs} is defined as in Equation 3.5, to be read considering the exclusion of all the photons used to reconstruct strips and all charged particles used to build the τ_h candidate from the p_T-sum. As already described in Section 3.6, the $\Delta\beta$ correction for taus is computed by summing the transverse momenta of charged particles that are within a cone of size $\Delta R = 0.8$ around the τ_h direction, with the requirement that the longitudinal distance between each of these tracks and the τ_h production vertex has to be greater than 0.2 cm. Given the requirement of at least three hits per charged track (cfr. Section 4.2.1) and the $\Delta\beta$ correction applied (cfr. Section 3.6), cut-based isolation discriminators are in jargon called "HPS $\Delta\beta$ 3-hits" isolation discriminators[2]. Three working points are defined for the cut-based isolation: loose ($I^{\tau}_{abs} < 2.0$ GeV), medium ($I^{\tau}_{abs} < 1.0$ GeV) and tight ($I^{\tau}_{abs} < 0.8$ GeV).

MVA-BASED The most updated MVA-based isolation mainly exploits the fact that tau lepton has a small lifetime[3], but still significant with respect to the CMS transverse impact parameter and secondary vertex resolutions [10]. The observables used as input variables to train the BDT used in this context are:

- the charged and the neutral particle isolation p_T-sums, treated as separated inputs;
- the reconstructed decay mode;
- the transverse impact parameter d_0 of the leading track of the τ_h candidate and its significance d_0/σ_{d_0} (where σ_{d_0} is the uncertainty associated with the d_0);
- the distance between tau production and decay vertices $|\vec{r}_{SV} - \vec{r}_{PV}|$ and its significance $|\vec{r}_{SV} - \vec{r}_{PV}|/\sigma_{|\vec{r}_{SV}-\vec{r}_{PV}|}$. The tau production vertex and its decay vertex are reconstructed with the Adaptive Vertex Fitting algorithm [126] (the position of the tau production vertex is refitted excluding the tracks related to the τ_h candidate);

[2] An older version of the isolation discriminator is the "HPS $\Delta\beta$ 8-hits", which is based on the requirement of having at least eight hits per charged track.

[3] The MVA isolation discriminator based on tau lifetime information represents the most updated version of the MVA-based isolation discriminator developed during the LHC Run I. However, at the moment of the work done in the context of the analyses presented in this thesis, the legacy MVA-based isolation discriminator recommended by CMS was a version not yet including the lifetime information. A quantitative comparisons between MVA isolation discriminator including and not-including the lifetime information is provided in Figure 4.11.

- a flag to indicate the successfully reconstruction of the τ_h decay vertex.

Different working points are defined depending on the requirement imposed on the BDT output variable. This requirement is properly tuned in order to make the tau identification efficiency almost independent of the hadronic tau p_T.

4.2.3 Light lepton rejection discriminators

Electrons and muons originating from W and Z bosons decays have high probability to satisfy isolation requirements and they might be misidentified as hadronically decaying taus. Therefore, dedicated discriminators to reduce the electron and muon misidentification rates have been developed. In particular, both electrons and muons can be reconstructed in the $h^{\pm}$ decay mode and, additionally, an electron can also be reconstructed in the $h^{\pm}\pi^0$ decay mode in case it radiates a bremsstrahlung photon that subsequently converts. The HPS algorithm includes cut-based and MVA-based discriminators for both electrons and muons rejection.

CUT-BASED anti-*e* DISCRIMINATOR It is just for historical reasons that this discriminator is referred to as the anti-*e* cut-based discriminator. In reality, in fact, it consists in a "cut" applied on the output of the BDT that has been trained to separate electrons from charged pions in the particle-flow algorithm (cfr. Section 3.5.1). Three working points are provided: the loose, the medium and the tight. The loose (medium) working point corresponds to a particle-flow electron MVA discriminator value smaller than 0.6 (-0.1). Concerning the tight working point, τ_h candidates pass this discriminator in case they satisfy the anti-*e* medium requirement and in case the electron is not reconstructed near the boundary of adjacent ECAL barrel modules ($|\eta| < 0.018$, $0.423 < |\eta| < 0.461$, $0.770 < |\eta| < 0.806$, $1.127 < |\eta| < 1.163$) or within the transition region between ECAL barrel and endcap ($1.460 < |\eta| < 1.558$). The identification of electrons entering these regions is difficult to be performed due to inactive detector material, which distorts the shower shape variables used by the particle-flow electron MVA algorithm.

MVA-BASED anti-*e* DISCRIMINATOR The cut-based anti-*e* discriminator is fully based on the particle-flow electron MVA identification, which is optimized to reject charged pions. Therefore, a more optimal discriminator, has been developed to be used in the context of the tau identification. The discriminating variables used to train a BDT are:

- the ratio between the energy deposits in the ECAL and the sum of the ECAL and HCAL energy deposits, for both the charged particles and photons constituting the reconstructed τ_h candidate;
- the ratio between the energy deposits in the ECAL and the reconstructed momentum, both related to the highest-p_T charged particle of the τ_h candidate. The corresponding quantity measured in HCAL is also considered;
- the root-mean-square of the distances, in η and ϕ, between the photons and the leading charged particle of the tau signal cone. These two quantities are reweighted by the scalar sum of the considered photons p_T;
- the fraction of the τ_h energies carried by photons;
- the difference between the electron GSF track momentum measured by the curvature of the track at the innermost and outermost position, divided by the one measured at the innermost position;
- the ratio between the total ECAL energy and the inner track momentum;
- the ratio between the bremsstrahlung photon energy measured in the ECAL and in the tracker;
- the mass of the tau candidate;
- $(N_{hits}^{GSF} - N_{hits}^{KF})/(N_{hits}^{GSF} + N_{hits}^{KF})$, where N_{hits}^{GSF} (N_{hits}^{KF}) is the number of inner tracker hits related to the track reconstructed by the GSF (Kalman filter) algorithm (cfr. Sections 3.1.1, 3.3 and 3.5.1).

Four working points are defined (loose, medium, tight and very tight) by varying the cut on the BDT outcome.

CUT-BASED anti-μ discriminator Two working points, loose and tight, are provided for the anti-μ discriminator. The sufficient condition to not satisfy the loose one is either finding track segments in at least two muon stations within a cone of size $\Delta R = 0.5$ centered on the τ_h direction, or having a leading track in the tau signal cone with associated calorimetric energy deposits (ECAL plus HCAL) smaller than the 2% of the track momentum itself. If a tau candidate satisfies the loose working point requirements and no muon hits within a cone of size $\Delta R = 0.5$ around the τ_h direction are found in the CSC, DT or RPC chambers located in the two outermost muon stations, such a candidate passes also the tight working point.

MVA anti-μ discriminator This discriminator is the output of a BDT. The observables used as inputs to the BDT to separate hadronic taus from muons are: the calorimeter energy associated with the leading charged particle of the τ_h candidate, the calorimeter energy associated with any charged particle or photon constituting the τ_h candidate, the fraction of the hadronic-tau p_T carried by the leading charged particle in the signal cone, the number of track segments in the muon system reconstructed within a cone of size $\Delta R = 0.5$ around the τ_h direction and the number of muon stations with at least one hit detected within a cone of size $\Delta R = 0.5$ centered on the τ_h direction. Loose, medium and tight working points are defined for this discriminator.

4.3 Expected performance

The expected performance of the HPS tau identification algorithm have been studied in terms of p_T, η and N_{vtx} (the multiplicity of reconstructed primary vertices in an event is used as a measure of pileup) using simulated events. The simulated samples used for this study are given in Table. 4.2.

Dataset Description	Generator
$Z/\gamma^* \to LL$ ($L = e, \mu, \tau$)	MADGRAPH [115]
$Z' \to \tau\tau$	PYTHIA [114]
W+jets	MADGRAPH
QCD-multijets	PYTHIA

Table 4.2: Simulated samples used to study the expected tau identification performance listed with the corresponding MC generators. PYTHIA has been also used for the parton showering and hadronization, and TAUOLA [119] for simulating tau decays. Pileup conditions characterizing the 2012 data-taking have been also generated with PYTHIA and added to all simulated samples.

Tau identification efficiencies and misidentification rates for electrons and muons have been computed using $Z/\gamma^* \to LL$ (with $M_{LL} > 50$ GeV) simulated events. The estimate of the tau identification efficiency has been performed also at higher p_T exploiting $Z' \to \tau\tau$ (with the generator-level Z' mass equal to 2.5 TeV) events. Because the jet fragmentation functions depend on the initiating parton (light/heavy quark or gluon), the fake rate from jets is flavor dependent. For this reason, given the different compositions of quarks and gluon jets in the two samples, both W+jets and QCD-multijets events (the latter produced considering a flat p_T distribution in the range between 15 and

3000 GeV) have been used for the computation of the jet$\rightarrow \tau_h$ misidentification rates. In order to remove the contamination from real electrons, muons and taus from W decays, those leptons have been removed from the jet before proceeding with the τ_h reconstruction in the selection of W+jets events.

To better understand the difference between taus and jets kinematics, crucial for the expected performance studies, the reader can look at the transverse momentum distributions of the visible tau decay products at generator-level, obtained in simulated $Z/\gamma^* \rightarrow \tau\tau$ and $Z' \rightarrow \tau\tau$ events, and shown in Figure 4.1. The p_T distributions of reconstructed jets in QCD-multijets and W+jets events are also reported in Figure 4.1.

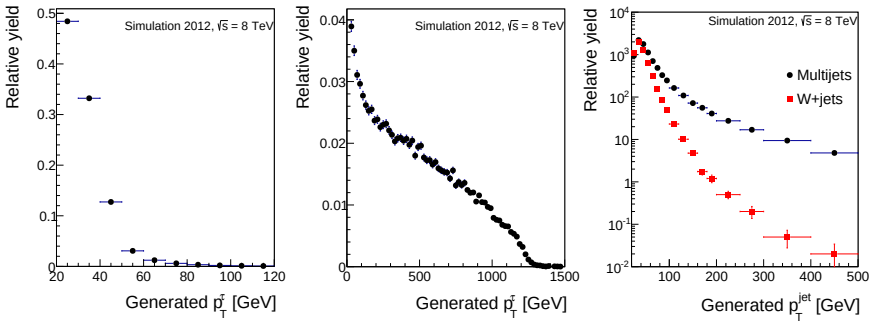

Figure 4.1: Transverse momentum distributions of the visible decay products of hadronic tau decays at generator-level in simulated $Z/\gamma^* \rightarrow \tau\tau$ (left) and $Z' \rightarrow \tau\tau$ (middle) events. The p_T distributions of reconstructed jets in QCD-multijets (black) and W+jets (red) events are also shown (right). No true electrons, muons and taus have been included in the reconstructed jet for W+jets events [25] [A].

The performance of the decay mode finding algorithm is studied by quantifying the correlation between reconstructed and generator-level decay modes, as represented in Figure 4.2 (left) in the form of a 3×3 matrix. Reconstructed taus are required to match, in a given decay mode, an hadronic tau at the generator-level within a cone of size $\Delta R < 0.3$, for which the visible decay products satisfy $p_T > 20$ GeV and $|\eta| < 2.3$. Additionally, the reconstructed candidates have to satisfy the loose working point of the cut-based isolation discriminator. Each entry (i,j) of the matrix represents the fraction of taus generated in the decay channel i (where i can be equal to one of the three decay modes considered in the HPS algorithm, $h^\pm$, $h^\pm\pi^0$s and $h^\pm h^\mp h^\pm$) and reconstructed in the decay mode j (where j can correspond to one charged hadron, one charge hadron plus strip(s) and three charged hadrons). The diagonal entries exceed 80%, reaching almost 100% for the $h^\pm h^\mp h^\pm$ decay mode. These

results confirm that the algorithm reconstructs the correct decay mode for the bulk of the events. Additionally, Figure 4.2 (right) demonstrates that the true tau decay mode is reconstructed almost independently of the number of reconstructed vertices (N_{vtx}), therefore of pileup conditions. Concerning the $h^{\pm}$ decay mode, the few percent decrease observed in the probability is due to events in which particles from pileup, which release energy in the ECAL near the tau, are reconstructed as π^0s, causing the tau to be reconstructed in the $h^{\pm}\pi^0$s decay mode.

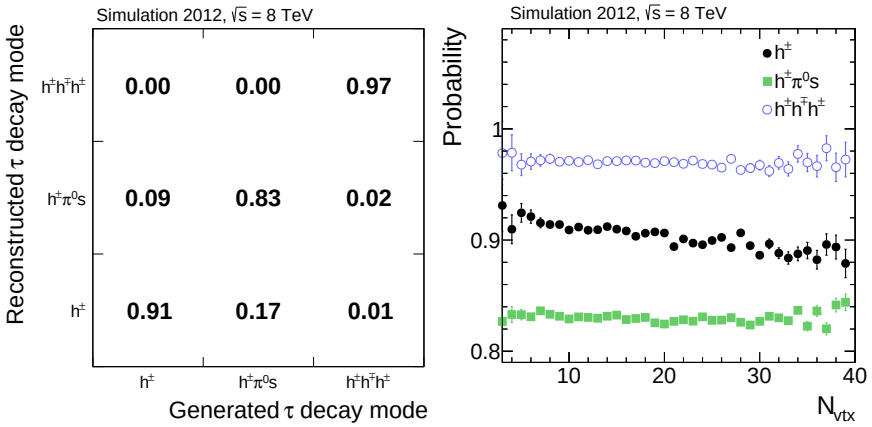

Figure 4.2: Correlation between generated and reconstructed decay mode for hadronic tau decay in the simulated $Z/\gamma^* \to \tau\tau$ events (left). Probability to correctly reconstruct the tau in the decay mode it was generated in as a function of the pileup (right). Reconstructed τ_h candidates are required to match, in a given decay mode, an hadronic tau at the generator-level within a cone of size $\Delta R < 0.3$, for which the visible decay products satisfy $p_T > 20$ GeV, $|\eta| < 2.3$ and the loose working point of the cut-based isolation discriminator [25] [A].

The performance of the HPS algorithm is also studied in terms of energy *response* and *relative resolution*, defined as the mean and the standard deviation, respectively, of the distribution of the ratio between the reconstructed and the generator-level momentum of the visible tau decay products. These distributions, which have been estimated in simulated $Z' \to \tau\tau$ events for different pileup conditions ($N_{vtx} \leqslant 12$, $13 \leqslant N_{vtx} \leqslant 17$ and $N_{vtx} \geqslant 18$) and decay modes ($h^{\pm}$, $h^{\pm}\pi^0$s, $h^{\pm}h^{\mp}h^{\pm}$), are shown in Figure 4.3. Also in this case, reconstructed taus are required to satisfy the loose working point of the cut-based isolation discriminator. The p_T^{rec}/p_T^{gen} distributions result to be negligibly affected by the pileup and are more populated in the region where $p_T^{rec}/p_T^{gen} < 1$. The conclusion drawn is that the majority of the events is characterized by either

reconstructed particle-flow particles falling outside of the tau signal cone, or by reconstructed tau candidates that fail the requirements on the tau mass. As a consequence, the energy response (cfr. Figure 4.4) results to be below the unity in the whole generator-level visible tau p_T spectrum considered. It is crucial, then, estimating corrections to the scale of the tau energy, as it will be explained in Section 4.5. Almost negligible is the pileup effect on the energy reconstruction.

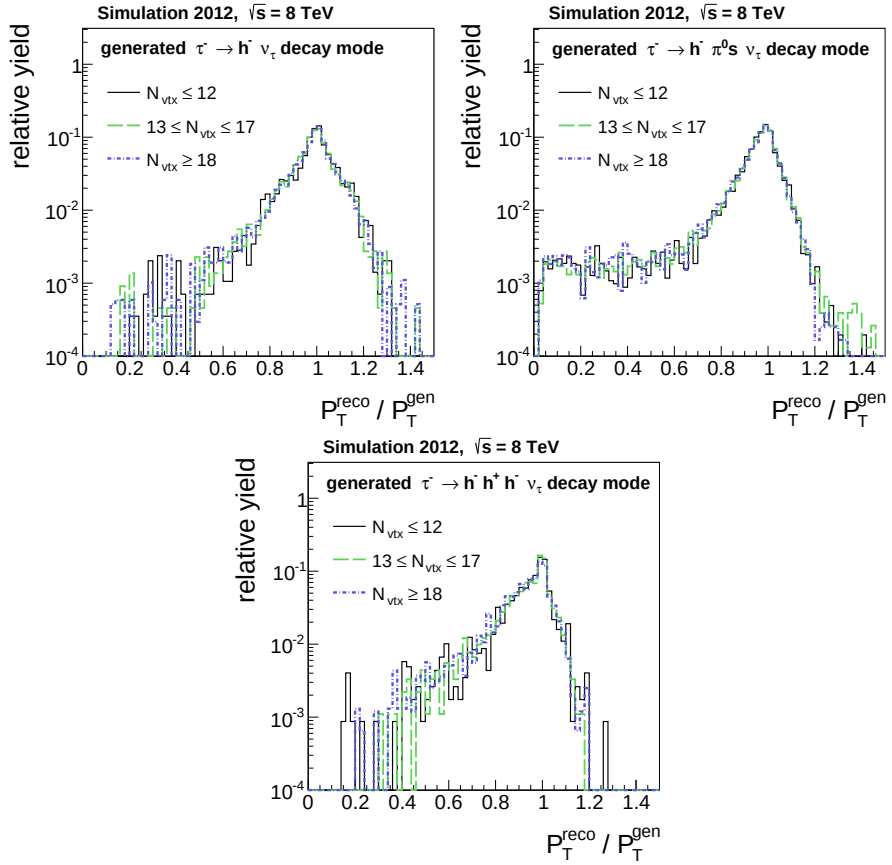

Figure 4.3: Ratio of reconstructed to generator level visible tau p_T estimated in simulated $Z' \to \tau\tau$ events, for different pileup conditions and for different decay modes: $h^{\pm}$ (top left), $h^{\pm}\pi^0$s (top right) and $h^{\pm}h^{\mp}h^{\pm}$ (bottom). Reconstructed τ_h candidates are required to match, in a given decay mode, an hadronic tau at the generator-level within a cone of size $\Delta R < 0.3$, for which the visible decay products satisfy $p_T > 20$ GeV, $|\eta| < 2.3$ and the loose working point of the cut-based isolation discriminator [A].

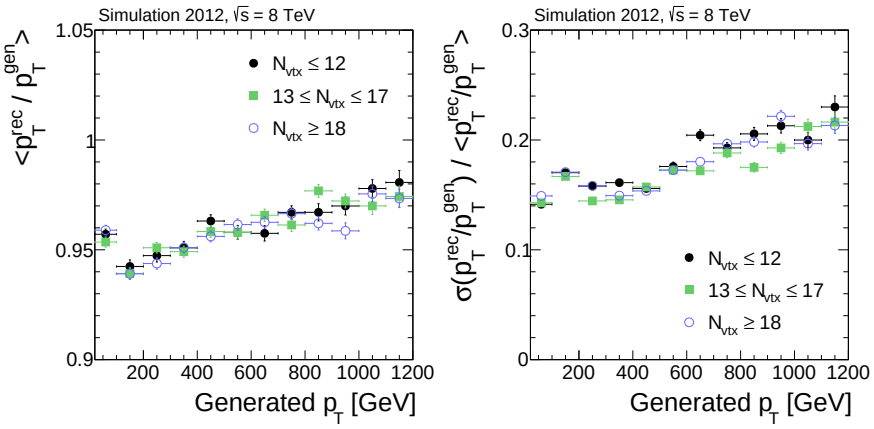

Figure 4.4: Tau energy response (left) and relative resolution (right) as function of generator-level visible tau p_T in simulated $Z' \to \tau\tau$ events and for different pileup conditions ($N_{vtx} \leqslant 12$, $13 \leqslant N_{vtx} \leqslant 17$ and $N_{vtx} \geqslant 18$). Reconstructed τ_h candidates are required to match, in a given decay mode, an hadronic tau at the generator-level within a cone of size $\Delta R < 0.3$, for which the visible decay products satisfy $p_T > 20$ GeV, $|\eta| < 2.3$ and the loose working point of the cut-based isolation discriminator [25] [A].

4.3.1 Tau identification efficiency

The probability of reconstructing and identifying an isolated tau decaying hadronically is defined as:

$$\epsilon = \frac{p_T^{rec} > 20\,\text{GeV}\,,\, |\eta_{rec}| < 2.3\,,\, \text{DM-finding}\,,\, \text{tau isolation}}{p_T^{gen} > 20\,\text{GeV}\,,\, |\eta_{gen}| < 2.3}. \tag{4.2}$$

The denominator of Equation 4.2 is populated by generator-level tau candidates, whose visible decay products satisfy $p_T > 20$ GeV and $|\eta| < 2.3$ requirements. Reconstructed taus are those that match the generator-level ones within $\Delta R < 0.3$ and satisfy the same kinematical requirements applied at the denominator. Moreover, taus have to be reconstructed in one of the hadronic decay modes studied and pass a given isolation discriminant. Tau identification efficiency computed in simulated samples of $Z/\gamma^* \to \tau\tau$ and $Z' \to \tau\tau$ are reported in Figure 4.5 for the cut-based and MVA-based tau isolation discriminant. The expected efficiencies range between 40% and 70%, depending on the approach considered (cut-based or MVA-based) and the working point adopted. Generally, the efficiencies are higher in $Z' \to \tau\tau$ compared to $Z/\gamma^* \to \tau\tau$ events. Tau identification efficiencies as a function of the

generator-level tau pseudorapidity and the reconstructed vertex multiplicity are reported in Appendix A.

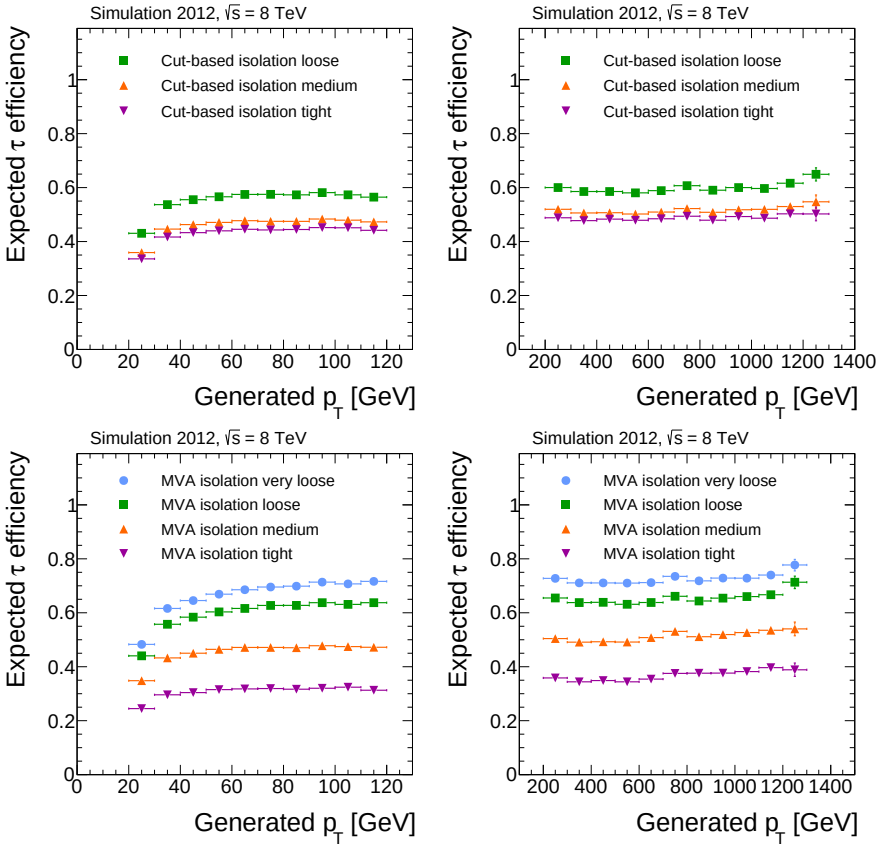

Figure 4.5: Efficiencies for hadronic tau decays in simulated $Z/\gamma^* \to \tau\tau$ (left) and $Z' \to \tau\tau$ (right) events to be reconstructed in one of the decay modes $h^\pm$, $h^\pm\pi^0$s and $h^\pm h^\mp h^\pm$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$, and to pass different working points corresponding to the cut-based (top) and MVA-based (bottom) tau isolation discriminators. Efficiencies, computed according to the Equation 4.2, are shown as function of the generator-level p_T of the visible decay products for τ_h within the geometric acceptance $|\eta| < 2.3$ [25] [A].

In correspondence of the loose cut-based isolation working point, the expected efficiency is about 40%, 60% and 70% for the $h^\pm h^\mp h^\pm$, $h^\pm\pi^0$s and $h^\pm$ decay modes, respectively. The corresponding numbers for the tight working point of the MVA-based isolation discriminators are 30%, 40% and 60%.

The efficiencies of the discriminators against electrons and muons are defined as:

$$\epsilon = \frac{\text{anti-}\ell\text{ discriminant}}{p_T^{rec} > 20\text{ GeV}\,,\ |\eta_{rec}| < 2.3\,,\ \text{DM-finding}\,,\ \text{loose CB isolation}}. \quad (4.3)$$

According to the Equation 4.3, these efficiencies are determined relatively to reconstructed τ_h candidates (matched with generator-level taus within $\Delta R < 0.3$) with $p_T > 20$ GeV, $|\eta| < 2.3$, satisfying the decay mode and the loose working point of the cut-based tau isolation requirements. On top of the denominator selection criteria, tau candidates have to satisfy the anti-e/μ discriminator in order to populate the numerator. The discrimination against electrons (muons) efficiencies range between 60% and 95% (95% and 99%).

4.3.2 Misidentification rates

The probability of misreconstructing quark and gluon jets as hadronic taus (indicated in the following as either misidentification or fake rate) is defined as follows:

$$P_{fr} = \frac{p_T^\tau > 20\text{ GeV}\,,\ |\eta_\tau| < 2.3\,,\ \text{DM-finding}\,,\ \text{tau isolation}}{p_T^{jet} > 20\text{ GeV}\,,\ |\eta_{jet}| < 2.3}. \quad (4.4)$$

To be labelled as a fake tau, a reconstructed jet with $p_T > 20$ GeV and $|\eta| < 2.3$ has to match, within $\Delta R < 0.3$, a tau satisfying the standard identification criteria, $p_T > 20$ GeV and $|\eta| < 2.3$ requirements. It is important to notice that, in Equation 4.4, p_T^{jet} and η_{jet} are computed by summing the momenta of all particle constituents of the jet, while p_T^τ and η_τ refer to only the charged particles and photons included in the tau decay mode reconstruction. Besides, jet energies are calibrated [127] and corrected for pileup effects [140, 133], whereas no calibration or pileup correction is applied to τ_h candidates.

The jet$\rightarrow \tau_h$ misidentification rates, shown in Figure 4.6, are computed in simulated QCD-multijets and W+jets events, as explained in Section 4.3). Depending on the working point considered for the cut-based and MVA-based discriminators, misidenfication rates range between approximately 0.03% and 1%. It can be observed that, in general, the misidentification rates are higher in W+jets compared to QCD-multijets events given the higher fraction of quark jets in W+jets events. Quark jets typically have a lower particle multiplicity and are more collimated than gluon jets, increasing their probability to be misidentified as τ_h candidates. Moreover, another source of difference is

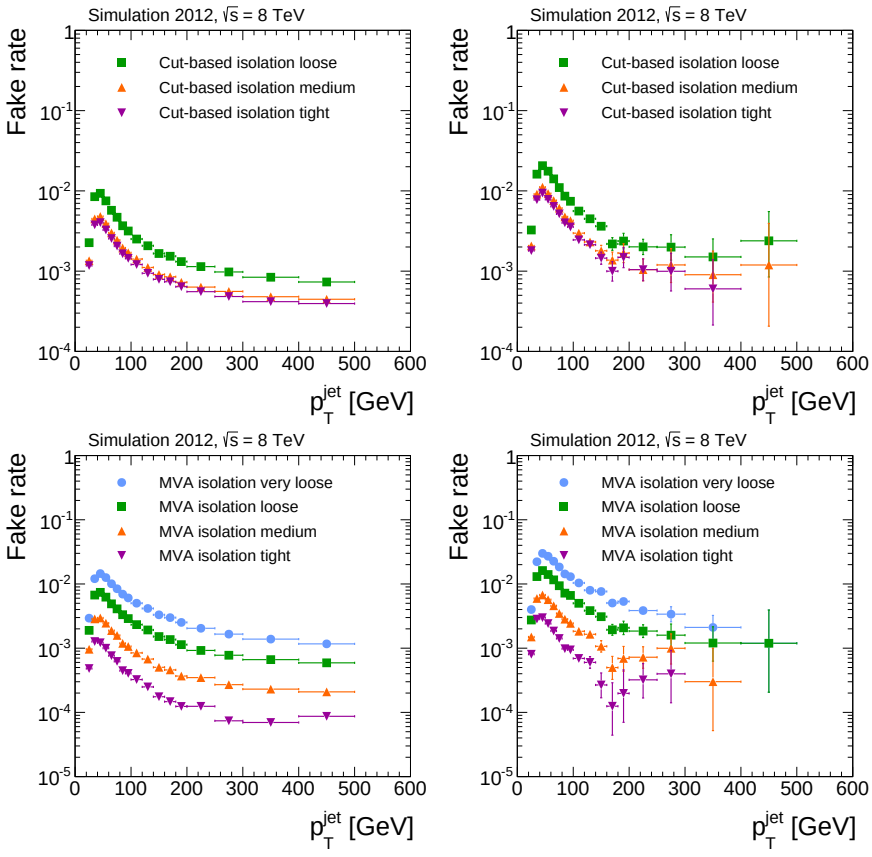

Figure 4.6: Probabilities for quark and gluon jets, computed in simulated QCD-multijets (left) and W+jets (right) events, to pass the cut-based (top) ant the MVA-based (bottom) tau identification discriminators as function of the p_T of reconstructed jets within the acceptance $|\eta| < 2.3$. No true electrons, muons and taus have been included in the reconstructed jet for W+jets events. All probabilities shown are computed according to Equation 4.4 [A].

the jet p_T spectrum, which actually plays a role, given the dependency of the jet$\rightarrow \tau_h$ misidentification rates on jet p_T: the jet$\rightarrow \tau_h$ misidentification probabilities for quark jets, as well as gluon jets, typically decrease as a function of the jet p_T, as the particle multiplicity increases for higher energetic jets. More plots on the jet$\rightarrow \tau_h$ misidentification rates computed as a function of the reconstructed jet η and reconstructed vertex multiplicity are reported in Appendix A.

The misidentification rates for $e \to \tau_h$ and $\mu \to \tau_h$ are determined in the following way:

$$P_{fr} = \frac{p_T^\tau > 20\,\text{GeV}\,,\, |\eta_\tau| < 2.3\,,\, \text{DM-finding}\,,\, \text{loose CB iso.}\,,\, \text{anti-}\ell\ \text{discr.}}{p_T^\ell > 20\,\text{GeV}\,,\, |\eta_\ell| < 2.3}. \tag{4.5}$$

Electrons and muons are wrongly reconstructed as tau if, at generator-level, they pass $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$, and if there are matched τ_h candidates reconstructed within a cone of size $\Delta R < 0.3$, which are properly identified and isolated. Typical $e \to \tau_h$ misidentification probabilities, computed in simulated $Z/\gamma^* \to ee$ events, range from a few per mille to a few percent. The rates for $\mu \to \tau_h$ misidentification are expected to be at or below the per mille level in $Z/\gamma^* \to \mu\mu$ events. Tau identification efficiencies and misidentification rates averaged over p_T and η and for pileup conditions characteristic of the 2012 data-taking period are summarized in Table 4.3.

Tau isolation discriminants				
working point	Efficiency		jet$\rightarrow \tau_h$ misidentification rate	
	$Z/\gamma^* \rightarrow \tau\tau$	$Z'(2.5\ \text{TeV}) \rightarrow \tau\tau$	W+jets	QCD-multijets
Cut-based				
Loose	49.0%	58.9%	$9.09 \cdot 10^{-3}$	$3.86 \cdot 10^{-3}$
Medium	40.8%	50.8%	$5.13 \cdot 10^{-3}$	$2.06 \cdot 10^{-3}$
Tight	38.1%	48.1%	$4.38 \cdot 10^{-3}$	$1.75 \cdot 10^{-3}$
MVA-based				
Very loose	55.9%	71.2%	$1.29 \cdot 10^{-2}$	$6.21 \cdot 10^{-3}$
Loose	50.7%	64.3%	$7.38 \cdot 10^{-3}$	$3.21 \cdot 10^{-3}$
Medium	39.6%	50.7%	$3.32 \cdot 10^{-3}$	$1.30 \cdot 10^{-3}$
Tight	27.3%	36.4%	$1.56 \cdot 10^{-3}$	$4.43 \cdot 10^{-4}$

Discrimination against electrons			
working point	Efficiency		$e \rightarrow \tau_h$ misidentification rate
	$Z/\gamma^* \rightarrow \tau\tau$	$Z'(2.5\ \text{TeV}) \rightarrow \tau\tau$	$Z/\gamma^* \rightarrow ee$
Very loose	94.3%	89.6%	$2.38 \cdot 10^{-2}$
Loose	90.6%	81.5%	$4.43 \cdot 10^{-3}$
Medium	84.8%	73.2%	$1.38 \cdot 10^{-3}$
Tight	78.3%	65.1%	$6.21 \cdot 10^{-4}$
Very tight	72.1%	60.0%	$3.54 \cdot 10^{-4}$

Discrimination against muons			
working point	Efficiency		$\mu \rightarrow \tau_h$ misidentification rate
	$Z/\gamma^* \rightarrow \tau\tau$	$Z'(2.5\ \text{TeV}) \rightarrow \tau\tau$	$Z/\gamma^* \rightarrow \mu\mu$
Cut-based			
Loose	99.3%	96.4%	$1.77 \cdot 10^{-3}$
Tight	99.1%	95.0%	$7.74 \cdot 10^{-4}$
MVA-based			
Loose	99.5%	99.4%	$5.20 \cdot 10^{-4}$
Medium	99.0%	98.8%	$3.67 \cdot 10^{-4}$
Tight	98.0%	97.7%	$3.18 \cdot 10^{-4}$

Table 4.3: Expected efficiencies and misidentification rates corresponding to various tau identification discriminants. All the numbers are obtained for pileup conditions characteristic of the 2012 data-taking period and by averaging over p_T and η ranges considered for the computation.

4.4 Performance validation with 2012 data

4.4.1 Tau identification efficiency

The HPS algorithm performance have been also measured in $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events selected in 2012 data. The measurement has been performed by considering several background processes: $Z/\gamma^* \to \ell\ell$ ($\ell = e, \mu$), dibosons (WW, WZ, ZZ), W+jets, $t\bar{t}$ and QCD-multijets. Background contributions are reduced by requiring the selection criteria reported in the following. The events are required to fire a single muon trigger. The reconstructed muon is required to satisfy $p_T > 25$ GeV and $|\eta| < 2.1$, to pass the particle-flow tight identification criteria, and the isolation requirement $I^{PF}_{rel} < 0.1$ (cfr. Chapter 3). A loose τ_h candidate selection is applied: the τ_h candidate has to be associated with a reconstructed jet with $|\eta| < 2.3$, to be separated from the muon by $\Delta R > 0.5$ and containing at least one track of $p_T > 5$ GeV. The leading track and the muon are required to be opposite in charge and to originate from the same primary vertex. In case more than one jet satisfies the criteria just listed, the highest-p_T jet is chosen. Additionally, the τ_h candidate is required to pass the loose tau identification discriminator against electrons, the tight discriminator against muons and having $p_T > 20$ GeV. If more than one tau passes the preselection criteria, the one with highest transverse momentum is chosen. The $t\bar{t}$ background is reduced by vetoing all the events with at least one b jet. Dibosons and $Z/\gamma^* \to \ell\ell$ background contamination is reduced by vetoing all those events with either a second loosely identified and isolated electron ($p_T > 15$ GeV and $|\eta| < 2.4$) or muon ($p_T > 5$ GeV and $|\eta| < 2.4$). The W+jet background is reduced by requiring the transverse mass computed between the muon and the missing energy in the event, defined as

$$m_T = \sqrt{2 p_T^{\ell} \not{E}_T (1 - \cos\Delta\phi)}\,, \tag{4.6}$$

to be less than 40 GeV. Specific requirements are imposed on a topological discriminator, introduced by the CDF experiment collaboration [141], to increase the purity of the $Z/\gamma^* \to \tau\tau$ sample: this discriminator is based on the fact that the angle between the neutrinos produced in the tau decay and the visible tau decay products is small, leading to a missing transverse energy vector oriented in the direction of the visible tau decay product. This topology is unlikely to be observed, for instance, in $t\bar{t}$ and W+jets events.

The hadronic tau identification efficiency ϵ_τ is measured via the *tag-and-probe* method [142]. In particular, the efficiency is defined as the ratio of the number

of genuine hadronic taus passing the tau identification discriminator under study over the total number of genuine hadronic taus:

$$\epsilon_\tau = \frac{N^\tau_{\text{pass}}}{N^\tau_{\text{pass}} + N^\tau_{\text{fail}}} \,. \tag{4.7}$$

The number of hadronic taus populating the *pass* region (N^τ_{pass}) and the *fail* region (N^τ_{fail}) is obtained via a template fit of some observables, performed simultaneously in the two regions and in which the tau identification efficiency is treated as the parameter of interest that has to maximize the likelihood function. Two observables have been considered: the multiplicity of tracks within a cone of size $\Delta R = 0.5$ centered around the τ_h candidate direction (N_{tracks})[4] and, for sake of cross-checking, the visible mass of the $\mu\tau_h$ pair (m_{vis}). The signal and background templates have been taken from simulation, except for the QCD background, whose shapes have been extracted from data. Signal and background normalizations have been taken from simulation too, except for W+jet and QCD-multijets background processes.

The efficiencies measured using N_{track} and m_{vis} give compatible results within the estimated uncertainties for both cut-based and MVA isolation discriminators. The tau identification efficiencies measured as a function of the reconstructed tau p_T, and using the variable N_{tracks} as observable, are reported in Figure 4.7 for the cut-based and the MVA-based isolation discriminators. Below the efficiency trends, the plots in Figure 4.7 report also the data-over-simulation ratio, which is useful to extract simulation-to-data corrections to account for differences observed between observed and expected results. In this case, scale factors are compatible with unity within uncertainties, which typically amount to 4.5%.

The efficiency for τ_h decays in $Z/\gamma^* \to \tau\tau$ events to pass the discriminators against electrons and muons are also measured via the tag-and-probe method, based on a template fit of the m_{vis} distribution. The efficiencies measured in data are in agreement with the expectations from simulation. The tau identification efficiency has been also measured in $t\bar{t} \to b\bar{b}\mu\tau_h$ events in order to validate the HPS algorithm performance in an environment characterized by higher hadronic activity than the one characterizing $Z/\gamma^* \to \tau\tau$ events. The measured efficiencies of all tau identification discriminators are found to be

[4] Hadronic taus in signal events are typically characterized by fewer tracks than τ_h candidates in background events. The variable N_{tracks} allows the tau identification efficiency to be measured as a function of the tau transverse momentum, as the track multiplicity is only weakly correlated to the p_T of the τ_h candidate.

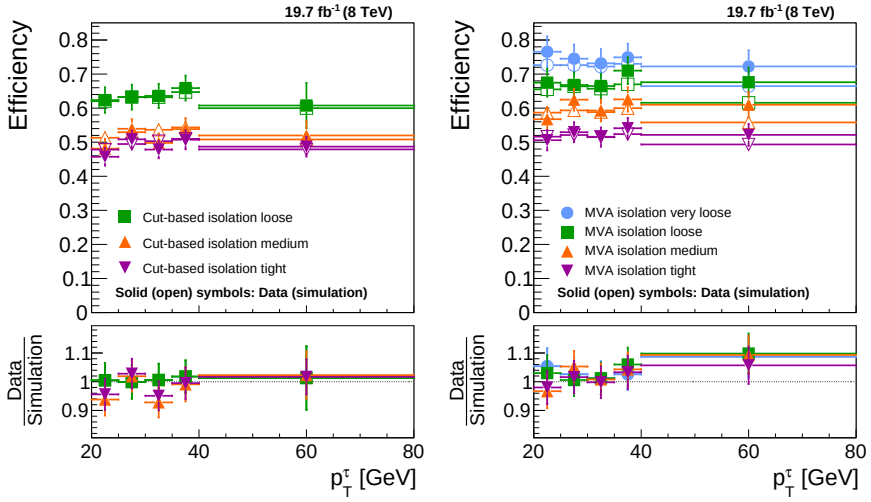

Figure 4.7: Efficiencies for hadronic tau decays measured in $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ data events and compared with expectations from simulations. Results are obtained for the cut-based (left) and the MVA-based (right) tau isolation discriminators as function of the p_T of reconstructed tau within the acceptance $|\eta| < 2.3$ [25].

compatible with the expectations within the uncertainty of the measurement, which amounts to 9-11%. More details can be found in [25].

4.4.2 jet$\to \tau_h$ misidentification rate

Data samples of W+jets and QCD-multijets events are used to validate the rates with which quark and gluon jets get misidentified as τ_h candidates. Events selected for the W+jet sample are required to pass the single muon trigger and to contain a muon selected as described in Section 4.4.1. The requirement on the transverse mass m_T (cfr. Equation 4.6) is $m_T > 50$ GeV. Additionally, a jet of $p_T > 20$ GeV and $|\eta| < 2.3$, such that $\Delta R(\mu, \text{jet}) > 0.5$, is required. Concerning the selection for the QCD-multijets events, at least two jets are required in the event such that at least one jet has fired a single-jet trigger with a p_T threshold of 320 GeV, and the other jet(s) the $p_T > 20$ GeV and $|\eta| < 2.3$ requirements. If only one jet satisfies the trigger requirement, that jet is excluded from the computation of the misidentification rate. If two or more jets satisfy the trigger requirement, all jets with $p_T > 20$ GeV and $|\eta| < 2.3$ are included in the computation of the misidentification rate. This procedure ensures that the rate for jet$\to \tau_h$ measured in QCD-multijet events

is not biased by trigger requirements. The jet $\rightarrow \tau_h$ misidentification rate is measured according to Equation (4.4), whose denominator is populated by particle-flow jets that have to originate from the primary vertex. The misidentification rate measured in data, and compared with the expectations from simulation, are reported in Figure 4.8 for QCD-multijets and W+jets events. The results, reported as a function of the reconstructed jet p_T for both the cut-based and the MVA-based isolation discriminators, confirm that, as already seen in Section 4.3.2, the misidentification rates are higher in W+jets than in QCD-multijets events. It can be additionally observed that, while the rates for jet$\rightarrow \tau_h$ measured in data exceed the expectations at low p_T, the rates measured at high p_T are lower than expected, causing an increase of the data-over-simulation ratio to approximately 20%. This behavior is observed for all τ_h identification discriminators, and it is of similar magnitude in the whole acceptance range considered for jets ($|\eta| < 2.5$).

The misidentification rates for electrons and muons have been measured in $Z/\gamma^* \rightarrow ee$ and $Z/\gamma^* \rightarrow \mu\mu$ events, by using the tag-and-probe technique, as detailed explained in [25]. It has been measured that, for electrons in the barrel region ($|\eta| < 1.460$), the measured misidentification rates exceed the prediction by up to a factor of 1.7. With a slightly smaller magnitude, a similar trend is observed for electrons in the endcap region ($|\eta| < 1.558$). Concerning the $\mu \rightarrow \tau_h$ misidentification rates, those measured in data exceed the expectations by up to a factor 1.9. The data-over-simulation ratio is higher for muons within $|\eta| > 1.2$ compared to those within $|\eta| < 1.2$ and it increases as the muon rejection criteria are tightened.

4.5 Tau energy scale

In the ideal case in which the particle-flow algorithm allows the charged particles and photons energy to be measured with infinite precision, the reconstructed tau energy is expected to be equal to the true energy of the visible tau decay products. However, in reality, it has been already shown in Figure 4.4 that the energy response is below the unity. This effect, affecting the characteristic scale of the tau energy, the tau energy scale τ-ES, is induced by different motivations. First of all, it can happen that spurious particles, like those belonging to pileup events, can be accidentally identified as signal constituents of the tau candidate, causing an extra contribution to the tau momentum. Looking at Figure 4.3, one can consider these events populating the region of the p_T^{rec}/p_T^{gen} distribution in which such a ratio is above the unity. Conversely, if a tau decay product is reconstructed by the particle-flow algo-

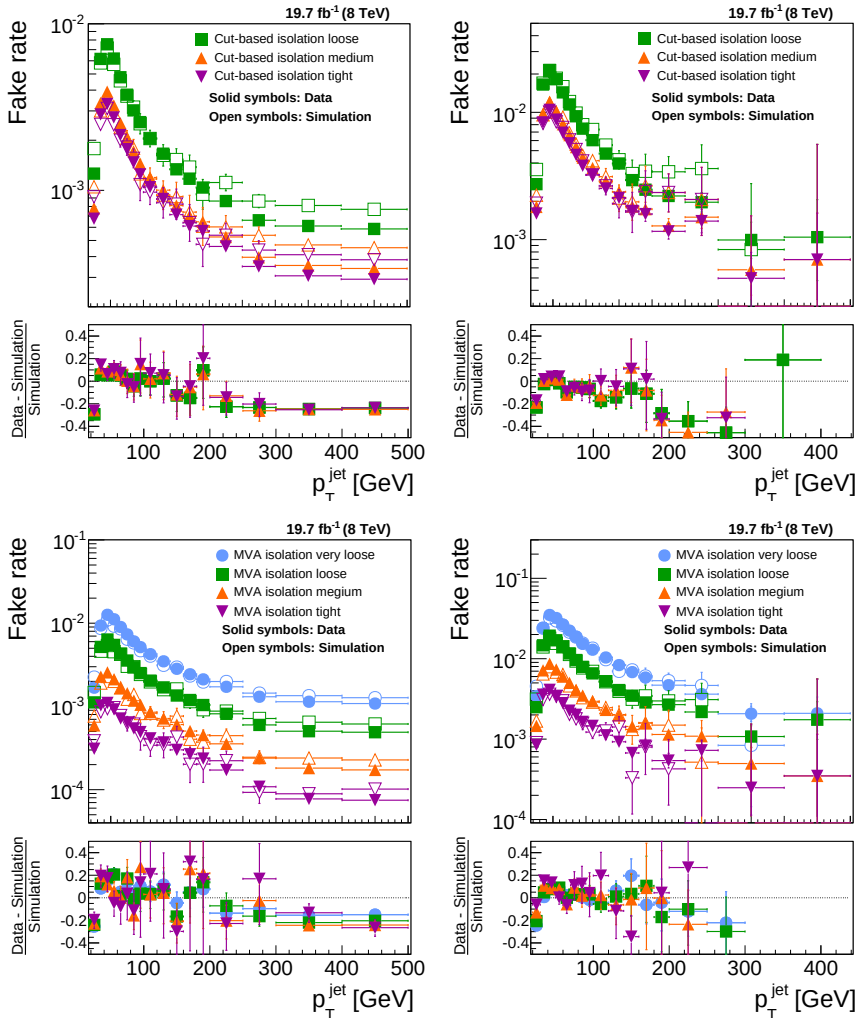

Figure 4.8: Probabilities for quark and gluon jets to be reconstructed as hadronic taus measured in data samples of QCD-multijets (left) and W+jets (right) events. The trends are computed as a function of the p_T of reconstructed jets within the acceptance $|\eta| < 2.3$, for both the cut-based (top) ant the MVA-based (bottom) tau isolation discriminators [25].

rithm, but not recognized by the HPS algorithm as a constituent of the tau signal cone, it will be probably treated as an isolation particle. In this case, whether the tau fails the isolation cut, implying a decrease of the reconstruction efficiency or, if the isolation selection is passed, the resulting tau candi-

date will be deprived of part of its original visible momentum: these events populate the region in which $p_T^{rec}/p_T^{gen} < 1$ (cfr. Figure 4.3) and they represent the majority of the cases. Moreover, miscalibration of the tracker and non linearities in the calorimeter energy response can alterate the tau energy scale.

The energy scale of hadronic taus is an important systematic uncertainty for all physics analyses with tau leptons in the final state. The common strategy adopted is applying data-to-simulation τ-ES correction factors, which are determined as described in the following. The quality of the description of the tau energy scale in simulation can be inferred and constrained by measurements performed in $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ data events. Two observables have been used for the tau energy scale measurement: the visible mass of the $\mu\tau_h$ pair (m_{vis}), sensitive to the τ-ES only given the high precision with which the muon momentum scale is known, and the mass of reconstructed τ_h candidate (m_τ). The method based on the m_{vis} has the advantage of exploiting the $Z \to \tau\tau$ peak purity reconstruction for all the tau decay modes ($h^\pm$, $h^\pm\pi^0$s and $h^\pm h^\mp h^\pm$), but the disadvantage of being contaminated by the presence of a resonance $\phi \to \tau\tau$ [29]. The measurement using the m_τ, instead, can be performed exploiting only two decay modes ($h^\pm\pi^0$s and $h^\pm h^\mp h^\pm$), but it is fully independent of the particle from which the tau originates.

The background processes considered are the same used in the tau identification efficiency measurement (cfr. Section 4.4.1). However, in this case, all of them, except the QCD-multijets background, which has been obtained directly from data, have been estimated from simulation. The events are required to satisfy a muon plus hadronic tau trigger selection. The reconstructed muon is required to satisfy the same requirements described in Section 4.4.1. An hadronic tau is required to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$, and the medium working point of the MVA isolation discriminator. The muon and the hadronic tau are required to be opposite in charge and to originate from the same primary vertex. In case the requirements just described are satisfied by more than one $\mu\tau_h$ pair, the one with the highest $p_T^\mu + p_T^{\tau_h}$ is chosen. The W+jet background is reduced by requiring the transverse mass between the muon and the missing energy in the event to be less than 40 GeV (cfr. Equation 4.6). Additionally, events with either a second loosely identified and isolated muon with $p_T > 15$ GeV and $|\eta| < 2.4$ are vetoed to suppress the $Z/\gamma^* \to \mu\mu$ background.

Several mass templates have been considered for both signal and background processes by discretely varying the tau energy scale between -6% and $+6\%$[5]

[5]Previous studies [143] have confirmed that the range -6% to $+6\%$ is sufficient to contain the measured τ-ES value as well as the $\pm1\sigma$ uncertainty interval in each case. The step size of 0.1%

in step of 0.1%, for each of the decay modes under study (depending on the mass observable) and in three bins of τ_h transverse momentum ($20 < p_T < 30$ GeV, $30 < p_T < 40$ GeV, $p_T > 45$ GeV). Data mass templates have been compared with those obtained from simulation, in correspondence of each τ-ES variation and fitted with shape templates for signal plus background processes. The parameter of interest of the maximum likelihood fit is the τ-ES. The ratio between data and simulation mass templates, corresponding to the best fit τ-ES value, represents the correction to the τ-ES. These corrections are reported in the three tau p_T bins and for different decay modes in Figure 4.9, relatively to m_{vis} and m_τ. It can be observed that, within the uncertainties, τ-ES corrections are not dependent on the tau p_T. Moreover, given the smaller uncertainties, the variable m_{vis} is observed to be less sensitive to the τ-ES than m_τ.

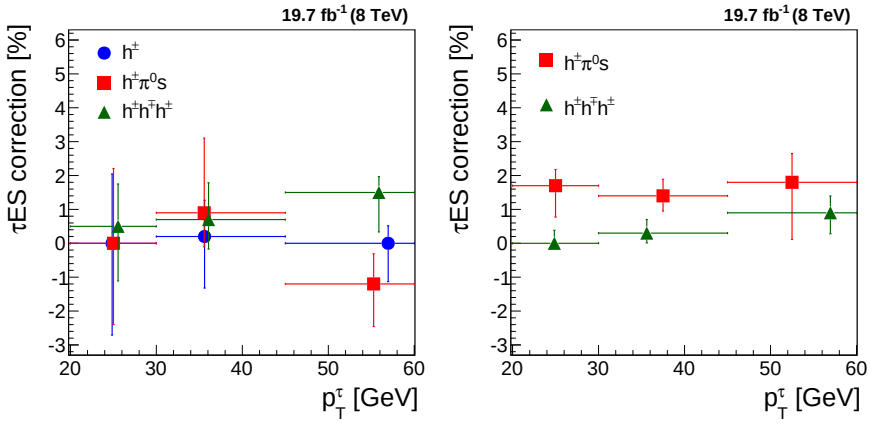

Figure 4.9: Tau energy scale corrections measured in $Z/\gamma^* \to \tau\tau$ events, considering the $\mu\tau_h$ final state. The results, obtained separately for $h^\pm$, $h^\pm\pi^0$s and $h^\pm h^\mp h^\pm$ decay modes, correspond to the value of the τ-ES that has maximized the likelihood fit performed on the visible mass of the tau pair, m_{vis} (left) and on the mass of reconstructed τ_h candidates, m_τ (right) [25].

4.6 The Secondary Vertex fit algorithm

In the search for a narrow resonance ϕ decaying to tau leptons, with a mass close to the mass of the Z boson, the $Z \to \tau\tau$ process constitutes an important

has been chosen as it is smaller than the precision that is expected to be achieved with the τ-ES measurement.

source of irreducible background. Therefore, the τ-pair mass becomes a powerful discriminating variable. The Secondary Vertex fit method (SVfit) [144] is the technique currently used in the CMS collaboration to reconstruct the τ-pair mass including also the contribution of the tau invisible decay products. The description of this algorithm follows.

The first step is understanding which are the parameters needed to fully constrain a tau lepton decay. In case of an hadronically decaying tau, six parameters are needed: the polar and azimuthal angles of the visible decay product system in the τ rest frame, the three boost parameters from the τ rest frame to the laboratory frame, and the invariant mass of the visible decay. In case the tau decays leptonically, the invariant mass of the two neutrinos[6] produced in the final state becomes the seventh kinematic parameter. The number of the unknown parameters, however, is reduced because of constraints imposed by the measurement of some observables: the components of the four-momentum of the τ visible decay products system, measured in the laboratory frame, and the two components $(\not{E}_x, \not{E}_y)$ of the missing transverse energy of the event. In the end, then, the unknown parameters left unconstrained are two for the τ decaying hadronically and three for the τ decaying leptonically. These parameters are:

- $\mathbf{X}$, the fraction of the tau lepton energy in the laboratory frame carried by the visible decay products;
- $\boldsymbol{\phi}$, the azimuthal angle of the tau momentum vector with respect to the one of the visible decay products;
- $\mathbf{m_{\nu\nu}}$, in case of leptonic decay only, is the mass of the neutrino-antineutrino system.

Despite the τ-pair kinematics is left unconstrained, the SVfit method[7] allows the τ-pair mass $m_{\tau\tau}$ to be reconstructed by means of a maximum likelihood fit. The model is based on the likelihood function $p(\vec{x}|\vec{y}, \vec{a}_1, \vec{a}_2)$, which is, however, the product of three likelihood functions: the first two functions model the parameters $\vec{a}_1$ and $\vec{a}_2$ of the τ lepton decay, and the third one quantifies

[6] Considering that the leptonic decay of a tau candidate is $\tau_\ell \to \ell \bar{\nu}_\ell \nu_\tau$ and that the hadronic tau one is $\tau_h \to 1(\text{or } 3)$ charged hadrons $+ \geqslant 0$ neutrals $+ \nu_\tau$, the invisible momentum system produced in the leptonic tau decay is characterized by the invariant mass of the two neutrinos, which is in general non-zero. On the contrary, the invisible decay system of the hadronically decaying tau is massless.

[7] The current version of the SVfit algorithm does not yet include the tau decay vertex information provided by the tracking detector. This additional information, which is planned to be included in the algorithm in the future, is actually what the SVfit owes its name to.

the compatibility of a τ-pair decay hypothesis with the measured $(\not{E}_x, \not{E}_y)$. Then, the likelihood function $p(\vec{x}|\vec{y}, \vec{a}_1, \vec{a}_2)$ represents the probability to measure the values $\vec{x} = (\not{E}_x, \not{E}_y)$, given that the unknown parameters specifying the kinematics of the two tau decays have values $\vec{a}_1 = (X^1_\ell, \phi^1_\ell, m^1_{\nu\nu})$ and $\vec{a}_2 = (X^2_\ell, \phi^2_\ell, m^2_{\nu\nu})$, and that the four-momenta of the tau visible decay products have measured values $\vec{y} = (p^1_{vis}, p^2_{vis})$. Of course, depending on the tau decay, the parameters $m^{1,2}_{\nu\nu}$ can be set to zero in the fit. The probability $P(m_{\tau\tau})$ is then obtained as:

$$P(m_{\tau\tau}) = \int \delta\left(m_{\tau\tau} - m_{\tau\tau}(\vec{y}, \vec{a}_1, \vec{a}_2)\right) p(\vec{x}|\vec{y}, \vec{a}_1, \vec{a}_2) d\vec{a}_1 d\vec{a}_2 \,, \tag{4.8}$$

where the integral takes into account all the possible configurations of the parameters $\vec{a}_1$ and $\vec{a}_2$, compatible with the measured $\vec{y}$. The integration is performed numerically with the VEGAS program [145]. For each event, the best estimate $\hat{m}_{\tau\tau}$ corresponds to the ditau pair mass value that has maximized the probability $P(m_{\tau\tau})$.

In the context of the SM $h(125\ \text{GeV}) \to \tau\tau$ search [15], it has been observed that the SVfit mass improves the separation between the signal and the irreducible background $Z/\gamma^* \to \tau\tau$, with respect to the ditau pair visible mass, as can be observed in Figure 4.10. The overall improvement to the expected significance is sizable and accounts to about 40%. The improvement yielding to the expected significance will be also shown in the context of the SM $Zh \to \ell\ell\tau\tau$ analysis (included in the SM $h(125\ \text{GeV}) \to \tau\tau$ search performed by the CMS collaboration), which will be described in the next chapter.

4.7 Final considerations and prospects

The HPS algorithm performance have improved steadily during the LHC Run I, thanks to the experience gained with the detector and the LHC running conditions. In Figure 4.11, the improvement over past years of the HPS algorithm performance is quantified in terms of expected efficiencies, which are estimated in $Z/\gamma^* \to \tau\tau$ and $Z' \to \tau\tau$ events, and expected jet$\to \tau_h$ misidentification probabilities, which are measured in simulated QCD-multijets events. Four different isolation discriminators are compared. From 2011 to 2012, the change in favor to three hits (green curve) required for the charged track reconstruction from the eight (cyan curve) implied a slight reduction in both efficiency and fake rate. In 2013, the development of MVA-based isolation discriminators (magenta and red curves) led to a sharp reduction in the fake

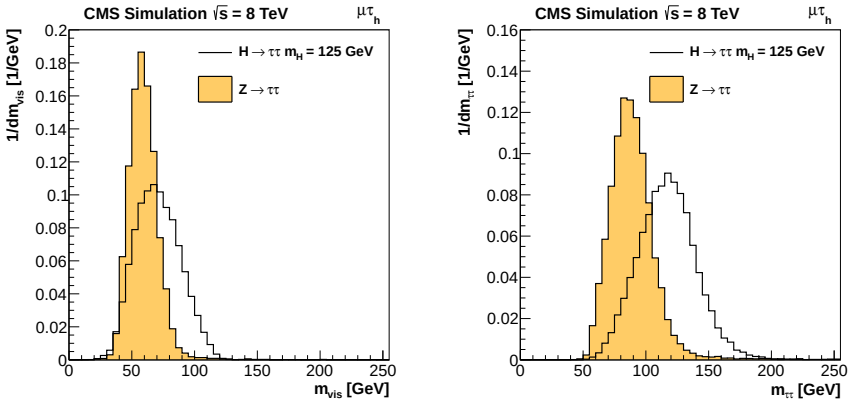

Figure 4.10: Comparison between the SM signal $h(125\ \text{GeV}) \to \tau\tau$ and the irreducible background $Z/\gamma^* \to \tau\tau$ in terms of the normalized ditau visible mass spectrum (left) and the SVfit one (right), obtained in the $\mu\tau_h$ channel [15].

rate (about 50%) and a significant increase in the tau identification efficiency (about 10%), especially at high p_T.

Intense improvements on the HPS algorithm performance for LHC Run II are ongoing. Preliminary studies on the expected performance for Run II have shown a decrease of the tau identification efficiency as a function of the generated tau p_T due to photon conversions (affecting mostly the $h^\pm\pi^0$s decay mode), nuclear interactions of charged pions in the tracking detector (affecting mostly $h^\pm$ and $h^\pm h^\mp h^\pm$), and tails in the ECAL of hadronic showers of charged pions, that failed to get linked to the track by the particle-flow algorithm. In all three cases, energy deposits in the ECAL, that are due to tau decay products, get reconstructed as either electrons or photons, spoiling the isolation of the tau candidate if they are not included in the strip reconstruction. The decrease in tau identification efficiency is mitigated by dynamically varying the size of the strip depending on the e/γ p_T, in order to be enlarged for low-p_T particles (taking into account that the bending of charged particles in the magnetic field increases inversely proportional to the p_T). Preliminary results show that the dynamic strip-size reconstruction brings 10% of improvement on the tau identification efficiency (estimated in simulated vector-boson fusion $h(125\ \text{GeV}) \to \tau\tau$ events), with almost similar performance on the jet$\to \tau_h$ misidentification rate (estimated in simulated QCD-multijets events).

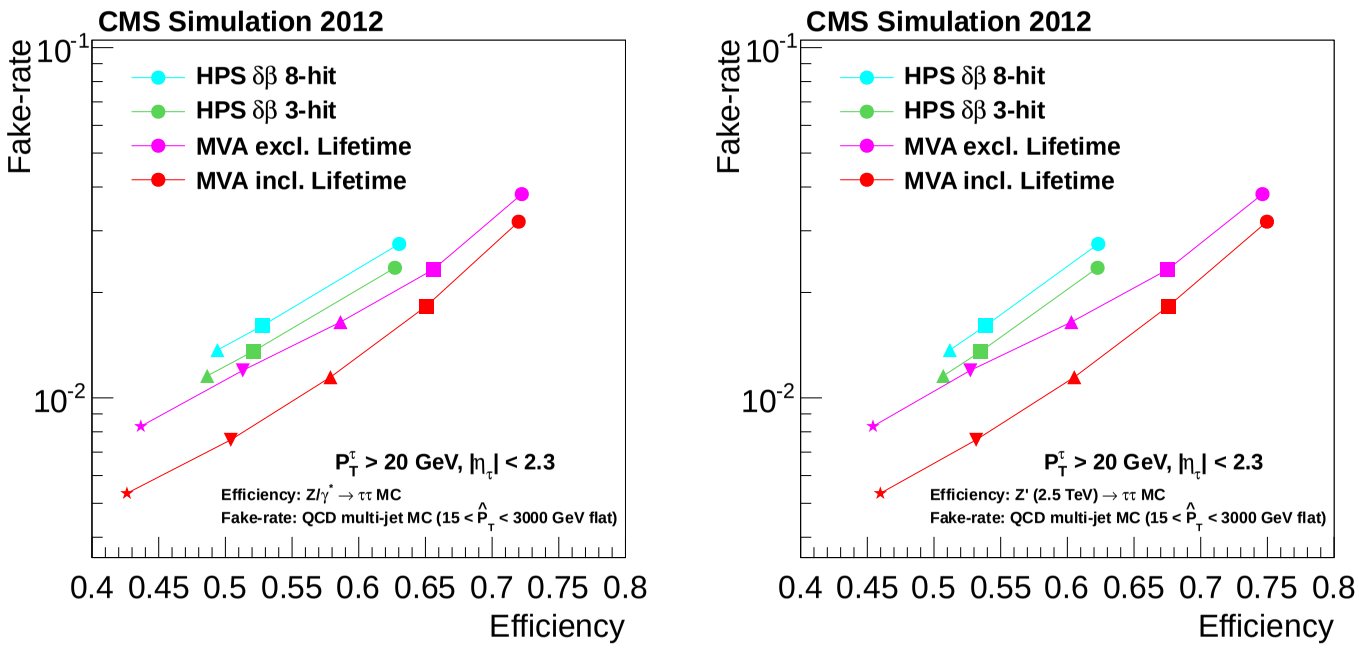

Figure 4.11: Jet$\to \tau_h$ misidentification rate *versus* tau identification efficiency (ROC curve) estimated in simulated $Z/\gamma^* \to \tau\tau$ (left) and $Z'(2.5\,\text{TeV}) \to \tau\tau$ (right) events for different tau isolation discriminators. From 2011, when the cut-based isolation discriminator based on eight hits per charged track, (cyan curve), there has been a slight reduction in the misidentification rate in 2012, when a smaller number of hits (three) for the charged track reconstruction was required for the cut-based isolation (green curve). The development of MVA-based isolation discriminators (magenta and red curves) in 2013 has led to significant improvements in both efficiency and misidentification rate [146].

5

Search for the SM Higgs boson decaying to tau pairs in the process $pp \to Zh$

The CMS search for the SM Higgs boson decaying to tau leptons, and produced in association with a Z *boson decaying to light leptons (electrons or muons), is presented in this chapter. The results reported refer to the main CMS analysis performed in the* $Zh \to \ell\ell\tau\tau$ *channel (where* ℓ *is a generic symbol, used from now on, to label electrons or muons. The symbol* L *will be used to indicate either electrons, muons or taus) by exploiting data collected by the CMS experiment during 2011 and 2012. In the context of the Ph.D. work, an independent and identical analysis has been performed to officially validate and scrutinize the main one (a summary of the main results obtained in the context of the synchronization process will be shown in Section 5.10) after having performed studies on the measurement of the background from jets misidentified as leptons and on the systematic uncertainties. The work presented in this chapter has been carried out during the first half of the doctoral work and has been included in paper documenting the first evidence of the* 125 GeV *Higgs boson decaying to a pair of tau leptons* [15].

5.1 Introduction

The search for the SM Higgs boson decaying to tau leptons is crucial to directly probe the Higgs coupling to fermions, therefore confirming that the Higgs boson discovered in the $\gamma\gamma$, ZZ and WW channels [147] is the scalar particle predicted by the SM. The $h \to \tau\tau$ process, despite having a branching ratio of an order of magnitude lower compared to the $h \to b\bar{b}$ process (cfr. Figure 1.9), can be more efficiently selected out of the background.

As described in Section 1.2.2, the SM process $pp \to h + X$ receives the largest contributions from gluon fusion, the vector-boson fusion, and from the associated production with vector bosons (Vh) or heavy quarks ($t\bar{t}h$). The analysis described in this chapter is specifically optimized for the Higgs boson production in association with the vector boson Z decaying in a pair of electrons or muons, which will be generally indicated as light leptons in the following. Requiring the presence of two additional light leptons not only enhances the signal over background ratio significantly (the sensitivity of the Vh production mechanism, then, becomes comparable to the gluon/vector-boson fusion ones), but also makes the $\ell\ell\tau\tau$ signature to be efficiently triggered.

Eight different final states are studied in this analysis, corresponding to the combination of two leptonic decays of the Z (ee and $\mu\mu$) and four decays of the Higgs boson τ-pair ($e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$): $eee\mu$, $eee\tau_h$, $ee\mu\tau_h$, $ee\tau_h\tau_h$, $\mu\mu e\mu$, $\mu\mu e\tau_h$, $\mu\mu\mu\tau_h$, and $\mu\mu\tau_h\tau_h$. The $\ell\ell ee$ and $\ell\ell\mu\mu$ final states are excluded because already considered in the search for the SM Higgs in the four leptons channel ($h \to ZZ \to 4\ell$) [12].

5.2 Datasets and simulated samples

As already anticipated in Section 2.2.7, all events satisfying either a set of HLT paths, characterized by requirements on a specific physics object, enter one specific dataset. Therefore, the choice of the trigger objects for the analysis defines which dataset has to be used. The datasets contain a range of CMS runs that were collected in a run period, which is typically the period between two LHC technical stops. Each run period might be characterized by different beam configurations, pileup rates and also trigger conditions. The 2011 period has been split into two runs: Run A, having an average of about 6 pileup events collected simultaneously with each hard scattering collision, and Run B, in which this number increased to about 10. During 2012 there have been four run periods labeled from A to D. The primary datasets used

in this analysis contain dielectron and dimuon triggered events, hence the names: `DoubleElectron` and `DoubleMu`. Both datasets correspond to a total integrated luminosity of $\mathcal{L} = 4.9 \pm 0.11$ fb^{-1} and 19.7 ± 0.5 fb^{-1} at the center-of-mass energy of $\sqrt{s} = 7$ TeV and 8 TeV, respectively. The names of analyzed datasets, split in run ranges, are summarized in Table 5.1.

Dataset	Run Range	$\mathcal{L}$ (fb^{-1})
$\sqrt{s} = 7$ TeV		
/%/Run2011A-16Jan2012-v1/AOD	160431 - 173692	2.241
/%/Run2011B-16Jan2012-v1/AOD	175832 - 180296	2.741
$\sqrt{s} = 8$ TeV		
/%/Run2012A-22Jan2013-v1/AOD	190456 - 193621	0.889
/%/Run2012B-22Jan2013-v1/AOD	193834 - 196531	4.429
/%/Run2012C-22Jan2013-v1/AOD	198022 - 203742	7.152
/%/Run2012D-22Jan2013-v1/AOD	203777 - 208686	7.318

Table 5.1: 2011 and 2012 datasets used in the SM $Zh \to 2\ell 2\tau$ analysis reported together with the corresponding run range and integrated luminosity. The symbol % stands for `DoubleElectron` or `DoubleMu`, corresponding to the names of primary datasets containing dielectron and dimuon triggered events, respectively.

Signal and background processes are simulated with several Monte Carlo (MC) generators (cfr. Section 2.4). PYTHIA [114] is used to simulate the signal production of the Higgs boson in association with the Z boson (Zh) and in association with a generic vector boson V or a heavy-quark pair ($Vh + t\bar{t}h$). The process $Vh + t\bar{t}h$, with $h(125\text{ GeV}) \to WW$, has been considered as a background. For each of these production mechanisms, independent samples have been generated for different values of M_h, ranging from 90 to 145 GeV with a step of 5 GeV. Each of these samples is normalized to the NNLO cross section times branching ratio, as recommended by the LHC Higgs Cross Section Working Group [51] (cfr. Sections 1.2.2 and 1.2.3). The Drell-Yan $Z/\gamma^*(\to \ell\ell)$+jets sample is generated with MADGRAPH [115]. The NNLO cross sections have been calculated with FEWZ [148]. Both light (q = d, u, s) and heavy-flavor (q = c, b) jets are included in the Drell-Yan sample. At generator-level, the two-lepton invariant mass is required to be grater than 50 GeV. Top-pair production samples are simulated with MADGRAPH ($t\bar{t}Z$) and POWHEG [118] ($t\bar{t}$) and the corresponding cross sections have been normalized at the NLO. The diboson WZ sample is generated with MADGRAPH. The $q\bar{q} \to ZZ$ sample is generated with PYTHIA. The WZ and $q\bar{q} \to ZZ$ cross sections are based on MCFM computations [149]. The gluon-induced ZZ background, although technically of NNLO compared to the tree-level ZZ

production, amounts to a non-negligible fraction of the total physical background (whose definition is postponed to Section 5.6). The full NNLO calculation for the ZZ production, which would also take these gluon-induced diagrams into account, is not available: therefore, the contributions are estimated by using a dedicated tool (GG2ZZ [150]), which computes the $gg \to ZZ$ cross section. MADGRAPH, POWHEG and GG2ZZ have been interfaced with PYTHIA for the parton shower and hadronization. All generators are interfaced with TAUOLA [119] for the simulation of tau lepton decays (cfr. Section 2.4). Table 5.2 reports the signal and background simulated processes, specifying the MC generator used and the total cross sections.

Process	Generator	Cross-section (pb) $\sqrt{s} = 7$ TeV	Cross-section (pb) $\sqrt{s} = 8$ TeV	Sample name
Signal				
Zh, $h \to \tau\tau$	PYTHIA	-	$[9.2 - 0.7]$ fb	$M_h = 90$-145 GeV
$Vh + t\bar{t}h$, $h \to \tau\tau$	PYTHIA	$[54 - 3.6]$ fb	$[66 - 4.5]$ fb	$M_h = 90$-145 GeV
Background				
$Vh + t\bar{t}h$, $h(125\,\text{GeV}) \to WW$	PYTHIA	5.28 fb	6.50 fb	-
VV				
$WZ \to 3\ell\nu$	MADGRAPH	0.868	1.057	`WZJetsTo3LNu`
$q\bar{q} \to ZZ \to 4\ell$	PYTHIA	0.106	0.130	`ZZTo4L`
$gg \to ZZ \to 4\ell$	GG2ZZ	-	0.048	`GluGluToZZTo4L`
$gg \to ZZ \to 2\ell 2\ell'$	GG2ZZ	-	0.012	`GluGluToZZTo2L2L`
$t\bar{t}$+X				
$t\bar{t}$	POWHEG	17.32	23.64	`TTTo2L2Nu2B`
$t\bar{t}Z$	MADGRAPH	0.139	0.208	`TTZJets`
Z+jets				
$Z/\gamma^* \to \ell\ell + \leqslant 4$ jets	MADGRAPH	3048	3503.7	`DYJetsToLL`

Table 5.2: Monte Carlo samples used for the SM $Zh \to 2\ell 2\tau$ analysis. The names of the generators used for the event simulation are reported. MADGRAPH, POWHEG and GG2ZZ are interfaced with PYTHIA for the parton shower and hadronization. All generators are interfaced with TAUOLA [119] for the simulation of tau lepton decays. Cross sections are reported for both $\sqrt{s} = 7$ TeV and 8 TeV.

Pileup events are properly simulated and added to the generated hard processes. The distribution of pileup events in the simulated sample is different with respect to that in the real data, which is computed based on the luminosity profile and the inelastic p-p scattering cross section, and, generally, consists of less pileup events than those simulated. Therefore, simulated events are reweighted according to the number of pileup events in MC and the distribution of pileup events in data, trough the so-called *pileup reweighting* procedure. The effect of the pileup-reweighting is shown in Figure 5.1.

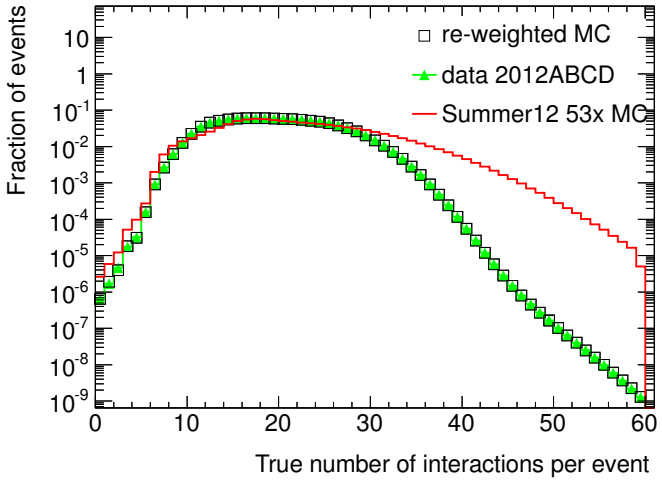

Figure 5.1: Comparison between the distribution of the true number interaction of events in 2012 data (green) and no-pileup reweighted one in simulation (red). After applying the pileup-reweighting procedure on simulated events, the corresponding distribution (empty squares) matches the one in data [AC].

5.3 Trigger selection

The Zh analysis strategy exploits light leptons coming from the Z decay in order to trigger the event. Double lepton triggers are used, as anticipated in Section 5.2. Muon and electron trigger efficiencies are measured, from data, using the tag-and-probe method [151]. A tight identification selection is applied to one object (the tag), and the efficiency is then measured with the second object (the probe), loosely selected. The efficiency is calculated as $\epsilon = N_{pass}/N_{tot}$, where N_{pass} is the number of probes passing the selection of which the efficiency is measured, and N_{tot} is the number of all probes. The tag does not enter the efficiency calculation, but ensures to select a high purity sample. The full description of the efficiency measurement is given in [151]. Given that the trigger algorithms emulated in the MC simulation differ from those running online, trigger efficiencies are estimated with the same method in simulation and, then, data-to-simulation correction factors, and the corresponding uncertainties, have been computed (cfr. Appendix D) and applied to simulated events. The HLT trigger paths used in the 7 and 8 TeV analyses are reported in Table 5.3. For every HLT path, the corresponding L1 trigger seeds (cfr. Sec-

tion 2.2.7), which consists of single or double lepton L1 triggers with p_T and η thresholds changing with time, are also reported in Table 5.3.

$\mu\mu$ channel		ee channel	
$\sqrt{s} = 7$ TeV			
HTL paths	L1 seeds	HTL paths	L1 seeds
DoubleMu7	L1_DoubleMu3	Ele17_X_Ele8_X*	L1_SingleEG12
Mu13_Mu8	L1_DoubleMu_5	Ele17_Y_Ele8_Y**	L1_SingleEG12
Mu17_Mu8	L1_DoubleMu_5	-	-
$\sqrt{s} = 8$ TeV			
HTL paths	L1 seeds	HTL paths	L1 seeds
Mu17_Mu8	L1_DoubleMu_10_3p5 OR L1_DoubleMu_12_5	Ele17_Y_Ele8_Y**	L1_DoubleEG_15_10
Mu17_TkMu8	L1_DoubleMu_10_3p5 OR L1_DoubleMu_12_5	-	-

* X = CaloIdL_CaloIsoVL
** Y = CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL

Table 5.3: 2011 and 2012 HLT paths used for the analysis, and their corresponding L1 seeds.

5.4 Object selection

The definition of the objects used in the analysis is described in the following. It is totally based on the algorithms described in Chapter 3 and 4.

PRIMARY VERTEX The DA algorithm (cfr. Section 3.1.2) is the method used in CMS to reconstruct primary vertices. Out of several good primary vertex candidates, only one single vertex is selected for each event: it has to be identified from a fit with at least four degrees of freedom, and the distance of its position from the beamspot has to be smaller than 24 cm in z and 2 cm in the transverse coordinate r. Out of all vertices satisfying these requirements, the one with the highest sum of the squares of the transverse momenta of tracks assigned to it ($\sum p_T^2$), is chosen as the primary vertex of the signal event.

ELECTRONS The electrons used in this analysis are not purely particle-flow reconstructed, as it will be the case for the analysis presented in the following chapter, but they are reconstructed according to the standard CMS methods (cfr. Section 3.3) after which the particle-flow electron identification and isolation criteria are applied (cfr. Section 3.5.1). The drawbacks of considering standard electron (and muon) collections is that in the reconstruction process, some low level detector information (e.g. energy deposits in the calorimeters)

might be used to reconstruct several candidates belonging to different collections. The resolution of these ambiguities, known as *cross-cleaning* process, is totally left to the analyzer. The cross-cleaning among electron, muon and tau collections has been performed "manually" in the context of this analysis (cfr. Section 5.5).

The three particle-flow electron identification working point selections (cfr. Table 3.1) are complemented by a conversion rejection cut, which consists of requiring:

a) a low probability ($< 10^{-6}$) for the electron track and any nearby track to originate from a same displaced vertex;

b) no missing hits in any of the pixel or tracker modules crossed by the electron track.

Additionally, the electron track impact parameter is required to be compatible with the primary vertex both in the longitudinal ($d_z < 0.1$ cm) and in the transverse plane ($d_{xy} < 0.02$ cm). The electron particle-flow relative isolation is computed as the ratio of the expression in Equation 3.5 to the electron transverse momentum. Additional requirements are applied in order to improve the performance of the isolation algorithm: every charged particle entering the computation of the isolation variable should be matched to the event vertex within $d_z < 2$ mm, and its direction should lie outside of a veto cone of radius 0.01 (0.015) around the electron momentum, if the electron is within the barrel (endcap) acceptance; no such additional requirements are applied to neutral hadrons; photons are, instead, required to be outside a veto cone of radius 0.08 around the electron rejection. Depending on the final state, different relative isolation values have been used, resulting from an optimization procedure that will be described later. The identification and isolation efficiencies for electrons have been measured using $Z \to ee$ data events, by using the tag-and-probe method [151]. Data-to-simulation correction factors, and their corresponding uncertainties, have been derived in p_T and η bins, for both 2011 and 2012 data, and used to rescale the simulated samples (cfr. Appendix D).

MUONS As for the electrons, also in this case the standard muon reconstruction (cfr. Section 3.4) has been used. The particle-flow muon identification (cfr. Section 3.5.2) and isolation criteria, however, can be still applied. The two particle-flow muon identification working points, loose and tight, have been summarized in Table 3.2. The muon particle-flow relative isolation, as for the electron one, is computed by dividing the expression in Equation 3.5 by the

muon transverse momentum. The difference with respect to the electron isolation resides in the definition of the veto cones: every charged particle entering the computation of the isolation variable should be matched to the event vertex within $d_z < 2$ mm, and its direction should lie outside of a veto cone of radius 0.0001 around the muon momentum; neutral hadrons and photons should have $p_T > 0.5$ GeV and point outside a veto cone of size 0.01 around the muon direction. Also for muons, the identification and isolation efficiencies have been measured by applying the tag-and-probe method to $Z \to \mu\mu$ data events [151]. Data-to-simulation correction factors, and their corresponding uncertainties, have been derived in p_T and η bins, for both 2011 and 2012 data, and used to rescale the simulated samples (cfr. Appendix D). Details on p_T and η selections, as well as on identification and isolation requirements applied depending on the final state, will be provided later.

TAUS Hadronically decaying taus are reconstructed with the HPS algorithm (cfr. Chapter 4). In order to separate genuine hadronically decaying taus from quark and gluon jets, the decay mode reconstruction and the cut-based isolation discriminators have been applied to all taus, regardless of the final state considered. In particular, as it will be explained later, the working point chosen for the cut-based isolation discriminator is the loose one for the $\ell\tau_h$ and the medium one for the $\tau_h\tau_h$ final states. Concerning the reduction of the probability to reconstruct genuine electrons and muons as taus, different working points for the anti-electron and anti-muon discriminators have been chosen for the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states, depending on whether the dominant background contributions arise from $Z \to ee$+jets ($e\tau_h$) or from $Z \to \mu\mu$+jets ($\mu\tau_h$) or elsewhere ($\tau_h\tau_h$).

The analysis strategy is based on the usage of the SVfit algorithm [144] for the reconstruction of the τ-pair mass (cfr. Section 4.6). During the first phase of this analysis, the visible mass of the τ-pair was considered as observable. The choice of using the SVfit mass has led to a better signal-to-backgrounds separation than the visible mass alone, yielding to an improvement in the final expected limit (considering all the eight channels combined and both 7 and 8 TeV data) of about 30%, as can be observed in Figure 5.2.

JETS AND MISSING TRANSVERSE ENERGY Jets are reconstructed with the Anti-k_T algorithm (cfr. Section 3.5.4) using the list of particle-flow particles as inputs for the clustering process. To properly reconstruct the jet transverse momentum, jet energy corrections and resolutions factors have been applied to account for spurious particle contamination and to correct for the response, non-linearity and inhomogeneity of the calorimeters (cfr. Section 3.5.4). In this analysis, jets are required to have $p_T > 30$ GeV, $|\eta| < 4.7$ and, in order

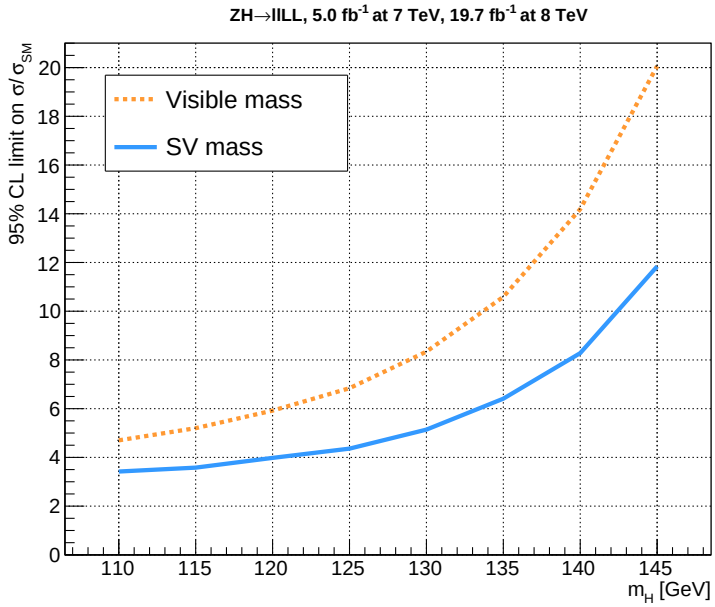

Figure 5.2: Comparison between expected limits computed considering all eight channels and both 7 and 8 TeV data, and using the ditau visible mass (orange) and the SVfit mass (blue) as observable. The usage of the SVfit mass leads to an improvement of the expected combined limit of about 30% [AC].

to separate real jets from detector noise, they have to satisfy the set of quality requirements summarized in Table 3.3. Concerning the pileup jet identification requirements, the loose working point has been used (cfr. Section 3.5.4). Jets originating from the fragmentation of b and c quarks are tagged according to the medium working point of the CSV discriminator, CSVM (cfr. Section 3.5.4). In particular, all jets passing the CSVM discriminator and with $p_T > 20$ GeV and $|\eta| < 2.4$ are tagged as b jets. The MVA particle-flow missing transverse energy has been used (cfr. Section 3.5.5).

5.5 Event selection

The signal selection consists of requiring four tightly identified leptons, as defined in Section 5.4. In particular a pair of electrons or muons is required to be compatible with a Z decay, while the other two leptons are selected to

build a Higgs boson candidate. Before going into the details of the selection criteria applied channel-by-channel, the additional topological requirements, which are common to all channels and whose aim is to further suppress the background arising from misreconstructed leptons, are listed below:

- all objects have to originate from the same primary vertex;
- cross-cleaning performed among electron, muon and tau collections, removing all those objects overlapping within $\Delta R < 0.3$;
- all four objects are required to be separated from each other by $\Delta R > 0.1$;
- events with b jets with a distance of $\Delta R > 0.4$ with respect to the electrons and muons in the event, are rejected in order to suppress the $t\bar{t}$ background;
- a veto is imposed on extra muons and electrons in the event, to suppress the overlap with $ZZ \to 4\mu/2\mu 2e/4e$ events. In particular, the event is rejected if there is any extra loosely identified muon or electron with $p_T > 10$ GeV and with a particle-flow relative isolation $I_{rel}^{PF} < 0.3$.

5.5.1 Z boson selection

A Z boson candidate is selected by requiring two opposite-sign electrons or muons, satisfying the following criteria:

$Z \to \mu\mu$

- two opposite-sign particle-flow loose muons (cfr. Table 3.2), both firing one of the HLT muon trigger paths listed in Table 5.3;
- $p_T > 20$ GeV (leading), 10 GeV (subleading) and $|\eta| < 2.4$;
- muon particle-flow relative isolation $I_{rel}^{PF} < 0.3$.

$Z \to ee$

- two opposite-sign particle-flow very loose electrons (cfr. Table 3.1), both firing one of the HLT electron trigger paths listed in Table 5.3;
- $p_T > 20$ GeV (leading), 10 GeV (subleading) and $|\eta| < 2.5$;
- electron particle-flow relative isolation $I_{rel}^{PF} < 0.3$.

The invariant mass of the reconstructed Z has to be in the range [60,120] GeV. If more than one Z candidate satisfies the requirements listed above, the one with the minimum absolute difference between its reconstructed mass value and the true one [29] is chosen.

5.5.2 h boson selection

The identification requirements applied on hadronically decaying taus have been chosen according to those used for the $h \to \tau\tau$ Run I analysis [152]. Concerning the decay mode finding requirements, the isolation and the anti-lepton rejection criteria used, the reader can refer to Sections 4.2.1, 4.2.2 and 4.2.3, respectively.

The $h \to \mu\tau_h$ and $h \to e\tau_h$ channels are mostly contaminated by Z+jets and WZ background events. In particular, the signal final state can be reproduced by a Z decaying into a pair of electrons or muons, an additional real muon coming from either a b jet or a W boson, and a jet that gets reconstructed as a hadronic τ. In order to reduce these backgrounds the following selection criteria are applied:

$h \to \mu\tau_h$

- one tightly identified particle-flow muon (cfr. Table 3.2) with $p_T > 10$ GeV and $|\eta| < 2.4$;
- muon particle-flow relative isolation $I^{PF}_{rel} < 0.3$;
- one HPS hadronic tau with $p_T > 15$ GeV and $|\eta| < 2.3$;
- the hadronic tau has to satisfy the decay mode finding test, the requirements on the anti-μ tight and the anti-e loose cut-based discriminators;
- tau particle-flow absolute isolation $I^{PF}_{abs} < 2$ GeV (corresponding to the cut-based loose isolation discriminator working point);

$h \to e\tau_h$

- one tightly identified particle-flow electron (cfr. Table 3.1) with $p_T > 10$ GeV and $|\eta| < 2.5$;
- electron particle-flow relative isolation $I^{PF}_{rel} < 0.2$;
- one HPS hadronic tau with $p_T > 15$ GeV and $|\eta| < 2.3$;

- hadronic tau has to satisfy the decay mode finding, the anti-μ loose cut-based and the anti-e tight MVA discriminators;
- tau particle-flow absolute isolation $I^{PF}_{abs} < 2$ GeV.

The fully hadronic final state $h \to \tau_h\tau_h$ is mostly affected by background due to jets being reconstructed as hadronic taus. The main background source comes from events events in which a Z is produced in association with two or more jets. This background is reduced by applying the following selection criteria:

$h \to \tau_h\tau_h$

- two HPS hadronic tau with $p_T > 15$ GeV and $|\eta| < 2.3$;
- hadronic taus have to satisfy the decay mode finding, the anti-μ and the anti-e loose cut-based discriminators;
- tau particle-flow absolute isolation $I^{PF}_{abs} < 1$ GeV (corresponding to the cut-based medium isolation discriminator working point).

The fully leptonic channel $h \to e\mu$ is very clean but it has the lowest branching ratio with respect to all the $h \to \tau\tau$ decays considered so far. The remaining small backgrounds come from Z+jets, WZ and $t\bar{t}Z$ events, which are reduced by requiring:

$h \to e\mu$

- one loosely identified particle-flow muon (cfr. Table 3.2) with $p_T > 10$ GeV and $|\eta| < 2.4$;
- one loosely identified particle-flow electron (cfr. Table 3.1) with $p_T > 10$ GeV and $|\eta| < 2.5$
- muon and electron particle-flow relative isolation $I^{PF}_{rel} < 0.3$.

For all the four $h \to \tau\tau$ decay channels considered, it has been additionally required that the τ-pair decay objects have opposite charge (OS). Furthermore, in order to reject the copious Z+jets background characterized by a softer lepton transverse momentum spectrum than the signal one, a requirement on the scalar sum of the τ-pair visible decay objects transverse momenta, labelled as L_T, is applied. The actual requirements on the L_T variable, and on the lepton isolation, depend on the channel and have been determined with an optimization procedure that will be explained in the following section.

5.5.3 Selection criteria optimization

The optimal values for the requirements on the lepton isolation and L_T variable are determined by minimizing the expected 95% CL upper limit (cfr. Appendix B) in correspondence of a 125 GeV SM Higgs boson. The lepton isolation and the L_T thresholds are scanned simultaneously over a range of possible optimal values, in order to find the optimal ones for each final state. The expected limits corresponding to different isolation and L_T thresholds are shown for each ditau final state in Figure 5.3.

As already anticipated in Section 5.5.2, the isolation values found to be optimal, and then used for the final event selection, are: $I^{PF}_{rel} < 0.3$ for the muon in the $e\mu$ and $\mu\tau_h$ final states; $I^{PF}_{rel} < 0.3$ and 0.2 for the electron in the $e\mu$ and $e\tau_h$ final states, respectively. The medium working point for the tau cut-based isolation, corresponding to $I^{PF}_{abs} < 1$ GeV, results to be optimal for the $\tau_h\tau_h$ final state. The tau isolation for $\ell\tau_h$ final states has not been optimized, but, among the three available working points for the tau cut-based isolation discriminator, the loosest one has been preferred. This choice is motivated by the fact that, being the $\ell\tau_h$ final states cleaner than the $\tau_h\tau_h$ one, a looser isolation allows to gain in signal acceptance while limiting the increase in the residual background. The L_T thresholds found to be optimal are summarized in Table 5.4.

	$\ell\ell\mu\tau_h$	$\ell\ell e\tau_h$	$\ell\ell\tau_h\tau_h$	$\ell\ell e\mu$
$L_T >$	45 GeV	30 GeV	70 GeV	25 GeV

Table 5.4: L_T thresholds optimized for each ditau final state.

5.6 Physical background

Those processes that are characterized by four genuine prompt leptons in the final state, hence reproducing the signal-like topology, and thus more difficult to separate from the signal, are referred as physical background (often these backgrounds are also-called "irreducible"). All the backgrounds belonging to this class are estimated from simulation. The $q\bar{q} \to ZZ$ processes constitute the most important physical background. The other contributions, though small, come from the processes $gg \to ZZ$, $t\bar{t}Z$, where the Z decays into an electron or a muon pair and both top quarks decay leptonically (to e, μ or τ). It has been checked that, despite the b-jet veto, about 4% of the $t\bar{t}Z$ background events satisfy the signal selection in the $e\mu$ channel. Also the processes $Vh +$

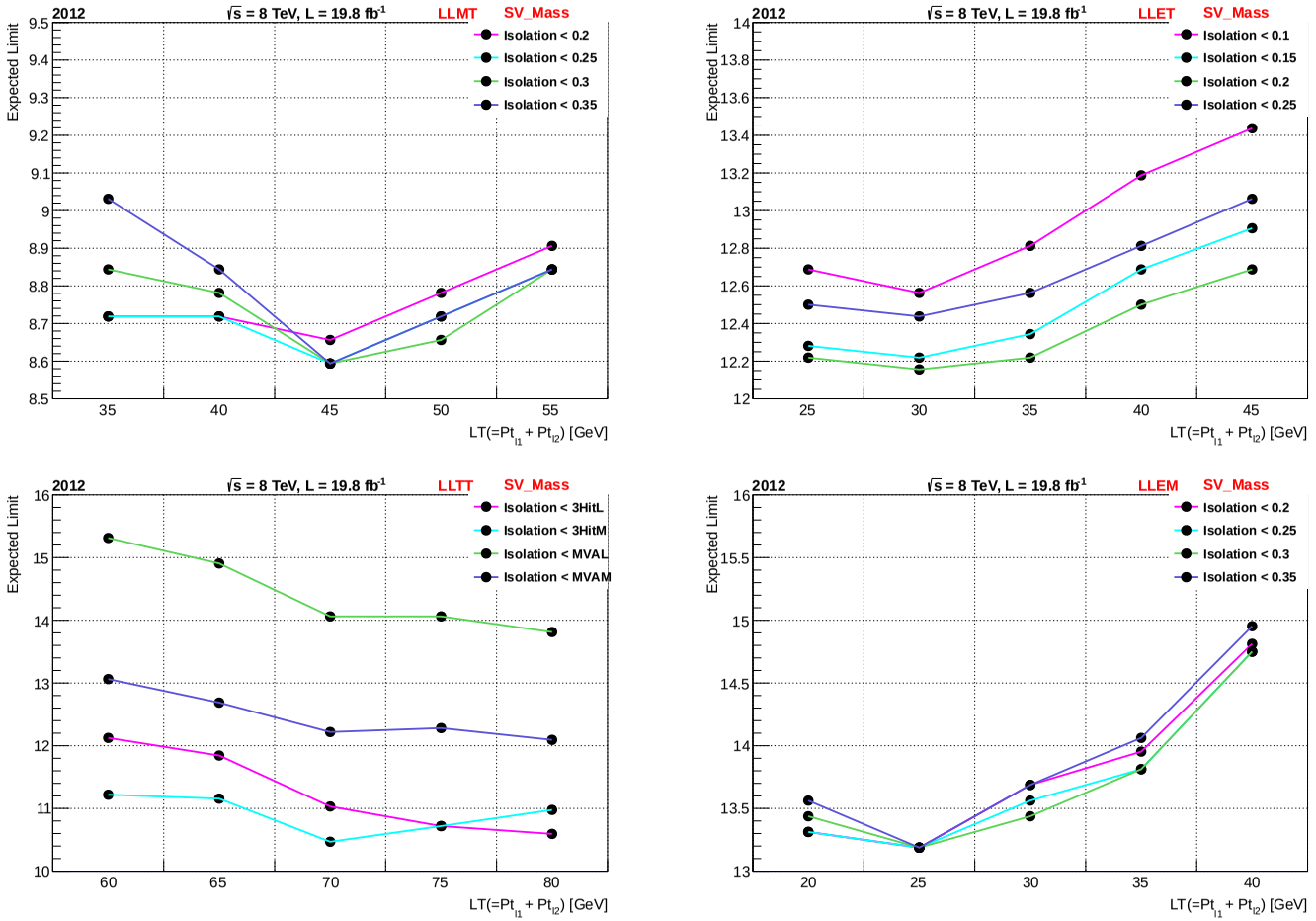

Figure 5.3: Expected 95% CL upper limit computed at $M_h = 125$ GeV for different lepton isolation and L_T values, in the $\mu\tau_h$ (top left), $e\tau_h$ (top right), $\tau_h\tau_h$ (bottom left) and $e\mu$ (bottom right) final states [AC].

$t\bar{t}h$ with $h(125\,\text{GeV}) \to WW$, which are considered as physical backgrounds, are found to be non negligible in the $e\mu$ channel.

5.7 Fake background

The background processes where at most three genuine prompt leptons are present in the final state are referred to as fake backgrounds. These processes can contribute with events in the final selection if one or more jets (or one of their component particles), or non-prompt genuine leptons arising from heavy-flavour hadron decays, are misidentified as isolated leptons, or if photons are misidentified as electrons. The main processes contributing to this jet-induced source of background are Z+jets and WZ+jets productions. Both processes are characterized by a pair of genuine leptons coming from the Z decay. While in the Z+jets process it is more likely to have two heavy or light flavor jets misreconstructed as genuine leptons, the so-called *fake* leptons, the WZ+jets production can more likely contribute with one jet reconstructed as a genuine lepton. Smaller contributions arise from W+jets (three fake leptons) and $t\bar{t}$ (two or three fake leptons) productions. The fake background is not well modelled in simulation, therefore it has been estimated using a method based on experimental data. This data-driven method, called *fake rate* method, is described in the following.

The fake rate method consists of a first step in which the measurement of the misidentification probability to reconstruct a jet as a genuine lepton, the fake rate, is performed. More precisely, the fake rate is the probability that a genuine jet satisfying loose lepton selection criteria, also satisfies tight selection criteria corresponding to those required in the final event selection of the analysis. In the following, unless differently specified, the adjectives "loose" and "tight" will be exclusively used with the general meaning that has been just described, not forgetting that the corresponding equivalence in terms of selection working points depends on the lepton flavor and on the final state.

The probability to reconstruct a genuine jet as a lepton, in the context of a given analysis, is assumed to be only dependent on the transverse momentum and pseudorapidity of the jet. The dependency on the rest of the event is reduced by choosing a control sample that is topologically similar to the final state required in the final analysis selection. Fake rates are measured for each lepton flavor independently by using a data control-sample in which a set of loose leptons is selected. The control data sample is selected such that the majority of these loose leptons are genuine jets. The one chosen for this

analysis is built by selecting a Z candidate, as described in Section 5.5, and two additional loose leptons having the same charge (SS). In the second step, the measured fake rates are used to extract the fake background contribution in the signal region. Both fake rate method steps are going to be described in the next section.

5.7.1 Lepton fake rates determination

TAU FAKE RATES are measured using two fake-object enriched control-regions defined in the following way:

- $\boldsymbol{\tau_h\tau_h}$ **topology:** Z+ two loose taus having the same charge and satisfying the same signal selection criteria applied in the $\tau_h\tau_h$ final state, but the isolation.
- $\boldsymbol{\ell\tau_h}$ **topology:** Z+ one electron or muon (selected as described in Section 5.5 for the $\ell\tau_h$ channels) + one loose tau having the same charge of the previously selected light lepton and satisfying the same selection criteria used in $\ell\tau_h$ final states, but the isolation.

By counting the number of loose taus that also satisfy the tight isolation requirements in the signal region in bins of p_T and η ($|\eta| < 1.4$ and $1.4 < |\eta| < 2.3$) of the closest particle-flow jet[1] (cfr. Section 3.5.4), the fake rates are obtained. The choice of parametrizing the fake rate also in bins of η has been driven by the fact that the tau fake rate is almost flat in the barrel, while it increases by about 50% in the endcaps. However, despite this observed trend, just a 1-2% difference with respect to an only p_T-dependent fake rate has been observed. In a given η bin, the p_T dependent measured fake rates are then fitted with the exponential function $f(p_T) = a + b \cdot \exp(c \cdot p_T)$. The misidentification probabilities measured in 2012 data for the loose working point of the tau cut-based isolation discriminator (used in the final selection of the $\ell\tau_h$ channels), with the corresponding fit overlaid, are reported in Figure 5.4 ($\ell\tau_h$ and $\tau_h\tau_h$ same-sign topologies results are shown on top and bottom, respectively). Comparisons with tau fake rates measured in 2011 data have shown that, for $|\eta| < 2.3$ and $p_T < 50$ GeV, the fake rate decreases by about 20% as the number of pileup decreases. In the same acceptance region and for $p_T > 50$ GeV, tau fake rates measured in 2011 and 2012 behave in the same way.

[1] Generally valid for each lepton flavor, the jet used in the parametrization must satisfy the minimum $\Delta R(\text{lepton}, \text{jet}) < 0.5$. If no jet is found, the lepton p_T and η are used.

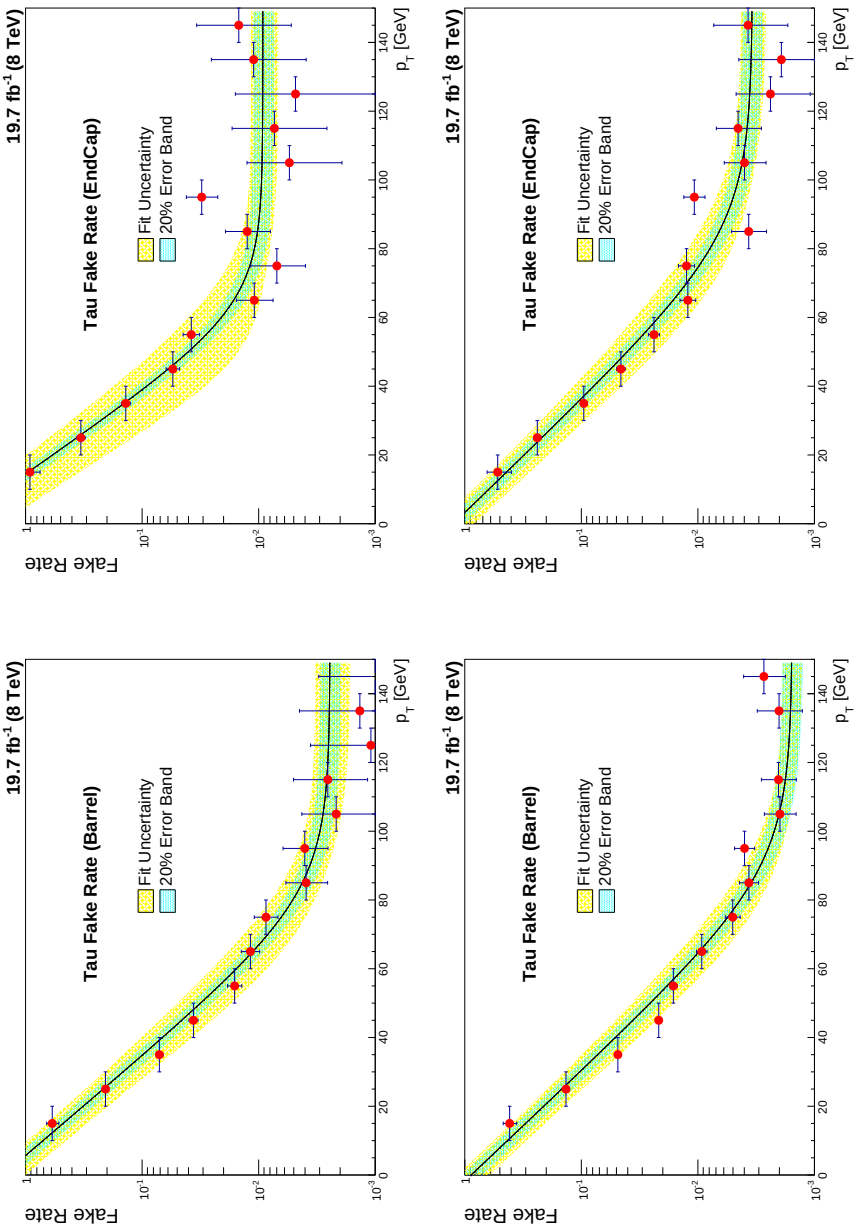

Figure 5.4: Tau fake rate measured as a function of the closest-jet-to-tau p_T using 2012 data with the corresponding exponential fits overlaid. Fake rates computed for $|\eta| < 1.4$ and $1.4 < |\eta| < 2.3$ are reported on left and right, respectively, and those measured in $\ell\tau_h$ and $\tau_h\tau_h$ same-sign topologies on top and bottom, respectively. Both statistical (yellow area) and systematic (azure area) uncertainty bands are shown. More details about the estimation of the fake background uncertainties will be provided in Section 5.8 [AC].

ELECTRON AND MUON FAKE RATES are measured in $Z + \ell\tau_h$ same-sign topology control-regions[2]. In particular, a Z and an hadronic tau are selected as described in Section 5.5 for each of the two $\ell\tau_h$ channels. Then, one loose lepton (electron or muon) having the same charge of the previously selected tau and satisfying the same selection criteria used in the final selection for the $\ell\tau_h$ final states, but identification and isolation, is required. The number of loose leptons passing the corresponding tight identification and isolation criteria used in the signal region is then measured as a function of the closest-jet-to-lepton p_T. In order to suppress the contamination to the fake rate measurement of genuine prompt leptons from the WZ process, are removed by requiring the transverse mass m_T, built from the lepton ℓ and the $\not{E}_T$ (cfr. Equation 4.6), to be less than 30 GeV. Events including extra identified and isolated electrons or muons (cfr. Section 5.5) are vetoed in order to remove genuine leptons from the ZZ process. Examples of electron and muon fake rates measured in 2012 data are shown in Figure 5.5. Such results correspond to the final selection criteria applied in the $e\mu$ channel (cfr. Section 5.5). Similarly to the tau fake rate, the measured muon and electron fake rates are fitted with exponential functions.

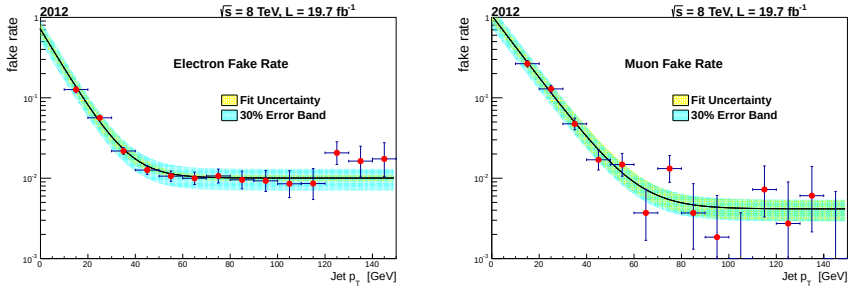

Figure 5.5: Electron fake rate (left) and muon fake rate (right) measured as a function of the closet-jet-to-lepton p_T in 2012 data. Tight electrons and muons are required to satisfy the same signal criteria applied in the $e\mu$ channel (cfr. Section 5.5). The corresponding exponential fits are overlaid. Both statistical (yellow area) and systematic (azure area) uncertainty bands are shown in the plots. More details about the estimation of the fake background uncertainties will be provided in Section 5.8 [AC].

[2]For lack of statistics, electrons and muons fake rates have not been computed in the $e\mu$ channel. However, having confirmed the universality of the lepton fake rate with the observation of comparable tau fake rates measured in $\ell\tau_h$ and $\tau_h\tau_h$ final states, it is expected that the electron and muon fake rates measured in $\ell\tau_h$ channels are valid also for the $e\mu$ channel.

5.7.2 Derivation of data-driven background rate estimate using fake rates

The measured fake rates are used to extract the fake background contributions in the signal region. The technique to perform this measurement is based on the definition of three control-regions in which a Z candidate and two leptons having opposite charge are required. The strategy consists in using the misidentification pro abilities to reweight all those events in which loose leptons fail the tight signal analysis requirements. The control regions are defined as follows:

- **Control Region 00 (CR_{00}):** both leptons pass the loose identification criteria but not the tight required in the signal region;
- **Control Region 01 (CR_{01}):** one lepton passes the tight identification requirements of the signal region, the other passes only the loose ones. In this control region the loose lepton is the highest-p_T tau in the $\tau_h\tau_h$, the tau in the $\ell\tau_h$ and the muon in the $e\mu$ final state,
- **Control Region 10 (CR_{10}):** as in the control region CR_{01}, but here the loose lepton is the lowest-p_T tau in the $\tau_h\tau_h$, the light lepton in the $\ell\tau_h$ and the electron in the $e\mu$ final state.

The analytical formula to estimate the fake background contributions to the signal regions is derived as follows. Let N^{Z+jets} indicates the number of Z+jets events in which two fake leptons satisfy loose selection requirements. This sample can be split in four subsamples, each containing the following numbers of events:

$$N_{00}^{Z+jets} = N^{Z+jets}(1-f_1)(1-f_2) \tag{5.1}$$

is the number of events populating the CR_{00} in which both loose leptons do not satisfy the tight requirements;

$$N_{01}^{Z+jets} = N^{Z+jets}(1-f_1)f_2 \tag{5.2}$$

is the number of events populating the CR_{01} in which only one loose lepton (whose flavor depends on the channel, as explained before) satisfies also the tight requirements;

$$N_{10}^{Z+jets} = N^{Z+jets}f_1(1-f_2) \tag{5.3}$$

is the number of events populating the CR_{10} in which only the other loose lepton satisfies the tight requirements;

$$N_{11}^{Z+jets} = N^{Z+jets} f_1 f_2 \tag{5.4}$$

is the number of events populating the signal region CR_{11} in which both loose leptons satisfy the tight requirements.

In Equations 5.1, 5.2, 5.3 and 5.4, f_1 and f_2 represent the fake rates of the two leptons involved, whose flavor depends on the channel considered. Given that it can be safely assumed that the CR_{00} region selected in data is indeed fully dominated by Z+jets events, then one can estimate the Z+jets background contribution to the signal region using Equation 5.1:

$$N_{11}^{Z+jets} = N_{00} \frac{f_1 f_2}{(1-f_1)(1-f_2)}, \tag{5.5}$$

where $N_{00} = N_{00}^{Z+jets}$ (for the sake of lightness of notation) is estimated directly from the yield in data in the control region CR_{00}. The number N_{00} can be used to estimate the Z+jets background contribution to the CR_{01} and CR_{10} as well, using Equations 5.2, 5.3 and 5.4:

$$N_{01}^{Z+jets} = N_{00} \frac{f_2}{(1-f_2)}, \qquad N_{10}^{Z+jets} = N_{00} \frac{f_1}{(1-f_1)} \, . \tag{5.6}$$

The WZ+jets process, instead, can be assumed to contribute only to the control regions CR_{01} and CR_{10}, as well as to the signal region CR_{11}, but not to CR_{00}. Therefore, given that the CR_{01} and CR_{10} control regions are a mixture of WZ+jets and Z+jets events ($N_{01} = N_{01}^{WZ+jets} + N_{01}^{Z+jets}$, $N_{10} = N_{10}^{WZ+jets} + N_{10}^{Z+jets}$), the number of events N_{01} and N_{10} estimated from data for each control-region and the already estimated Z+jets background, can be used to extract the contribution of the WZ+jets background in the signal region. In particular:

$$\begin{aligned} N_{11,01}^{WZ+jets} &= \left(N_{01} - N_{01}^{Z+jets} \right) \frac{f_1}{1-f_1} \\ &= \left(N_{01} - N_{00} \frac{f_2}{1-f_2} \right) \frac{f_1}{1-f_1} \\ &= N_{01} \frac{f_1}{1-f_1} - N_{00} \frac{f_1 f_2}{(1-f_1)(1-f_2)} \end{aligned} \tag{5.7}$$

is the contribution to the signal region, coming from the CR_{01}, while

$$\begin{aligned} N_{11,10}^{WZ+\text{jets}} &= \left(N_{10} - N_{10}^{Z+\text{jets}}\right) \frac{f_2}{1-f_2} \\ &= \left(N_{10} - N_{00}\frac{f_1}{1-f_1}\right) \frac{f_2}{1-f_2} \\ &= N_{10}\frac{f_2}{1-f_2} - N_{00}\frac{f_1 f_2}{(1-f_1)(1-f_2)} \end{aligned} \tag{5.8}$$

is the one coming from the control region CR_{10}. Therefore, the final contribution to the signal region will be:

$$\begin{aligned} N_{11} &= N_{11}^{Z+\text{jets}} + N_{11}^{WZ+\text{jets}} \\ &= N_{11}^{Z+\text{jets}} + N_{11,01}^{WZ+\text{jets}} + N_{11,10}^{WZ+\text{jets}} \end{aligned} \tag{5.9}$$

and, making use of all the equations defined above, the final formula used to estimate the fake background contribution to the signal region is:

$$N_{11} = N_{01}\underbrace{\frac{f_1}{1-f_1}}_{w_{01}} + N_{10}\underbrace{\frac{f_2}{1-f_2}}_{w_{10}} - N_{00}\underbrace{\frac{f_1 f_2}{(1-f_1)(1-f_2)}}_{w_{00}} . \tag{5.10}$$

The final analytical formula reported in Equation 5.10 summarizes that, practically, it is sufficient to count the number of events in each control region and "correct" this number with a proper weight w_{ij} depending on the lepton fake rates. It is important to notice that, in this analysis (and contrary to what done in the one described in Chapter 6), the choice has been made that when the contribution of the control region CR_{00} exceeds the sum of the other two (thus yielding to a negative estimate for the overall fake background), the background estimate is assumed to be the contribution from Z+jets only, i.e the expression in Equation 5.5. This pragmatic, though arguable, choice has the advantage of avoiding having to deal with negative background estimates and is by construction conservative. In other words, the fake background estimate is given by:

$$N_{11} = \max\left[(N_{01} \cdot w_{01} + N_{10} \cdot w_{10} - N_{00} \cdot w_{00}),\ (N_{00} \cdot w_{00})\right] . \quad (5.11)$$

As it will be explained in the context of the 2HDM analysis (cfr. Chapter 6), this conservative approach can be abandoned in favor of a more sophisticated statistical method which allow the negative background rates, possibly coming from Equation 5.10, to be properly treated. The fake rate method strategy has been summarized and schematically represented in Figure 5.6.

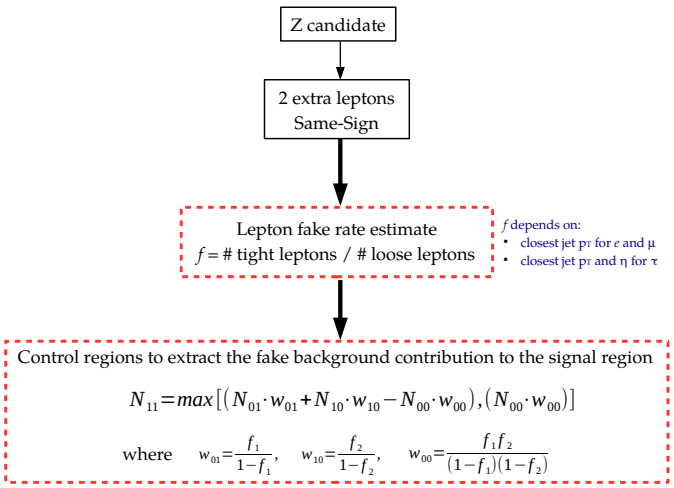

Figure 5.6: Schematic representation of the fake rate method strategy.

5.7.3 Results

Fake background yields are reported, for all final states, in Table 5.5 for 2011 and 2012 data.

The fake background ditau mass shapes have been obtained from data in signal-free regions, characterized by same-sign topologies. In order to reduce at the minimum the statistical fluctuations affecting the shapes, the bias on the mass template has been studied by loosening the lepton isolation and the requirement on the L_T quantity. A shift towards the higher ditau mass region has been observed by loosening the selection on the L_T, therefore it has

	Fake background estimates	
Channel	7 TeV	8 TeV
$\mu\mu e\mu$	0.13±0.09	1.55±0.27
$\mu\mu e\tau_h$	1.02±0.19	7.64±0.64
$\mu\mu\mu\tau_h$	0.88±0.19	4.79±0.47
$\mu\mu\tau_h\tau_h$	0.12±0.07	3.08±0.31
$eee\mu$	0.15±0.08	0.80±0.17
$eee\tau_h$	1.16±0.21	5.68±0.49
$ee\mu\tau_h$	0.34±0.10	2.87±0.31
$ee\tau_h\tau_h$	0.59±0.16	2.46±0.26

Table 5.5: Fake background yields estimated with the fake rate method for each of the eight final states. Results corresponding to both 2011 and 2012 data are reported.

been chosen to loosen the lepton isolation, given that, as can be observed from the ditau mass templates in Figure 5.7, the corresponding shapes are compatible, within the so-called *bin-by-bin uncertainties*[3], to those obtained requiring signal-region tight isolation. In particular, the particle-flow relative isolation for electrons and muons have been required to be less than 2.0 (see mass templates on top left, middle and right of Figure 5.7 for $e\mu$, $e\tau_h$ and $\mu\tau_h$ channels, respectively), and for the tau isolation it has been required that the output of the MVA-based isolation discriminator (with no tau lifetime information included, cfr. Section 4.2.2) to be grater than 0.4 (see mass templates on bottom left, middle and right of Figure 5.7 for $\tau_h\tau_h$, $e\tau_h$ and $\mu\tau_h$ channels, respectively).

5.7.4 Validation of the method

A closure test has been carried out to assess the validity of the data-driven approach to estimate the fake background contribution. The fake background estimation procedure has been validated in a dedicated signal-free region, where events are selected as explained in Section 5.5, with the exception of requiring ditau same-sign pairs. In order to get more statistics, no selection on L_T has been applied and requirements on the isolation values have been loosened. Background shapes have been thus obtained from side bands with the caveat of having used even looser isolation requirements in order to profit from larger statistics. As shown in Figure 5.8, a reasonable agreement is obtained between the direct estimation from data (represented in Figure 5.8 by

[3]In order to take into account possible fluctuations between one bin and another, an uncertainty labelled as *bin-by-bin* is applied to each bin of the mass distribution.

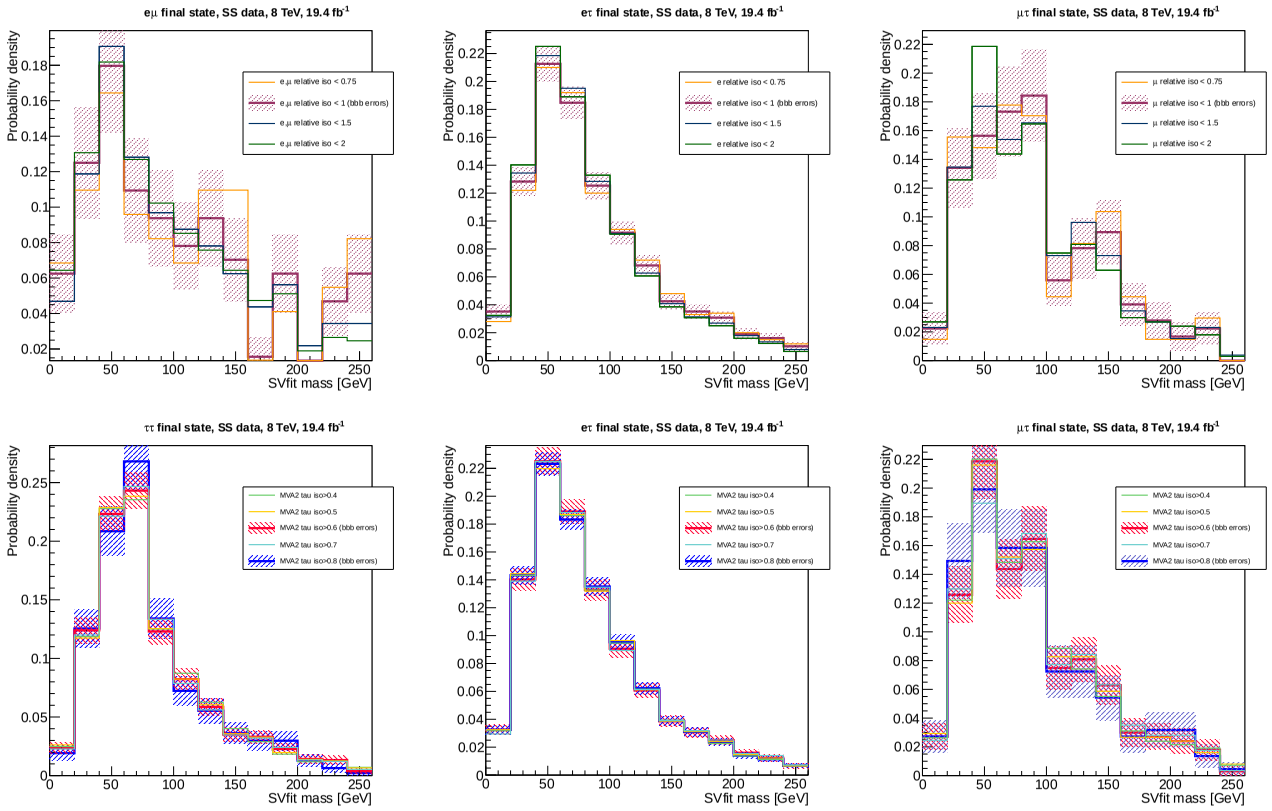

Figure 5.7: Comparison of fake background ditau mass shapes computed in 2012 data, considering same-sign topologies, for different values of the lepton isolation and for different channels. It can be seen that, by requiring $I_{rel}^{PF} < 2$ for electrons and muons (distributions on top left, middle and right for $e\mu$, $e\tau_h$ and $\mu\tau_h$ channels, respectively) and a BDT output grater than 0.4 for the tau MVA-based isolation discriminator (bottom left, middle and right for $\tau_h\tau_h$, $e\tau_h$ and $\mu\tau_h$ channels, respectively), the corresponding shapes are compatible, within bin-by-bin statistical errors, with those obtained requiring signal region tight isolation [AC].

black dots) and the one obtained by applying the fake rate method (solid red line) for both $\ell\tau_h$ and $\tau_h\tau_h$ final states.

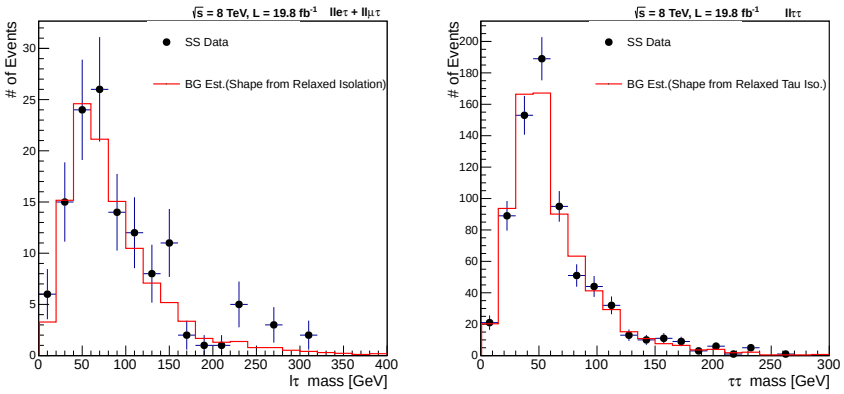

Figure 5.8: Validation of the fake rate method performed by comparing normalizations and shapes obtained directly from same-sign ditau topologies in data with what has been obtained using the fake rate method in the same regions [AC].

Furthermore, the main results obtained with the fake rate method based on four-lepton topology have been compared with those obtained using a three-lepton topology. In the latter case, for the estimation of the electron and muon fake rates, a Z boson is selected (as described in Section 5.5) and, only one single loose lepton is required. The fake rates have been computed by counting the number of loose leptons that satisfy tight selection criteria in bins of p_T of the jet closest to the lepton and the contamination of the genuine prompt leptons from W in WZ events is suppressed by requiring $m_T < 30$ GeV (cfr. Equation 4.6). These probabilities have then been used to extract the fake background normalizations and shapes, as described previously. The final estimate obtained with the three-lepton topology fake rate method has been found compatible with the one based on four-lepton topology within statistical errors.

5.8 Systematic uncertainties

All identified sources of systematic uncertainties are described below.

LUMINOSITY The p-p integrated luminosity uncertainty amounts to 2.2% and 2.6% for 2011 and 2012 data, respectively [153] [154]. It affects the signal and the background yields for which calculated cross sections have been considered. The background processes estimated from data or normalized to the CMS measured cross sections are not affected by this systematic source.

FAKE BACKGROUND ESTIMATION The uncertainties on the fake background is estimated by evaluating an individual uncertainty for each measured lepton fake rate and by propagating it to the full background calculation. The statistical uncertainties of the fit parameters have been used to compute upper and lower bound limits to the central value of the fit, which are shown as yellow bands in Figures 5.4 and 5.5. The uncertainty has been estimated to be 20% for the tau fake rate and 30% for electrons and muons. The higher uncertainty assessed for the light lepton fake rates takes into account also the topology dependence of such rates. The fake rate uncertainties have been then propagated through the background calculation to derive individual systematic uncertainties for each decay channel. The propagation leads to a total uncertainty between 10 to 30%, depending on the channel. The tau fake rate uncertainty in $\tau_h\tau_h$ and $\ell\tau_h$ final states and the electron/muon fake rate in the $\ell\tau_h$ and $e\mu$ final states are considered as uncorrelated, as they refer to different physics objects, and they have been measured independently in different control regions. However, the uncertainties in a given lepton fake rate are considered as fully correlated when they intervene in different final states.

MUON/ELECTRON EFFICIENCIES AND ENERGY SCALES As already explained previously, the efficiency of the muon and electron trigger, identification, and isolation selections, as well as the associated uncertainties, are measured from data with the tag-and-probe method [151]. By using the same method applied also on simulated samples, p_T and η correction factors are derived, with their corresponding uncertainties, and used to rescale the simulated events (cfr. Appendix D). The tag-and-probe method is also used to estimate correction factors and associated uncertainties for the light lepton energy scale. Each trigger efficiency has been varied within its relative uncertainty in order to estimate the effect of the $+1\sigma$ and -1σ shifts on the ZZ background yield. The final estimated uncertainties are 3% and 2% for the muon and electron trigger, respectively. Similarly, the global systematic uncertainties assessed for the electron and muon identification, isolation and energy scales are 2% and 1%, respectively. These uncertainties are considered as uncorrelated between 2011 and 2012 data runs.

HADRONIC TAU IDENTIFICATION AND ENERGY SCALE The hadronic tau identification uncertainty has been assessed to be 6% by CMS using the tag-and-probe method [143]. The uncertainty on the hadronic tau energy scale has been measured to be 3% (cfr. Section 4.5).

UNCERTAINTIES RELATED TO JETS It has been checked that the uncertainties assigned to the jet energy scale and resolution have a negligible impact on the analysis and, therefore, have not been taken into account. Concerning the b jets, the uncertainties related to the b-tagging efficiencies are p_T-dependent and vary from 3% up to 12% (for $p_T < 30$ GeV) [155, 135]. The impact of these uncertainties on the signal and background normalization is found to be of the order of 1% for each channel. The uncertainty on the probability to reconstruct a light quark or gluon jet as a b-jet (mistagging rate) is found to have a negligible impact.

SIMULATED SAMPLE NORMALIZATION The statistical uncertainty arising from the limited size of the different diboson background simulated samples is accounted for by a normalization uncertainty that varies between 3 and 10% and is uncorrelated between all the samples.

THEORETICAL UNCERTAINTIES The theoretical uncertainty on the signal cross section and PDF are 2.7% and 2.5% for 2011 and 2012, respectively [156]. The main uncertainty on the estimation of the ZZ background arises from the theoretical uncertainty on the ZZ production cross section. The Hessian method [157] has been used to compute the uncertainties on the PDFs (CTEQ 6.6 [158]). The method consists in diagonalizing a matrix with a dimension equal to the number of free parameters of the PDF. The CTEQ PDFs are characterized by $N_p = 20$ free parameters, therefore, the diagonalization results in 20 orthonormal eigenvector directions. Subsequently, each eigenvector is evaluated in the "plus" and "minus" direction, corresponding to the deviation of $+1\sigma$ and -1σ of the central value of the fit parameter, respectively, leading to a total of 40 "error" PDFs. The final systematic uncertainty on a measured variable X is then obtained by summing over the 20 couples of error PDFs in the following way:

$$\Delta X = \frac{1}{2}\left(\sum_{i=1}^{N_p}\left[X_i^+ - X_i^-\right]^2\right)^{1/2} \tag{5.12}$$

where the sum runs over the number of PDF parameters, which coincides with the number of eigenvectors, and X_i^+ (X_i^-) is the prediction of the observable X obtained considering the error PDF evaluated in correspondence of the eigenvector i in the plus (minus) direction.

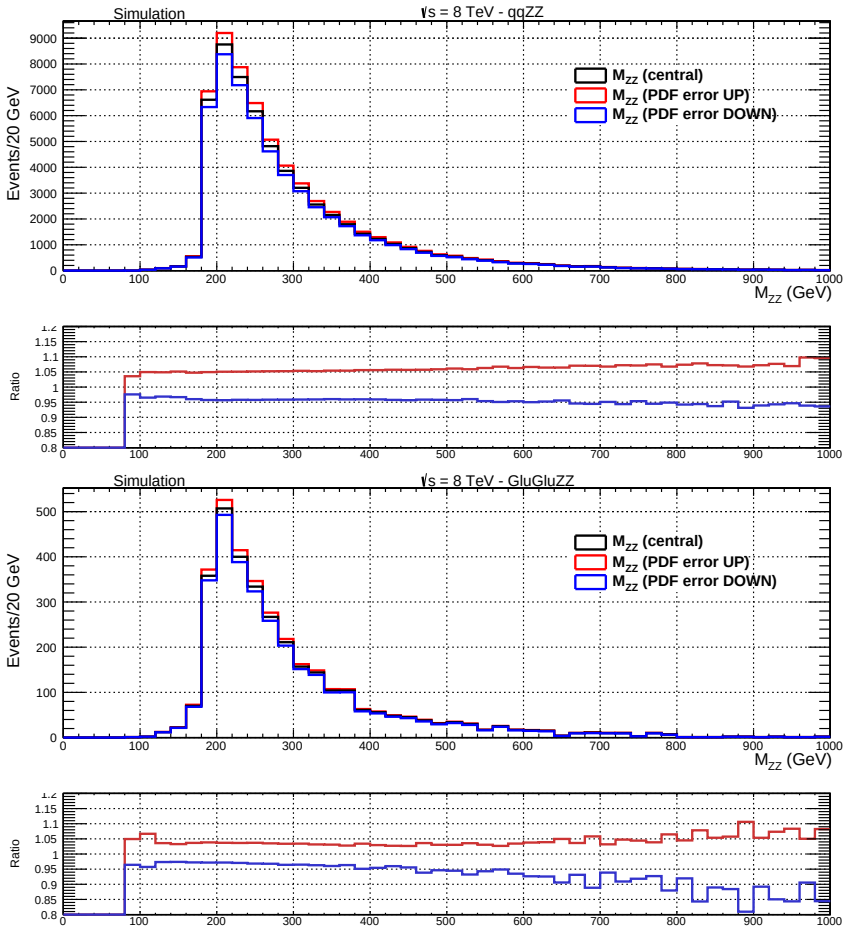

Figure 5.9: PDF uncertainties for (top) $q\bar{q} \to ZZ \to 4\ell$ and (bottom) $gg \to ZZ \to 4\ell$ processes. The red and the blue lines in the ratio pad represent the upward and the downward fluctuations with respect to the unweighted mass spectrum, respectively [A].

The results, obtained by considering the four-lepton invariant mass of the ZZ decay objects as observable, are summarized in Figure 5.9 both for the $q\bar{q} \to ZZ$ and the $gg \to ZZ$ production modes. The value of the uncertainty on

the PDF has been estimated to be 5% (10%) for the $q\bar{q} \to ZZ$ ($gg \to ZZ$) production mode. The α_s uncertainty effect has been neglected because of its small impact on the theoretical uncertainty on the ZZ production cross section. Concerning the QCD factorization scales uncertainties, those computed in the context of the $h \to ZZ \to 4\ell$ analysis [159] have been used. A detailed study on the $q\bar{q} \to ZZ$ sample, in fact, has confirmed that the QCD scale uncertainty has a very similar impact on the two analyses. The uncertainty on the QCD scale amounts to 2.6-6.7% (24-44%) for the $q\bar{q} \to ZZ$ ($gg \to ZZ$) production mode. The uncertainty assigned to the small background from $t\bar{t}Z$ production mode amount to 50% (as measured by CMS [160]).

5.9 Results

The distribution of the SVfit $M_{\tau\tau}$ variable has been used to discriminate between the signal and the different sources of background. The mass range considered for the histogram is $[0, 300]$ GeV, with a bin width of 20 GeV, yielding to fifteen bins in total. The choice of the binning is mainly driven by the SVfit mass resolution, which is indeed roughly 20 GeV at $M_h = 125$ GeV. The observed and expected SVfit mass distributions are reported in Figure 5.10 for all eight channels combined and for 2011 and 2012 data-taking periods separately. The mass distributions for each of the eight channels and data-taking periods separately are reported in Appendix C.

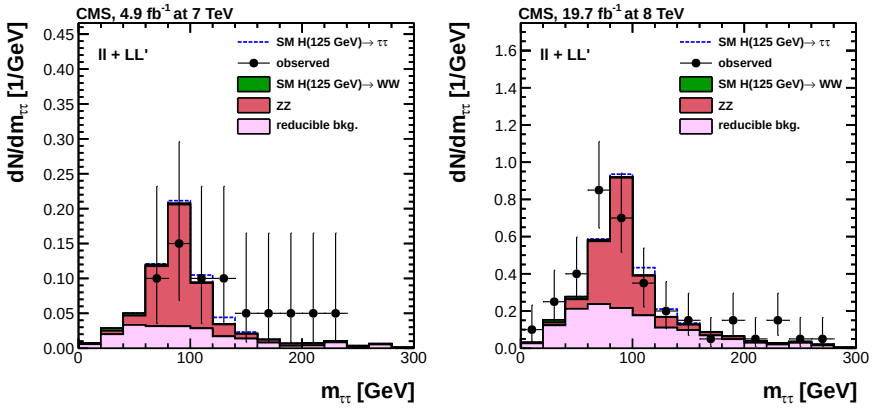

Figure 5.10: Observed and predicted SVfit ditau mass distributions, considering all eight channels combined, for $\sqrt{s} = 7$ (left) and 8 TeV (right) data [AC].

The binned mass distributions of the eight channels have been used for the statistical interpretation of the results, which is based on the methods described in Appendix B. The expected background and signal distributions are fitted to the data through a maximum likelihood fit in order to infer the value of the signal strength modifier μ, which is defined as the ratio between the observed cross section and one expected from the SM predictions. The expected background ditau mass distributions have been normalized according to the maximum likelihood fit outcome in the background-only hypothesis. The resulting distributions from the fit, indicated as *postfit* distributions (which differs from the *prefit* distributions shown in Figure 5.10), are then combined in such a way that each channel contributes by an amount proportional to its signal purity, which is evaluated as $S/(S+B)$. This variable denotes the ratio of the signal to the signal plus background yields in the central range of the $M_{\tau\tau}$ spectrum which is expected to contain the 68% of the signal events for $M_h = 125$ GeV. The integral of the observed data is not affected by this procedure. The SV-fit postfit ditau mass spectra are shown in Figure 5.11, considering all eight channels combined, for 2011 and 2012 data. The postfit ditau mass spectra are reported, for all eight channels and data-taking periods separately, in Appendix C.

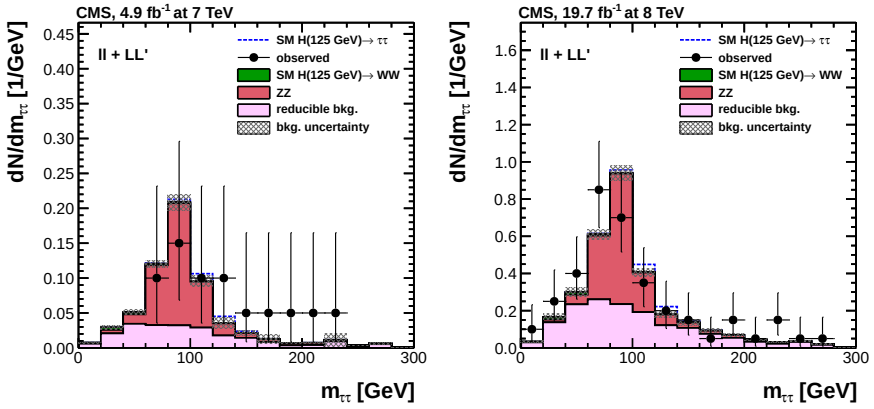

Figure 5.11: Observed and predicted ditau mass distribution, considering all eight channels combined, for 7 TeV (left) and 8 TeV (right) data analyses. The normalization of the predicted background distributions corresponds to the result of the fit. The signal distribution is, instead, normalized to the SM prediction ($\mu = 1$) [15, 161] [AC].

The observed and predicted event yields for all final states are reported in Table 5.6. In this table, the event yields of the predicted backgrounds correspond

to the result of the global fit, while, instead, the signal yields are normalized to the SM prediction. The $S/(S+B)$ values are also reported.

Channel	Signal		Background		Data		$S/(S+B)$	
	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV
$\mu\mu e\mu$	0.051 ± 0.01	0.202 ± 0.008	1.0 ± 0.1	5.1 ± 0.3	3	9	0.100	0.105
$\mu\mu e\tau_h$	0.078 ± 0.004	0.293 ± 0.014	2.2 ± 0.1	12.2 ± 0.6	1	8	0.092	0.081
$\mu\mu\mu\tau_h$	0.111 ± 0.005	0.427 ± 0.021	2.4 ± 0.3	10.5 ± 0.6	2	12	0.103	0.092
$\mu\mu\tau_h\tau_h$	0.073 ± 0.006	0.285 ± 0.022	0.8 ± 5.8	5.8 ± 0.4	0	4	0.195	0.150
$eee\mu$	0.045 ± 0.002	0.185 ± 0.007	1.0 ± 0.0	4.0 ± 0.2	1	4	0.077	0.082
$eee\tau_h$	0.075 ± 0.004	0.279 ± 0.013	2.2 ± 0.1	10.2 ± 0.5	4	13	0.092	0.081
$ee\mu\tau_h$	0.087 ± 0.004	0.385 ± 0.018	1.5 ± 0.1	7.6 ± 0.4	2	11	0.135	0.149
$ee\tau_h\tau_h$	0.061 ± 0.004	0.260 ± 0.020	1.1 ± 0.1	4.8 ± 0.4	1	9	0.127	0.148

Table 5.6: Observed and predicted event yields for all final states, considering the full $M_{\tau\tau}$ region. The event yields of the predicted background distributions correspond to the result of the global fit. The signal yields are normalized to the SM prediction ($\mu = 1$). The $S/(S+B)$ variable denotes the ratio of the signal to the signal plus background yields in the central range of the $M_{\tau\tau}$ spectrum which contains the 68% of the expected signal events for $M_h = 125$ GeV.

5.9.1 Analysis of the fit

In order to know how well the data compare to the model, a study on the output of the maximum likelihood fit has been performed. In the maximum likelihood fit, each source of systematic uncertainty has its associated nuisance parameter, as explained in Appendix B. The difference between the postfit values and the prefit values of a given nuisance parameter (Δz), divided by its prefit uncertainty (σ_{in}) has been analyzed. This quantity, $\Delta z/\sigma_{in}$, is known as *pull*. The maximum likelihood fit also evaluates the a posteriori variance of each nuisance parameter, defined as σ_{out}/σ_{in}, which allows to estimate how much the postfit uncertainties are allowed to vary with respect to the prefit ones. The signs of the pull values is a possible indication that the uncertainty related to the corresponding nuisance parameter has been underestimated or overestimated, assuming that the hypothesis of the fit (background-only) is correct. Small a posteriori variances might indicate that data are constraining a certain variable better than the a priori knowledge. The largest pulls observed in the maximum likelihood fit to data, with the corresponding a posteriori variances, are summarized in Table 5.7. The most affected systematic is the one related to the modeling of the single bin fluctuations in the fake background shape (bin-by-bin uncertainties), especially in those bins with limited

statistics (one can compare these results with the postfit ditau mass distributions in Appendix C).

Channel	$\sqrt{s}$(GeV)	Nuisance	$\Delta z/\sigma_{in}$	σ_{out}/σ_{in}
$eee\mu$	8	Fake bkg. bin-by-bin #8	+0.10	0.15
$eee\tau_h$	7	Fake bkg. bin-by-bin #6	+0.10	0.25
$eee\tau_h$	7	Fake bkg. bin-by-bin #7	+0.10	0.21
$ee\mu\tau_h$	7	Fake bkg. bin-by-bin #8	+0.25	0.20
$ee\mu\tau_h$	8	Fake bkg. bin-by-bin #7	−0.12	0.22
$ee\mu\tau_h$	8	Fake bkg. bin-by-bin #8	+0.13	0.21
$ee\tau_h\tau_h$	7	Fake bkg. bin-by-bin #7	+0.25	0.24
$ee\tau_h\tau_h$	8	Fake bkg. bin-by-bin #10	+0.42	0.21
$ee\tau_h\tau_h$	8	Fake bkg. bin-by-bin #5	−0.12	0.22
$ee\tau_h\tau_h$	8	Fake bkg. bin-by-bin #7	−0.12	0.15
$ee\tau_h\tau_h$	8	Fake bkg. bin-by-bin #8	−0.10	0.15
$\mu\mu e\tau_h$	7	Fake bkg. bin-by-bin #9	+0.27	0.24
$\mu\mu\mu e$	7	Fake bkg. bin-by-bin #11	−0.33	0.19
$\mu\mu\mu e$	8	Fake bkg. bin-by-bin #13	+0.21	0.15
$\mu\mu\mu\tau_h$	7	Fake bkg. bin-by-bin #12	+0.36	0.23
$\mu\mu\mu\tau_h$	8	Fake bkg. bin-by-bin #2	+0.43	0.23
$\mu\mu\mu\tau_h$	8	Fake bkg. bin-by-bin #7	+0.44	0.21
$\mu\mu\mu\tau_h$	8	Fake bkg. bin-by-bin #9	−0.13	0.25
$\mu\mu\tau_h\tau_h$	8	Fake bkg. bin-by-bin #9	+0.18	0.21

Table 5.7: Pulls ($\Delta z/\sigma_{in}$) and a posteriori variances (σ_{out}/σ_{in}) related to the bin-by-bin uncertainties of the fake background shapes.

An additional way to quantify the discrepancy between the observed values and the expected ones under a given statistical model, is computing the so-called *goodness of fit* (GOF). In particular, the GOF test used in this analysis is based on the minimization of a given likelihood-ratio [162]. The general idea is the following: starting from the expected distribution of the background, one generates pseudo-data sets (the toys) allowing the nuisance parameters to vary within their associated uncertainties. For each toy, the fitting procedure with the same background model used for the "real" analysis is performed and the result of the minimization of the likelihood-ratio is treated as an entry of the GOF distribution. Examples of the GOF distributions obtained in the context of this analysis for the channels $\ell\ell + e\mu, e\tau_h, \mu\tau_h, \tau_h\tau_h$, for the 8 TeV analysis, are reported in Figure 5.12. These are the distributions of the minima of the likelihood-ratios obtained for each toy. The arrow is the result obtained using real data. By comparing the observed result with the expectation from the toys, the used statistical model can be judged (the closer is the arrow to the mean of the distribution, the better is the fit). Any possible indications of underestimation or overestimation of the uncertainties, however, have to

be carefully scrutinized in order to exclude any possible bias due to possible statistical fluctuations in data, rather than a flaw in the modeling.

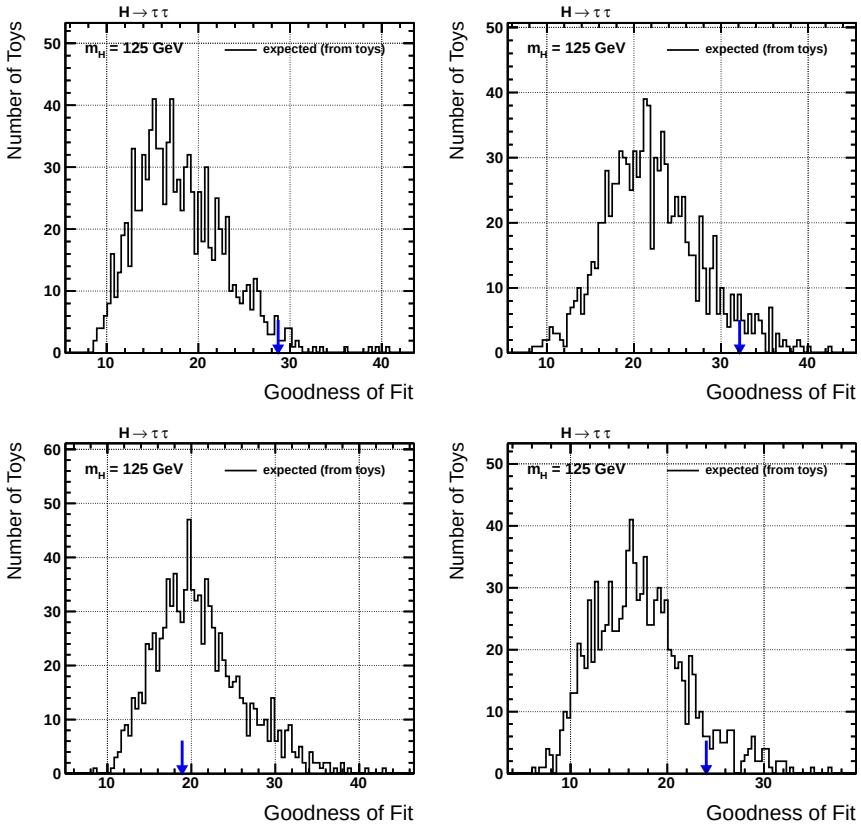

Figure 5.12: Goodness of fit as computed with pseudo-data for $\ell\ell e\mu$ (top left), $\ell\ell e\tau_h$ (top right), $\ell\ell\mu\tau_h$ (bottom left), and $\ell\ell\tau_h\tau_h$ (bottom right) channels for the $\sqrt{s} = 8$ TeV analysis. The arrow represents the value observed in data [AC].

5.9.2 Exclusion limits

As can be seen from Figure 5.11, there are no significant excesses over the SM backgrounds, therefore exclusion limits at 95% CL (cfr. Appendix B) on the signal strength modifier $\mu = \sigma_{95\%}/\sigma_{TH}$ have been computed. In particular, exclusion limits are computed for a SM Higgs (SVfit) mass ranging between 90 GeV and 145 GeV. The values of the observed upper limits at 95% CL on the signal strenght modifier are listed in Table 5.8 for the eight hypothesized

values of M_h, considering all eight channels combined and for $7 + 8$ TeV data analysis. The expected exclusion limits in the absence of signal (background-only hypothesis) are also reported together with the corresponding ± 1 and $\pm 2\sigma$ confidence level bands.

M_h (GeV)	Exp. -2σ	Exp. -1σ	Exp.	Exp. $+1\sigma$	Exp. $+2\sigma$	Obs.
90	1.51	2.05	2.93	4.25	5.94	2.61
95	1.61	2.17	3.1	4.5	6.29	2.73
100	1.64	2.22	3.17	4.63	6.49	2.62
105	1.63	2.23	3.2	4.7	6.65	2.74
110	1.58	2.17	3.14	4.63	6.61	3.14
115	1.64	2.25	3.27	4.84	6.93	3.51
120	1.8	2.46	3.58	5.33	7.61	4.53
125	1.92	2.67	3.89	5.83	8.43	5.44
130	2.24	3.11	4.58	6.88	9.99	7.08
135	2.8	3.9	5.77	8.68	12.7	9.47
140	3.66	5.1	7.53	11.3	16.5	13.2
145	5.21	7.26	10.7	16.2	23.6	19.8

Table 5.8: $7 + 8$ TeV expected and observed (with ± 1 and $\pm 2\sigma$ uncertainty bands) 95% CL upper limit on the signal strenght modifier μ as a function of M_h, considering all channels combined and including all statistical and systematic uncertainties.

Expected and observed exclusion limits, considering the contribution of all channels, are graphically summarized in Figure 5.13 for 7 TeV, 8 TeV and for 7 and 8 TeV combined datasets. $\pm 1\sigma$ and $\pm 2\sigma$ bands are reported in green and yellow, respectively.

The main conclusions drawn by looking at the results for $7 + 8$ TeV data (cfr. Table 5.8 and Figure 5.13), are: the observed exclusion limit is compatible with the background-only hypothesis in the whole mass range considered within $\pm 1\sigma$; the production of an Higgs boson with mass $M_h = 125$ GeV, at a rate 5.44 times larger than the SM prediction, is ruled out at 95% CL, while a production rate 3.89 times the SM one is expected to be excluded in the hypothesis of only background. Moreover, it has been checked that the increase of the limit in the high mass region is due to the treatment of the $Zh(125\text{ GeV}) \to \ell\ell WW$ process as a background, in correspondence of each signal mass point. The best-fit value for the signal strength modifier amounts to $\hat{\mu} = 1.61 \pm 1.85$ at $M_h = 125$ GeV.

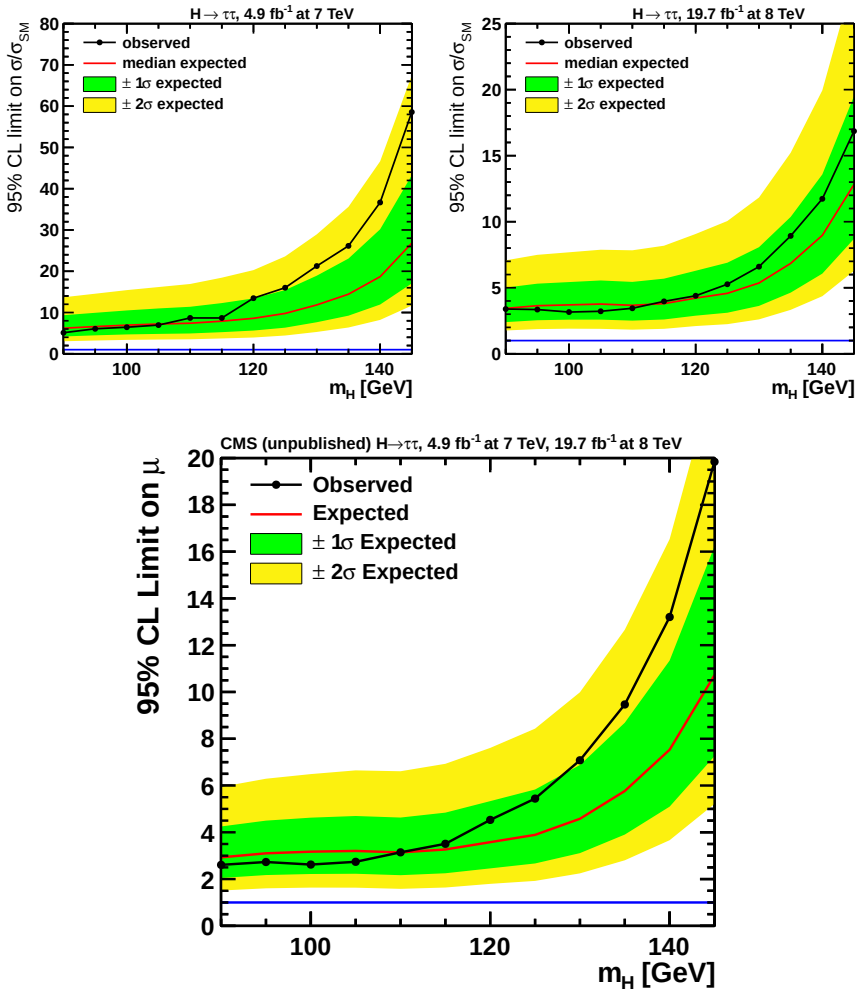

Figure 5.13: Expected (red line) and observed (black line) 95% CL exclusion limits on the signal strenght μ for 7 TeV (top left), 8 TeV (top right) and for 7 and 8 TeV combined datasets (bottom), considering all channels combined and including all statistical and systematic uncertainties. The benchmark points used for the computation of the exclusion limits are shown as a dot in the black line. The green and yellow bands correspond to $\pm 1\sigma$ and $\pm 2\sigma$ intervals on the expected limit, respectively [161] [AC].

5.10 Analysis validation

As anticipated in the opening of the chapter, all the results shown have been independently validated during the Ph.D. work. It is customary within the

CMS collaboration, in fact, having two or more groups performing the same search by using independent analysis frameworks. This is what has happened, in the context of this analysis, for the ULB[4] and the UCL groups. The former has carried out the official CMS analysis, the latter, which the author of the manuscript belongs to, has cross-checked the main analysis in all its aspects. This section aims to summarize the main results of such synchronization, concerning data collected at 8 TeV.

Perfect agreement (100%) has been obtained concerning the signal and the physical background selection: an example is provided in Figure 5.14 in which the SVfit ditau mass distributions are reported for data, signal $Zh(125\,\text{GeV}) \to 2\ell 2\tau$ and ZZ background processes estimated from simulation, considering the $\mu\mu e\tau_h$ channel. The agreement reached, as far as concern the estimate of the fake background from data, is within 1% (cfr. Figure 5.14), fully compatible within statistical uncertainties.

The agreement on the expected and observed combined 95% CL limits is within 1% and $1 - 8$%, respectively, as can be observed from Figure 5.15. Further studies carried out to understand the differences in the observed limit have demonstrated that such a disagreement is due to some event migration by few GeV to the next bin of the SVfit mass spectrum. It happens, in fact, that in case of very boosted objects (like the one coming the Higgs decay), where the momentum is very large compared to the mass, even a small difference in the inputs (rounded errors, for instance) provided to the SVfit mass calculator can lead to shifts in the mass of the (light) input particles (taus in particular). As the mass defines the integration region, then a slight change in mass affects significantly the result.

5.11 Run I $h \to \tau\tau$ combination

The $pp \to Zh$ process results have been combined with those obtained studying the Higgs associated production with a *W* boson (performed considering four channels $\ell + L\tau_h = \mu + \mu\tau_h$, $e + \mu\tau_h/\mu + e\tau_h$, $\mu\tau_h\tau_h$ and $e + \tau_h\tau_h$) and the main production mechanisms gluon fusion and vector-boson fusion (performed considering six final states $LL' = \mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$, $e\mu$, $\mu\mu$ and ee) [15].

The inclusion of the Vh analyses has led to an increase of the expected *p*-value at $M_h = 125$ GeV from 3.62σ to 3.73σ. In correspondence of the same mass point, the observed *p*-value has changed from 3.36σ to 3.20σ with the inclusion of the Vh results. The final observed *p*-value computed included all three

[4] Université libre de Bruxelles, Bruxelles (Belgium).

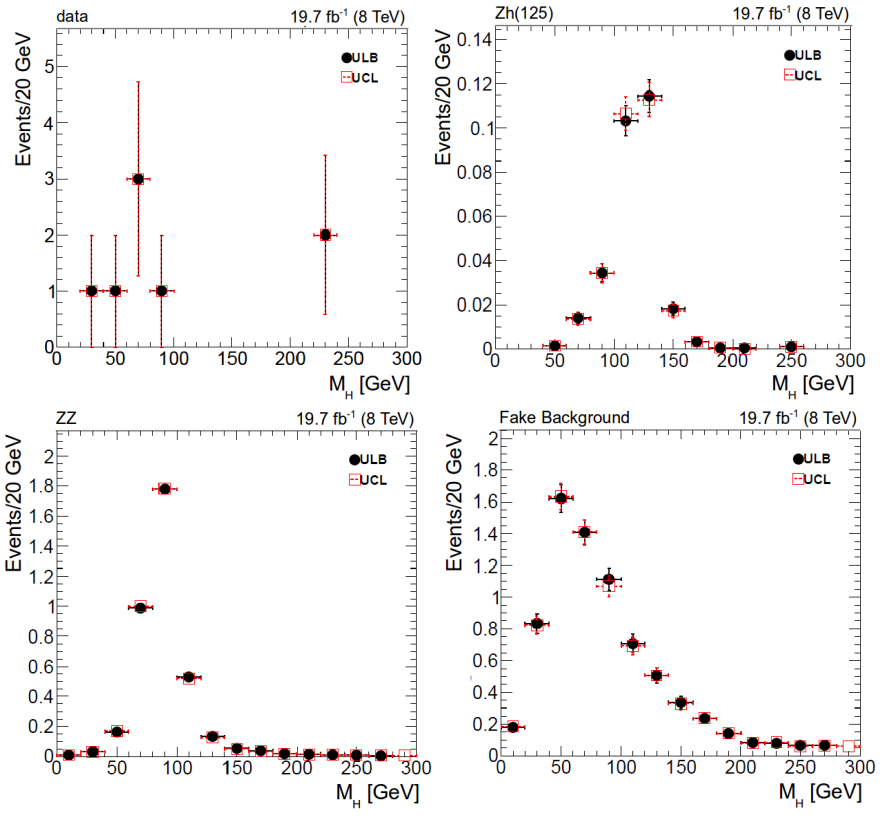

Figure 5.14: Representative SVfit ditau mass distributions obtained in the context of the validation of the main CMS analysis presented in this chapter. The spectra correspond to the $\mu\mu e\tau_h$ channel and refer to data (top left), simulated signal $Zh(125\ \text{GeV}) \to 2\ell 2\tau$ (top right) and ZZ background (bottom right), and data-driven estimated fake background (bottom left). In each plot, the black and the red dots correspond to the results of the main and the cross-check analyses, respectively. The agreement in data, signal and physical background selection is 100%, while the agreement in the fake background estimate (cfr. Section 5.7) is within 1% [AC].

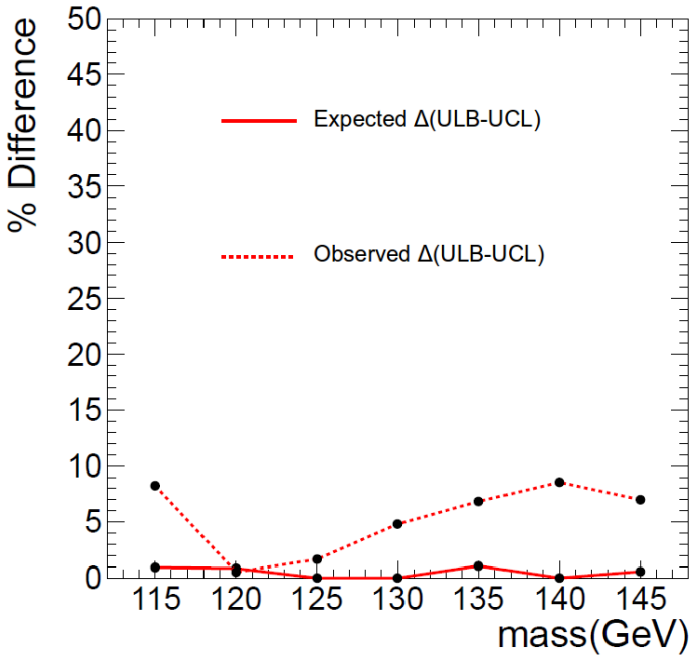

Figure 5.15: Percentage difference of the expected (solid line) and observed (dashed line) combined limits between the ULB and the UCL groups. The differences observed amount to 1% and $1-8$% for the expected and the observed limits, respectively [AC].

analyses is shown in Figure 5.16 (right) as a function of the Higgs boson mass. The observed p-value is minimal for $M_h = 120$ GeV with a significance of 3.3σ. The observed significance is larger than three standard deviations for $115 < M_h < 130$ GeV. The best-fit value of the signal strength modifier, combining all three production mechanisms, is $\hat{\mu} = 0.78 \pm 0.27$ at $M_h = 125$ GeV. Figure 5.16 (left) summarizes the results of the fit performed in each decay channels studied in the (gluon fusion plus vector-boson fusion) $h \to \tau\tau$ analysis, and the one obtained by combining Wh and Zh analyses, which amounts to $\hat{\mu} = -0.33 \pm 1.02$ at $M_h = 125$ GeV. Despite the positive value of the signal strength modifier measured in the Zh analysis, the one related to the Vh combination is negative. Such result is due to a background underfluctuation observed in the Wh analysis.

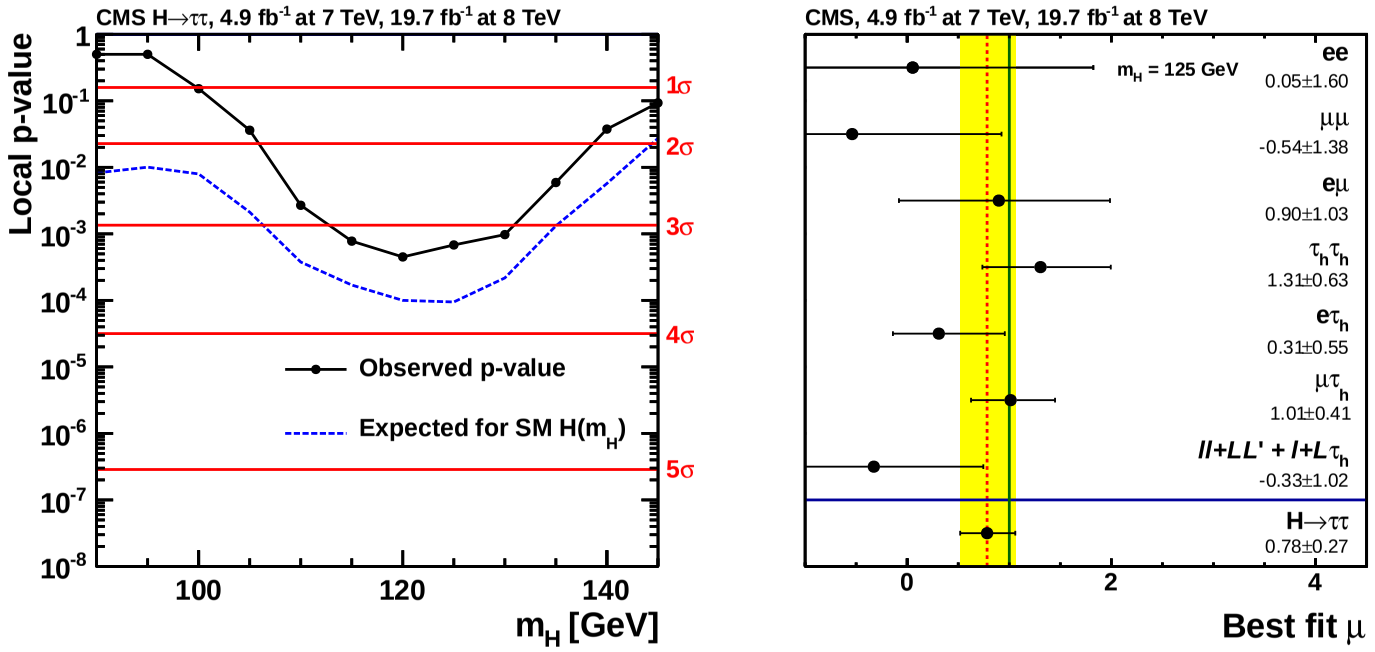

Figure 5.16: Right: combined local p-value and significance in number of standard deviations as a function of the SM Higgs boson mass hypotheses for the the combination of all production mechanisms and channels studied in the CMS $h \to \tau\tau$ search. The observation (solid line) is compared to the expectation (dashed line) for a SM Higgs boson with mass M_h. Left: best-fit signal strength values computed for independent channels for $M_h = 125$ GeV. The red dashed line indicates the combined best-fit signal strength [15] [AC].

6

Search for Higgs bosons decaying to tau pairs in the 2HDM

A search for a new resonance decaying to a Z *boson and a lighter resonance, in the context of the 2HDM, is described in this chapter. Being the analysis strategy independent of the assumed theoretical model and spin-parity (cfr. Chapter 1, both the processes* $pp \to H \to ZA$ *and* $pp \to A \to ZH$ *(whose corresponding Feynman diagrams have been shown in Figure 1.13) will be studied. In particular, the* Z *boson is identified via its decay into either a pair or electrons or a pair of muons, while the* H/A *resonance will be identified through its decay into a pair of tau leptons. The analysis has been performed using data collected by the CMS experiment during 2012. All the aspects of the analysis, starting from the development of the analysis framework and up to the extraction of the final limits, have been carried out during the last two years of the doctoral work. The majority of the results presented in this chapter have been included in a CMS Physics Analysis Summary* [22] *together with a different analysis performed in the* $b\bar{b}$ *channel. An even more sophisticated statistical description of the background estimation has lead to an update of the results shown in* [22]*, which, together with the full* $\tau\tau + b\bar{b}$ *combination, are going to be included in the final paper of the analysis [23], to be submitted to Physics Letter B* [163]*. As in Chapter 5, the following notations will be used:* $\ell = e/\mu$ *and* $L = e/\mu/\tau$*.*

6.1 Introduction

A completely different and independent analysis framework, with respect to that used for the SM Zh analysis (cfr. Chapter 5), has been developed for the 2HDM analysis presented in this chapter. However, apart from this technical aspect, there are some elements in common and some notable differences between the two analyses, which is worth underlying for two main reasons: a) to make easier the reading of the text given that, since some expertise and technicalities developed while performing the SM analysis have been used also in the 2HDM one, there is no need to describe and contextualize those again; b) to emphasize the upgraded strategies oriented to the achievement of a high efficiency in the selection of the signal events and in a better modelling of the background.

What is in common?

a) The eight final states $eee\mu$, $ee e\tau_h$, $ee\mu\tau_h$, $ee\tau_h\tau_h$, $\mu\mu e\mu$, $\mu\mu e\tau_h$, $\mu\mu\mu\tau_h$, and $\mu\mu\tau_h\tau_h$. As explained in Section 5.1, the $\ell\ell ee$ and $\ell\ell\mu\mu$ final states have been not included in the SM Higgs search in order to avoid overlaps with the search for the SM $h \to ZZ \to 4\ell$ [12]. This strategy has been adopted also in this case more for historical reasons than to avoid overlaps with other analyses. However, the inclusion of the $\ell\ell ee$ and $\ell\ell\mu\mu$ channels is planned Run II;

b) the topological cuts applied event-by-event and listed in Section 5.5, except for the cross-cleaning among lepton collections as it will be explained later;

c) the trigger paths used for 2012 analysis are the same listed in Table 5.3;

d) accordingly, `DoubleElectron` and `DoubleMu` 8 TeV datasets are the same listed in Table 5.1, corresponding to an integrated luminosity of 19.8 fb^{-1};

e) data-to-simulation correction factors related to the lepton identification and isolation, which are estimated with the tag-and-probe method [151], have been used to rescale the simulated samples (cfr. Section D). Data-to-simulation differences related to trigger efficiencies and lepton energy scales, which have been estimated to be of the order of 1-3% overall, have been taken directly into account as systematic uncertainties.

What differs?

a) Excluding electrons, muons and the missing transverse energy, the reconstruction and identification criteria used in this analysis are the same described in Section 5.4. The intrinsic properties of the reconstructed electron and muon collections used in both analyses are different as explained in Chapter 3. Standard reconstructed electrons and muons have been used in the SM Zh analysis, while pure particle-flow reconstructed and identified objects have been used in the context of this analysis. As a consequence, as already anticipated, no cross-cleaning among lepton collections has been performed, being the particle-flow algorithm able, by construction, to unambiguously identify every single particle in the event. Additionally, in this analysis, the particle-flow $\not{E}_T$ has been used instead of the MVA particle-flow $\not{E}_T$ (cfr. Section 3.5.5);

b) the signal, which will be described in Section 6.2 is different for obvious reasons. Concerning the simulated background processes, those already listed in Table 5.2 can be somehow considered as a subset of the simulates samples used in the 2HDM analysis, which are listed in Table 6.1. Also for this analysis, for all the samples generated with MADGRAPH [115] and POWHEG [118], these generators have been interfaced with PYTHIA for the parton shower and hadronization. The tau lepton decays have been simulated with TAUOLA [119] (cfr. Section 2.4). The effect of the pileup-reweighting on simulated samples (cfr. Section 5.2) is shown in Figure 6.1;

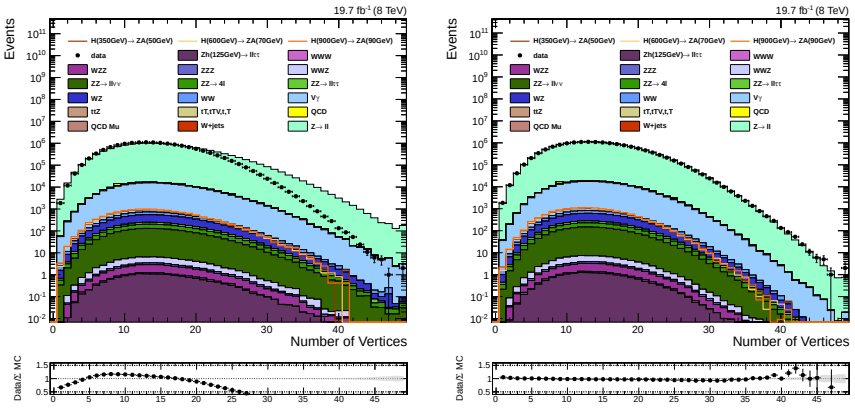

Figure 6.1: Expected and observed vertex multiplicity distributions before (left) and after (right) having applied the pileup-reweighting on simulated events. Reference signal distributions for the process $H \to ZA$ ($M_H = 350$ GeV $- M_A = 50$ GeV, $M_H = 600$ GeV $- M_A = 70$ GeV and $M_H = 900$ GeV $- M_A = 90$ GeV) are superimposed to the stacked background distributions, all estimated from simulation. The signal $\sigma \times BR$ has been normalized to 1 pb [*A*].

c) some differences in the selection have been introduced and a different strategy for the optimization of the final selection has been chosen (more details in Section 6.3);

d) the fake rate method (cfr. Section 5.7) presents some differences and has been more deeply scrutinized from the statistical point of view.

6.2 Signal samples

The signal samples have been generated with MADGRAPH5 [115] using a 2HDMC calculator [164, 165], which allows the matrix element calculations in a general, CP-conserving two-Higgs-doublet model to be performed. The program allows the user to choose not only a particular parametrization of the 2HDM potential but also different Z_2 symmetries. The output, which is a file containing the generated events in the *Les Houches Event* (LHE) format [166], is given as input to PYTHIA and TAUOLA for the parton showering, hadronization, and for letting the tau decaying either leptonically or hadronically (cfr. Section 2.4).

As already explained in Chapter 1, in the minimal formulation of the 2HDM Lagrangian, the mass terms for the scalars are related to the parameter α and β, which determine the interactions of the various Higgs fields with the vector bosons and with the fermions. Considering the parameter space still favoured by direct searches (cfr. Section 1.5.2), the chosen values for the couplings and their mixing are $\cos(\beta - \alpha) = 0.01$ and $\tan\beta = 1.5$. In this scenario, the following assumptions have been made:

- the lightest scalar coincides with the SM Higgs boson, whose mass M_h is fixed to 125 GeV;
- $M_H^{\pm} = M_H$ in case the process $H \to ZA$ is considered. If the process $A \to ZH$ is under study, then $M_H^{\pm} = M_A$. This assumption preserves the degeneracy $M^2_{H^{\pm}} \sim M^2_{H/A}$, avoiding large mass splittings that would break the custodial symmetry;
- $M^2_{12} = M^2_{H^{\pm}} \tan\beta/(1 + \tan^2\beta)$ in order to consider the Z_2 symmetry softly broken;
- $\lambda_{6,7} = 0$ to avoid CP-violation at tree-level.

Process	Generator	Cross-section (pb)	Sample name
Zh (SM)			
Zh(125 GeV), $h \to \tau\tau$	PYTHIA	$2.65 \cdot 10^{-3}$	ZH_HToTauTau_M-125_lepdecay
VVV			
WWW+jets	MADGRAPH	$8.22 \cdot 10^{-2}$	WWWJets
WWZ+jets	MADGRAPH	$6.33 \cdot 10^{-2}$	WWZJets
WZZ+jets	MADGRAPH	$1.92 \cdot 10^{-2}$	WZZJets
ZZZ+jets	MADGRAPH	$4.59 \cdot 10^{-2}$	ZZZJets
VV			
$WZ \to 3\ell\nu$	MADGRAPH	1.057	WZJetsTo3LNu
$WW \to 2\ell 2\nu$	MADGRAPH	5.995	WWJetsTo2L2Nu
$ZZ \to 2\ell\nu$	MADGRAPH	0.320	ZZJetsTo2L2Nu
$ZZ \to 4\ell$	MADGRAPH	0.320	ZZJetsTo4L
Vγ			
$W\gamma \to \ell\nu\gamma$	MADGRAPH	461.6	WGToLNuG
$Z\gamma \to X$	MADGRAPH	123.9	ZG_Inclusive
$t\bar{t}$+X			
$t\bar{t}$ (fully leptonic)	MADGRAPH	23.89	TTJets_FullLeptMGDecays
$t\bar{t}W$	MADGRAPH	0.232	TTWJets
$t\bar{t}Z$	MADGRAPH	0.208	TTZJets
Single top			
$\bar{t}$ tW-channel	POWHEG	11.2	Tbar_tW-channel-DR
t tW-channel	POWHEG	11.2	T_tW-channel-DR
$\bar{t}$ t-channel	POWHEG	30.7	Tbar_t-channel
t t-channel	POWHEG	56.4	T_t-channel
$\bar{t}$ s-channel	POWHEG	1.76	Tbar_s-channel
t s-channel	POWHEG	3.89	T_s-channel
QCD electromagnetic-enriched			
$30 < p_T < 80$ GeV	PYTHIA	4615893	QCD_Pt_30_80_EMEnriched
$80 < p_T < 170$ GeV	PYTHIA	183294.9	QCD_Pt_80_170_EMEnriched
$170 < p_T < 250$ GeV	PYTHIA	4586.52	QCD_Pt_170_250_EMEnriched
$250 < p_T < 350$ GeV	PYTHIA	556.75	QCD_Pt_250_350_EMEnriched
$p_T > 350$ GeV	PYTHIA	89.1	QCD_Pt_350_EMEnriched
QCD μ-enriched			
$p_T > 20$ GeV	PYTHIA	134680	QCD_Pt_350_EMEnriched
Z+jets			
$Z/\gamma^* \to \ell\ell+ \leqslant 4$ jets	MADGRAPH		DYJetsToLL
$Z/\gamma^* \to \ell\ell + 1$ jet	MADGRAPH		DY1JetsToLL
$Z/\gamma^* \to \ell\ell + 2$ jets	MADGRAPH	3533	DY2JetsToLL
$Z/\gamma^* \to \ell\ell + 3$ jets	MADGRAPH		DY3JetsToLL
$Z/\gamma^* \to \ell\ell + 4$ jets	MADGRAPH		DY4JetsToLL
W+jets			
$W \to \ell\nu+ \leqslant 4$ jets	MADGRAPH		WJetsToLNu
$W \to \ell\nu + 1$ jet	MADGRAPH		W1JetsToLNu
$W \to \ell\nu + 2$ jets	MADGRAPH	36257	W2JetsToLNu
$W \to \ell\nu + 3$jets	MADGRAPH		W3JetsToLNu
$W \to \ell\nu + 4$ jets	MADGRAPH		W4JetsToLNu

Table 6.1: Simulated background processes used for the 2HDM $H/A \to ZA/H$ analysis. The names of the generators used for the event simulation are reported. The MADGRAPH and POWHEG tools have been interfaced with PYTHIA for the parton shower and hadronization. TAUOLA [119] has been used to simulate tau lepton decays. Cross sections are also reported for $\sqrt{s} = 8$ TeV.

The production cross section used in the normalization of the signal samples have been computed at the approximate NNLO accuracy with SUSHI [167] using the PDF sets provided by the MSTW2008 collaboration at LO, NLO and NNLO level of accuracy [48]. Figure 6.2 shows the theoretical cross sections times branching ratios for the process $H \to ZA \to \ell\ell\tau\tau$ in the left triangle, and the process $A \to ZH \to \ell\ell\tau\tau$ in the right triangle. The $\sigma \cdot BR$ for the signal samples used in the analysis correspond to values of $M_{H/A}$ and $M_{A/H}$ varying in the ranges $[200, 1000]$ GeV and $[15, 900]$ GeV, respectively, with the constraint $M_{H/A} > M_{A/H} + M_Z$. The region where M_H is smaller than M_h is forbidden by the theoretical model. More than one-hundred benchmark points corresponding to different mass values for the process $H \to ZA \to \ell\ell\tau\tau$ have been fully simulated using the GEANT4 toolkit [121]. Just few fully-simulated samples, instead, have been produced for the $A \to ZH \to \ell\ell\tau\tau$ process, in order to verify that the selection efficiency in the signal region does not depend on the H or A being the heavy resonance. The study has been motivated by the need of reducing the computing-time arising from the heavy process of the full simulation. The check has confirmed the CP-independence of the signal efficiency, allowing a safe "mirroring" of all the results obtained with the signal $H \to ZA$, to obtain those related to the signal $A \to ZH$, with a proper normalization to the corresponding $\sigma \cdot BR$.

Figure 6.3 (left) shows different $\Delta R(\tau, \tau)$ distributions for five different signal samples ($M_H = 350$ GeV and $15 < M_A < 90$ GeV). It is evident that as the M_A decreases, the ΔR distributions move to lower values corresponding to the cases where the ditau pairs are highly boosted, i.e. where the $\Delta R < 0.5$ and the two tau jets tend to overlap. A more extended view is provided in Figure 6.3 (right), where a two-dimensional $\Delta R(\tau, \tau)$ map is shown as a function of the M_H and M_A masses. Two extreme regions can be identified: the non-boosted region for $M_H/M_A < 5$, in which taus are well separated and properly reconstructed by the standard HPS algorithm and the boosted region for $M_H/M_A > 10$, in which, instead, the two taus cannot be separated and properly reconstructed by the HPS algorithm. Dedicated HPS algorithms for reconstructing boosted taus have been developed and are currently under physics validation studies. In this analysis, no algorithms able to reconstruct boosted topologies have be used, therefore, improved results are expected in the low M_A region when dedicated algorithms will become available.

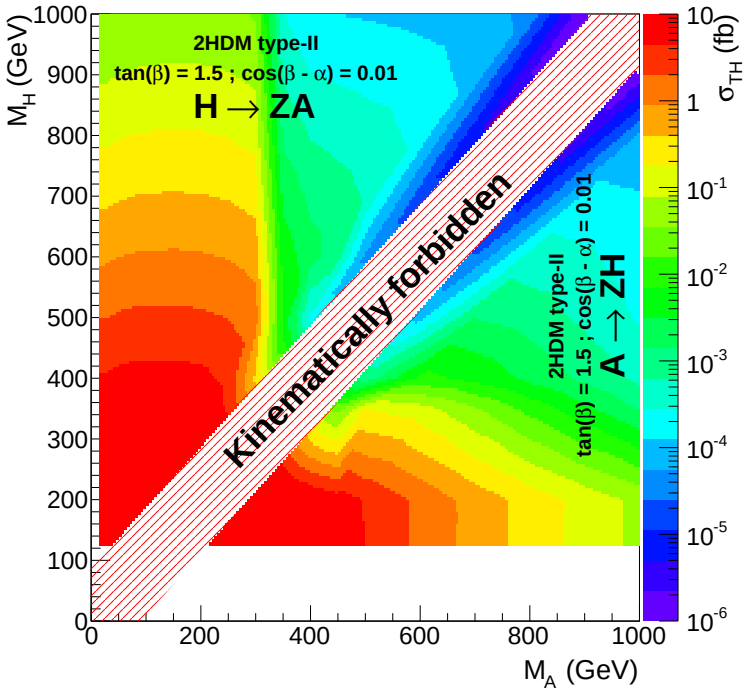

Figure 6.2: Approximate NNLO SusHi [167] cross section times branching ratio values, expressed in fb, for the processes $H \to ZA \to \ell\ell\tau\tau$ (left triangle) and $A \to ZH \to \ell\ell\tau\tau$ (right triangle). The region "Kinematically forbidden" is the one constrained by requiring $M_{H/A} > M_{A/H} + M_Z$, while the white region in the right triangle reflects the constraint $M_H > M_h$ imposed by the theoretical model. The chosen model parameter values, $\cos(\beta - \alpha) = 0.01$ and $\tan\beta = 1.5$, are reported [A].

6.3 Event selection

The first step of the event selection is the reconstruction of the a Z boson candidate by selecting a pair of well identified and isolated opposite-sign electrons or muons, which are expected to trigger the event. Secondly, two more opposite-sign leptons are required to build the ditau pair.

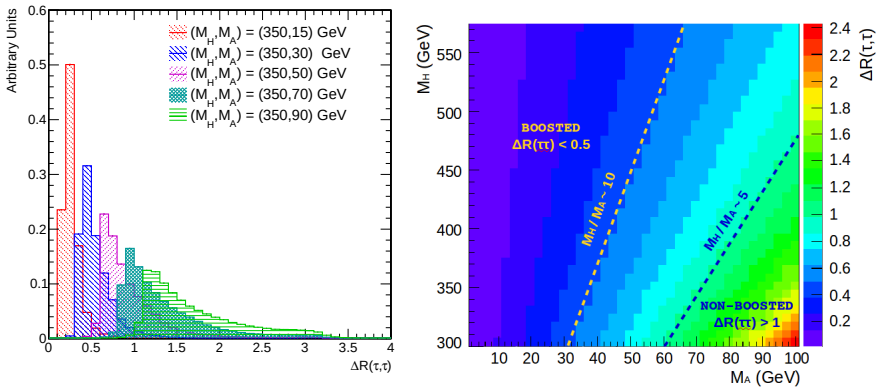

Figure 6.3: Left: one-dimensional $\Delta R(\tau,\tau)$ distributions for for five different signal samples ($M_H = 350$ GeV and $15 < M_A < 90$ GeV). As M_A decreases, the ΔR distributions move to lower values, towards the region in which the ditau pairs are highly boosted. Right: two-dimensional $\Delta R(\tau,\tau)$ as a function of M_H and M_A. Ditau non-boosted (boosted) topologies are expected for $M_H/M_A < 5$ ($M_H/M_A > 10$) [A].

6.3.1 Z boson selection

Few changes have been introduced in the context of the Z boson selection with respect to the one adopted in the SM Higgs search described in Chapter 5. The description in Section 5.5.1 is still valid with the exception of the following changes. The threshold required for the subleading lepton p_T has now been increased from 10 GeV to 20 GeV. This choice, driven by the need of harmonizing the Z boson selection in both the $\ell\ell\tau\tau$ and $\ell\ell b\bar{b}$ final states, has been found to have a negligible impact on the expected signal significance. Moreover, the invariant mass of the reconstructed Z has to satisfy the requirement $|M_{\ell\ell} - M_Z| < 15$ GeV, which is tighter than that of the SM Zh analysis, to further reduce the background (especially $V\gamma$ and single-top processes). The Z boson invariant mass distribution is reported in Figure 6.4 (left). In order to further increase the sensitivity in the high ditau mass region, a $p_T > 20$ GeV is required for the Z boson candidate. This criterion allows the dominant Z+jets background to be reduced, as can be seen from Figure 6.4 (right). If more than one Z candidate satisfies the requirements listed above, the one with the minimum absolute difference between the reconstructed mass value and the true one [29] is chosen.

Furthermore, looking at the distribution of the ϕ difference between the Z and $\not{E}_T$ directions, i.e. $|\Delta\phi(Z, \not{E}_T)|$, it can be seen from Figure 6.5 that the Z+jets background events are uniformly distributed in the whole ϕ range, while in

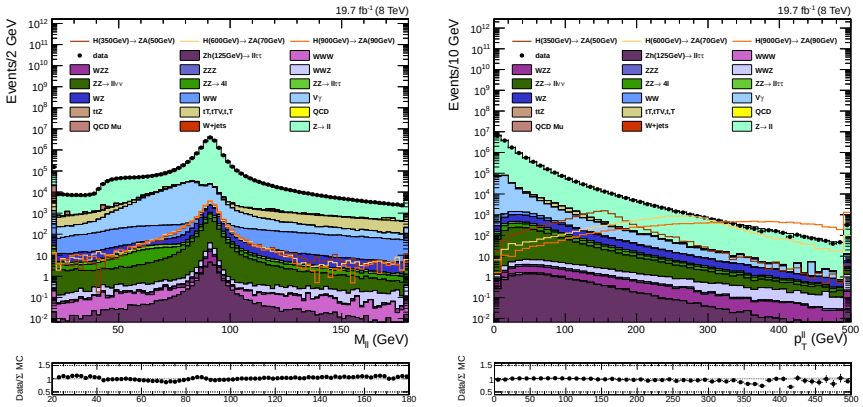

Figure 6.4: Expected and observed invariant mass (left) and p_T (right) of the reconstructed Z boson candidate obtained after having selected the Z candidate only. Reference signal distributions for the process $H \to ZA$ ($M_H = 350$ GeV $- M_A = 50$ GeV, $M_H = 600$ GeV $- M_A = 70$ GeV and $M_H = 900$ GeV $- M_A = 90$ GeV) are superimposed to the stacked background distributions, all estimated from simulation. The signal $\sigma \times BR$ has been normalized to 1 pb [A].

the signal events the Z boson is more likely forming with the $\not{E}_T$ an angle of π. Thus, in order to suppress the Z+jets background, only those events satisfying the requirement $|\Delta\phi(Z, \not{E}_T)| > 1.5$ are accepted.

6.3.2 A/H boson selection

The strategy to select the A/H boson candidate is the same as in the SM analysis (cfr. Section 5.5.2) except some changes. In particular, some selection criteria have been tuned with the optimization process explained in Section 6.3.3. The identification criteria for the muon in the $\mu\tau_h$ final states have been loosened in order to enhance the sensitivity to the signal in an already highly pure channel. Concerning the requirements on lepton identification and isolation, the reader can refer to Chapters 3 and 4.

$A/H \to \mu\tau_h$

- One loose identified particle-flow muon (cfr. Table 3.2) with $p_T > 10$ GeV and $|\eta| < 2.4$;
- muon particle-flow relative isolation $I^{PF}_{rel} < 0.3$;
- one HPS hadronic tau with $p_T > 20$ GeV and $|\eta| < 2.3$;

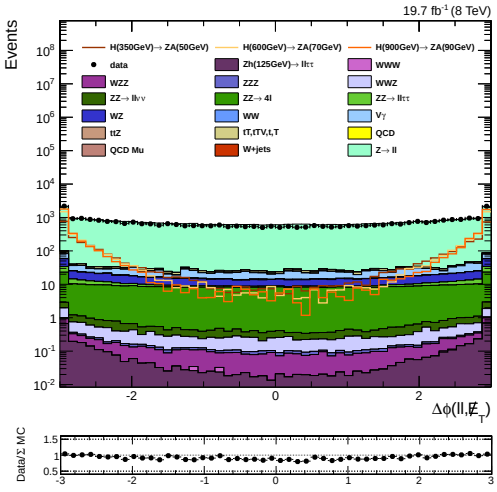

Figure 6.5: Expected and observed $|\Delta\phi(Z, \not{E}_T)|$ distributions obtained after having selected the Z candidate only. Reference signal distributions for the process $H \to ZA$ ($M_H = 350$ GeV $- M_A = 50$ GeV, $M_H = 600$ GeV $- M_A = 70$ GeV and $M_H = 900$ GeV $- M_A = 90$ GeV) are superimposed to the stacked background distributions, all estimated from simulation. The signal $\sigma \times BR$ has been normalized to 1 pb [A].

- the hadronic tau has to satisfy the decay mode finding, the anti-μ tight and the anti-e loose cut-based discriminators;
- tau particle-flow absolute isolation $I^{PF}_{abs} < 2$ GeV (corresponding to the loose cut-based discriminator working point).

$A/H \to e\tau_h$

- One tightly identified particle-flow electron (cfr. Table 3.1) with $p_T > 10$ GeV and $|\eta| < 2.5$;
- electron particle-flow relative isolation $I^{PF}_{rel} < 0.3$;
- one HPS hadronic tau with $p_T > 20$ GeV and $|\eta| < 2.3$;
- the hadronic tau has to satisfy the decay mode finding, the anti-μ loose cut-based and the anti-e tight MVA discriminators;
- tau particle-flow absolute isolation $I^{PF}_{abs} < 2$ GeV.

$A/H \to \tau_h\tau_h$

- Two HPS hadronic tau with $p_T > 20$ GeV and $|\eta| < 2.3$;
- hadronic taus have to satisfy the decay mode finding, the anti-μ and the anti-e loose cut-based discriminators;
- tau particle-flow absolute isolation $I^{PF}_{abs} < 2$ GeV.

$A/H \to e\mu$

- One loosely identified particle-flow muon with $p_T > 10$ GeV and $|\eta| < 2.4$;
- one loosely identified particle-flow electron with $p_T > 10$ GeV and $|\eta| < 2.5$;
- muon and electron particle-flow relative isolation $I^{PF}_{rel} < 0.3$.

In each channel, the Higgs dilepton pairs are required to have opposite charges. A requirement on the scalar sum of the visible ditau pair decay objects transverse momentum (L_T) has been imposed. As for the isolation requirements, also the selection applied on the L_T variable has been optimized as described in Section 6.3.3.

The same SM Zh analysis topological requirements are applied commonly to all channels to mainly reduce further the fake backgrounds (cfr. Section 5.5): all four leptons have to come from the same primary vertex and be separated from each other by $\Delta R > 0.1$; all events with at least one b-tagged jet, or with more than four leptons, are rejected.

6.3.3 Selection criteria optimization

In order to find the optimal selection requirements on the lepton isolation and L_T thresholds of the ditau pair decay objects, an optimization procedure has been executed using the a priori expected signal significance computed from an Asimov dataset of signal plus background events (cfr. Appendix B) and considering the signal samples normalized to $\sigma \times BR = 1$ fb. This procedure differs from that adopted for the SM Higgs search where the expected exclusion limit was minimized. The a priori expected significance does not depend on the observed data, and therefore it respects the blind nature of the search. The rationale behind the optimization procedure is that for any event selection threshold on a physical quantity to be optimized, the chosen value is the one

that maximizes the expected signal significance. The details of the method are explained in the following.

For each final state and signal sample (i.e. a pair of M_H and M_A mass values), various values for the thresholds to be optimized have been scanned:

- $I^{PF}_{rel} < 0.1, 0.2, 0.3$ for electrons and muons;
- $I^{PF}_{abs} < 2, 1$ GeV for taus (the two thresholds correspond to the loose and medium working points for the cut-based tau isolation, respectively);
- $L_T > 0, ..., 200$ GeV in steps of 20 GeV,

for a total of 198 possible combinations, which will be referred to as *selection points*. In correspondence of each selection point, the expected significance is computed and, at the end, all the selection points obtained are sorted in ascending order of significance. The first ranked selection point, corresponding to the maximum significance, is chosen as optimal. This procedure guarantees to obtain the best discovery potential for each final state and for each signal sample considered. However, the procedure involves dealing with a large number of selections points.

In order to make the procedure lighter from a computational point of view, a simplification has been introduced. First of all, a smaller set of selection points valid for all the mass pairs has been identified. In order to achieve this goal, different combinations of possible selection requirements have been tested by fixing the isolation for taus and light leptons and varying the L_T threshold only in the range $[0, 200]$ GeV in steps of 20 GeV: in this way, it is possible to check how far the expected significance can drift for each cut combination and for each mass pair and channel, with respect to the most optimal one obtained considering the full set of 198 selection points. In Figure 6.6, the expected signal significances are reported, as a function of the M_A and for a fixed value of M_H, for the four final states.The black line corresponds to the best significance obtained considering 198 selection points, while all the other coloured lines correspond to different combinations of selections where only the lepton isolation requirements have been fixed. The conclusion is that, among all the mass combinations and final states, the requirement on the isolation for electrons and muons can be fixed to $I^{PF}_{rel} < 0.3$. Moreover, the loose working point for the cut-based tau isolation can be chosen without loosing significantly in potential discovery. The selection on the L_T is still scanned in the range $[0, 200]$ GeV in steps of 20 GeV, for a total of eleven selection points optimized for each final state and signal sample.

Having performed both a cut-and-count and a shape-based analysis, as it will be explained in Section 6.7, the optimization procedure has been repeated independently for each analysis strategy, possibly leading to different L_T selections in correspondence of the signal sample.

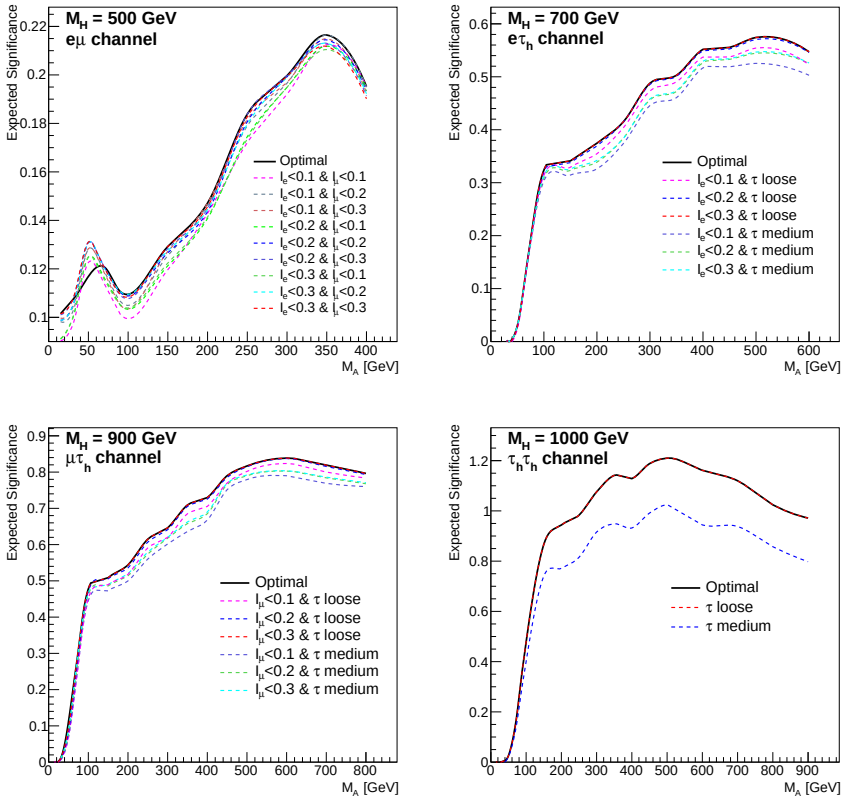

Figure 6.6: Expected significance comparisons obtained during the optimization of the analysis selection. The black lines correspond to the best significance obtained considering 198 selection points. The coloured lines correspond to different combinations of possible selections where only the lepton isolation requirements have been fixed. Across all the mass combinations and final states, the requirement on the isolation for electrons and muons can be fixed to $I_{rel}^{PF} < 0.3$ and the loose working point for the cut-based tau isolation can be chosen without loosing significantly in potential discovery. For illustration purpose only, different M_H values have been considered for each final state: $M_H = 500$ GeV for $e\mu$ (top left), $M_H = 700$ GeV for $e\tau_h$ (top right), $M_H = 900$ GeV for $\mu\tau_h$ (bottom left) and $M_H = 1000$ GeV for $\tau_h\tau_h$ (bottom right) [A].

6.4 Physical background

Normalizations and shapes of the background processes belonging to this category are entirely estimated from simulation, as already explained in Section 5.6. The main contribution is provided by ZZ and the SM $Zh(125\,\text{GeV}) \to \ell\ell\tau\tau$ processes. Another contribution comes from the WWZ process where one Z decays into either a dielectron or a dimuon pair and the two W bosons decay leptonically. The fully leptonic $t\bar{t}Z$ process contribution has been also considered as physical background. The (almost negligible) ZZZ and the WZZ contributions have also been estimated from simulation. These processes can be considered physical backgrounds as they have at least four prompt leptons in the final state and they can contribute to the signal region in case the additional ones escape detection because of detector acceptance.

6.5 Fake background

As already explained in Section 5.7, background processes with no more than three genuine leptons are estimated from data with the fake rate method. The fake rate method idea used in this analysis is the same as the one of the SM analysis. Few important differences exist, however. Comparing Figure 6.7, in which a summary of the fake rate method used in this analysis is schematically represented, with Figure 5.6, in which the corresponding scheme for the SM analysis is reported, the reader can realize that there are differences both in the computation of the fake rates and in their applications. In particular, the fake rate method used in the 2HDM analysis, is based on the measurement of the misidentification probabilities in background-enriched regions that target the $Z + 1$ jet topology. The reason for this choice is the reduction of the statistical uncertainties in the measurement of the fake rate and the observation that the event activity accompanying the jet that is misidentified as a lepton does not have a significant impact on the measurement. Subsequently, the fake background estimation is obtained by applying Equation 5.10 instead of Equation 5.11, which means that the possibility of having negative fake-background estimates is not excluded, thus avoiding any systematic bias.

6.5.1 Lepton fake rates determination

TAU FAKE RATE This fake rate is measured by requiring a Z boson, selected as explained in Section 6.3.1, and only one loose tau. The loose tau is required

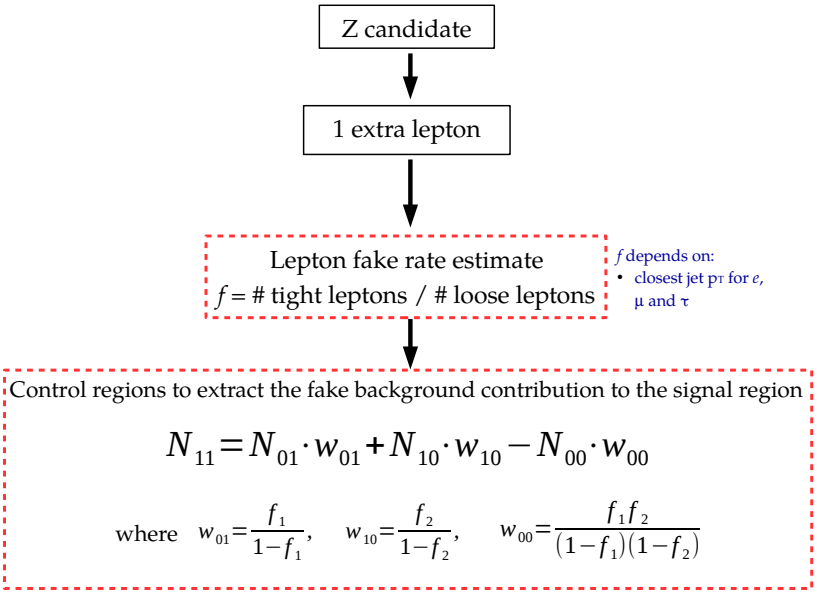

Figure 6.7: Schematic representation of the fake rate method strategy in the 2HDM analysis.

to satisfy the selections listed in Section 6.3.2, which differ in terms of the anti-lepton rejection depending on the final state, but the isolation. The WZ and ZZ background contributions are subtracted in order to not include genuine taus in the computation of the fake rate. Given that the description of genuine leptons in simulation is trustable, these contributions have been directly estimated from simulation. This subtraction is another important difference with respect to the method adopted in Chapter 5. By counting the number of loose taus that also satisfy the isolation requirements (being then promoted to tight taus) in bins of p_T of the reconstructed jet closest (within a $\Delta R < 0.5$) to the loose tau, the misidentification probability $f(p_T)$ is obtained. If no jet is found close-by the tau, the p_T of the tau itself is considered. In Figure 6.8, tau fake rate corresponding to requiring in the final selection the loose anti-lepton rejection and the loose cut-based isolation, is reported.

ELECTRON AND MUONS FAKE RATES These fake rates are measured according to the same strategy used for taus. On top of a Z boson, only one lepton is required to satisfy all the kinematical requirements listed in Section 6.3.2 with the exception of requiring looser identification and isolation requirements. The loose selection requirement correspond to particle-flow very loose and no identification for electrons and muons respectively and a particle-flow rel-

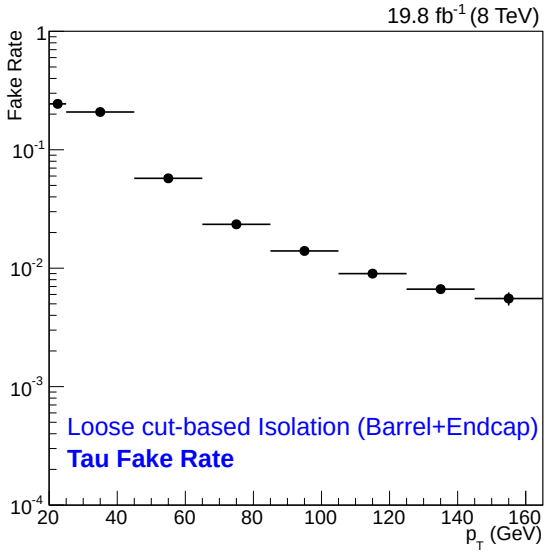

Figure 6.8: Tau fake rate measured as a function of the closest-jet-to-tau p_T. The tight requirements correspond to loose anti-lepton rejections and loose cut-based isolation working points [A].

ative isolation less than 0.5 for both. The lepton fake rates have been obtained in bins of the closest-jet-to-lepton p_T. Also in this case, the contributions of genuine leptons have been estimated from simulation and subtracted. Examples of electron and muon fake rates are reported in Figure 6.9. This figure reports the fake rates corresponding to a final selection in which the particle-flow loose identification and $I^{PF}_{rel} < 0.3$ for electrons (particle-flow loose identification and $I^{PF}_{rel} < 0.3$ for muons) are required.

Contrary to the analysis in Chapter 5, no tau fake rate measurement has been attempted in bins of η as it has been checked that a separate description in the barrel and in the endcap leads to compatible results within the statistical uncertainties. Moreover, it has been verified that a negligible difference exists between the results obtained by fitting the fake rate histograms with an exponential function and those obtained by using directly the histograms. Thus, also in this case, the simplest approach has been chosen and no fits have been performed.

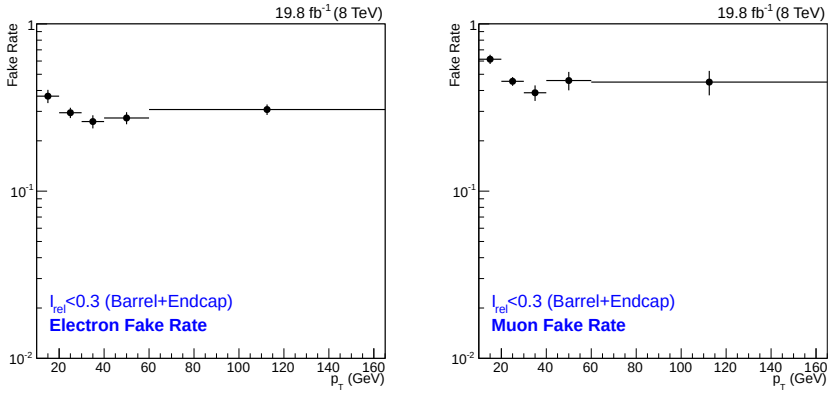

Figure 6.9: Electron (left) and muon (right) fake rates measured as a function of the closest-jet-to-lepton p_T. The fake rate represents the probability that a particle-flow very loose (no) identified electron (muon), with $I^{PF}_{rel} < 0.5$, satisfies the particle-flow loose identification and $I^{PF}_{rel} < 0.3$ [A].

6.5.2 Derivation of data-driven estimate using fake rates

The strategy used to estimate the fake background contribution is formally expressed in Equation 5.10. The analytical derivation of the formula, provided in Section 5.7.2, is based on the definition of three control regions (CR_{01}, CR_{10} and CR_{00}) in which at least one or both leptons selected on top of a Z boson, satisfy the loose identification requirements, but not the tight ones. To estimate the fake background in the signal region, the number of events counted in each control region has to be properly reweighted by using the lepton fake rates measured as previously explained. The weights, in particular, depend on the p_T of the misidentified lepton and its flavor. The contamination of genuine leptons tagged wrongly as fakes has been estimated from simulation and subtracted from each control region. In particular, to have an idea about the amount of the physical background contamination, one can look at the closest-jet p_T distributions for each final state in each control region. It has been found that mainly the ZZ, and in minor fractions WWZ, WZZ, ZZZ, Zh, and $t\bar{t}Z$ processes need to be subtracted. Two examples of closest-jet p_T distributions are reported in Figure 6.10 for the CR_{00} of the $e\tau_h$ final state (left) and for the CR_{01} of the $\tau_h\tau_h$ final state (right).

Moreover, it would seem that this method assumes that the prompt rate, i.e. the probability for a prompt lepton to pass both the loose criteria and the tight ones, is equal to unity. In reality, this assumption might be not necessarily true,

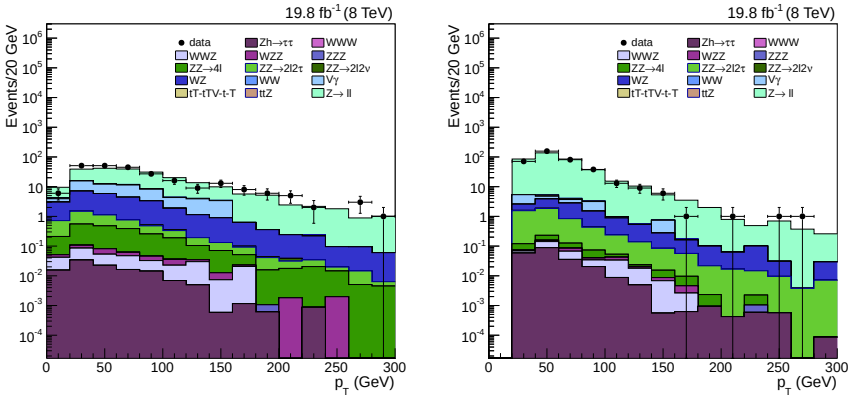

Figure 6.10: Closest-jet p_T distributions related the CR_{00} of the $e\tau_h$ final state (left) and the CR_{01} of the $\tau_h\tau_h$ final state (right). The data-driven contribution is superimposed to all the other stacked contributions estimated from simulation. It can be seen that the main contamination from genuine leptons to the fake-background estimate in the signal region comes from the ZZ process and this contribution, together with other minor ones (*WWZ*, *WZZ*, ZZZ, Zh, and $t\bar{t}Z$), has been subtracted [A].

with the drawback of introducing additional systematic uncertainties. It has been numerically proved (cfr. Appendix E) that the formula used to estimate the fake background (cfr. Equation 5.10), combined with the subtraction of the physical background, is totally equivalent with a more complex formula, which takes into account the prompt rate [168].

The same control regions used for the estimate of the fake background yields, have been used to extract the relative shapes. In the latter case, however, in order to reduce the statistical uncertainty, both same-sign and opposite-sign dilepton pairs have been considered, whereas the shape normalization has been obtained from just opposite-sign events.

For the estimate of the fake-background yields, Equation 5.10 has been applied, allowing negative results in the case in which the contribution of the control-region CR_{00} is greater than the sum of the other two ($CR_{01} + CR_{10}$). Negative results can also be obtained in the case in which no events are counted in data and a non null physical background contamination is estimated from simulation. Given the presence of zero or very small counts, a proper treatment of the statistical uncertainties needs to be performed in order to avoid to be biased from these "pathological" cases. The final background estimate, with its associated statistical uncertainty, has thus been obtained with two methods: the "analytical error propagation" and the "toy-based".

ANALYTICAL ERROR PROPAGATION A statistical error equal to 1.8 [169] has been assigned in case zero events are counted in data. Instead, a standard Poisson error ($\sqrt{N}$) has been assigned in case of non zero counts[1], for both the events counted in data and simulation. Subsequently, the standard analytical error propagation [170] has been performed to extract the statistical error to be assigned to the results of Equation 5.10. In case a negative background estimate is obtained from Equation 5.10, the Feldman-Cousins procedure [171] has been adopted to infer a less biased best estimate and associated uncertainty: the final estimate has been considered as zero and its associated statistical uncertainty has been corrected with the Feldman-Cousins 68% CL upper limit. The results obtained with this approach are those documented in [22]. However, the standard analytical error propagation is only valid under the assumption of having quantities normally distributed and for which the approximation of having Poisson errors is valid, which occurs in the limit of a high number of events. This procedure is thus expected to lead to biased results in the case of this analysis.

TOY-BASED A more accurate method for the estimate of the errors has been therefore studied. The toy based method is a numerical approach that consists in generating several pseudo-experiments to model the real one under given assumptions. The starting point is rewriting the final formula (Equation 5.10) for the estimate of the fake background as:

$$N_{11}^{obs} = N_{01}^{obs} \cdot w_{01} + N_{10}^{obs} \cdot w_{10} - N_{00}^{obs} \cdot w_{00} - X^{MC} \,, \tag{6.1}$$

where, as already explained, N_{ij}^{obs} is the number of data events counted in a control region CR_{ij}, w_{ij} is the associated weight depending on the lepton fake rate and X^{MC} is a term formally introduced here for the first time, that represents the physical background contribution estimated from simulation (and subtracted from data). In each pseudo-experiment four quantities are randomly generated:

- **three N_{ij}^{toy}** (one for each control region) extracted according to the Bayesian approach [170], by performing a random sampling from a posterior prob-

[1] The probability distribution of a Poisson random variable X with parameter $N > 0$ is given by the formula $P(X) = e^{-N} N^k/k!$, where $k = 0, 1, 2, \ldots$. The mean $E(X)$ and the variance $V(X)$ of a Poisson distribution are both equal to N. Therefore, the standard deviation σ, which represent the statistical error assigned to the measurement N distributed as a Poisson, is: $\sigma = \sqrt{V(X)} = \sqrt{N}$.

ability density function $p(N_{ij}^{toy}|N_{ij}^{obs})$, whose expression is given by:

$$\begin{aligned} p(N_{ij}^{toy}|N_{ij}^{obs}) &= \text{Poisson}(N_{ij}^{toy}|N_{ij}^{obs}) \\ &\equiv e^{-N_{ij}^{obs}} \frac{(N_{ij}^{obs})^{N_{ij}^{toy}}}{N_{ij}^{toy}!} , \end{aligned} \tag{6.2}$$

assuming a flat prior on N_{ij}^{toy}.

- **One X^{toy}**. Given the large number of counts from the simulated samples, this quantity has been extracted randomly by assuming a Gaussian *pdf* with mean X^{MC} and sigma σ_X^{MC}.

For a given final state, L_T selection, and mass bin[2], 10^5 pseudo-experiments have been generated and, all the 10^5 outcomes obtained by applying Equation 6.1, have been collected in a histogram, which the final fake background estimate has been extracted from. Equation 6.1 has been applied by considering lepton-flavor related weights w_{ij} averaged on the closest-jet p_T range (see Figures 6.8 and 6.9) and considering any negative outcome as zero. Following the recommended prescriptions of the LHC Combination group [172], the final estimate has been extracted by fitting the final histogram with a log-normal distribution (cfr. Equation B.2). Two examples of the final distributions obtained, with the log-normal fit superimposed, are shown in Figure 6.11 for the $\mu\mu\mu\tau_h$ and the $\mu\mu\tau_h\tau_h$ channel, having considered $L_T > 20$ GeV in both cases and only one single mass bin (cut-and-count analysis).

6.5.3 Results

Representative fake backgrounds yields are reported for two L_T selections ($L_T > 20$ GeV and $L_T > 180$ GeV) in Table 6.2, where the analytical error propagation and the toy-based method are also compared. A comparison of the fake background shapes is reported in Figure 6.12 for the $\mu\mu\mu\tau_h$ and the $\mu\mu\tau_h\tau_h$ channels, in correspondence of the selection $L_T > 20$ GeV.

The conclusion that can be drawn by having a look at these representative results, but that is still valid for almost all the L_T selections and channels, is that the toy-based best estimates are higher than those of the analytical method. On the other hand, the associated uncertainties are smaller. As a

[2]Pseudo-experiments have been generated in one single bin of the ditau SVfit mass spectra in the context of the cut-and-count analysis and in eight different mass bins for the shape-based analysis.

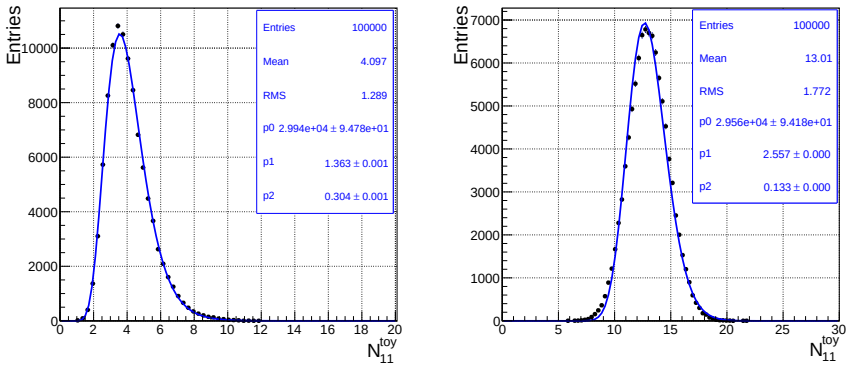

Figure 6.11: Examples of the final toy-based distributions (dots) obtained for the estimate of the fake background, fitted with a log-normal function (line). The distribution on the left is related to the $\mu\mu\mu\tau_h$ channel, the one on the right to the $\mu\mu\tau_h\tau_h$ channel. In both cases, the $L_T > 20$ GeV has been considered and the estimate has been performed in one single ditau mass bin (cut-and-count analysis) [A].

	$L_T > 20$ GeV		$L_T > 180$ GeV	
Channel	Analytical	Toy-based	Analytical	Toy-based
$eee\mu$	0.146 ± 1.85	0.685 ± 1.1	0 ± 1.96	0.628 ± 1.02
$eee\tau_h$	1.09 ± 1.21	1.52 ± 1.24	0.073 ± 0.756	0.413 ± 0.44
$ee\mu\tau_h$	0.482 ± 1.82	1.19 ± 1.12	0.080 ± 1.67	0.628 ± 0.955
$ee\tau_h\tau_h$	11.7 ± 2.58	11.7 ± 1.74	0.262 ± 0.263	0.357 ± 0.244
$\mu\mu e\mu$	0.252 ± 1.9	1.29 ± 1.68	0 ± 1.96	0.626 ± 1.02
$\mu\mu e\tau_h$	0.259 ± 0.933	0.827 ± 0.885	0 ± 0.754	0.32 ± 0.405
$\mu\mu\mu\tau_h$	3.24 ± 1.18	3.91 ± 1.38	0.072 ± 1.67	0.616 ± 0.921
$\mu\mu\tau_h\tau_h$	12.9 ± 2.84	12.9 ± 1.84	0.057 ± 0.025	0.221 ± 0.038

Table 6.2: Fake background yields estimated for $L_T > 20$ GeV and $L_T > 180$ GeV. Results obtained with the analytical error propagation and the toy-based methods are compared.

consequence, it has been observed that the expected 95% CL upper limits on the signal strength modifier μ are greater in the case of the toy-based approach than in the analytical one. Concerning the observed limits, however, it can be noticed an opposite behavior: the observed 95% CL upper limits on the signal strength modifier μ are smaller in the case of the toy-based approach than

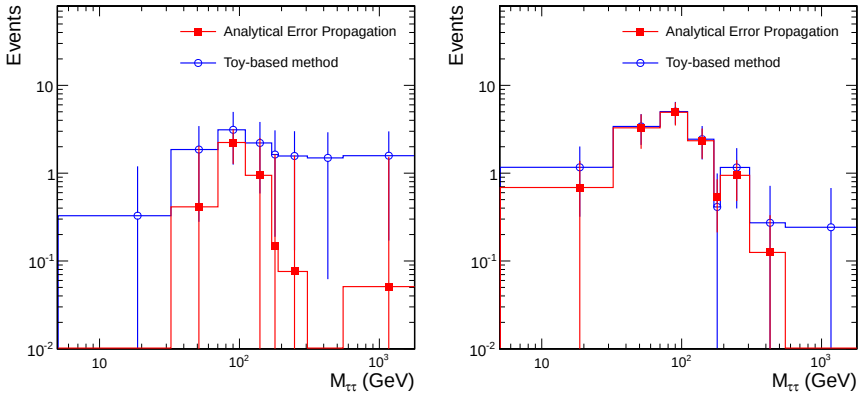

Figure 6.12: Fake background shapes estimated for $L_T > 20$ GeV for the $\mu\mu\mu\tau_h$ (left) and the $\mu\mu\tau_h\tau_h$ (right) channels. Results obtained with the analytical error propagation and the toy-based methods are compared [A].

in the analytical one. This happens because the counts observed in data lie, in general, in between of the level of background estimated with the analytical error approach and the level of background estimated with the toy based approach. As a result, in the latter case, slightly tighter limits are estimated and wider regions are excluded than in the former case. It is true that, in the observed limits, also the uncertainty bands can play a role, but, in this case, the results can be fully explained by simply reasoning on the yields only. All the results presented in Section 6.7 have been obtained by estimating the fake background with the toy-based approach. These results are also those documented in the analysis corresponding paper [23]. However, the reader can compare these results with those documented in [22], which, instead, are based on the analytical-error-propagation approach.

In order to evaluate the validity of the method just explained, a closure test has been performed on simulated fake background processes. The fake rate method has been applied on those simulated samples that represent exactly that kind of background topology which needs to be rejected (mainly Z+jets and WZ+jets processes). The misidentification probabilities have also been obtained from the same simulated samples. The comparison between the results of this procedure and the true yields obtained in the signal region has provided not only a reasonable measure of the validity of the method in data, but it has also allowed to assess a systematic uncertainty of 40%, for all channels and all L_T thresholds, on the estimates of the fake backgrounds obtained with this method.

6.6 Systematic uncertainties

All systematic uncertainties considered in the context of this analysis are summarized in the Tab. 6.3.

Source	Uncertainty [%]	Type
Luminosity	2.6	rate
Pile-up effects	$1-3$	shape
Lepton ID/Iso/ES	2	rate
Lepton trigger efficiency	1	rate
Tau ID/Iso	6	rate
Tau ES	3	shape
Jet ES/resolution	3	shape
Fake bkg. estimate	40	shape
b-jet veto	1	rate
Bkg. normalization (ZZ)	11	rate

Table 6.3: Summary of the systematic uncertainties considered in this analysis. The last column quotes whether the source of uncertainty is assigned as rate (normalization) uncertainty or as a shape (probability distribution function) uncertainty. The statistical treatment of these uncertainties, depending on their type, has been extensively described in Appendix B.

The majority of the sources of systematic uncertainties are in common with those already listed for the SM Zh analysis (cfr. Section 5.8) as well as the CMS recommended methods that have been used to estimate their impact. This is the case for the uncertainties related to: luminosity; light lepton trigger, light lepton and tau identification, isolation, energy scales (indicated as ID, Iso and ES, respectively); the b-jet veto. The limited knowledge of the total inelastic cross section used to estimate the expected number of interactions for pileup-reweighting has been considered as an additional source of uncertainty in the context of this analysis.

Moreover, differently with respect to the SM Zh analysis, the uncertainties related to the jet energy scale and resolution (cfr. Section 3.5.4) have been taken into account and treated according to the prescription provided by CMS [173]. The effect of the jet energy scale and resolution have been estimated by varying the nominal jet energy scale (or resolution) by $\pm 1\sigma$. The systematic uncertainties estimated on leptons, taus and jets have been then propagated on the $\not{E}_T$ and, therefore, on the mass shapes. Given the definition of the missing en-

ergy, the propagation on the $\not{E}_T$ estimator of the effect due to the uncertainties previously listed, has been done in two steps:

- adding leptons, taus and jets contribution to the $\not{E}_T$ estimator
- subtracting the corresponding object collection contributions, once the nominal energy scale (or resolution) have been varied by $\pm 1\sigma$ (for electrons, muons, taus and jets).

As already explained, the uncertainty to be assigned to the fake background estimate has been estimated in the context of the closure test. In particular the uncertainty has been computed as $|A - B|/A$ where A represents the "data-driven estimate" of the simulated fake background events and B the pure expected prediction of the same events in the signal region. This computation has been performed for all the channels and for different L_T cut values, considering only those cases with small enough statistical uncertainties on simulation. As can be seen from Figure 6.13, where two examples of the outcome of this procedure are reported for the $\mu\tau_h$ and the $\tau_h\tau_h$ channels, the three last highest L_T thresholds probed in the optimization procedure (160, 180, 200 GeV) have not been included because for those thresholds the simulated fake background samples have not enough statistics to be reliable on the prediction of the fake background using the data-driven method. Figure 6.13 shows that the quantity $|A - B|/A$ is around 40% (dashed line) for all the channels and for all the possible analysis selections within the statistical errors. This value has thus been considered as the systematic uncertainty on the fake background estimation.

An uncertainty of 11% has been assigned to the ZZ background normalization. This value comes from the propagation of the uncertainties from the CMS cross section measurements of the ZZ process [174]. Uncertainties on additional background processes, such as WZ, $t\bar{t}$, Z+jets, do not play any role in this analysis because they are estimated from data control regions. From the statistical point of view, the uncertainties have been modelled according to the statistical methods described in Appendix B.

6.7 Results

As already anticipated, both a cut-and-count and a shape-based analysis, which uses as observable the distribution of the $M_{\tau\tau}$ mass reconstructed with the SV-fit algorithm [144] (cfr. Section 4.6), have been performed. The cut-and-count

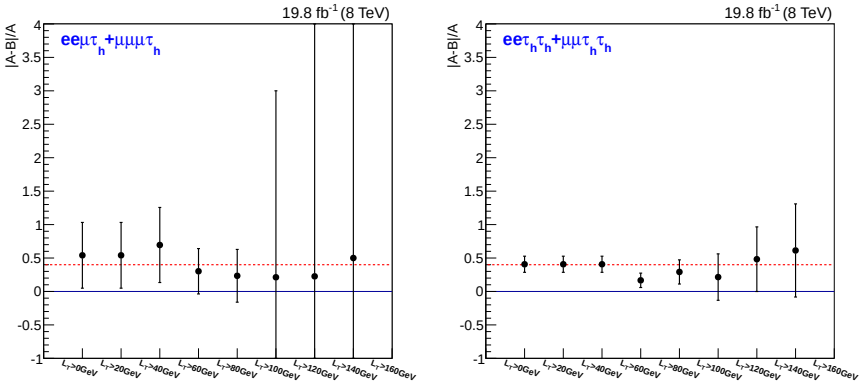

Figure 6.13: Fake background systematic uncertainties estimated for $\mu\tau_h$ channel (left) and $\tau_h\tau_h$ channel (right). The y-axis reports the absolute difference between the "data-driven estimate" results obtained using simulated fake background events (A) and the pure expectation of the same events in the signal region (B), divided by the the data-driven estimate ($|A - B|/A$). Each bin on the x-axis refers to a given L_T threshold [A].

approach, which is based only on the number events passing a selection dependent on the values of M_H and M_A, is more model-independent than the shape-based search, which is however more sensitive to the benchmark signal. The expected number of signal and background events, together with those observed in data and obtained after having applied the full optimized signal selection, are summarized in Table 6.4 for the signal $H(350\ \text{GeV}) \to ZA(150\ \text{GeV})$ and in Table 6.5 for the signal $H(500\ \text{GeV}) \to ZA(15\ \text{GeV})$. The optimized L_T thresholds are:

- for the signal $H(350\ \text{GeV}) \to ZA(150\ \text{GeV})$, $L_T > 40$ GeV for $e\mu$, $L_T > 80$ GeV for $e\tau_h$, $L_T > 60$ GeV for $\mu\tau_h$, $L_T > 100$ GeV for $\tau_h\tau_h$;
- for the signal $H(500\ \text{GeV}) \to ZA(15\ \text{GeV})$, $L_T > 80$ GeV for $e\mu$, $L_T > 20$ GeV for $\mu\tau_h$ and $L_T > 60$ GeV for $e\tau_h$ and $\tau_h\tau_h$.

In both cases, the signal has been normalized to $\sigma \times BR = 1$ fb and, concerning the fake background yields estimated with the data-driven method, statistical and systematic uncertainties have been explicitly shown. All the other systematic uncertainties affecting the rate, which are included in the final limit computation, are not explicitly shown. The two illustrative signals have been chosen because of the different boosts that characterize the ditau pair (cfr. Figure 6.3).

Concerning the shape-based analysis, Figures 6.14 and 6.15 show, for the same benchmark signal samples, the $M_{\tau\tau}$ distributions for each of the eight channels separately. The optimization performed in the context of the shape-based analysis has led, in general, to different L_T threshold with respect to those used in the cut-and-count analysis:

- for the signal $H(350\,\text{GeV}) \to ZA(150\,\text{GeV})$, $L_T > 60$ GeV for $e\mu$ and $\mu\tau_h$, $L_T > 80$ GeV for $e\tau_h$ and $L_T > 120$ GeV for $\tau_h\tau_h$;
- for the signal $H(500\,\text{GeV}) \to ZA(15\,\text{GeV})$, $L_T > 60$ GeV for $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$ and $L_T > 180$ GeV for $e\mu$.

In order to know how well the data compares to the background modeling and to the systematic uncertainty, a preliminary study on the pulls and the a posteriori variances has been performed also in the context of this analysis (cfr. Section 5.9.1). The study performed on a significant number of signal benchmark points has revealed that, in the context of the shape-based analysis, the bin-by-bin statistical uncertainties related to the fake background estimation are, in general, affected by $\sigma_{out}/\sigma_{in} > 10\%$ and $\Delta z/\sigma_{in} \approx -0.4$. Pulls and the a posteriori variances related to the systematic uncertainty on the fake background estimates are reported in Table 6.6 for both the cut-and-count and the shape-based analysis and for both the benchmark signal processes considered so far. The results obtained, under the assumption of excluding any possible statistical fluctuation in data, might reveal an over-estimation of the a priori value assigned to the fake-background systematic.

As can observed from the cut-and-count and shape-based analyses results, in the current data sample no significant evidence of a signal over the considered backgrounds has been observed. Therefore the CLs criterion has been used to determine 95% CL limits on the signal cross section times branching ratio. The statistical interpretation of the results follows the methods described in Appendix B.

Figure 6.16 shows the 95% CL expected and observed limits on the $\sigma \times BR$ for the cut-and-count and the shape-based analyses. Limits related to the lower-right triangle, which corresponds to the signal process $A \to ZH$, have been obtained by simply mirroring the results obtained in the higher-left triangle (cfr. Section 6.2). As can be seen from Figure 6.16, cross sections of about 5 fb are excluded in the majority of the M_A-M_H plane for both cut-and-count and shape based analyses.

$H(350\,\text{GeV}) \to ZA(150\,\text{GeV})$				
Processes	$eee\mu$	$eee\tau_h$	$ee\mu\tau_h$	$ee\tau_h\tau_h$
$Zh(M_h = 125\,\text{GeV})$	0.074 ± 0.006	0.067 ± 0.006	0.20 ± 0.01	0.106 ± 0.008
WZZ	0.008 ± 0.004	0.007 ± 0.004	0.007 ± 0.004	0.004 ± 0.003
ZZZ	0.0017 ± 0.0009	–	0.0022 ± 0.0009	$(0.008 \pm 0.006) \times 10^{-1}$
WWZ	0.20 ± 0.03	0.03 ± 0.01	0.09 ± 0.02	–
$ZZ \to 4\ell$	0.020 ± 0.005	0.11 ± 0.01	0.09 ± 0.01	0.050 ± 0.008
$ZZ \to 2\ell 2\tau$	1.37 ± 0.04	0.66 ± 0.03	2.44 ± 0.06	0.89 ± 0.03
$t\bar{t}Z$	0.08 ± 0.04	–	–	–
Fakes	$0.7 \pm 1.1 \pm 0.3$	$0.3 \pm 0.4 \pm 0.1$	$0.8 \pm 1.0 \pm 0.3$	$0.7 \pm 0.2 \pm 0.3$
Total Bkg.	$2.5 \pm 1.1 \pm 0.3$	$1.2 \pm 0.4 \pm 0.1$	$3.7 \pm 1.0 \pm 0.3$	$1.7 \pm 0.2 \pm 0.3$
Signal	0.088 ± 0.006	0.094 ± 0.006	0.191 ± 0.009	0.153 ± 0.008
Data	1	0	3	4
Processes	$\mu\mu e\mu$	$\mu\mu e\tau_h$	$\mu\mu\mu\tau_h$	$\mu\mu\tau_h\tau_h$
$Zh(M_h = 125\,\text{GeV})$	0.117 ± 0.008	0.102 ± 0.008	0.25 ± 0.01	0.139 ± 0.009
WZZ	0.013 ± 0.005	0.003 ± 0.002	0.008 ± 0.004	0.004 ± 0.003
ZZZ	$(0.005 \pm 0.005) \times 10^{-1}$	0.0014 ± 0.0008	0.003 ± 0.001	0.0020 ± 0.0009
WWZ	0.26 ± 0.04	0.08 ± 0.02	0.08 ± 0.02	0.02 ± 0.01
$ZZ \to 4\ell$	0.029 ± 0.006	0.11 ± 0.01	0.11 ± 0.01	0.063 ± 0.009
$ZZ \to 2\ell 2\tau$	1.84 ± 0.05	0.71 ± 0.03	3.29 ± 0.07	1.19 ± 0.04
$t\bar{t}Z$	0.13 ± 0.05	–	0.04 ± 0.03	–
Fakes	$1.4 \pm 1.3 \pm 0.6$	$1.0 \pm 0.8 \pm 0.4$	$1.9 \pm 1.0 \pm 0.7$	$0.7 \pm 0.2 \pm 0.3$
Total Bkg.	$3.8 \pm 1.3 \pm 0.6$	$2.0 \pm 0.8 \pm 0.4$	$5.6 \pm 1.0 \pm 0.7$	$2.1 \pm 0.2 \pm 0.3$
Signal	0.107 ± 0.007	0.103 ± 0.007	0.27 ± 0.01	0.162 ± 0.008
Data	4	1	2	2

Table 6.4: Expected number of signal and background events compared to the number of observed events in data. The signal $M_H = 350\,\text{GeV} - M_A = 150\,\text{GeV}$ is normalized to $\sigma \times BR = 1$ fb. The yields correspond to $L_T > 40$ GeV for $e\mu$, $L_T > 80$ GeV for $e\tau_h$, $L_T > 60$ GeV for $\mu\tau_h$ and $L_T > 100$ GeV for $\tau_h\tau_h$. The statistical uncertainties are reported for all the signal and background processes, while only systematic uncertainties related to the fake background estimated from data are shown.

$H(500\,\text{GeV}) \to ZA(15\,\text{GeV})$				
Processes	$eee\mu$	$eee\tau_h$	$ee\mu\tau_h$	$ee\tau_h\tau_h$
$Zh(M_h = 125\,\text{GeV})$	0.021 ± 0.003	0.101 ± 0.008	0.27 ± 0.01	0.25 ± 0.01
WZZ	0.006 ± 0.003	0.009 ± 0.004	0.009 ± 0.004	0.009 ± 0.004
ZZZ	$(0.005 \pm 0.004) \times 10^{-1}$	$(0.004 \pm 0.004) \times 10^{-1}$	0.0024 ± 0.0009	0.003 ± 0.001
WWZ	0.16 ± 0.03	0.04 ± 0.01	0.09 ± 0.02	$-$
$ZZ \to 4\ell$	0.018 ± 0.005	0.13 ± 0.01	0.12 ± 0.01	0.15 ± 0.01
$ZZ \to 2\ell 2\tau$	0.30 ± 0.02	1.26 ± 0.04	4.88 ± 0.08	3.67 ± 0.07
$t\bar{t}Z$	0.04 ± 0.03	$-$	$-$	$-$
Fakes	$0.6 \pm 1.0 \pm 0.2$	$0.4 \pm 0.5 \pm 0.1$	$1.2 \pm 1.1 \pm 0.5$	$6.0 \pm 0.9 \pm 2.4$
Total Bkg.	$1.1 \pm 1.0 \pm 0.2$	$1.9 \pm 0.5 \pm 0.1$	$6.6 \pm 1.1 \pm 0.5$	$10.1 \pm 0.9 \pm 2.4$
Signal	0.061 ± 0.005	$(0.009 \pm 0.006) \times 10^{-1}$	0.002 ± 0.001	0.0011 ± 0.0007
Data	0	1	3	10
Processes	$\mu\mu e\mu$	$\mu\mu e\tau_h$	$\mu\mu\mu\tau_h$	$\mu\mu\tau_h\tau_h$
$Zh(M_h = 125\,\text{GeV})$	0.047 ± 0.005	0.138 ± 0.009	0.35 ± 0.01	0.33 ± 0.01
WZZ	0.013 ± 0.005	0.007 ± 0.003	0.011 ± 0.004	0.012 ± 0.005
ZZZ	$(0.005 \pm 0.005) \times 10^{-1}$	0.002 ± 0.001	0.004 ± 0.001	0.005 ± 0.001
WWZ	0.22 ± 0.04	0.09 ± 0.02	0.09 ± 0.02	0.02 ± 0.01
$ZZ \to 4\ell$	0.017 ± 0.005	0.14 ± 0.01	0.16 ± 0.01	0.16 ± 0.01
$ZZ \to 2\ell 2\tau$	0.32 ± 0.02	1.51 ± 0.04	6.58 ± 0.09	4.84 ± 0.08
$t\bar{t}Z$	0.12 ± 0.05	0.02 ± 0.02	0.06 ± 0.04	$-$
Fakes	$0.6 \pm 1.0 \pm 0.2$	$1.2 \pm 0.9 \pm 0.5$	$3.9 \pm 1.4 \pm 1.6$	$4.0 \pm 0.8 \pm 1.6$
Total Bkg.	$1.3 \pm 1.0 \pm 0.2$	$3.1 \pm 0.9 \pm 0.5$	$11.2 \pm 1.4 \pm 1.6$	$9.4 \pm 0.8 \pm 1.6$
Signal	0.061 ± 0.005	$(0.010 \pm 0.007) \times 10^{-1}$	0.002 ± 0.001	$(0.002 \pm 0.002) \times 10^{-1}$
Data	2	2	5	7

Table 6.5: Expected number of signal and background events compared to the number of observed events in data. The signal $M_H = 500\,\text{GeV} - M_A = 15\,\text{GeV}$ is normalized to $\sigma \times BR = 1$ fb. The yields correspond to $L_T > 80$ GeV for $e\mu$, $L_T > 20$ GeV for $\mu\tau_h$ and $L_T > 60$ GeV for $e\tau_h$ and $\tau_h\tau_h$. The statistical uncertainties are reported for all the signal and background processes, while only systematic uncertainties related to the fake background estimated from data are shown.

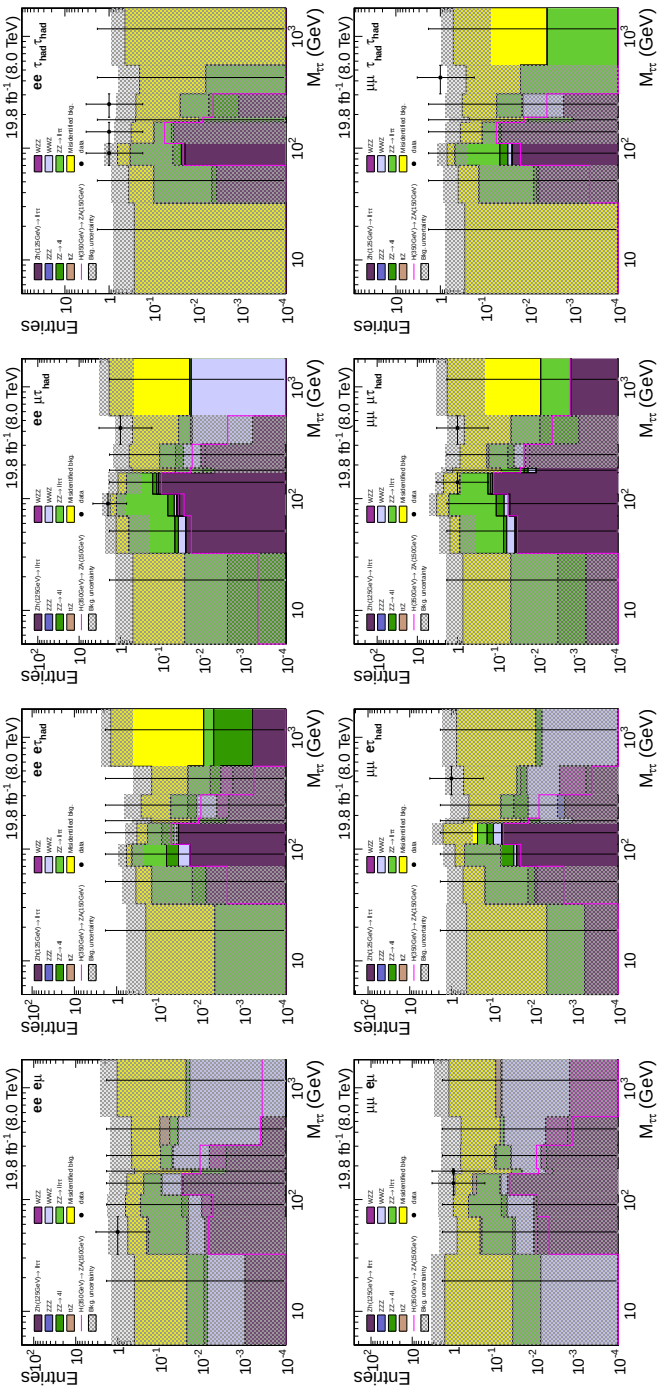

Figure 6.14: SVfit $M_{\tau\tau}$ distributions for the signal H(350 GeV) $\rightarrow$ ZA(150 GeV) with corresponding $\sigma \times BR = 1$ fb. The mass templates correspond to the following optimized L_T thresholds: $L_T > 60$ GeV for $e\mu$ and $\mu\tau_h$, $L_T > 80$ GeV for $e\tau_h$ and $L_T > 120$ GeV for $\tau_h\tau_h$. Garwood intervals for data points are reported [A].

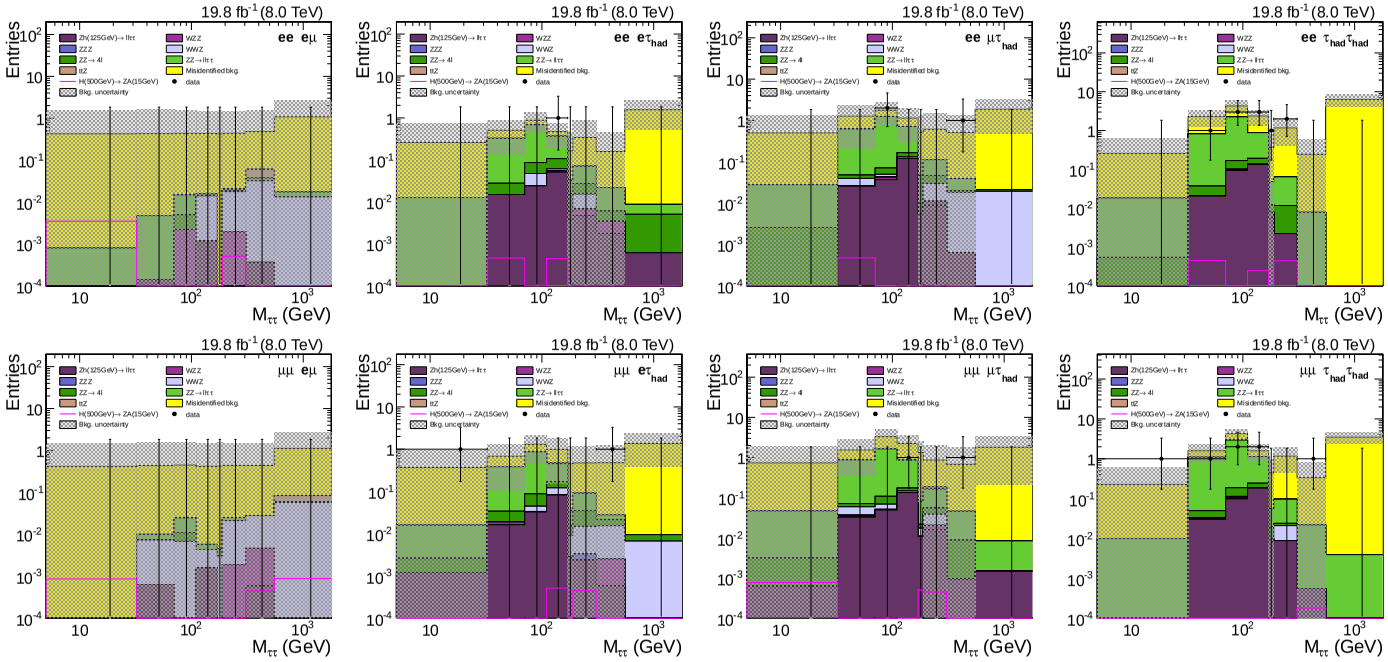

Figure 6.15: SVfit $M_{\tau\tau}$ distributions for the signal $H(500\ \text{GeV}) \to ZA(15\ \text{GeV})$ with corresponding $\sigma \times BR = 1$ fb. The mass templates correspond to the following optimized L_T thresholds: $L_T > 60$ GeV for $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$ and $L_T > 180$ GeV for $e\mu$. Garwood intervals for data points are reported [A].

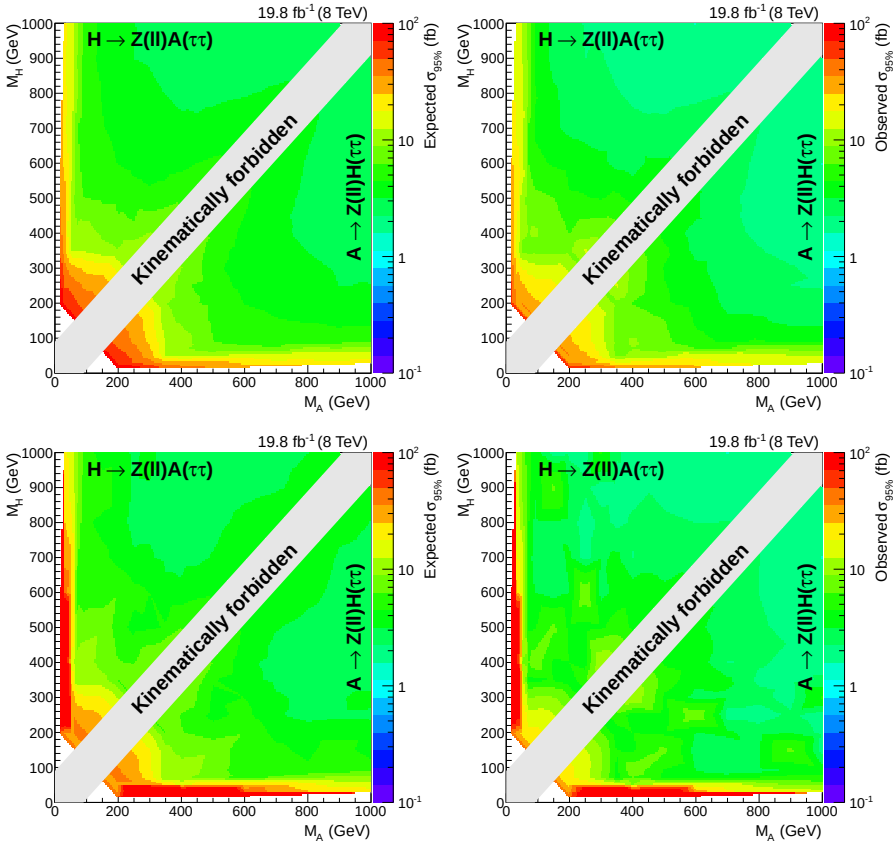

Figure 6.16: Limits on $\sigma \times BR$ for the process $H/A \to ZA/H \to \ell\ell\tau\tau$ as a function of M_A and M_H. Expected (left) and observed (right) limits are shown for the cut-and-count (top) and the shape-based (bottom) analyses [A].

$H(350\ GeV) \to ZA(150\ GeV)$			
$\Delta z/\sigma_{in}$	σ_{out}/σ_{in}	$\Delta z/\sigma_{in}$	σ_{out}/σ_{in}
cut-and-count		shape-based	
-0.30	0.89	-1.59	0.68
$H(500\ GeV) \to ZA(15\ GeV)$			
$\Delta z/\sigma_{in}$	σ_{out}/σ_{in}	$\Delta z/\sigma_{in}$	σ_{out}/σ_{in}
cut-and-count		shape-based	
-0.63	0.76	-2.20	0.63

Table 6.6: Pulls ($\Delta z/\sigma_{in}$) and a posteriori variances (σ_{out}/σ_{in}) related to the systematic uncertainty on the fake background estimation. Results for the signal samples $H(350\ GeV) \to ZA(150\ GeV)$ and $H(500\ GeV) \to ZA(15\ GeV)$ for both the cut-and-count and the shape-based analyses.

6.7.1 Interpretation in the Type-II 2HDM

The results shown so far can be interpreted in the context of the Type-II 2HDM (cfr. Section 1.5.1). The assumption made on the theoretical model are those listed in the context of the signal generation description (cfr. Section 6.2). Expected and observed limits on the signal strength $\mu = \sigma_{95\%}/\sigma_{TH}$, where the σ^{TH} is the approximate NNLO 2HDM theoretical prediction for the cross section as given by SUSHI, are shown in the Fig. 6.17, on left and right, respectively. On top and bottom of the same figure, cut-and-count and shape-based analyses results are reported, respectively. The 95% CL excluded regions are delimited by solid black lines. The dashed lines, instead, delimit those areas that represent the expected 68% of probability to exclude the region that have been actually excluded at the 95% CL.

The region with no exclusion power is the one in which $M_H > 300$ GeV and $M_A > 300$ GeV: in this region, in fact, there is drop in the signal cross section due to the enhancement of the $A/H \to t\bar{t}$ process [75]. As expected, the shape-based approach allows to exclude a wider region with respect to the cut-and-count approach, especially in the triangle related to the process $H \to ZA$. In the region in which $M_H > M_A + M_Z$, a portion of the phase space is excluded in the ranges $200 \lesssim M_H \lesssim 400$ GeV and $15 \lesssim M_A \lesssim 200$ GeV. In the mirrored region, the ranges $100 \lesssim M_H \lesssim 250$ GeV and $200 \lesssim M_A \lesssim 500$ GeV are excluded at the 95% CL. It has to be specified that such exclusion limits are valid under the assumption of a narrow-width approximation for the heavier Higgs involved in the process. It has been checked [22] that such approximation well describes what is theoretically modelled in a small region of the parameter space ($M_H < 400$ GeV and $M_A < 300$ GeV). Outside this region,

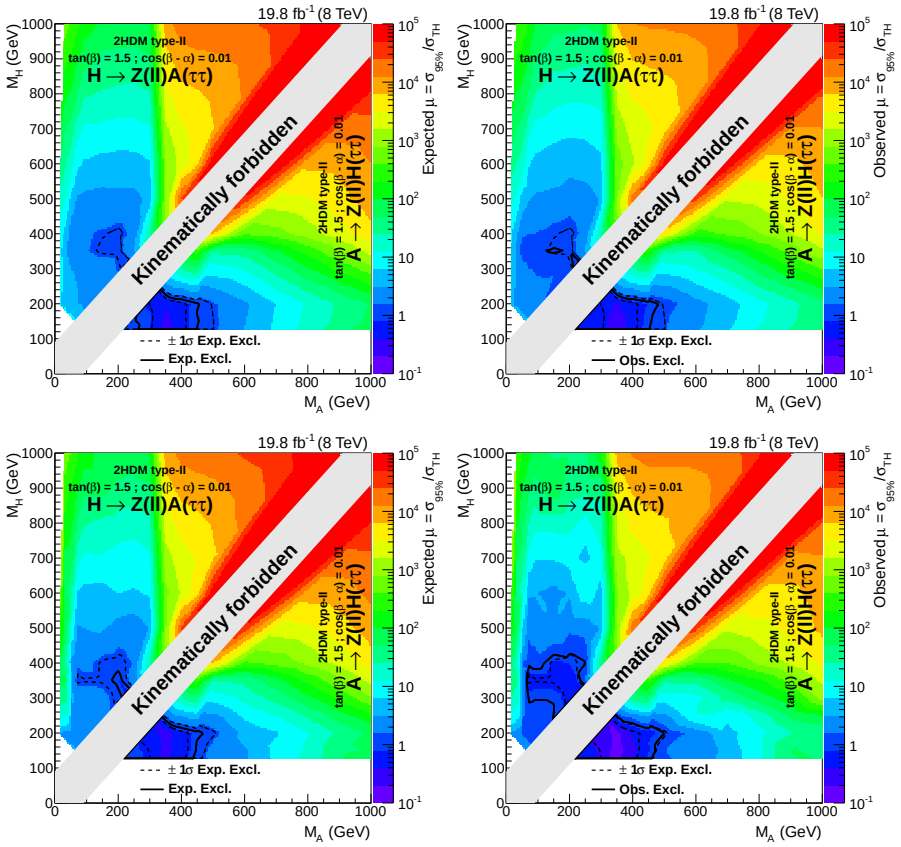

Figure 6.17: Limits on $\mu = \sigma_{95\%}/\sigma_{TH}$ for the $\ell\ell\tau\tau$ final state as a function of M_A and M_H. Cross sections are normalized to the approximate NNLO SuSHi predictions. Expected (left) and observed (right) limits are shown for the cut-and-count (top) and shape-based (bottom) analyses. $\pm 1\sigma$ expected exclusion contours are drawn (dashed lines), together with the actual excluded regions delimited with solid lines [A].

correction factors have to be taken into account for a more rigorous reinterpretation of the results.

6.8 $\tau\tau$ and $b\bar{b}$ combination

As already mentioned at the beginning of this chapter, the combined results of the $\ell\ell\tau\tau$ and $\ell\ell b\bar{b}$ final state analyses, which are, instead, documented separately in [22], are now in the reviewing process (to be submitted to the Physics Letter B journal) [23]. Expected and observed upper limits on the signal strength modifier μ have been derived for the combination of $\ell\ell\tau\tau$ and $\ell\ell b\bar{b}$ final states. In Figure 6.18, in which the observed limits are shown, the dashed (solid) contour represents the expected (observed) excluded area. It can be seen that a significant portion of the phase space is excluded: the ranges $200 \lesssim M_H \lesssim 700$ GeV and $20 \lesssim M_A \lesssim 200$ GeV are excluded for the process $H \to ZA$, while the ranges $200 \lesssim M_A \lesssim 700$ GeV and for $100 \lesssim M_H \lesssim 300$ GeV are excluded for the process $A \to ZH$.

It is worth mentioning that, in the $\ell\ell b\bar{b}$ final state, two moderate excesses are visible in the region around $(M_H, M_A) = (286, 93)$ GeV and $(M_H, M_A) = (662, 575)$ GeV, corresponding to a local significance of 2.6 and 2.85σ [22]. Going beyond the Type-II scenario, these results can be also explained in the context of the Type-III scenario, in which for $\tan\beta \geqslant 1$ the ratio of the A branching fraction into $\tau\tau$ to the one into $b\bar{b}$ is expected to be smaller than 10% [75] (cfr. Section 1.5.1).

The limits on the signal strength modifier have been also computed as a function of the 2HDM parameters $\tan\beta$ and $\cos(\beta - \alpha)$ for a given signal. These results are reported in Figure 6.19 in correspondence of the benchmark signal $H(378\text{ GeV}) \to ZA(188\text{ GeV})$. Both plots reported in Figure 6.19 show the same results in terms of the excluded region for the 2HDM parameters, which correspond to $0.5 \lesssim \tan\beta \lesssim 2.0$ and $-0.7 \lesssim \cos(\beta - \alpha) \lesssim 0.3$. In addition, the plot on the left shows that, the size of the excluded region is actually driven by the more sensitive $b\bar{b}$ final state (red contour), given that the region excluded by the $\tau\tau$ final state only (blue contour) is relatively smaller ($0.7 \lesssim \tan\beta \lesssim 1.7$ and $-0.3 \lesssim \cos(\beta - \alpha) \lesssim 0.1$). Moreover, looking at the right plot in Figure 6.19, the reader can appreciate how the signal strenght modifier varies in the $\tan\beta$-$\cos(\beta - \alpha)$ plane.

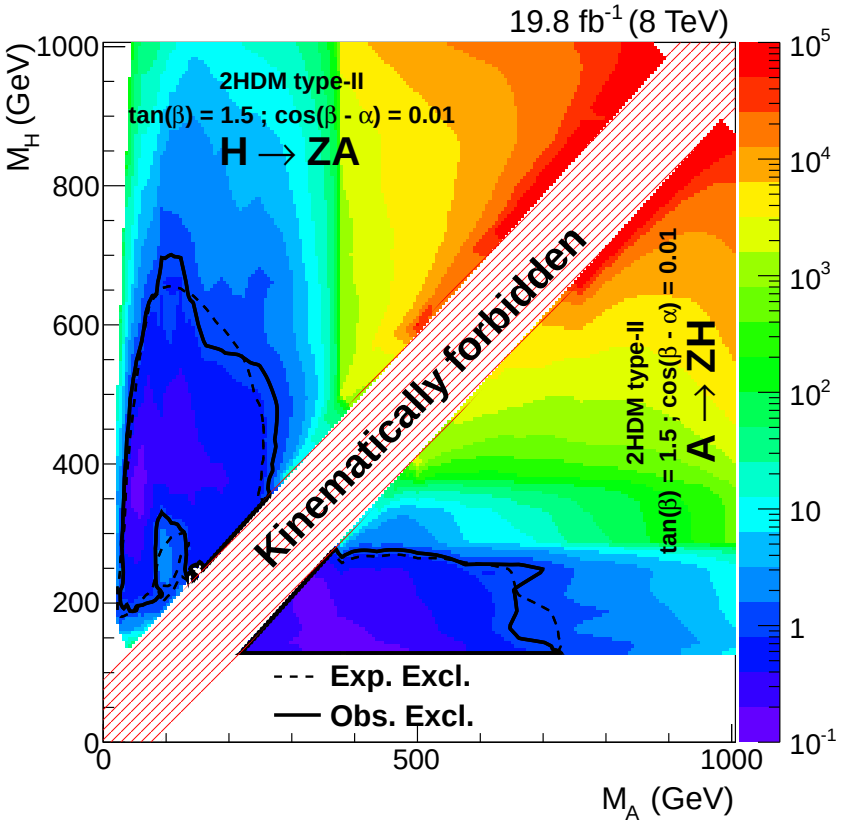

Figure 6.18: Observed limits on the signal strength modifier $\mu = \sigma_{95\%}/\sigma_{TH}$ for the 2HDM benchmark after the combination of $\ell\ell\tau\tau$ and $\ell\ell b\bar{b}$ final states. Cross sections are normalized to the approximate NNLO SUSHI predictions. The dashed contour delimits the expected excluded region, the solid contour, instead, the observed excluded region [AC].

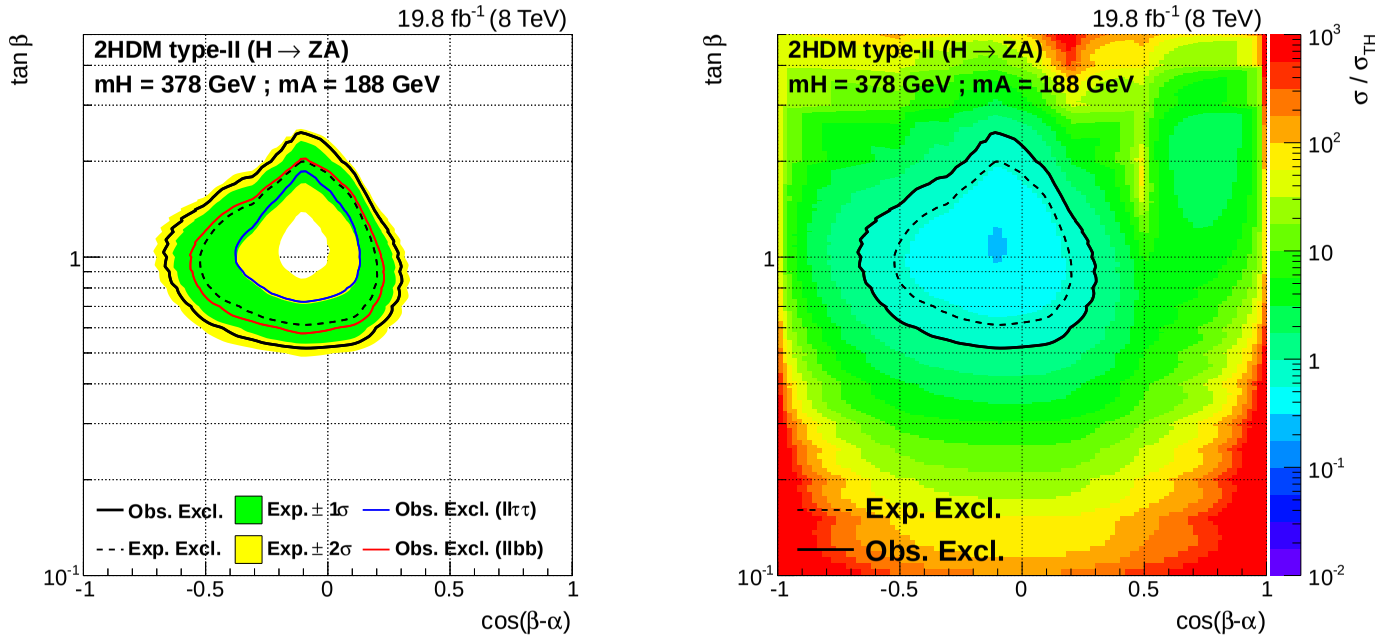

Figure 6.19: Combined ($\ell\ell\tau\tau$ and $\ell\ell b\bar{b}$) upper limits on the signal strength modifier $\mu = \sigma_{95\%}/\sigma_{TH}$ as a function of the 2HDM parameters $\tan\beta$ and $\cos(\beta - \alpha)$, for the signal $H(378\,\text{GeV}) \to ZA(188\,\text{GeV})$. In both plots expected (observed) limits are delimited by black dashed (solid) lines. Top: observed upper limits are also shown separately for the $\ell\ell\tau\tau$ (blue line) and the $\ell\ell b\bar{b}$ (red line) final states. Bottom: a map of the observed limits values on $\mu = \sigma_{95\%}/\sigma_{TH}$ is reported in the $\tan\beta$-$\cos(\beta - \alpha)$ plane [AC].

Conclusions and outlook

On March 30, 2010, the first proton-proton collisions produced by the LHC, the world's largest and most powerful particle accelerator, at a center-of-mass energy $\sqrt{s} = 7$ TeV, marked the beginning of a long-awaited new era in particle physics. The LHC did not disappoint the expectations: on July 4, 2012, the ATLAS [11] and the CMS [12] experiments announced the discovery of a new boson with a mass of 125 GeV. A tremendous and thrilling research work followed to understand if the newly discovered particle was *the* particle predicted by the Standard Model (SM), the theory that describes the ordinary matter and three of the four fundamental interactions. During the LHC Run I, ended in 2012 with p-p collisions at a center-of-mass energy of 8 TeV, the SM has been further confirmed.

In such a context, the search for the SM Higgs boson h decaying into a pair of tau leptons and produced in association with a Z boson decaying into either a dielectron or dimuon pair, belongs to those searches aiming to demonstrate that the newly discovered boson was the SM Higgs boson. This search has been presented in this thesis, as it has been performed during the first half of the Ph.D. period. In particular, the main contributions given to the publication of these deeply scrutinized results by the CMS collaboration [15] have been a study on the measurement of the background from jets misidentified as leptons, the estimation of all systematic uncertainties affecting the results and the development of an independent software code (event selection and background estimates) aimed at verifying the correctness of the analysis in view of the publication of the evidence for the decay of the 125 GeV Higgs boson to tau leptons. The analysis has exploited data collected by the CMS experiment during 2011 and 2012, corresponding to a total integrated luminosity of about 25 fb^{-1}. The analysis also has given no significant evidence of the SM Higgs boson and 95% CL exclusion limits were set, excluding a SM Higgs boson production rate above 5.4 time the expected SM value for a mass of 125 GeV. This result has however been combined with other searches for

the SM Higgs boson in the $\tau\tau$ decay channel, yielding the first evidence for the decay of the new boson into a ditau pair corresponding to a significance of 3.2σ. The combined signal strength and best fit mass are both compatible with the SM expectations. From this combination the couplings of the Higgs boson to fermions and vector bosons have been measured, obtaining results in agreement with the SM.

However, despite the experimental success of the SM, it is objective that this theory is not able to explain the whole Universe. It does not contemplate, for instance, the description of the dark matter and the gravity interaction, and it has also strong theoretical limitations at very high energies ($\mathcal{O}(10^{19}$ GeV$)$), as the hierarchy problem [21]. Among all the possible theories beyond the SM, the two-Higgs-doublet model (2HDM) provides the simplest extensions of the SM scalar sector, which not only aims to solve the hierarchy problem, but is also supported by additional strong motivations [19, 76, 79].

In the context of the 2HDM, a search for a new resonance decaying to a Z boson and a lighter resonance has been performed during the second Ph.D. working period, by exploiting 8 TeV data only, for a total integrated luminosity of about 20 fb^{-1}. The original contribution presented in this thesis covers all the aspects of the analysis, starting from the generation of the signal processes in MADGRAPH, until the statistical interpretation of the results. Both the processes $pp \to H \to ZA$ and $pp \to A \to ZH$ have been used as benchmark signals, considering the decay of the Z boson into either a pair or electrons or a pair of muons, and the H/A resonance decaying into a pair of tau leptons. With no significant excess observed over the expected backgrounds, 95% CL exclusion limits on the product of the cross section and branching ratio have been set at the level of about 5 fb in the majority of the $M_A - M_H$ plane. The results have also been interpreted in the context of a specific Type-II 2HDM formulation, characterized by the parameters $\tan\beta = 1.5$ and $\cos(\beta - \alpha)$. For this benchmark scenario, which is not yet excluded by previous searches [89, 91, 92], exclusion limits on the signal strength μ have been computed. In particular, in the region in which $M_H > M_A + M_Z$, a portion of the phase space is excluded in the ranges $200 \lesssim M_H \lesssim 400$ GeV and $15 \lesssim M_A \lesssim 200$ GeV. In the mirrored region in which $M_A > M_H + M_Z$, the ranges $[100, 250]$ GeV for M_H and $[200, 500]$ GeV for M_A are excluded at the 95% CL. Such exclusion limits have been also derived in the $\tan\beta - \cos(\beta - \alpha)$ parameter space, leading to exclude the following regions: $0.7 \lesssim \tan\beta \lesssim 1.7$ and $-0.3 \lesssim \cos(\beta - \alpha) \lesssim 0.1$. The combination with a similar analysis performed in the $\ell\ell b\bar{b}$ final state has allowed to exclude a wider portion of the parameter space: $0.5 \lesssim \tan\beta \lesssim 2.0$ and $-0.7 \lesssim \cos(\beta - \alpha) \lesssim 0.3$. All these results are going to be included in the final paper of the analysis [23].

Both analyses presented in this thesis are characterized by tau leptons in the final state. Such leptons, the heaviest existing in Nature, decay hadronically in 65% of the cases, leading to collimated jets in the final state, which are characterized by a low particle multiplicity. Distinguishing hadronically decaying taus from quark and gluon jets is one of the most important challenge for analyses dealing with taus, especially at the LHC, where the cross section for QCD-multijet production exceeds by many orders of magnitude the rate at which tau leptons are produced. For this reason, in parallel with the work carried on the searches mentioned above, the entire Ph.D. working period has been devoted to the validation of the algorithm used in CMS to reconstruct and identify hadronically decaying taus and to the study of its performance. The main results obtained, in the context of the expected performance, have been reported in this thesis and constitute an important section of the paper containing the most updated results on tau reconstruction and identification in CMS during LHC Run I, which is going to be published in JINST [25].

After the LHC Run I, the LHC has stopped its operations for about two years, during the so-called Long Shutdown (LS1). During the LS1 the LHC has been upgraded to be able to produce higher center-of-mass energy collisions. The CMS subdetectors have been upgraded too and, in parallel, a tremendous effort has also been put in the improvements of the reconstruction algorithms. In April 2015, the LHC Run II has started and since June the LHC has been able to provide stable p-p collisions at a center-of-mass-energy of 13 TeV. Despite a complicated restart of the CMS magnet due to problems with the cryogenic system in providing liquid Helium, CMS has been able to record 83.51 pb^{-1} of integrated luminosity out of the total 106.08 pb^{-1} delivered by the LHC [100]. The Run II is scheduled to end in 2018, after having reached a center-of-mass-energy of 14 TeV and a total integrated luminosity of 150 fb^{-1}.

The observation of the SM Higgs boson decay into tau pairs, with a statistical significance of 5σ, together with a more accurate SM Higgs characterization in the fermionic channels, is for sure one of the first goals that are expected to be achieved in the Run II. The increase of the partonic luminosity, especially for gluon initiated processes, will increase the production cross section by a factor two (cfr. Figure 6.20). The improved physics objects performance achieved [175] will contribute to enhance the sensitivity to the SM signal. Moreover, the increased center-of-mass energy and luminosity will be profitable for searches for heavy resonances, thanks to the sharp increase of the partonic luminosity especially for pseudoscalar and scalar resonances with a mass grater than 300 GeV (cfr. Figure 6.20). Such motivations support the case of the beyond SM search for the $gg \to H/A \to ZA/H$ signal. Moreover, as already discussed in the thesis, additional technical improve-

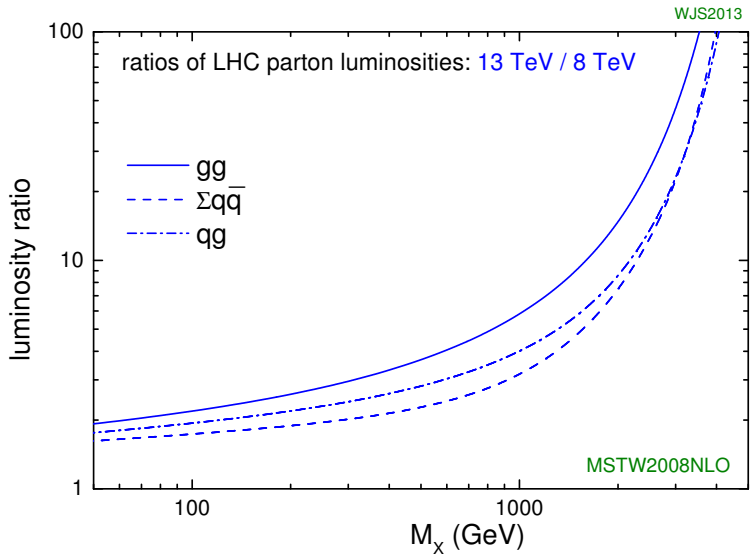

Figure 6.20: Parton luminosity ratios (13 TeV/8 TeV) calculated as a function of a the mass of a singly produced heavy object using MSTW2008 (NLO) PDF [95].

ments are planned to be made: the inclusion of two additional channels ($\ell\ell ee$ and $\ell\ell\mu\mu$) and the usage of the recently developed reconstruction algorithms for boosted tau topologies, which will allow the sensitive region to be further extended where $M_{H/A} \approx 10 \cdot M_{A/H}$.

A

Additional HPS expected performance plots

This appendix contains additional plots related to the HPS algorithm expected performance, which were not fully reported in the main body of the thesis to ease the reading process.

A.1 Tau identification efficiency

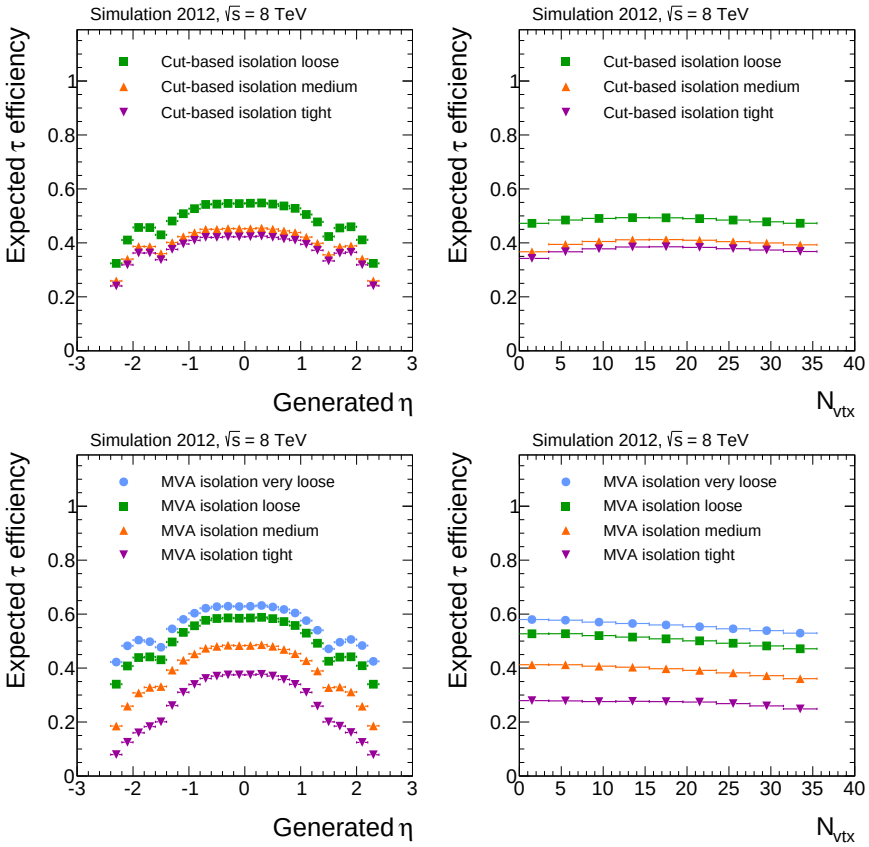

Figure A.1: Efficiencies for hadronic tau decays in simulated $Z/\gamma^* \to \tau\tau$ events to be reconstructed in one of the decay modes $h^\pm$, $h^\pm\pi^0$s and $h^\pm h^\mp h^\pm$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$, and to pass different working points corresponding to the cut-based (top) and MVA-based (bottom) tau isolation discriminant. The efficiencies, computed according to the Equation 4.2, are shown as functions of the generator-level tau η (left) and vertex multiplicity N_{vtx} (right) [A].

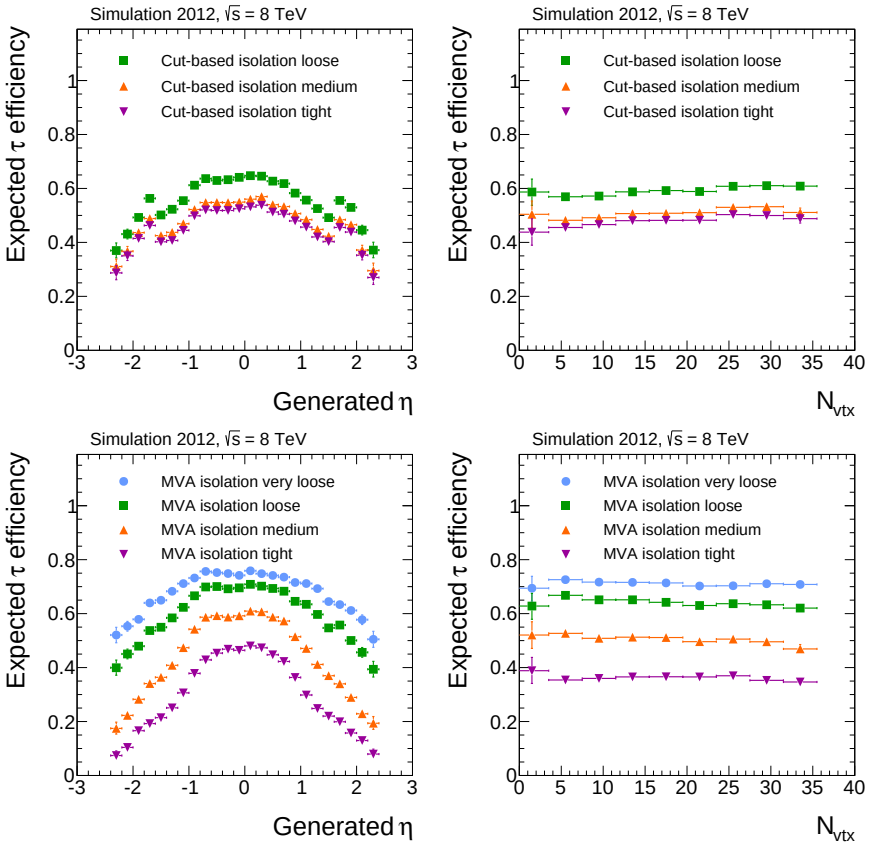

Figure A.2: Efficiencies for hadronic tau decays in simulated $Z' \to \tau\tau$ events to be reconstructed in one of the decay modes $h^{\pm}$, $h^{\pm}\pi^0$s and $h^{\pm}h^{\mp}h^{\pm}$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$, and to pass different working points corresponding to the cut-based (top) and MVA-based (bottom) tau isolation discriminant. The efficiencies, computed according to the Equation 4.2, are shown as functions of the generator-level tau η (left) and vertex multiplicity N_{vtx} (right) [A].

A.2 jet $\rightarrow \tau_h$ fake rates

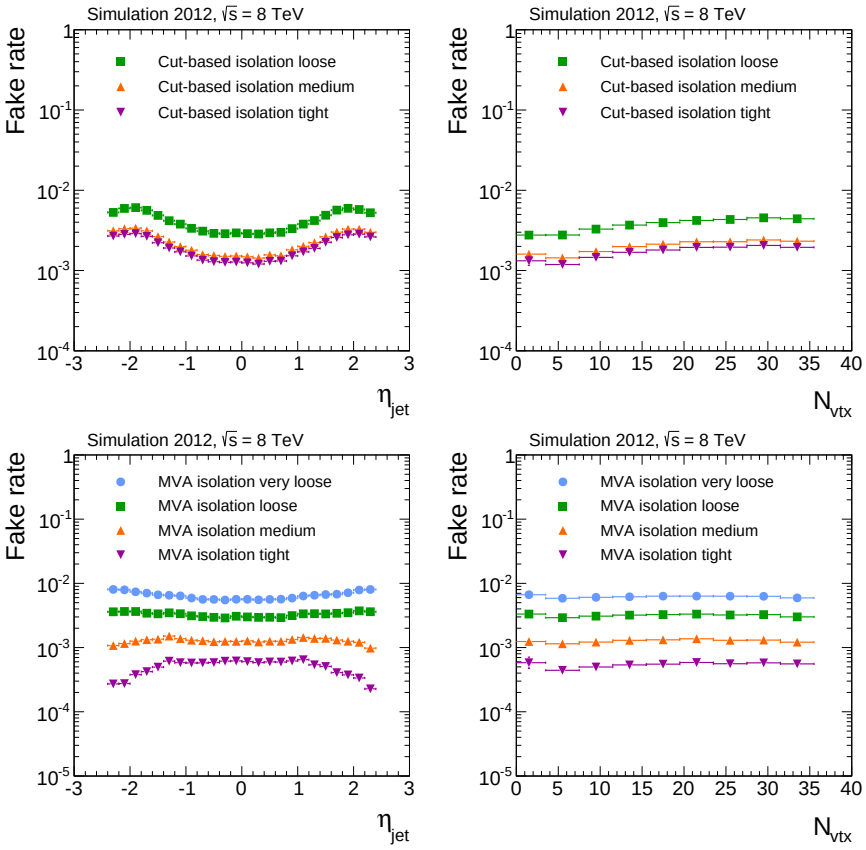

Figure A.3: Probabilities for quark and gluon jets, computed in simulated QCD-multijets events, to be reconstructed in one of the decay modes $h^{\pm}$, $h^{\pm}\pi^0$s and $h^{\pm}h^{\mp}h^{\pm}$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$, and to pass different working points corresponding to the cut-based (top) and MVA-based (bottom) tau isolation discriminant. The misidentification rates, computed according to the Equation 4.4, are shown as functions of the reconstructed jet η (left) and vertex multiplicity N_{vtx} (right) [A].

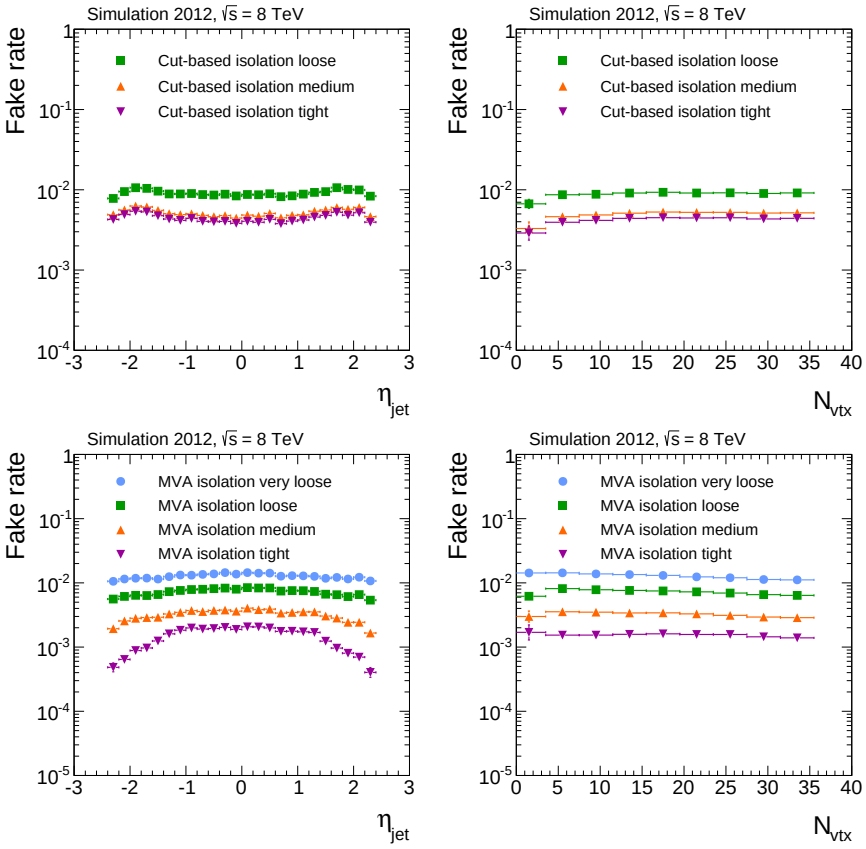

Figure A.4: Probabilities for quark and gluon jets, computed in simulated W+jets events, to be reconstructed in one of the decay modes $h^{\pm}$, $h^{\pm}\pi^0$s and $h^{\pm}h^{\mp}h^{\pm}$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$, and to pass different working points corresponding to the cut-based (top) and MVA-based (bottom) tau isolation discriminant. The misidentification rates, computed according to the Equation 4.4, are shown as functions of the reconstructed jet η (left) and vertex multiplicity N_{vtx} (right) [A].

B

Statistical methods

In the context of both analyses presented in Chapters 5 and 6, the binned SVfit $M_{\tau\tau}$ distributions of each channel have been used for the statistical interpretation of the results. Moreover, in the search for Higgs bosons decaying to tau pairs in the 2HDM, presented in Chapter 6, a cut-and-count analysis has been also performed, considering, as observable, the number of signal and background events counted in one single bin. As a general approach, one can indicate with n_i, where the index i ranges on the number of bins considered in the analysis ($i > 1$ and $i = 1$ for a shape-based and a cut-and-count analysis, respectively), the total number of data yields measured. The expected event yields ν_i, in a given bin i, is parametrized as:

$$\nu_i = \mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta}) \,, \tag{B.1}$$

where s_i and b_i are the signal and background yields expected in that bin, respectively. Signal and background expectations are assumed to be dependent on a set of *nuisance parameters* $\vec{\theta}$ and, in this case, the signal expected yield is fixed to its theoretical prediction. The multiplicative parameter μ, called *signal strenght modifier*, quantifies the strenght of the observed signal with respect to the value expected from the theory. In the statistical interpretation of data observation, every source of systematic uncertainty has its own associated nuisance parameter. The nuisance parameters $\vec{\theta}$ are *a priori* known from auxiliary

measurements constraining the parameters themselves. In case of systematic uncertainties related to the overall normalization, the prior *pdf* $\rho(\theta_j|\hat{\theta}_j)$ chosen for the unknown θ_j, given the observation $\hat{\theta}_j$, is a *Log-normal* distribution:

$$\rho(\theta_j, \hat{\theta}_j) = \frac{1}{\theta_j \sqrt{2\pi} \ln(\kappa)} e^{-\left(\frac{\ln(\theta_j/\hat{\theta}_j)}{\sqrt{2}\ln\kappa}\right)^2}, \tag{B.2}$$

where κ is related to the multiplicative "relative uncertainty" $\kappa - 1 \equiv \sigma_{rel} \equiv \delta\theta_j/\theta_j$. Some uncertainties, e.g. those affecting the energy scale, have the property of scaling the kinematical variable correlated with the mass, thus altering its shape. These uncertainties, called "shape uncertainties" are modeled using three distributions governed by a nuisance parameter f: the nominal distribution ($f = 0$) and those obtained by scaling the nominal template up ($f = +1$) and down ($f = -1$) by one standard deviation of the nuisance f. A family of distributions is obtained starting from these three templates using the "vertical morphing" technique [176]. The morphing function interpolates quadratically the contents of each bin for $|f| < 1$ and extrapolate the result of the quadratic interpolation linearly for $|f| > 1$. A gaussian distribution with mean 0 and $\sigma = 1$ is assigned, then, to the single morphing parameter.

Once defined, the *pdfs* assigned to each systematic uncertainty are incorporated into the likelihood model $\mathcal{L}$, written as:

$$\begin{aligned}\mathcal{L}(obs|\mu, \vec{\theta}) &= \text{Poisson}\left(obs|\mu \cdot s(\vec{\theta}) + b(\vec{\theta})\right) \cdot p(\hat{\vec{\theta}}|\hat{\theta}) \\ &= \prod_{i=1}^{N} \frac{[\mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})]^{n_i}}{n_i!} e^{-[\mu \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})]} \cdot p(\hat{\vec{\theta}}|\hat{\theta}) \,.\end{aligned} \tag{B.3}$$

Equation B.3 represents the product of the Poissonian distributions corresponding to the observation of n_i events for all the final states, in each of the bins of the mass distribution (or in one single bin, in the case of a cut-and-count analysis), given the expectation for the signal s_i and the background b_i. Additionally, $p(\hat{\vec{\theta}}|\vec{\theta})$ are the reinterpretation of the *pdfs* $\rho(\vec{\theta}|\hat{\vec{\theta}})$ as posteriors (Bayesian theorem) assuming flat priors $\vec{\theta}$. Starting from $\mathcal{L}$, it is now possible to perform a test statistics which is used to assess how compatible the observed mass distributions are with either the background-only expectation (*null* hypothesis H_0) and the signal+background expectation (*alternative* hypothesis H_1). The test statistic that is going to be explained, and the treatment of the

nuisance parameters just described, follow the recommendations of the LHC Higgs Combination group [172].

In the computation of limits and significance for Higgs boson searches at LHC, it has been decided to follow preferentially the frequentist paradigm, rather than the bayesian one [1], by means of a test statistic. Practically, a test statistic is a function which summarizes the information on the observed data, expected signal and background, and all the associated uncertainties. Quantitatively, the test statistic q_μ is defined through the profile likelihood $\lambda(\mu)$ depending on the parameter of interest μ:

$$\lambda(\mu) = -2 \ln \frac{\mathcal{L}(\text{obs}|\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\text{obs}|\hat{\mu}, \hat{\vec{\theta}})} \,. \tag{B.4}$$

Is in the process of fitting the expected background and signal distributions to data, to infer the value of the parameter of interest μ, that each single term in Equation B.4 can be more properly framed: $\hat{\hat{\vec{\theta}}}_\mu$ is the value that, in the fit, has maximized $\mathcal{L}$ for a given value of μ, while $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the maximum likelihood estimators. Equation B.4 is constrained by the boundary condition $0 \leqslant \hat{\mu} \leqslant \mu$ that ensures that the expected signal yield is positive and that one-sided confidence intervals are obtained when setting limits on μ ($\mu \geqslant \hat{\mu}$). These conditions are imposed by forcing $\hat{\mu} = 0$ if negative best fit values are found and by setting the value of the test statistic to 0 when $\mu < \hat{\mu}$.

$$q(\mu) = \begin{cases} -2 \ln \frac{\mathcal{L}(\text{obs}|\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\text{obs}|0, \hat{\hat{\vec{\theta}}}(0))} & , \hat{\mu} < 0 \\ -2 \ln \frac{\mathcal{L}(\text{obs}|\mu, \hat{\hat{\vec{\theta}}}_\mu)}{\mathcal{L}(\text{obs}|\hat{\mu}, \hat{\vec{\theta}})} & , 0 \leqslant \hat{\mu} \leqslant \mu \\ 0 & , \mu < \hat{\mu} \end{cases} \tag{B.5}$$

By construction, the test statistic q_μ acquires small values if the observation is consistent with the signal+background hypothesis H_1, large values in case of the null hypothesis H_0 (background only). To quantify the agreement between what observed in data and a given hypothesis, the *p*-value is computed.

[1] In the frequentist paradigm, the probability is seen as the asymptotic value for the frequency of the outcomes of a large number of identical experiments, and therefore frequentist inference is used to make statements about the probability of an experimental outcome for a given model. In the Bayesian paradigm, the probability is interpreted as a subjective degree of belief in the validity of a theory. In this context, statements are made about the probability of a model to be true given the observed experimental outcome and the subjective prior assumption about that probability before the experiment was performed

Generally, the *p*-value is the probability of finding, under a given hypothesis H_i, an incompatibility between what is observed and what the hypothesis H_i predicts. The measurement of this incompatibility can be based on the test statistic q_μ and, in this case, the *p*-value p_0 quantifies the probability that the value q_0 is at least as large as the one observed in data, under the background only hypothesis. Equivalently, the probability to obtain a value q_μ at least as large as the one observed in data, under the signal+background hypothesis is the *p*-value p_μ. The computation of these *p*-values, by means of a test statistic, becomes quite straightforward in the limit of large event statistics, the so-called *asymptotic limit* [177]. Asymptotically, the test statistic q_μ becomes a well-defined analytical function $f(q_\mu|\mu s + b)$ [177], which can be numerically integrated to obtain the *p*-values given an observed value of the test statistic q_μ^{obs}:

$$p_\mu = \int_{q_\mu^{obs}}^{\infty} f(q_\mu|\mu s + b) dq_\mu, \qquad p_0 = \int_{q_\mu^{obs}}^{\infty} f(q_\mu|b) dq_\mu \,. \tag{B.6}$$

If an excess is found with respect to the background-only hypothesis, it is quantified in terms of the equivalent significance Z, defined such that the upper-tail probability of a gaussian distribution, computed from Z standard deviations above its mean up to infinity, is equal to p_0. The significance of $Z = 5$ required for the Higgs boson discovery correspond to $p_0 = 2.87 \times 10^{-7}$.

The signal strenght modifier μ is then said to be excluded at a confidence level (CL_s) $1 - \alpha$ if the ratio:

$$CL_s(\mu) = \frac{CL_{s+b}}{CL_b} = \frac{p_\mu}{p_0} \equiv \alpha \,. \tag{B.7}$$

The observed 95% CL exclusion limit on μ is then obtained by varying μ until $\alpha = 0.05$. Historically, the sensitivity of an experiment is quantified in terms of the median expected exclusion limit on μ in the background-only hypothesis H_0, together with the intervals in which this median is expected to lie in the 68% (1σ) and 95% (2σ) of the cases. In order to perform the estimation, the ensemble of simulated datasets is replaced by a single representative one, known in literature as the *Asimov* dataset [177].

C

Ditau mass spectra for the SM $ZH \to \ell\ell\tau\tau$ analysis

In this appendix, the ditau mass distributions before the maximum likelihood fit process (*prefit*) are directly compared with the same distribution after having performed the fit (*postfit*). Ditau mass distributions for all the eight channels considered in the SM $ZH \to \ell\ell\tau\tau$ and for $\sqrt{s} = 7$ and 8 TeV analyses separately. The postfit mass spectra can be directly compared with the yields reported in Table 5.6. In the postfit distributions, the normalization of the predicted background contributions corresponds to the result of the global fit. The signal distributions, instead, are normalized to the SM prediction.

C.1 eeeμ channel

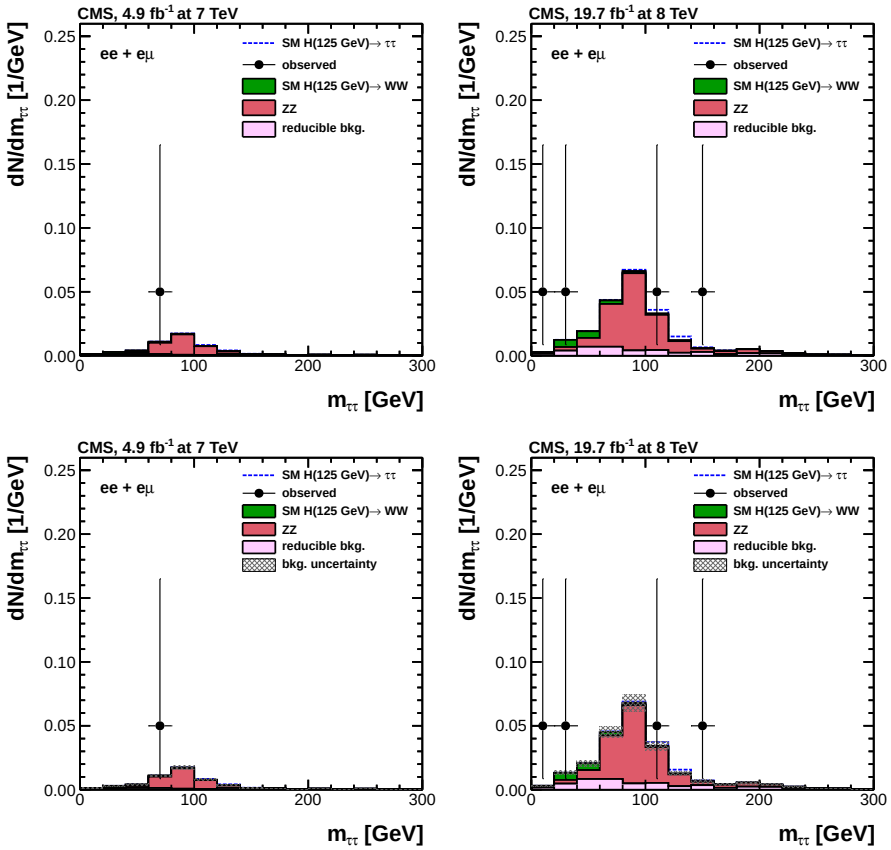

Figure C.1: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the eeeμ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.2 eeeτ_h channel

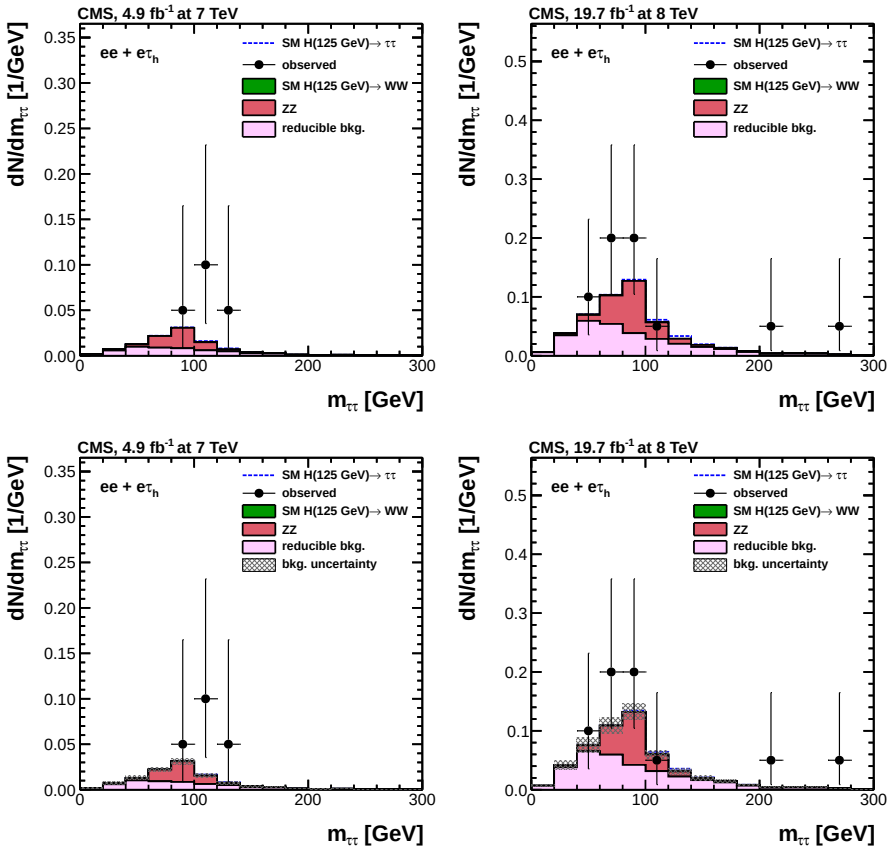

Figure C.2: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the eeeτ_h channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.3 $ee\mu\tau_h$ channel

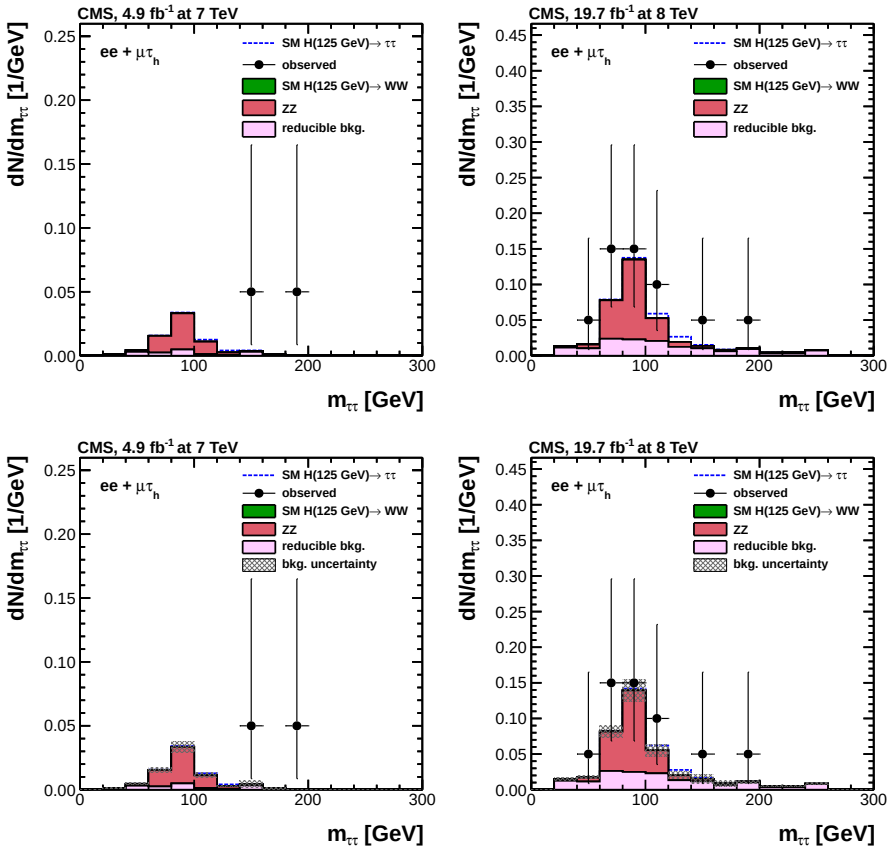

Figure C.3: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the $ee\mu\tau_h$ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.4 $ee\tau_h\tau_h$ channel

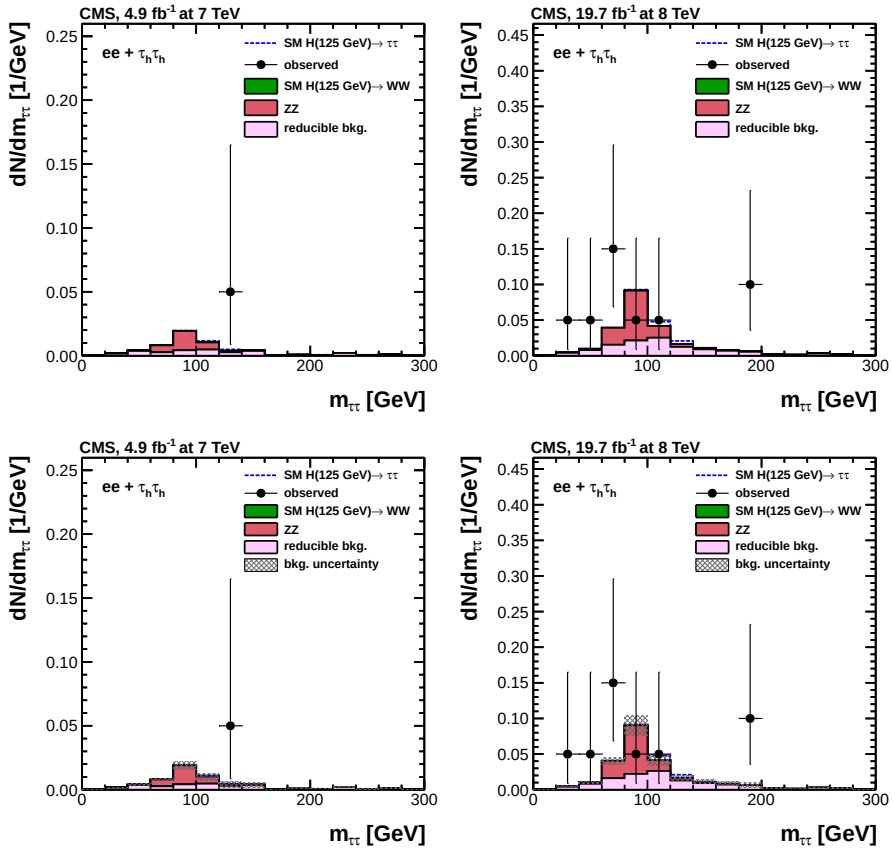

Figure C.4: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the $ee\tau_h\tau_h$ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.5 μμeμ channel

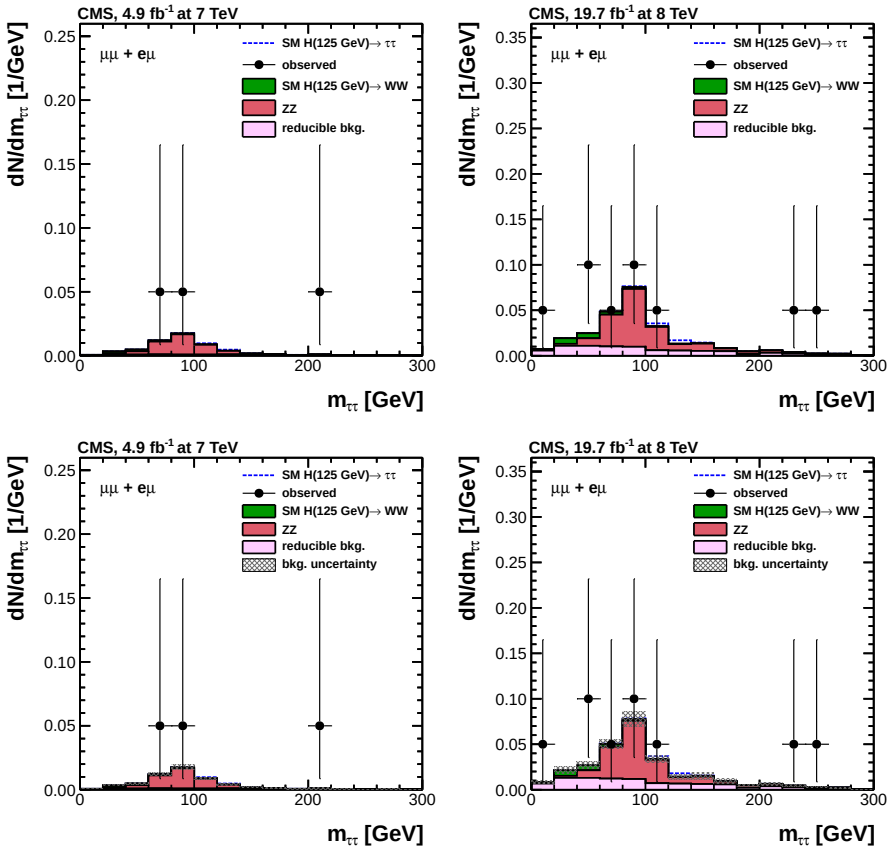

Figure C.5: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the μμeμ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.6 $\mu\mu e\tau_h$ channel

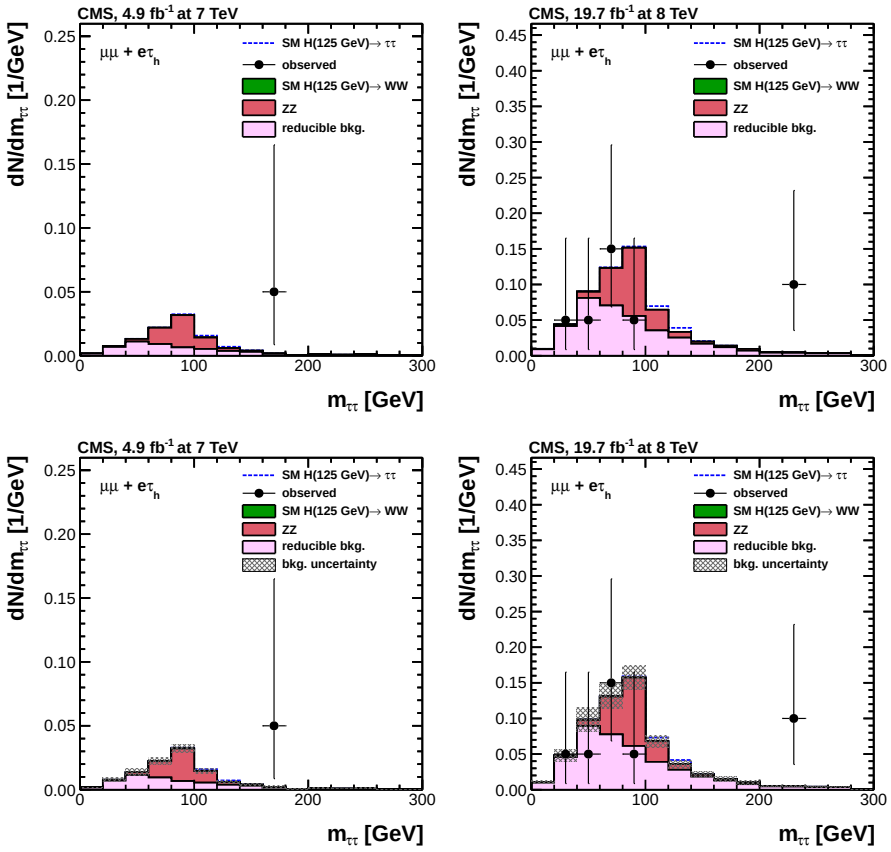

Figure C.6: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the $\mu\mu e\tau_h$ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.7 $\mu\mu\mu\tau_h$ channel

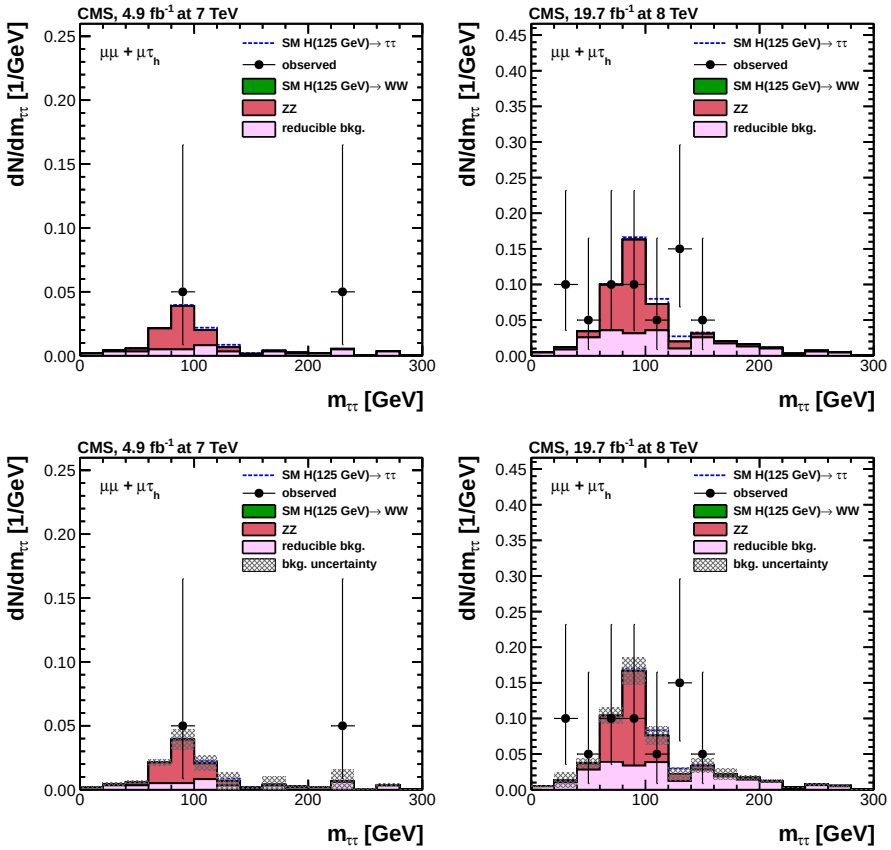

Figure C.7: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the $\mu\mu\mu\tau_h$ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

C.8 $\mu\mu\tau_h\tau_h$ channel

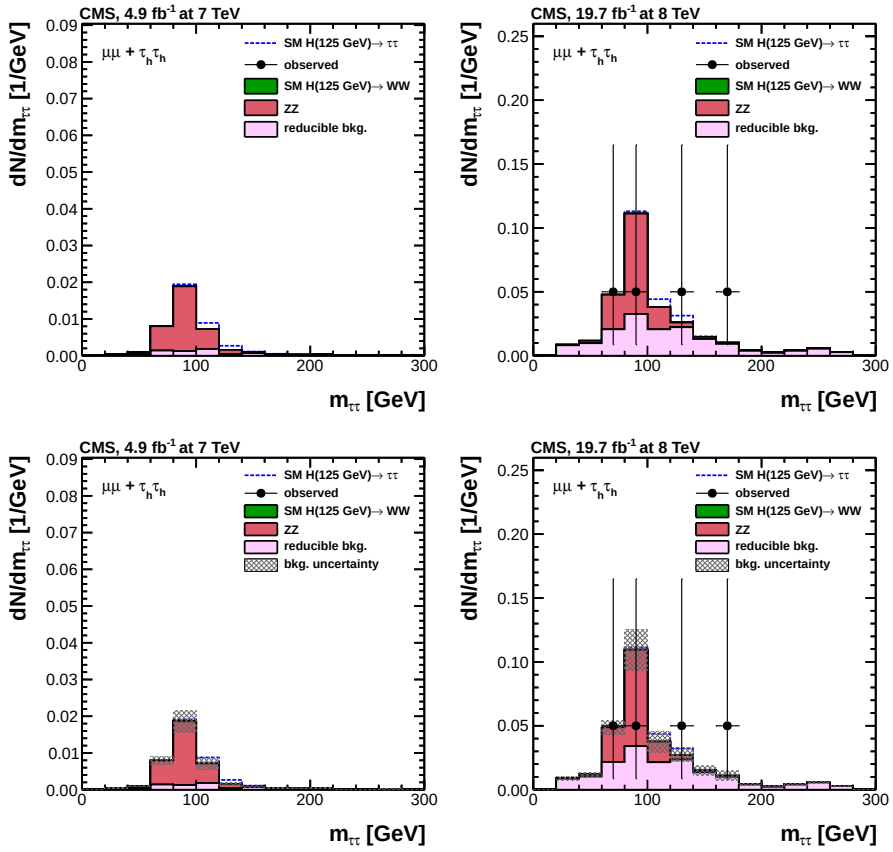

Figure C.8: SVfit ditau mass prefit (top) and postfit (bottom) [161] distributions for the $\mu\mu\tau_h\tau_h$ channel for $\sqrt{s} = 7$ (left) and 8 TeV (right) [AC].

D

Trigger and lepton scale factors

This appendix contains data-to-simulation scale factors used to correct simulated processes in the context of the SM $Zh \to \ell\ell\tau\tau$ (cfr. Chapter 5) and the 2HDM $H/A \to ZA/H \to \ell\ell\tau\tau$ (cfr. Chapter 6) analyses. Concerning the SM analysis, trigger-efficiencies scale factors are reported in Tables D.1 and D.2 for 2011 and 2012 data, respectively. The scale factors related to lepton identification and isolation efficiencies are, instead, reported in Tables D.3 and D.4 for muons and electrons, respectively. Also in this case, results for 2011 and 2012 data taking are provided.

Data-to-simulation scale factors used in the context of the 2HDM analysis are listed in Tables D.5 and D.6. These tables report the 2012 scale factors related to electron and muon identification and isolation efficiencies, respectively. Trigger-efficiency scale factors have not been used in the context of the 2HDM analysis. As already explained in Section 6.1, data-to-simulation differences related to trigger efficiencies have been taken directly into account as systematic uncertainties. All the errors quoted in the tables are statistical only.

2011 Data/MC SF for Trigger Efficiencies				
DoubleMuon				
$\lvert\eta\rvert/p_T$	$10-15$ GeV	$15-20$ GeV	$20-30$ GeV	$\geqslant 30$ GeV
$0-1.5$	1.01 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.992 ± 0.006
$\geqslant 1.5$	1.03 ± 0.05	1.07 ± 0.05	1.04 ± 0.03	1.06 ± 0.02
DoubleElectron				
$\lvert\eta\rvert/p_T$	$10-15$ GeV	$15-20$ GeV	$20-30$ GeV	$\geqslant 30$ GeV
$0-1.479$	0.98 ± 0.01	1.00 ± 0.01	1.001 ± 0.004	1.003 ± 0.002
$\geqslant 1.479$	0.97 ± 0.05	1.05 ± 0.02	1.00 ± 0.01	1.008 ± 0.008

Table D.1: Data-to-simulation correction factors for 2011 DoubleMuon and DoubleElectron trigger (cfr. Table 5.3) efficiencies.

2012 Data/MC SF for Trigger Efficiencies					
DoubleMuon					
$\lvert\eta\rvert/p_T$	$10-15$ GeV	$15-20$ GeV	$20-25$ GeV	$25-30$ GeV	$\geqslant 30$ GeV
$0-0.8$	0.9818 ± 0.0058	0.9781 ± 0.0056	0.9873 ± 0.0060	0.9755 ± 0.0074	0.9956 ± 0.0087
$0.8-1.2$	0.9713 ± 0.0124	0.9782 ± 0.0171	0.9532 ± 0.0116	0.9818 ± 0.0187	0.9644 ± 0.0165
$1.2-2.1$	0.9675 ± 0.0176	0.9587 ± 0.0179	0.9605 ± 0.0159	0.9732 ± 0.0251	0.9530 ± 0.0209
DoubleElectron					
$\lvert\eta\rvert/p_T$	$10-15$ GeV	$15-20$ GeV	$20-25$ GeV	$25-30$ GeV	$\geqslant 30$ GeV
$0-0.8$	0.9639 ± 0.0197	0.9762 ± 0.0115	0.9683 ± 0.0087	0.9756 ± 0.0084	1.0035 ± 0.0078
$0.8-1.479$	0.8898 ± 0.0205	0.9647 ± 0.0113	0.9666 ± 0.0083	0.9896 ± 0.0083	0.9977 ± 0.0073
$1.479-2.3$	0.9228 ± 0.0470	0.9199 ± 0.0212	0.9679 ± 0.0149	0.9473 ± 0.0162	0.9885 ± 0.0135

Table D.2: Data-to-simulation correction factors for 2012 DoubleMuon and DoubleElectron trigger (cfr. Table 5.3) efficiencies.

2011 Muon ID/Isolation SF			
Loose			
$\|\eta\|/p_T$	$10-15$ GeV	$15-20$ GeV	$20-100$ GeV
$0-1.5$	0.9895 ± 0.0141	1.0168 ± 0.0065	1.0061 ± 0.0004
$1.5-2.1$	1.0303 ± 0.0177	1.0247 ± 0.0100	1.0144 ± 0.0010
Tight			
$\|\eta\|/p_T$	$17-20$ GeV	$20-30$ GeV	$\geqslant 30$ GeV
$0-1.6$	0.9272 ± 0.0029	0.9721 ± 0.0023	1.0403 ± 0.0039
$1.6-2.1$	0.9160 ± 0.0025	0.9702 ± 0.0010	0.9870 ± 0.0023
2012 Muon ID/Isolation SF			
Loose			
$\|\eta\|/p_T$	$10-20$ GeV	$20-30$ GeV	$\geqslant 30$ GeV
$0-0.8$	0.9678 ± 0.0070	0.9792 ± 0.0008	0.9847 ± 0.0002
$0.8-1.2$	0.9583 ± 0.0036	0.9818 ± 0.0012	0.9798 ± 0.0001
$\geqslant 1.2$	0.9912 ± 0.0022	0.9898 ± 0.0008	0.9889 ± 0.0001
Tight			
$\|\eta\|/p_T$	$20-30$ GeV	$\geqslant 30$ GeV	
$0-0.8$	0.9542 ± 0.0015	0.9731 ± 0.0003	
$0.8-1.2$	0.9629 ± 0.0021	0.9730 ± 0.0004	
$\geqslant 1.2$	0.9871 ± 0.0015	0.9912 ± 0.0005	

Table D.3: Data-to-simulation correction factors for 2011 and 2012 loose and tight muon identification and isolation efficiencies.

2011 Electron ID/Isolation SF			
Loose			
$\|\eta\|/p_T$	$10-15$ GeV	$15-20$ GeV	$20-100$ GeV
$0-1.479$	1.039 ± 0.0056	0.9622 ± 0.0670	0.9849 ± 0.0021
$1.479-2.1$	0.9758 ± 0.1644	1.1483 ± 0.0681	1.0117 ± 0.0020
Tight			
$\|\eta\|/p_T$	$20-30$ GeV	$\geqslant 30$ GeV	
$0-1.479$	0.9359 ± 0.0079	1.0273 ± 0.0028	
$1.479-2.1$	0.9070 ± 0.0001	0.9663 ± 0.0013	
2012 Electron ID/Isolation SF			
Loose			
$\|\eta\|/p_T$	$10-20$ GeV	$20-30$ GeV	$\geqslant 30$ GeV
$0-0.1479$	0.9503 ± 0.0060	0.9429 ± 0.0025	0.9629 ± 0.0003
$1.479-2.1$	0.9278 ± 0.0149	0.9515 ± 0.0022	0.9697 ± 0.0006
Tight			
$\|\eta\|/p_T$	$24-30$ GeV	$\geqslant 30$ GeV	
$0-0.1479$	0.8616 ± 0.0024	0.9322 ± 0.0004	
$1.479-2.1$	0.7903 ± 0.0060	0.9212 ± 0.0002	

Table D.4: Data-to-simulation correction factors for 2011 and 2012 loose and tight muon identification and isolation efficiencies.

2012 Electron ID/Isolation SF				
Loose				
$\|\eta\|/p_T$	< 30 GeV	$30 - 40$ GeV	$40 - 50$ GeV	> 50 GeV
$0 - 0.8$	1.005 ± 0.004	1.004 ± 0.003	1.008 ± 0.002	1.008 ± 0.002
$0.8 - 1.442$	0.981 ± 0.006	0.991 ± 0.001	0.994 ± 0.002	0.999 ± 0.003
$1.442 - 1.556$	1.044 ± 0.012	0.998 ± 0.005	0.989 ± 0.005	0.994 ± 0.005
$1.556 - 2.0$	0.980 ± 0.007	0.992 ± 0.003	1.004 ± 0.003	1.006 ± 0.004
> 2.0	1.017 ± 0.008	1.019 ± 0.004	1.005 ± 0.010	1.009 ± 0.003
Tight				
$\|\eta\|/p_T$	< 30 GeV	$30 - 40$ GeV	$40 - 50$ GeV	> 50 GeV
$0 - 0.8$	0.973 ± 0.004	0.979 ± 0.003	0.984 ± 0.002	0.983 ± 0.001
$0.8 - 1.442$	0.948 ± 0.007	0.961 ± 0.006	0.972 ± 0.002	0.977 ± 0.005
$1.442 - 1.556$	0.983 ± 0.020	0.983 ± 0.010	0.957 ± 0.004	0.978 ± 0.008
$1.556 - 2.0$	0.957 ± 0.010	0.962 ± 0.004	0.985 ± 0.002	0.986 ± 0.005
> 2.0	1.001 ± 0.009	1.002 ± 0.005	0.999 ± 0.004	0.995 ± 0.004

Table D.5: Data-to-simulation correction factors for 2012 loose and tight electron identification and isolation efficiencies.

2012 Muon ID/Isolation SF				
Loose				
$p_T/\|\eta\|$	$0 - 0.9$	$0.9 - 1.2$	$1.2 - 2.1$	> 2.1
< 20 GeV	0.9927 ± 0.0058	0.9957 ± 0.0067	1.0034 ± 0.0006	0.9989 ± 0.01
$20 - 25$ GeV	0.9922 ± 0.0015	0.9976 ± 0.0022	1.0027 ± 0.0011	
$25 - 30$ GeV	0.9961 ± 0.0005	1.0014 ± 0.0009	0.9991 ± 0.0005	
$30 - 35$ GeV	0.9984 ± 0.0003	0.9986 ± 0.0006	0.9985 ± 0.0004	
$35 - 40$ GeV	0.9985 ± 0.0002	0.9995 ± 0.0003	0.9985 ± 0.0003	
$40 - 50$ GeV	0.9987 ± 0.0001	0.9986 ± 0.0002	0.9985 ± 0.0001	
$50 - 60$ GeV	0.9970 ± 0.0003	0.9979 ± 0.0006	0.9967 ± 0.0005	
$60 - 90$ GeV	0.9937 ± 0.0007	0.9923 ± 0.0013	0.9904 ± 0.0011	
$90 - 140$ GeV	1.0015 ± 0.0014	1.0017 ± 0.0021	1.0020 ± 0.0018	
> 140 GeV	1.0134 ± 0.0122	0.9644 ± 0.0329	1.0007 ± 0.0142	
Tight				
$p_T/\|\eta\|$	$0 - 0.9$	$0.9 - 1.2$	$1.2 - 2.1$	> 2.1
< 20 GeV	0.9471 ± 0.0027	0.9518 ± 0.0032	0.9800 ± 0.0016	1.0031 ± 0.01
$20 - 25$ GeV	0.9749 ± 0.0015	0.9884 ± 0.0023	0.9973 ± 0.0012	
$25 - 30$ GeV	0.9971 ± 0.0009	1.0008 ± 0.0016	1.0078 ± 0.0009	
$30 - 35$ GeV	0.9939 ± 0.0006	0.9985 ± 0.0012	1.0068 ± 0.0007	
$35 - 40$ GeV	0.9934 ± 0.0004	0.9991 ± 0.0007	1.0037 ± 0.0005	
$40 - 50$ GeV	0.9941 ± 0.0002	0.9982 ± 0.0003	1.0020 ± 0.0002	
$50 - 60$ GeV	0.9955 ± 0.0004	0.9987 ± 0.0006	1.0012 ± 0.0004	
$60 - 90$ GeV	0.9990 ± 0.0005	0.9991 ± 0.0008	1.0006 ± 0.0005	
$90 - 140$ GeV	1.001 ± 0.0008	0.9995 ± 0.0014	0.9999 ± 0.0096	
> 140 GeV	1.000 ± 0.0018	0.9968 ± 0.0030	0.9999 ± 0.0023	

Table D.6: Data-to-simulation correction factors for 2012 loose and tight muon identification and isolation efficiencies.

E

Analytical validation of the fake rate formula

```
#include <iostream>
using namespace std;

int frxchecks(){
int NZevents=10000;
int NWZ_ej_events=500;
int NWZ_mj_events=300;
int NZZevents=100;

double fE = 0.1,  fM = 0.2;
double pE = 0.8, pM = 0.94;
double epsE = fE/(1-fE);
double epsM = fM/(1-fM);
double etaE = (1-pE)/pE;
double etaM = (1-pM)/pM;

////////////////////////////////////// CR00
double Nz00  = NZevents*(1-fE)*(1-fM);
double Nwz_mj00 = NWZ_mj_events*(1-fE)*(1-pM);
double Nwz_ej00 = NWZ_ej_events*(1-pE)*(1-fM);
double Nzz00 = NZZevents*(1-pE)*(1-pM);
double N00tot = Nz00+Nwz_mj00+Nwz_ej00+Nzz00;

////////////////////////////////////// CR01
```

```
double Nz01  = NZevents*(1-fE)*fM;
double Nwz_mj01 = NWZ_mj_events*(1-fE)*pM;
double Nwz_ej01 = NWZ_ej_events*(1-pE)*fM;
double Nzz01 = NZZevents*(1-pE)*pM;
double N01tot = Nz01+Nwz_mj01+Nwz_ej01+Nzz01;

////////////////////////////////////// CR10
double Nz10  = NZevents*fE*(1-fM);
double Nwz_mj10 = NWZ_mj_events*fE*(1-pM);
double Nwz_ej10 = NWZ_ej_events*pE*(1-fM);
double Nzz10 = NZZevents*pE*(1-pM);
double N10tot = Nz10+Nwz_mj10+Nwz_ej10+Nzz10;

////////////////////////////////////// CR11
double Nz11  = NZevents*fE*fM;
double Nwz_mj11 = NWZ_mj_events*fE*pM;
double Nwz_ej11 = NWZ_ej_events*pE*fM;
double Nzz11 = NZZevents*pE*pM;
double N11tot = Nz11+Nwz_mj11+Nwz_ej11+Nzz11;

////////////////////////////////////// No subtraction, prompt rate != 1
        ////////////////////////////////////// (CMS AN-2010/261)
double k = 1 / ( (1-(epsE*etaE)) * (1-(epsM*etaM)) );
double Npf = epsM*k*( -N00tot*epsE + N10tot + epsE*etaM*N01tot - etaM*N11tot );
double Nfp = epsE*k*( -epsM*N00tot + etaE*epsM*N10tot + N01tot - etaE*N11tot );
double Nff = epsE*epsM*k*( N00tot - etaE*N10tot - etaM*N01tot + etaM*etaE*N11tot );
double Ntot_AN2010 = Npf+Nfp+Nff;

////////////////////////////////////// Subtraction, prompt-rate == 1
        ////////////////////////////////////// (CMS AN-2014/265, PAS 15-001)
double CR01 = epsE*(N01tot-Nzz01);
double CR10 = epsM*(N10tot-Nzz10);
double CR00 = epsM*epsE*(N00tot-Nzz00);
double Ntot_AN2014 = CR01+CR10-CR00;

return 0;
}
```

Bibliography

[1] J. Thomson, "Cathode rays", *Philosophical Magazine* **44** (1987) 293–316.

[2] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider", *Phys. Lett.* **B126** (1983) 398–410. doi:10.1016/0370-2693(83)90188-0.

[3] UA2 Collaboration, "Evidence for $Z^0 \to e^+e^-$ at the CERN anti-p p Collider", *Phys. Lett.* **B129** (1983) 130–140. doi:10.1016/0370-2693(83)90744-X.

[4] UA1 Collaboration, "Experimental Observation of Events with Large Missing Transverse Energy Accompanied by a Jet Or a Photon(s) in p anti-p Collisions at $\sqrt{s} = 540$ GeV", *Phys. Lett.* **B139** (1984) 115. doi:10.1016/0370-2693(84)90046-7.

[5] 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.

[6] SLAC-SP-017 Collaboration, "Discovery of a Narrow Resonance in e^+e^- Annihilation", *Phys. Rev. Lett.* **33** (1974) 1406–1408. [Adv. Exp. Phys.5,141(1976)]. doi:10.1103/PhysRevLett.33.1406.

[7] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, arXiv:hep-ex/9503002. doi:10.1103/PhysRevLett.74.2626.

[8] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, arXiv:hep-ex/9503003. doi:10.1103/PhysRevLett.74.2632.

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

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

[11] 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, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

[12] 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, arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021.

[13] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323. doi:10.1103/PhysRevLett.13.321.

[14] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509. doi:10.1103/PhysRevLett.13.508.

[15] CMS Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *JHEP* **05** (2014) 104, arXiv:1401.5041. doi:10.1007/JHEP05(2014)104.

[16] LHCb Collaboration, "First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$", *Phys. Rev. Lett.* **110** (2013), no. 2, 021801, arXiv:1211.2674. doi:10.1103/PhysRevLett.110.021801.

[17] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", *Astron. Astrophys.* **571** (2014) A16, arXiv:1303.5076. doi:10.1051/0004-6361/201321591.

[18] A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi et al., "Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster", *Phys. Rev. Lett.* **113** (2014) 251301, arXiv:1402.4119. doi:10.1103/PhysRevLett.113.251301.

[19] M. Trodden, "Electroweak baryogenesis: A Brief review", arXiv:hep-ph/9805252.

[20] R. P. Feynman, "Feynman lectures on gravitation", *Reading, USA: Addison-Wesley (1995) 232 p. (The advanced book program)* (1996).

[21] E. Gildener and S. Weinberg, "Symmetry Breaking and Scalar Bosons", *Phys. Rev.* **D13** (1976) 3333. doi:10.1103/PhysRevD.13.3333.

[22] CMS Collaboration, "Search for H/A decaying into Z+A/H, with Z to $\ell\ell$ and A/H to fermion pair", *CMS-PAS-HIG-15-001* (2015).

[23] CMS Collaboration, "Search for resonances in final states with a Z boson and a pair of b jets or a pair of tau leptons", *To be submitted to Physics Letter B*.

[24] CMS Collaboration, "Performance of tau-lepton reconstruction and identification in CMS", *JINST* **7** (2012) P01001, arXiv:1109.6034. doi:10.1088/1748-0221/7/01/P01001.

[25] CMS Collaboration, "Tau reconstruction and identification in CMS during LHC run 1", *To be submitted to Journal of Instrumentation*.

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

[27] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266. doi:10.1103/PhysRevLett.19.1264.

[28] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367–377.

[29] Particle Data Group Collaboration, "Review of Particle Physics (RPP)", *Chin. Rev.* **C38** (2014) 090001. doi:10.1088/1674-1137/38/9/09000.

[30] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215. doi:10.1016/S0031-9163(64)92001-3.

[31] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking", *CERN-TH-401* (1964).

[32] R. L. Jaffe, "Multi-Quark Hadrons. 1. The Phenomenology of $Q^2\bar{Q}^2$ Mesons", *Phys. Rev.* **D15** (1977) 267. doi:10.1103/PhysRevD.15.267.

[33] D. Strottman, "Multi - Quark Baryons and the MIT Bag Model", *Phys. Rev.* **D20** (1979) 748–767. doi:10.1103/PhysRevD.20.748.

[34] H. Hogaasen and P. Sorba, "The Systematics of Possibly Narrow Quark States with Baryon Number One", *Nucl. Phys.* **B145** (1978) 119. doi:10.1016/0550-3213(78)90417-0.

[35] H. J. Lipkin, "New Possibilities for Exotic Hadrons: Anticharmed Strange Baryons", *Phys. Lett.* **B195** (1987) 484. doi:10.1016/0370-2693(87)90055-4.

[36] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, arXiv:1507.03414. doi:10.1103/PhysRevLett.115.072001.

[37] H. Weyl, "Gruppentheorie und Quantenmechanik", *Zeit. für Physik* (1929).

[38] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346. doi:10.1103/PhysRevLett.30.1343.

[39] B. Andersson, "The Lund fragmentation of a multigluon string state", *Acta Phys. Polon.* **B32** (2001) 3993–4011.

[40] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys.Rev.Lett.* **81** (1998) 1562–1567, arXiv:hep-ex/9807003. doi:10.1103/PhysRevLett.81.1562.

[41] 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, arXiv:hep-ex/0104049. doi:10.1103/PhysRevD.64.112007.

[42] C. S. Wu, E. Ambler, R. W. Hayward et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.* **105** (1957) 1413–1414. doi:10.1103/PhysRev.105.1413.

[43] 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, arXiv:hep-ex/0509008. doi:10.1016/j.physrep.2005.12.006.

[44] P. Sikivie, M. B. Voloshin, and V. I. Zakharov *Nucl. Phys.* **B173** (1980) 189.

[45] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys. Rev.* **127** (1962) 965–970. doi:10.1103/PhysRev.127.965.

[46] 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.

[47] H. Yukawa, "On the Interaction of Elementary Particles I", *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57. doi:10.1143/PTPS.1.1.

[48] A. Martin, W. Stirling, R. Thorne et al., "Parton distributions for the LHC", *Eur. Phys. J.* **C63** (2009) 189–285, arXiv:0901.0002. doi:10.1140/epjc/s10052-009-1072-5.

[49] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, arXiv:hep-ph/0409313. doi:10.1142/9789814503266_0001.

[50] C. Anastasiou, C. Duhr, F. Dulat et al., "Higgs Boson Gluon-Fusion Production in QCD at Three Loops", *Phys. Rev. Lett.* **114** (2015) 212001, arXiv:1503.06056. doi:10.1103/PhysRevLett.114.212001.

[51] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", arXiv:1101.0593. doi:10.5170/CERN-2011-002.

[52] S. Mandelstam, "Determination of the pion - nucleon scattering amplitude from dispersion relations and unitarity. General theory", *Phys. Rev.* **112** (1958) 1344–1360. doi:10.1103/PhysRev.112.1344.

[53] P. Bolzoni, F. Maltoni, S. Moch et al., "Higgs production via vector-boson fusion at NNLO in QCD", *Phys. Rev. Lett.* **105** (2010) 011801, arXiv:1003.4451. doi:10.1103/PhysRevLett.105.011801.

[54] M. Wiesemann, R. Frederix, S. Frixione et al., "Higgs production in association with bottom quarks", *JHEP* **02** (2015) 132, arXiv:1409.5301. doi:10.1007/JHEP02(2015)132.

[55] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", arXiv:1307.1347. doi:10.5170/CERN-2013-004.

[56] "LEP design report". CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.

[57] 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, arXiv:hep-ex/0306033. doi:10.1016/S0370-2693(03)00614-2.

[58] TeVI Group, "Design Report Tevatron 1 project", *FERMILAB-DESIGN-1984-01* (1984).

[59] 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, arXiv:1207.6436. doi:10.1103/PhysRevLett.109.071804.

[60] ATLAS and CMS Collaborations, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* **114** (2015) 191803, arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.

[61] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. J.* **C75** (2015), no. 5, 212, arXiv:1412.8662. doi:10.1140/epjc/s10052-015-3351-7.

[62] CMS Collaboration, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV", arXiv:1411.3441.

[63] ATLAS Collaboration, "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector", arXiv:1506.05669.

[64] ATLAS Collaboration, "Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data", *ATLAS-CONF-2014-009, ATLAS-COM-CONF-2014-013* (2014).

[65] ATLAS Collaboration, "Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector", arXiv:1503.01060.

[66] CMS Collaboration, "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs", *Phys. Lett.* **B736** (2014) 64, arXiv:1405.3455. doi:10.1016/j.physletb.2014.06.077.

[67] 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, arXiv:1310.3687. doi:10.1103/PhysRevD.89.012003.

[68] ATLAS Collaboration, "Evidence for Higgs boson decays to the $\tau^+\tau^-$ final state with the ATLAS detector", *ATLAS-CONF-2013-108, ATLAS-COM-CONF-2013-095* (2013).

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

[70] ATLAS Collaboration, "Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the top quark-Higgs boson Yukawa coupling using data taken at 7 TeV and 8 TeV with the ATLAS detector", *ATLAS-CONF-2014-043, ATLAS-COM-CONF-2014-059* (2014).

[71] D. Buttazzo, G. Degrassi, P. P. Giardino et al., "Investigating the near-criticality of the Higgs boson", *JHEP* **12** (2013) 089, arXiv:1307.3536. doi:10.1007/JHEP12(2013)089.

[72] S. Forte et al., "The Standard Model from the LHC to future colliders: a contribution to the Workshop "What Next" of INFN", arXiv:1505.01279.

[73] P. H. Frampton, "Vacuum Instability and Higgs Scalar Mass", *Phys. Rev. Lett.* **37** (1976) 1378. doi:10.1103/PhysRevLett.37.1378.

[74] G. Degrassi, S. Di Vita, J. Elias-Miro et al., "Higgs mass and vacuum stability in the Standard Model at NNLO", *JHEP* **08** (2012) 098, arXiv:1205.6497. doi:10.1007/JHEP08(2012)098.

[75] G. Branco, P. Ferreira, L. Lavoura et al., "Theory and phenomenology of two-Higgs-doublet models", *Phys. Rept.* **516** (2012) 1–102, arXiv:1106.0034. doi:10.1016/j.physrep.2012.02.002.

[76] H. E. Haber and G. L. Kane, "The search for supersymmetry: Probing physics beyond the Standard Model", *Phys. Rept.* **117** (1985) 75–263. doi:doi:10.1016/0370-1573(85)90051-1.

[77] J. E. Kim, "Light Pseudoscalars, Particle Physics and Cosmology", *Phys. Rept.* **150** (1987) 1–177. doi:10.1016/0370-1573(87)90017-2.

[78] D. Lopez-Val, T. Plehn, and M. Rauch, "Measuring extended Higgs sectors as a consistent free couplings model", *JHEP* **10** (2013) 134, arXiv:1308.1979. doi:10.1007/JHEP10(2013)134.

[79] A. Broggio, E. J. Chun, M. Passera et al., "Limiting two-Higgs-doublet models", *JHEP* **11** (2014) 058, arXiv:1409.3199. doi:10.1007/JHEP11(2014)058.

[80] F. Jegerlehner and A. Nyffeler, "The Muon g-2", *Phys. Rept.* **477** (2009) 1–110, arXiv:0902.3360. doi:10.1016/j.physrep.2009.04.003.

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

[82] J. F. Gunion and H. E. Haber, "The CP conserving two Higgs doublet model: The Approach to the decoupling limit", *Phys. Rev.* **D67** (2003) 075019, arXiv:hep-ph/0207010. doi:10.1103/PhysRevD.67.075019.

[83] D. M. Asner et al., "ILC Higgs White Paper", in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*. 2013. arXiv:1310.0763.

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

[85] S. de Visscher, J. M. Gerard, M. Herquet et al., "Unconventional phenomenology of a minimal two-Higgs-doublet model", *JHEP* **08** (2009) 042, arXiv:0904.0705. doi:10.1088/1126-6708/2009/08/042.

[86] R. V. Harlander, S. Liebler, and T. Zirke, "Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-Doublet Model", *JHEP* **02** (2014) 023, arXiv:1307.8122. doi:10.1007/JHEP02(2014)023.

[87] G. Dorsch, S. Huber, K. Mimasu et al., "Echoes of the Electroweak Phase Transition: Discovering a second Higgs doublet through $A_0 \to ZH_0$", *Phys. Rev. Lett.* **113** (2014), no. 21, 211802, arXiv:1405.5537. doi:10.1103/PhysRevLett.113.211802.

[88] B. Hespel, F. Maltoni, and E. Vryonidou, "Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM", *JHEP* **06** (2015) 065, arXiv:1503.01656. doi:10.1007/JHEP06(2015)065.

[89] CMS Collaboration, "Search for a pseudoscalar boson decaying into a Z boson and the 125 GeV Higgs boson in $\ell\ell b\bar{b}$ final states", *Phys. Lett.* **B748** (2015) 221–243, arXiv:1504.04710. doi:10.1016/j.physletb.2015.07.010.

[90] CMS Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$", *To be submitted to Physics Letter B*.

[91] CMS Collaboration, "Searches for heavy Higgs bosons in two-Higgs-doublet models and for $t \to ch$ decay using multilepton and diphoton final states in pp collisions at 8 TeV", *Phys. Rev.* **D90** (2014) 112013, arXiv:1410.2751. doi:10.1103/PhysRevD.90.112013.

[92] ATLAS Collaboration, "Search for a CP-odd Higgs boson decaying to Zh in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Phys. Lett.* **B744** (2015) 163–183, arXiv:1502.04478. doi:10.1016/j.physletb.2015.03.054.

[93] L. Evans, "The Large Hadron Collider (LHC)", *New J. Phys.* **9** (2007) 335. doi:10.1088/1367-2630/9/9/335.

[94] 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.

[95] "J. Stirling, Private Communication - HEP, Imperial". http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html.

[96] CMS Collaboration, "CMS Physics Technical Design Report, Volume I: Detector Performance and Software", *CERN-LHCC-2006-001* (2006).

[97] F. Marcastel, "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN", *OPEN-PHO-CHART-2013-001* (2013). General Photo.

[98] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005.

[99] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.

[100] "CMS Luminosity - Public results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[101] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. arXiv:1405.6569. CERN-PH-EP-2014-070. CMS-TRK-11-001, P10009. 80 p.

[102] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", *CERN-LHCC-97-033* (1997).

[103] P. Adzic et al., "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST* **2** (2007) P04004. doi:10.1088/1748-0221/2/04/P04004.

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

[105] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report", *CERN-LHCC-97-031* (1997).

[106] CMS and LHCb Collaborations, "Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data", *Nature* (2015) arXiv:1411.4413. doi:10.1038/nature14474.

[107] CMS Collaboration, "CMS, the Compact Muon Solenoid. Muon technical design report", *CERN-LHCC-97-32* (1997).

[108] 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, arXiv:1306.6905. doi:10.1088/1748-0221/8/11/P11002.

[109] CMS Collaboration, "CMS TriDAS project, Technical Design Report, Volume 1: The Trigger Systems", *CERN-LHCC-2000-038* (2000).

[110] CMS Collaboration, "CMS TriDAS project, Technical design report, Volume 2: Data acquisition and high-level trigger", *CERN-LHCC-2002-026* (2002).

[111] P. R. Klabbers, R. C. Aggleton, M. D. J. Baber et al., "Installation and Commissioning of the CMS Level-1 Calorimeter Trigger Upgrade", Technical Report CMS-CR-2014-423, CERN, Geneva.

[112] CMS Collaboration, "CMS: The computing project. Technical design report", *CERN-LHCC-2005-023* (2005).

[113] I. Bird, K. Bos, N. Brook et al., "LHC computing Grid. Technical design report", *CERN-LHCC-2005-024* (2005).

[114] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026.

[115] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph5 : Going Beyond", *JHEP* **06** (2011) 128, arXiv:1106.0522. doi:10.1007/JHEP06(2011)128.

[116] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph", *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156. doi:10.1088/1126-6708/2003/02/027.

[117] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, arXiv:hep-ph/0204244. doi:10.1088/1126-6708/2002/06/029.

[118] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, arXiv:0709.2092. doi:10.1088/1126-6708/2007/11/070.

[119] S. Jadach, Z. Was, R. Decker et al., "The Tau Decay Library Tauola: Version 2.4", *Comput. Phys. Commun.* **76** (1993) 361.

[120] N. Davidson, G. Nanava, T. Przedzinski et al., "Universal Interface of TAUOLA Technical and Physics Documentation", *Comput. Phys. Commun.* **183** (2012) 821–843, arXiv:1002.0543. doi:10.1016/j.cpc.2011.12.009.

[121] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.

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

[123] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth.* **A262** (1987) 444–450. doi:10.1016/0168-9002(87)90887-4.

[124] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, arXiv:1405.6569. doi:10.1088/1748-0221/9/10/P10009.

[125] K. Rose, "Deterministic Annealing for Clustering, Compression, Classification, Regression and related Optimisation Problems", *Proceedings of the IEEE* **Vol. 86, Issue 11** (1998).

[126] W. Waltenberger, R. Frühwirth, and P. Vanlaer, "Adaptive vertex fitting", *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 12, N343.

[127] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277. doi:10.1088/1748-0221/6/11/P11002.

[128] W. Adam, R. Frühwirth, A. Strandlie et al., "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", *CMS-NOTE-2005-001* (2005), no. CMS-NOTE-2005-001,.

[129] 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, arXiv:1502.02701. doi:10.1088/1748-0221/10/06/P06005.

[130] B. P. Roe, H.-J. Yang, J. Zhu et al., "Boosted decision trees, an alternative to artificial neural networks", *Nucl. Instrum. Meth.* **A543** (2005), no. 2-3, 577–584, arXiv:physics/0408124. doi:10.1016/j.nima.2004.12.018.

[131] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/ψ and W events at 7 TeV", *CMS-PAS-PFT-10-003* (2010).

[132] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[133] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, arXiv:0707.1378. doi:10.1016/j.physletb.2007.09.077.

[134] CMS Collaboration, "Pileup Jet Identification", *CMS-PAS-JME-13-005* (2013).

[135] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events", *CMS-PAS-BTV-13-001* (2013).

[136] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, arXiv:1411.0511. doi:10.1088/1748-0221/10/02/P02006.

[137] CMS Collaboration, "MET performance in 8 TeV data", *CMS-PAS-JME-12-002* (2013).

[138] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, arXiv:1106.5048. doi:10.1088/1748-0221/6/09/P09001.

[139] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531–533. doi:10.1103/PhysRevLett.10.531.

[140] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) 005, arXiv:0802.1188. doi:10.1088/1126-6708/2008/04/005.

[141] CDF Collaboration, "Search for neutral MSSM Higgs bosons decaying to tau pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **96** (2006) 011802, arXiv:hep-ex/0508051. doi:10.1103/PhysRevLett.96.011802.

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

[143] CMS Collaboration, "Performance of τ-lepton reconstruction and identification in CMS", *JINST* **7** (2012) 1001, arXiv:1109.6034. doi:10.1088/1748-0221/7/01/P01001.

[144] L. Bianchini, J. Conway, E. K. Friis et al., "Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques", *J. Phys. Conf. Ser.* **513** (2014) 022035. doi:10.1088/1742-6596/513/2/022035.

[145] G. P. Lepage, "A New Algorithm for Adaptive Multidimensional Integration", *J. Comput. Phys.* **27** (1978) 192. doi:10.1016/0021-9991(78)90004-9.

[146] "Tau ID Performance Plots", *CMS Detector Performance Note* **CMS-DP-2014-015** (2014).

[147] 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, arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021.

[148] R. Gavin, Y. Li, F. Petriello et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388–2403, arXiv:1011.3540. doi:10.1016/j.cpc.2011.06.008.

[149] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, arXiv:1105.0020. doi:10.1007/JHEP07(2011)018.

[150] T. Binoth, N. Kauer, and P. Mertsch, "Gluon-induced QCD corrections to $pp \to ZZ \to \ell\bar{\ell}\ell'\bar{\ell}'$", arXiv:0807.0024. doi:10.3360/dis.2008.142.

[151] M. Bachtis, C. Calabria, S. Chhibra et al., "Search for a Standard Model Higgs boson produced in the decay channel $H \to ZZ \to 2\ell 2\tau$ final states with CMS detector at $\sqrt{s} = 7$ TeV", CMS Analyis Note 2011/402, CERN, (2011).

[152] CMS Collaboration, "Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV", *CMS-PAS-HIG-13-004* (2013).

[153] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update", *CMS-PAS-SMP-12-008* (2011).

[154] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting Summer 2013 Update", *CMS-PAS-LUM-13-001* (2013).

[155] "Recommendation for using b-tag objects in physics analyses". https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation.

[156] S. Dittmaier, S. Dittmaier, C. Mariotti et al., "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", arXiv:1201.3084.

[157] J. Pumplin, D. R. Stump, J. Huston et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195. doi:10.1088/1126-6708/2002/07/012.

[158] Nadolsky, P.M. et al., "Implications of CTEQ global analysis for collider observables",. *Phys. Rev. Lett.* **D78** (2008) 013004. doi:10.1103/PhysRevD.78.013004.

[159] N. Amapane et al., "Updated results on the new boson discovered in the search for the standard model Higgs boson in the $H \to ZZ \to 4\ell$ channel in pp collisions at $\sqrt{s}$ =7 and 8 TeV", CMS Analyis Note 2012/367, CERN, (2012).

[160] CMS Collaboration, "Measurement of associated production of vector bosons and top quark-antiquark pairs at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **110** (2013) 172002, arXiv:1303.3239. doi:10.1103/PhysRevLett.110.172002.

[161] "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons". https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig13004PubTWiki.

[162] R. D. Cousins, "Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms". http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[163] "Physics Letters B - Particle Physics, Nuclear Physics and Cosmology". http://www.journals.elsevier.com/physics-letters-b/.

[164] A. Alloul, N. D. Christensen, C. Degrande et al., "FeynRules 2.0 - A complete toolbox for tree-level phenomenology", *Comput. Phys. Commun.* **185** (2014) 2250–2300, arXiv:1310.1921. doi:10.1016/j.cpc.2014.04.012.

[165] D. Eriksson, J. Rathsman, and O. Stal, "2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual", *Comput. Phys. Commun.* **181** (2010) 189–205, arXiv:0902.0851. doi:10.1016/j.cpc.2009.09.011.

[166] J. Alwall, A. Ballestrero, P. Bartalini et al., "A Standard format for Les Houches event files", *Comput. Phys. Commun.* **176** (2007) 300–304, arXiv:hep-ph/0609017. doi:10.1016/j.cpc.2006.11.010.

[167] R. V. Harlander, S. Liebler, and H. Mantler, "SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM", *Comput. Phys. Commun.* **184** (2013) 1605–1617, arXiv:1212.3249. doi:10.1016/j.cpc.2013.02.006.

[168] H. Bakhshian, A. Fahim, A. Calderon et al., "Computing the contamination from fakes in leptonic final states", CMS Analysis Note 2010/261, CERN, (2010).

[169] F. Garwood, "Fiducial Limits for the Poisson Distribution", *Biometrika* **28** (1936) 437–442. doi:10.1093/biomet/28.3-4.437.

[170] G. D. Cowan, "Statistical Data Analysis". Oxford University Press, 1988.

[171] G. J. Feldman and R. D. Cousins, "A unified approach to the classical statistical analysis of small signals", *Phys.Rev.* **D57** (1998) 3873–3889, arXiv:physics/9711021. doi:10.1103/PhysRevD.57.3873.

[172] ATLAS Collaboration, "Procedure for the LHC Higgs boson search combination in summer 2011", *ATL-PHYS-PUB-2011-011* (2011).

[173] H. Held, P. Schieferdecker, and M. Voutilainen, "Update of the Measurement of the Jet p_T Resolution in $\sqrt{s} = 7$ TeV Collision Data with the Asymmetry Method", CMS Analyis Note 2010/371, CERN, (2010).

[174] CMS Collaboration, "Measurement of the production cross section and anomalous trilinear gauge couplings in lll'l' decays at sqrt(s) = 8 TeV at the LHC", *CMS-PAS-SMP-13-005* (2013).

[175] CMS Collaboration, "Additional Performance Results for EPS 2015",.

[176] J. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", arXiv:1103.0354.

[177] G. Cowan, K. Cranmer, E. Gross et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, arXiv:1007.1727. doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z.