



Search for the Standard Model Higgs in Tau Decays in Association with a W Boson

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Mauro Verzetti

aus

Italien

Promotionskomitee

Prof. Dr. V. Chiochia

Prof. Dr. F. Canelli

Dr. C. Veelken

Dr. S. Taroni

Zürich, 2014

Abstract

Diese Dissertation stellt die Ergebnisse der Suche nach dem Standardmodell-Prozess $pp \rightarrow WH \rightarrow \ell \nu \tau_\ell \tau_h$, in dem das Higgs-Boson in Tau-Leptonen zerfällt, wobei eines der Tau-Leptonen hadronisch zerfällt, dar. Die Anwesenheit eines zusätzlichen Leptons aus dem W-Boson-Zerfall führt zu einer starken Reduzierung des Untergrunds in diesem Kanal.

Das Higgs-Boson, welches ursprünglich von P. Higgs, F. Englert et al. 1964 vorhergesagt wurde, war lange Zeit das einzig fehlende, vom Standardmodell (SM) vorhergesagte Teilchen und war daher Kern vieler großer Anstrengungen in der experimentellen Teilchenphysik. Suchen nach diesem schwer zu entdeckendem Teilchen am LEP und auch am Tevatron waren erfolglos. Der Large Hadron Collider (LHC) und die beiden Hauptexperimente (ATLAS und CMS) wurden hauptsächlich gebaut, um Fortschritte in der Suche nach dem Higgs-Boson zu erzielen.

Die Anstrengungen machten sich am 4. Juli 2012 bezahlt, als beide Experimente die Entdeckung einer Resonanz um 125 GeV in den Spektren der $\gamma\gamma$ und der ZZ Endzustände bekannt gaben. Obwohl diese Nachricht enthusiastisch aufgenommen wurde, mussten fermionische Zerfälle dieser Resonanz erst noch entdeckt werden. Die Suche nach dem Zerfallsprozess $H \rightarrow \tau\tau$ war der vielversprechendste Kanal um diesen Zerfall zu entdecken und stand im direkten Wettbewerb mit dem Zerfall $H \rightarrow b\bar{b}$, der zwar ein größeres Verzweigungsverhältnis besitzt, aber dafür auch mit größerem Untergrund zu kämpfen hat.

In dieser Arbeit konzentrieren wir uns auf die assoziierte Produktion eines Higgs-Bosons und eines W-Bosons, um möglichst kleine Untergrundprozesse zu haben. Dieser Endzustand ist Teil eines größeren Suchprogramms des CMS-Experiments, welches im ersten Kapitel dieser Arbeit kurz beschrieben wird.

Der Hauptuntergrund in diesem Endzustand setzt sich aus einer Reihe von Prozessen (hauptsächlich Top-Quark- und W-Boson-Produktion zusammen mit Jets) zusammen, in denen mindestens eines der Leptonen im Endzustand aus der Fehlidentifizierung eines Quark- oder Gluonjets stammt. Um diese Untergrundquelle zu modellieren wird eine datengetriebene Methode entwickelt, die sogenannte k-Nearest Neighbors Classifier verwendet.

In dieser Suche wurde der Produktionswirkungsquerschnitt des Higgs-Bosons mit einer

kombinierten Signalstärke von $\mu = -2.6 \pm 1.7$ gemessen, die unterhalb der SM-Erwartung liegt, jedoch mit dieser innerhalb eines Signifikanzniveaus von 2.1σ kompatibel ist. Das Ergebnis dieser Analyse wurde mit anderen Suchen nach Tau-Lepton-Zerfällen des Higgs-Bosons im CMS-Experiment kombiniert und resultierte damit in der ersten Beobachtung dieses Prozesses.

Abstract

This thesis presents the results of the search for the Standard Model (SM) process $pp \rightarrow WH \rightarrow \ell \nu \tau \ell \tau_h$, where the Higgs boson decays to tau leptons and one of them decays hadronically. The presence of the additional lepton from the decay of the W boson largely reduces the backgrounds in this channel.

Originally predicted by P. Higgs, F. Englert et al. in 1964, the Higgs boson has been for long time the only missing particle predicted by the SM and has been therefore the target of major efforts in experimental particle physics. Searches for this elusive particles have been conducted without any success both at LEP and Tevatron. The Large Hadron Collider (LHC) and two major experiments (ATLAS and CMS) have been built with the main purpose of casting some light in this longstanding mystery.

These efforts culminated on July 4th 2012, when both the experiments announced the observation of a resonance at 125 GeV in the $\gamma\gamma$ and ZZ spectra combined. While this announcement generated a well deserved enthusiasm, fermionic decays of this resonance were still to be observed. The search for the decay process $H \rightarrow \tau\tau$ has been the most promising channel to first observe the decay of the SM Higgs boson to fermions, as the other main competitor, $H \rightarrow b\bar{b}$, has larger branching fraction but much higher backgrounds.

In this work we focus on the associated production of a Higgs boson and a W boson as a low-background environment to observe such decay. This final state is part of a wider $H \rightarrow \tau\tau$ search program performed by CMS which is briefly outlined in the first chapter of this work.

The main source of background in this final state comes from a wide range of processes (mainly top quark production and W boson production in association with jets) in which at least one of the leptons in the final state comes from a misidentified quark or gluon jet. To model this source of backgrounds a fully data-driven method is developed.

The search observes a combined signal strength of $\mu = -2.6 \pm 1.7$, below the SM expectations but still compatible with it at 2.1σ level. This result has been combined with the other searches for Higgs decays into tau pairs performed at the CMS experiment, leading to the first observation of such process.

Acknowledgements

I first of all want to thank my advisor Prof Vincenzo Chiochia, for his trust in my abilities and his mentoring, which is of prime importance in such a large collaboration.

I owe special gratitude towards to Dr. Evan Friis, who sharpened my programming skills as well as my physics ones, and was a real pleasure to work in close contact with him, and to Dr. Hella Snoek for her support and for her wake-up call, leading me to understand that I was able to take matters in my own hands.

I want to thank Prof. Sridhara Dasu for his guidance in the Higgs to tau group, especially in the hard periods before the final approval of the analysis and the Tau Physics Object Group, and in particular its two conveners Dr. Monica Acosta and Dr. Christian Veelken, which acted like a second institution for me and helped me throughout my studies.

A special thank also to my fiancée Jessica and my parents for their help and support in these four years, which included some difficult periods, like every doctoral study, that would have been much harder without the presence of a caring companion and supportive relatives.

I finally want to thank the University Zurich group: Silvia, Carlotta, Barbara, William and Mirena for sharing with me this journey, our secretaries Carmelina, Ruth and Monica because I'm really unable to process bureaucracy, and in general all the friends that I've made during these four years, who helped making my working and private life better.

*There is a theory which states that if ever
anybody discovers exactly what the Universe is
for and why it is here, it will instantly
disappear and be replaced by something even
more bizarre and inexplicable. There is another
theory which states that this has already
happened.*

Douglas Adams, The Restaurant at the End
of the Universe

Contents

1	The Higgs boson	1
1.1	The Standard Model	2
1.1.1	Elementary particles	2
1.1.2	Elementary forces	3
1.1.3	A gauge theory of particle interactions	4
1.1.4	Discrete and continuous symmetries	6
1.2	The electroweak force	7
1.3	The Higgs mechanism	10
1.3.1	Spontaneous symmetry breaking	10
1.3.2	The Higgs mechanism	12
1.3.3	Couplings to gauge bosons	15
1.3.4	Couplings to fermions	15
1.4	The Higgs boson at the LHC	16
1.4.1	Higgs production and decays	16
1.5	Higgs boson searches	20
2	The Large Hadron Collider and the CMS experiment	25
2.1	The Large Hadron Collider	25
2.2	The CMS detector	29
2.2.1	Inner Tracker	31
2.2.2	Electromagnetic Calorimeter	32
2.2.3	Hadronic Calorimeter	33
2.2.4	Muon System	34
2.2.5	Trigger	36

2.3	Data storage and processing	37
3	Event reconstruction in CMS	39
3.1	Track reconstruction	39
3.1.1	Reconstruction of muon tracks	41
3.2	Vertex reconstruction	42
3.3	Electron and photon reconstruction	43
3.4	Particle Flow	44
3.4.1	Muon identification	45
3.4.2	Isolation	45
3.4.3	Jet clustering	46
3.4.4	Missing transverse energy	46
3.5	Tau reconstruction and identification	47
3.5.1	Reconstruction	47
3.5.2	Identification	48
3.5.3	Light lepton rejection	49
3.5.4	Performance	51
4	$pp \rightarrow WH$	59
4.1	Event Selection	59
4.2	Background estimation	63
4.2.1	Irreducible backgrounds	64
4.2.2	Reducible backgrounds	64
4.2.3	Fake lepton method	66
4.2.4	Validation of the fake lepton method	71
4.2.5	Fake tau method	71
4.2.6	Electron charge mis-identification	80
4.2.7	Control regions	80
4.3	Systematics uncertainties and category optimization	81
4.3.1	Category optimization	81
4.3.2	Systematic uncertainties	86
4.4	Results	87

4.5	Limit setting procedure and results	89
4.5.1	Statistical framework	89
4.5.2	Fit results	95
4.5.3	Exclusion limits	97
5	Conclusions and possible future developments	107
A	Additional plots	109
A.1	Background estimation plots	109
A.2	Background control regions	109

Chapter 1

The Standard Model of particle physics and the Higgs boson

In the current understanding of particle physics, elementary particles are described as excitations of their peculiar fields, a description derived from quantum mechanics, and their interaction are governed by a Lagrangian. In this scheme each fundamental force is associated to a field and its particle(s), which have integer spin value (gauge bosons). Conserved physical observables (particle's quantum numbers) are preserved by dedicated gauge symmetries introduced in the Lagrangian. Back in the 1960's, the behavior of the weak force, responsible for the β decays and, more generally, for all the decays with long lifetime observed in nature (e.g. the decay of the muon), could not fit in this description. Due to its long lifetime and short interaction distance, the weak force mediator had to be massive, but the direct introduction of a mass term for a bosonic field would have lead to a disruption of the Lagrangian gauge symmetry.

In this period the work of several physicists, including Peter Higgs and Francois Englert, led to a different formulation of the origin of gauge boson masses, compliant with the requirements of a gauge-invariant Lagrangian. This breakthrough consolidated the model by predicting a new elementary particle: the Higgs boson.

In the first part of this chapter the mathematical formulation of the elementary particles and interactions, the so-called *Standard Model* (SM) of particle physics, is reviewed. In the second part a detailed review of the status of experimental Higgs boson studies is presented.

1.1 The Standard Model

1.1.1 Elementary particles

Elementary particles composing matter are *fermions*, i.e. particles with spin $1/2$ which obey the Fermi-Dirac statistics. These particles can be divided into two categories according to their interaction with the strong force (described later): leptons and quarks. The former are neutral with respect to the strong force, while the latter carry a strong charge, the so-called "color". Quarks and leptons are divided into three families (or generations). Each family contains a doublet of either quarks or leptons which carry the same quantum numbers as their homologous in the other families. The only distinctive characteristic between families is the mass of their components.

In addition, for each particle there is a corresponding antiparticle. Each anti-matter particle carries the same mass, spin and lifetime as its counter-parts, but has opposite quantum charges.

Leptons Leptons comprise three charged particles, the electron (e^-), the muon (μ^-), and the tau (τ^-), all carrying the same electric charge as the electron: $Q = -e = -1.602 \times 10^{-19}$ C. Each of the charged leptons is associated to a *neutrino*, a particle with neutral charge assumed to be massless in the SM. The recent evidence of neutrino oscillations [1] proved that neutrinos do have a mass, albeit extremely small (< 2 eV [2]). The precise values of the neutrino masses are not yet measured. The existence of a fourth lepton family with neutrinos below 45 GeV has been excluded by the fit of the Z to invisible production at LEP [3].

Quarks Quarks carry fractional electric charge with respect to the charge of the electron. They also carry a quantum number, called *color*, which allows them to interact with the strong force. The peculiar behavior of the strong force *confines* the quarks within aggregates of multiple constituents called *hadrons*, making impossible the observation of "independent" (or "naked") quarks. Quarks and anti-quarks cluster in groups of two or three particles, forming *mesons* and *baryons*, respectively. The existence of aggregates of four and five (anti-)quarks, called *tetraquarks* and *pentaquarks*, has been long postulated and only recently confirmed by the BELLE [4] and LHCb collaborations [5].

Each quark family is composed by a quark with charge $+2/3$ and one of charge $-1/3$ also named "up-type" and "down-type" respectively from the name of the constituents of the first family (which form protons and neutrons). The presence of a fourth generation of leptons and quarks has not been completely excluded and is object of dedicated searches that are beyond the scope of this work.

Table 1.1 summarizes the properties of the quarks and the leptons currently known.

Table 1.1: Fundamental particles of the SM, listed with their main properties and classification. For each particle a corresponding anti-particle exists, with same mass and opposite electric charge.

	Generation	Name	Symbol	Charge [e]	Mass [MeV]
leptons	1	electron	e	-1	0.511
		electronic neutrino	ν_e	0	$< 2 \times 10^{-6}$
	2	muon	μ	-1	105.7
		muonic neutrino	ν_μ	0	< 0.19
	3	tau	τ	-1	1.78×10^3
		tauonic neutrino	ν_τ	0	< 18.2
quarks	1	up	u	$+2/3$	$2.3^{+0.7}_{-0.5}$
		down	d	$-1/3$	$4.8^{+0.5}_{-0.3}$
	2	charm	c	$+2/3$	$(1.275 \pm 0.025) \times 10^3$
		strange	s	$-1/3$	95 ± 5
	3	top	t	$+2/3$	$(173.07 \pm 0.52 \pm 0.72) \times 10^3$
		bottom/beauty	b	$-1/3$	$(4.18 \pm 0.03) \times 10^3$

1.1.2 Elementary forces

Elementary forces are mediated by bosons, particles with integer spin. The forces considered fundamental in the SM are:

The *strong* interaction It is responsible for confining the quarks within the hadrons. It also binds protons and neutrons inside the atomic nuclei. The color charge, linked to the strong interaction, can be defined in a three dimensional group. Unitary charges for the strong interaction are therefore named according to the three primary colors: red, green, and blue. Anti-colors can be defined as negative values of the fundamental color charges. The strong interaction is mediated by eight electrically neutral and massless bosons called *gluons*, each carrying a pair of color and anti-color charges. The peculiar characteristic of the strong interaction, which differentiates it from the other fundamental forces, is that its strength grows as the distance between the interacting particles increases, leading to the confinement of the quarks. A hadron, globally neutral with respect to strong interaction, it is often referred as “white” due to optical color analogy.

The *electromagnetic* (EM) interaction It is responsible for the reactions between electrically charged particles, and is empirically known to have an infinite range. The part of the SM that describes this interaction is called *Quantum Electrodynamics* (QED) and considers the interaction to be mediated by massless neutral particles called *photons*, which are the EM wave quanta.

Table 1.2: Fundamental interactions and their main properties.

Interaction	Effective coupling	Mediator	Mass [GeV]	Range [fm]
Gravitation	10^{-39}	graviton	0	inf
Electromagnetism	$1/137$	photon	0	inf
Weak force	10^{-5}	$W^\pm, Z$	80.4, 91.2	10^{-3}
Strong force	1	gluon	0	$< 1^1$

The *weak* interaction It is responsible for most of the nuclear decays that can be observed in nature (except the α and γ nuclear decays). Its short interaction range and relative long lifetime of particles decaying weakly, such as the muon, imply the mediators to be massive. The weak interaction allows for decays that are forbidden by the strong and EM interactions, such as flavor-changing decays (i.e. between different generations) and parity-violating decays. The weak interaction is mediated by two charged bosons, the $W^\pm$, of mass $m_W = 80.4$ GeV and a neutral boson, the Z^0 , of mass $m_Z = 91.2$ GeV.

The *gravitational* interaction It is responsible of the attraction between bodies. It is expected to be mediated by a spin 2 boson called the *graviton*. No experimental evidence has been found for the existence of such particle, however. The gravitational force is not included in the SM as its strength is so faint its effect on fundamental particles is not measurable with present-day experiments. A summary of the interactions is given in Table 1.2.

1.1.3 A gauge theory of particle interactions

As briefly outlined in the introduction, the interaction between fundamental particles is described by the Lagrangian formalism. The sum of the charges under a specific interaction is seen as a conserved quantity. In order to protect this quantity a dedicated symmetry is introduced in the form of a local *gauge symmetry*.

In analogy with the Lagrangian formalism in classical mechanics, the conserved quantity is the generator of the system symmetry. A Noether current is also associated to each charge and symmetry of the Lagrangian. The quantization of this symmetry leads to the presence of a vectorial field.

For each gauge symmetry exists an associated Lie group. Each of the fundamental particles and force mediators correspond to one of the fundamental representations of the Lie group, fully determining the behavior of the final theory.

¹The range of the nuclear force, not that of the quark-quark force

The behavior of a free fundamental particle of spin $1/2$ is described by the Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\mathcal{D} - m)\psi, \quad (1.1)$$

where $\mathcal{D} = \gamma^\mu \partial_\mu$ and ψ is the particle spinor.

From the Dirac Lagrangian it is possible to derive the Dirac equation:

$$i\partial_t\psi = H\psi, \quad (1.2)$$

where H is the free Hamiltonian associated to the Lagrangian. In this framework the electromagnetic interaction can be described by the U(1) Lie group, whose associated symmetry is a phase change of the spinorial field:

$$\psi'(x) = e^{iq\theta}\psi(x), \quad (1.3)$$

where q denotes the electric charge. A charged particle is represented by a complex field and the spin behavior is handled by the ψ spinor.

The behavior of a free photon, the gauge boson associated to the electromagnetic force, is described by the Lagrangian:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.4)$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (1.5)$$

and A_μ is the field associated to the photon. The minimal coupling between a spinor and a photon field is described by the interaction Lagrangian:

$$\mathcal{L} = iq\bar{\psi}\mathcal{A}\psi. \quad (1.6)$$

The full electromagnetic Lagrangian can therefore be written as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\mathcal{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.7)$$

where $\mathcal{D} = \partial_\mu + iqA_\mu$ is the extension of the four-dimensional derivative ∂ to ensure the covariance of the Lagrangian. This covariant derivative also embeds the interaction term between the fermions and the photon.

The strong interaction is described by the SU(3) group and its associated gauge symmetry, which is generated by the Gell-Mann matrices, τ_a . A fundamental property of

such group is that it is not *abelian*, i.e. the generator matrices do not commute. The commutation rules of the SU(3) group, as any other Lie group, are determined by its *structure constants* $[\tau_a, \tau_b] = iC_{abc}\tau^c$. This property leads to small modification of gauge bosons Lagrangian (Eq. 1.8) and of the coupling with the quarks (Eq. 1.9).

$$\mathcal{L}_{gluons} = -\frac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a, \text{ with } G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s C_{abc} A_\mu^b A_\nu^c \quad (1.8)$$

where A_ν^c is the field associated to the gluon.

$$\mathcal{L}_{quarks} = \bar{\psi}(i(\not{\partial} + ig_s \not{A} \tau_a) - m)\psi = \bar{\psi}(i\not{D} - m)\psi. \quad (1.9)$$

The presence of trilinear and quartic terms in Eq. 1.8 describe the self-interaction between gluons, which is allowed already at the first order of perturbative expansion (also referred as *tree level* or *Born approximation*). The field theory describing strong interactions is called *quantum chromodynamics* (QCD).

1.1.4 Discrete and continuous symmetries

In addition to the gauge symmetries described in the previous section, three discrete symmetries are of particular interest in quantum field theory: charge conjugation (C), parity (P), and time reversal (T). The properties of these symmetries are more easily described by their effect on a quantum state $|f(\vec{p}, h)\rangle$ containing a single particle f of momentum $\vec{p}$ and helicity $h = \vec{s} \cdot \vec{p}/p$, i.e. the projection of the spin along the direction of motion.

The charge conjugation operator C inverts the charge of such state transforming it in its anti-particle: $C|f(\vec{p}, h)\rangle = \eta_C |\bar{f}(\vec{p}, h)\rangle$, where η_C is a phase factor. The parity operator P inverts the spatial coordinates, it is often referred as the left-right symmetry: $P|f(\vec{p}, h)\rangle = \eta_P |f(-\vec{p}, -h)\rangle$, where η_P denotes the parity of the particle. The time inversion operator T inverts time leading to: $T|f(\vec{p}, h)\rangle = \eta_T |f(-\vec{p}, h)\rangle^*$, where η_T is a phase factor depending on the spin of the particle.

Both electromagnetic and strong interaction preserve these three symmetries, while the weak force was experimentally found to violate the charge and parity conjugations. Moreover in 1964 the first experimental evidence of CP (charge and parity combined) violation in K meson decays [6] was reported. The weak force was found to couple only to “left-handed” particles, i.e. particles with negative helicity. More recently the same effect has been found in B mesons decays and is still currently a lively field of study.

Up to now no evidence of CPT violation has been experimentally observed. An hypothetical observation of CPT violation would have very strong consequences in our understanding of the underlying structure of the fundamental interaction, as this symmetry forbids particles and anti-particles to have different mass or lifetime. The CPT conservation is linked to the statistics to which fermions and bosons obey.

1.2 The electroweak force

In the Standard Model Lagrangian presented so far, the weak force has not been discussed yet. The nature of this interaction has been in fact extremely difficult to model in a quantum field theory framework.

The first attempt to describe the weak interaction dates back to 1933 by Fermi, who suggested for the muon decay the following interaction Lagrangian:

$$\mathcal{L}_{Fermi} = \frac{1}{\sqrt{2}} G_F \bar{\nu}_\mu \gamma_\alpha (1 - \gamma^5) \mu \bar{e} \gamma^\alpha (1 - \gamma^5) \nu_e, \quad (1.10)$$

where $G_F = 1.166 \cdot 10^{-5} \text{ GeV}^{-2}$. This Lagrangian consists of a single point-like interaction among four fermions and correctly reproduces the muon decay at tree level. Increasing the degree of the perturbative expansion, though, several divergencies arise that cannot be cancelled or renormalized, leading to a violation of unitarity and a breakdown of the theory. It has also been demonstrated that this behavior can be immediately recognized by the dimensionality of the coupling constant.

It was later shown by Glashow in 1961 [7], and Salam and Ward [8] in the same year, that the weak interaction could be described in a quantum field theory with the SU(2) group and its related gauge bosons. The four-fermion interaction in Eq. 1.10 is divided into two interaction vertices connected by gauge vector bosons, the $W^\pm$ and the Z^0 . In order to obtain the same parity violating properties as observed in the experiments, the gauge group had to be extended to SU(2)×U(1), introducing a mixing between the gauge bosons. This theory showed to represent not only the weak interaction, but also the electromagnetic one. The new modeling of these fundamental interactions was finally completed by Weinberg in 1967 [9], who named it *electroweak theory*. As the weak interaction couples only to left-handed particles, the gauge group is sometimes noted as SU(2)_L, to remind this feature.

Since the generator of the fundamental representation of SU(2) are the Pauli matrices (noted as σ_i , $i \in [1, 3]$) it is possible to extend the spin formalism to the weak interaction. In this case the quantum number (or charge) carried by the fundamental particles is called *weak isospin*, noted with I. The gauge bosons linked to the generators of the symmetry will be represented as W^i . Left-handed fermions are considered to carry total isospin 1/2 and are grouped in doublets of the form:

$$L_f = \begin{pmatrix} f_{up} \\ f_{down} \end{pmatrix}. \quad (1.11)$$

Doublets can be divided in “up-type” and “down-type” fermions, where the former are considered to have the third component of the isospin $I_3 = +1/2$, while the latter have $I_3 = -1/2$. Up-type fermions are the neutrinos and the up, charm and top quarks, while down-type fermions are the charged leptons and the down, strange, and bottom

quarks.

The right-handed fermions were experimentally found not to interact weakly and therefore carry no charge. They are represented as singlets under the SU(2) group.

Left- and right-handed components of the fermions can be obtained by using the appropriate projector on the fermionic spinor: $\frac{1}{2}(1 - \gamma^5)$ and $\frac{1}{2}(1 + \gamma^5)$, respectively.

It is worth noting that quark flavor eigenstates under the weak interaction are not necessary also mass eigenstates, but rather a superimposition of them, giving raise to a mixing matrix. The mixing matrix has first been suggested by Cabibbo in 1963 [10] and by Kobayashi and Maskawa in 1973 [11].

The additional U(1) symmetry preserves a quantum number called *hypercharge*, Y , and is connected to a gauge boson, noted with B .

In this framework the derivative present in the Dirac Lagrangian, Eq. 1.2, is replaced by its covariant form:

$$\mathcal{D}_\mu^L = \partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i - ig' \frac{Y}{2} B_\mu. \quad (1.12)$$

Since the transformations under SU(2) and U(1) are independent, two coupling constants, g and g' , are assigned. The Lagrangian therefore becomes:

$$\mathcal{L} = \sum_f i \bar{L}_f \mathcal{D}^L L_f + \bar{f}_R^{up} \mathcal{D}^R f_R^{up} + \bar{f}_R^{down} \mathcal{D}^R f_R^{down}, \quad (1.13)$$

where f_R^{up} and f_R^{down} are the right-handed up and down fermion field spinors, respectively. Since f_R^{up} and f_R^{down} are singlets under SU(2) they are presented separately, but they still transform under U(1). To account for this effect we have defined $\mathcal{D}^R = \partial_\mu - ig' Y B_\mu / 2$, which is the same as $\mathcal{D}^L$, but lacks the couplings with the W fields.

We can now restrict ourselves to the electron family, without losing any generality, and identify in the Lagrangian a charged and a neutral current:

$$\mathcal{L}_{charged} = g \bar{L} \mathcal{W}_1 \frac{\sigma_1}{2} L + g \bar{L} \mathcal{W}_2 \frac{\sigma_2}{2} L, \quad (1.14)$$

$$\begin{aligned} \mathcal{L}_{neutral} = & \frac{g}{2} W_3^\mu (\bar{\nu}_{e,L} \gamma_\mu \nu_{e,L} - \bar{e}_L \gamma_\mu e_L) + \\ & + \frac{g'}{2} B^\mu (Y(L) (\bar{\nu}_{e,L} \gamma_\mu \nu_{e,L} + \bar{e}_L \gamma_\mu e_L) + \\ & + Y(\nu_R) \bar{\nu}_{e,R} \gamma_\mu \nu_{e,R} + Y(e_R) \bar{e}_R \gamma_\mu e_R) \end{aligned} \quad (1.15)$$

The charged part of the Lagrangian contains off-diagonal Pauli matrices that mix the neutrino and electron fields. By means of the substitution:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad (1.16)$$

it is possible to rewrite the charged part of the Lagrangian as:

$$\mathcal{L}_c = \frac{g}{\sqrt{2}}(\bar{L}W^+\sigma^+L + \bar{L}W^-\sigma^-L), \quad (1.17)$$

where:

$$\sigma^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \sigma^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (1.18)$$

In the neutral part of the Lagrangian, neither W^3 nor B can be directly identified with the photon, since they both interact with the neutrino which has no electric charge. As the two gauge bosons have the same quantum numbers ($Q = 0$ and $J^P = 1^-$), it is possible to create a linear combination of the two fields which represent the electromagnetic photon.

$$\begin{cases} B^\mu = A^\mu \cos \theta_W - Z^\mu \sin \theta_W \\ W_3^\mu = A^\mu \sin \theta_W + Z^\mu \cos \theta_W \end{cases} \quad (1.19)$$

This linear combination can be seen as a rotation of the fields by an angle θ_W , which is referred to as *Weinberg angle* in the literature. The neutral part of the Lagrangian therefore becomes:

$$\mathcal{L}_N = \bar{\Psi}\gamma_\mu(g \sin \theta_W T_3 + \frac{Y}{2}g' \cos \theta_W)\Psi A^\mu + \bar{\Psi}\gamma_\mu(g \cos \theta_W T_3 - \frac{Y}{2}g' \sin \theta_W)\Psi Z^\mu, \quad (1.20)$$

where:

$$\Psi = \begin{pmatrix} \nu_L \\ e_L \\ \nu_R \\ e_R \end{pmatrix}, \quad T_3 = \text{diag}\left(\frac{1}{2}, -\frac{1}{2}, 0, 0\right). \quad (1.21)$$

In this Lagrangian the boson A can be identified with the QED photon. Similarly to what happens to the fields, the electric charge can be written as a linear combination of the third component of the weak isospin and the hypercharge. The charge operator can be derived from the coupling of the photon field to the fermions:

$$Q = T_3 + \frac{Y}{2}. \quad (1.22)$$

In the case of the electron family this leads to the relation:

$$g \sin \theta_W = g' \cos \theta_W = e, \quad (1.23)$$

which links the coupling constants and the electron charge to the value of θ_W . The sine and cosine of the Weinberg angle can therefore be defined as:

$$\begin{aligned} \sin \theta_W &= \frac{e}{g} = \frac{g'}{\sqrt{g^2 + g'^2}}, \\ \cos \theta_W &= \frac{e}{g'} = \frac{g}{\sqrt{g^2 + g'^2}}. \end{aligned} \quad (1.24)$$

The identification of the A boson with the photon completely determines the choice of the quantum numbers of the fundamental particles and also defines their coupling to the Z boson. At the time of the formulation of the electroweak theory the Z boson has not yet been observed experimentally.

The interaction between the gauge boson fields is described by the Lagrangian:

$$\mathcal{L}_g = -\frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \quad (1.25)$$

where W and B represent the *field strength tensor* of the respective field. As in the QCD formalism, the $SU(2)$ generators do not commute with each other, creating additional trilinear and quartic terms in the Lagrangian that allow for interactions between the gauge fields.

The short range experimentally observed in the weak interaction necessitates the presence of massive gauge bosons. The naive introduction of a mass term in the Lagrangian is of the form:

$$\mathcal{L} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}m^2 A^\nu A_\nu. \quad (1.26)$$

It leads to a violation of the gauge invariance, which allows the theory to be perturbative and renormalizable.

1.3 The Higgs mechanism

1.3.1 Spontaneous symmetry breaking

Let us consider now a field theory comprising N scalar massless fields

$\vec{\phi}(x) = (\phi_1(x), \dots, \phi_N(x))$, and a Lagrangian of the form:

$$\mathcal{L}_\sigma = \frac{1}{2} \partial_\mu \vec{\phi}(x) \partial^\mu \vec{\phi}(x) - V(|\vec{\phi}(x)|). \quad (1.27)$$

The Lagrangian exhibits a global invariance under the field rotations $\phi_i'(x) = R_i^j \phi_j(x)$ which can be represented by the $O(N)$ group.

We now choose a potential of the type:

$$V(|\vec{\phi}(x)|) = -\frac{1}{2} \mu^2 |\vec{\phi}(x)|^2 + \frac{\lambda}{4} (|\vec{\phi}(x)|^2)^2, \quad (1.28)$$

with both λ and μ positive. It must be noted that with $\mu > 0$ the quadratic term has the wrong sign to be considered a mass for the field, and therefore the potential interpretation is consistent.

The perturbative expansion of the theory has to be performed around the classical vacuum configuration, which is the minimum of the potential. In this case the minimum is independent from the field configuration and is $|\vec{\phi}(x)|^2 = \mu^2/\lambda \neq 0$, which is degenerate and symmetric with respect to $O(N)$.

We can arbitrarily choose a vacuum configuration and parametrize the fields accordingly:

$$\phi_0^i = (0, \dots, v), \quad v = \frac{\mu}{\sqrt{\lambda}}, \quad \phi^i(x) = (\pi^k(x), v + \sigma(x)), \quad (1.29)$$

where $\pi^k(x)$ and $\sigma(x)$ are the transformed fields and the index k runs over $N - 1$ elements.

Substituting the new fields into the Lagrangian it is possible to obtain several components:

$$\begin{aligned} \mathcal{L}_{kin} &= \frac{1}{2} \partial_\mu \pi^k \partial^\mu \pi_k + \frac{1}{2} \partial_\mu \sigma \partial^\mu \sigma \\ \mathcal{L}_{\sigma mass} &= \frac{1}{2} \mu^2 \sigma^2 - \frac{6}{4} \lambda v^2 \sigma^2 = -\mu^2 \sigma^2 = -\frac{1}{2} (2\mu^2) \sigma^2 \\ \mathcal{L}_{\pi mass} &= 0 \\ \mathcal{L}_{int} &= -\frac{\lambda}{4} (\pi^k \pi_k)^2 - \frac{\lambda}{4} \sigma^4 - \frac{\lambda}{2} \sigma^2 \pi^k \pi_k - \sqrt{\lambda} \mu \sigma \pi^k \pi_k - \sqrt{\lambda} \mu \sigma^3 \\ \mathcal{L}_{vac} &= \frac{\mu^4}{4\lambda}. \end{aligned} \quad (1.30)$$

In this final Lagrangian the initial $O(N)$ symmetry is broken, as it is evident from the different couplings that the σ boson has with respect to the π bosons. A residual $O(N-1)$ symmetry is still present among the π^k fields. We can assign to the different fragments of the Lagrangian special meanings, related to the couplings:

- $\mathcal{L}_{kin}$ represents the kinetic term for N scalar bosons, as in the beginning of our model;
- $\mathcal{L}_{\sigma mass}$ arise after the spontaneous breaking of the $O(N)$ symmetry and represents a mass term for the σ boson. No mass term is assigned to the π fields, though;
- $\mathcal{L}_{int}$ contains triple and quartic couplings among the fields, in which σ gains a special status, breaking the initial $O(N)$ symmetry;
- $\mathcal{L}_{vac}$ represents the energy of the vacuum, the vacuum expectation value of the σ boson. Classically this term can be neglected, as only energy differences are physical observables.

This example can be generalized in the *Goldstone theorem* [12], which states that a theory containing K global symmetry generators, out of which d are broken by the choice of the potential and therefore of the vacuum, $K - d$ massless bosons arise, called Goldstone bosons.

In our case the initial $O(N)$ symmetry has $N(N - 1)/2$ generators of the symmetry, while the final $O(N-1)$ one has $(N - 1)(N - 2)/2$. The $N - 1$ broken generators protect the π fields from becoming massive.

This mechanism, called spontaneous symmetry breaking, is valid both in quantum field theory and in quantum mechanics, where it has significant impacts in the description of superconductivity [13].

1.3.2 The Higgs mechanism

The previously described mechanism has been studied in the context of a gauge field theory by several authors, with Higgs and Englert among them, reporting their results in 1964 [14, 15]. The combination of these studies with the $SU(2) \times U(1)$ formalism led to the electroweak theory by Weinberg and Salam in 1967.

If we now take, in the framework of the electroweak theory, a complex doublet under $SU(2)$:

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad (1.31)$$

and choose a potential of the kind:

$$V(\phi) = -\mu^2 |\phi|^2 + \lambda (|\phi|^2)^2, \quad (1.32)$$

with both $\lambda > 0$ and $\mu > 0$, we find that the minimum of the potential occurs at:

$$\phi^\dagger \phi = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \quad (1.33)$$

We can therefore define the vacuum state as:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \quad (1.34)$$

where $|v_1|^2 + |v_2|^2 = v^2$.

If we now impose that the vacuum state is neutral, i.e. $Q\phi_0 = 0$:

$$Q\phi_0 = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{2} + \frac{Y}{2} & 0 \\ 0 & -\frac{1}{2} + \frac{Y}{2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (1.35)$$

This leads to:

$$(1 + Y)v_1 = (1 - Y)v_2 = 0 \quad (1.36)$$

and we obtain two allowed field configurations:

- $v_1 = 0, v_2 = v, Y = 1$
- $v_1 = v, v_2 = 0, Y = -1$.

In the former case ϕ_1 is positive and ϕ_2 is neutral, while in the latter ϕ_1 is neutral and ϕ_2 has negative electric charge. It can be shown that the two configurations are equivalent.

We choose as a convention the first option and we decompose the ϕ_i fields in their real and imaginary components:

$$\phi(x) = \begin{pmatrix} \phi^{(+)}(x) \\ \phi^{(0)}(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \pi_1(x) + i\pi_2(x) \\ v + H(x) + i\pi_3(x) \end{pmatrix}, \quad (1.37)$$

where, as in the previous example, the π fields are the Goldstone bosons. To allow a proper renormalization, these fields have to be kept in the Lagrangian, but for our purpose we can apply the unitary gauge conditions to obtain:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (1.38)$$

where H denotes the Higgs boson field.

After this transformation the potential becomes:

$$V(\phi) = -\frac{\mu^4}{4\lambda} + \mu^2 H^2 + \lambda v H^3 + 2\lambda v^2. \quad (1.39)$$

From the potential a mass term for the Higgs boson arises: $m_H^2 = 2\mu^2 = 2\lambda v^2$. There are additional triple and quartic Higgs boson self-couplings that appear in the potential description. As previously mentioned, the additional $2\lambda v^2 = m_H^2$ term can be neglected when measuring physical observables. This term, though, gains importance once a coupling with gravity is introduced and it can be linked to the cosmological constant Λ of the theory of general relativity.

The importance of the spontaneous symmetry breaking mechanism becomes apparent when we compute the interaction Lagrangian for the ϕ field:

$$\mathcal{L}_{int} = (\mathcal{D}_\mu \phi)^\dagger \mathcal{D}^\mu \phi, \quad (1.40)$$

where

$$\begin{aligned} \mathcal{D}_\mu \phi &= (\partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i - ig' \frac{Y}{2} B_\mu) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \partial_\mu H(x) \end{pmatrix} - \frac{i}{2\sqrt{2}} (v + H(x)) \begin{pmatrix} g(W_{1\mu} - iW_{2\mu}) \\ -gW_{3\mu} + g'B_\mu \end{pmatrix}. \end{aligned} \quad (1.41)$$

Using the relations described in Eq. 1.24, 1.19, and 1.16 we can rewrite the spinor as function of the physical weak vector bosons:

$$\begin{pmatrix} g(W_{1\mu} - iW_{2\mu}) \\ -gW_{3\mu} + g'B_\mu \end{pmatrix} = \begin{pmatrix} \sqrt{2}gW_\mu^+ \\ -\sqrt{g^2 + g'^2}Z_\mu \end{pmatrix}. \quad (1.42)$$

The interaction Lagrangian becomes:

$$\mathcal{L}_{int} = (\mathcal{D}_\mu \phi)^\dagger \mathcal{D}^\mu \phi = \frac{1}{2} \partial_\mu H \partial^\mu H + (v + H)^2 \left[\frac{g^2}{4} W_\mu^+ W^{\mu-} + \frac{1}{8} (g^2 + g'^2) Z^\mu Z_\mu \right], \quad (1.43)$$

in which quadratic terms of the W and Z fields appear with the correct sign and are interpreted as a mass term.

In summary: with the Higgs mechanism it is possible to introduce a mass term for the gauge bosons without breaking the underlying gauge symmetry of the Lagrangian.

A common representation of this method arises from the computation of the helicity states of the gauge bosons: massless bosons have only two possible helicity states, while massive bosons have three. The Goldstone bosons can be viewed as the additional degree of freedom that is absorbed by the weak bosons to become massive.

The mass of the weak bosons is determined by their couplings to the Higgs boson:

$$M_W^2 = \frac{1}{4}g^2v^2, \quad M_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2. \quad (1.44)$$

Additionally, the following relation between the Z and W bosons masses holds:

$$\frac{M_W^2}{M_Z^2} = \frac{g^2}{g^2 + g'^2} = \cos^2 \theta_W. \quad (1.45)$$

1.3.3 Couplings to gauge bosons

The interaction Lagrangian contains additional terms describing the interaction of the Higgs boson with the weak gauge bosons. No direct coupling with the photon is predicted. The theory includes interaction vertices with two gauge bosons and one or two Higgs bosons. In addition, the vertices do not mix W and Z bosons. The coupling constants of these vertices are proportional to the mass of the gauge bosons that participate in the interaction and can be written as:

$$g_{HVV} = \frac{2m_V^2}{v}, \quad g_{HHVV} = \frac{2m_V^2}{v^2}, \quad (1.46)$$

where V can be either the W or the Z boson, and v is the vacuum expectation value of the Higgs field which can be derived from the Fermi constant by the relation $v = (\sqrt{2}G_F)^{-1/2} \approx 246$ GeV.

1.3.4 Couplings to fermions

The introduction of a mass term for fermions as in the Dirac equation:

$$\mathcal{L}_{m,Dirac} = -m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L), \quad (1.47)$$

would break the SU(2) gauge invariance of the Lagrangian as it directly mixes left- and right-handed components of the fermionic fields, which transform differently under SU(2). To introduce the mass term for the fermions without breaking the gauge invariance it is possible to resort again to the spontaneous symmetry breaking mechanism and the ϕ field. A *Yukawa*-type coupling ²:

$$\mathcal{L}_{Yuk} = -g_f(L_f\phi d_R^f + d_R^f\phi^\dagger L_f), \quad (1.48)$$

²This coupling was initially introduced by Yukawa [16] to describe the interactions between hadrons, postulating a new boson mediator, the *pion*

where d_R^f it's the right-handed singlet, respects the gauge invariance and can produce the desired mass term, which becomes:

$$m_f = \frac{g_f v}{\sqrt{2}}. \quad (1.49)$$

In addition to the masses of the different fermions, a term coupling the fermions with the Higgs boson appears, with a strength proportional to the fermion mass:

$$g_{Hff} = i \frac{m_f}{v}. \quad (1.50)$$

This procedure gives mass to the down-type fermions. The up-type fermions attain their masses from the coupling to the charge conjugate Higgs field:

$$\phi_C = i\sigma_2\phi^* = \frac{1}{\sqrt{2}} = \begin{pmatrix} v + H \\ 0 \end{pmatrix}. \quad (1.51)$$

As mentioned previously, the mass eigenstates are different from the eigenstates of the weak interaction. This difference is accounted for introducing a mixing matrix, the CKM matrix.

1.4 The Higgs boson at the LHC

1.4.1 Higgs production and decays

The Large Hadron Collider (LHC) is the largest and most advanced particle accelerator today. It is located at the Center for the European nuclear research, CERN (Geneva, Switzerland). The colliding center-of-mass energy of the two proton beams in the facility has been $\sqrt{s} = 7$ TeV in 2011 and $\sqrt{s} = 8$ TeV in 2012. A more detailed description of the accelerator and the experimental set-up is available in Chapter 2.

Protons are not elementary “point-like” particles, but are made of three valence quarks, two up quarks and a down one, bound by gluons. In addition to the valence quarks, the quantum vacuum and the binding energy among the quarks produce around them a cloud of virtual quarks and gluons that can be probed by a high energy collision process. The exact quark composition of the proton is impossible to compute from theoretical assumptions as the QCD is not perturbative at that energy scale. The momentum carried by each constituent, valence or virtual, is a continuous function that has to be measured experimentally.

The probability for a quark or a gluon to be observed within a proton, with a certain momentum $p = x \cdot p_{proton}$, with $x \in [0, 1]$, is embedded in the parton distribution func-

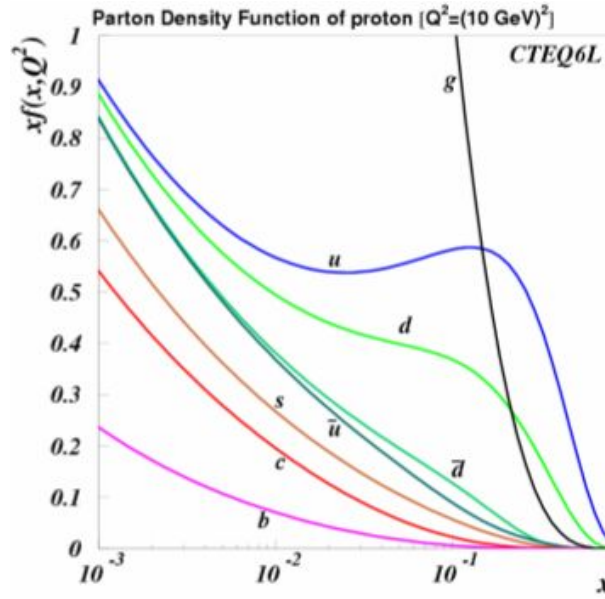


Figure 1.1: Parton distribution function of the proton as a function of the Bjorken variable x , for a $Q^2 = 10 \text{ GeV}^2$.

tions, measured experimentally. Figure 1.1 shows an example of a parton distribution function of the proton [17].

As shown in Figure 1.1, the majority of the partons present in the protons are gluons, especially at low x . This effects directly translate in a dominance of the gluon-fusion process in the Higgs boson production at the LHC. The Feynman diagram showing the gluon fusion process can be found in Figure 1.2a. The loop inside the gluon fusion process is dominated by the top quark, since it has the largest mass in the standard model, but may receive additional contribution from new physics.

The second largest contribution to the total Higgs production cross section comes from the vector boson fusion (VBF) process (Figure 1.2b) in which two quarks radiate two weak bosons that then interact producing a Higgs boson. The scattered quarks produce two jets in the forward/backward direction that can be used to experimentally tag the events.

The third most significant production mode is the associated production with a vector boson (VH, Figure 1.2c), also known as *Higgsstrahlung* process. The work presented in this thesis is focused on this production mechanism. Even though the production cross section is significantly lower than the previous two processes, the presence of a vector boson reduces greatly the background contamination.

The most rare Higgs production process is $t\bar{t}H$, the production in association with top quarks (Figure 1.2d).

The cross sections of the four production processes depends on the mass of the Higgs

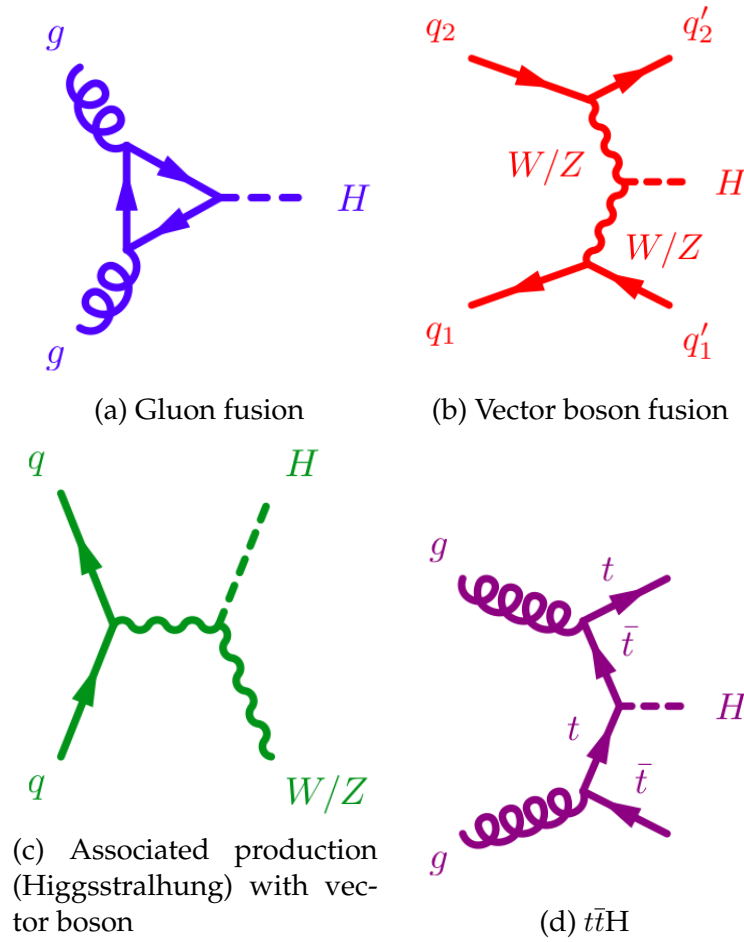


Figure 1.2: Tree-level Feynman diagrams for the four Higgs boson production mechanisms relevant at LHC.

boson, a free parameter of the SM, and on the center-of-mass energy of the two colliding beams. Figure 1.3 shows the production cross sections as a function of the Higgs mass at 7 and 8 TeV.

Even though the couplings of the Higgs boson are dictated only by the mass of the fundamental particles, the branching ratio of the Higgs boson to fundamental particles is also dependent on the Higgs mass, as a phase-space factor is added. The dependence of the branching fractions for the main Higgs decay modes as a function of its mass are shown in Figure 1.4.

Even though the Higgs boson has no direct coupling to the photon, a sizeable branching ratio is given by loop contributions. These loops, as in case of the gluon fusion, could contain contributions from new physics, making this process a viable channel to probe theories beyond the standard model. The branching fraction to down-type fermions, may be sensitive to several beyond the Standard Model (BSM) models, in which these

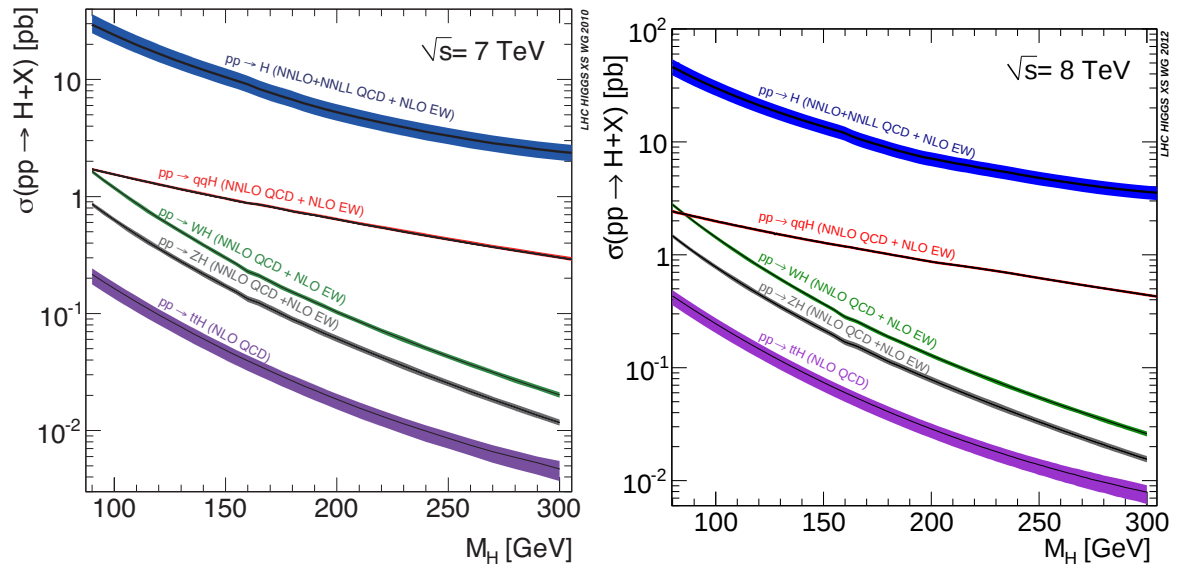


Figure 1.3: Higgs production cross section at 7 TeV (left) and 8 TeV (right). The bands represent the theoretical uncertainties [18].

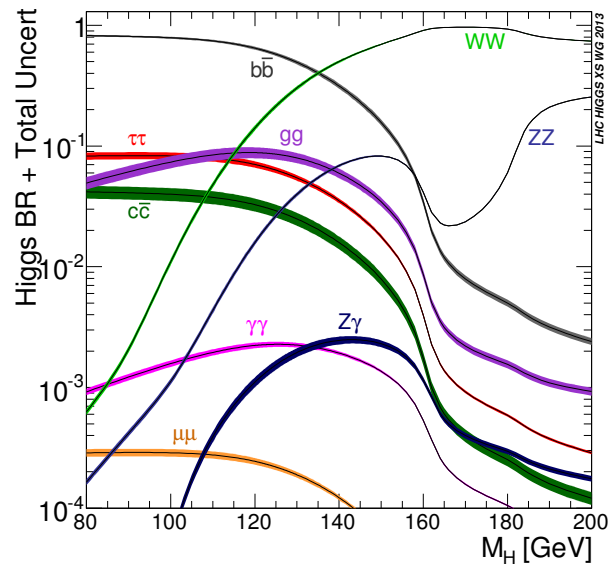


Figure 1.4: Higgs branching ratios for different decay channels as function of the Higgs mass [18].

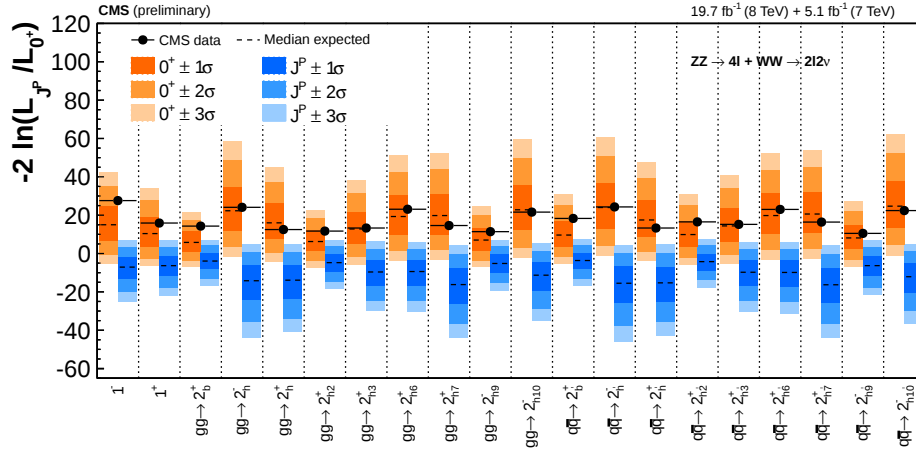


Figure 1.5: Summary of different J^P hypotheses tested against the 0^+ one using ZZ and WW Higgs candidate decays. The orange and blue bands represent the 1σ , 2σ , and 3σ around the median expected value for the SM Higgs boson hypothesis and the alternative hypothesis, respectively. The black points represent the observed values.

branching fractions are either enhanced or suppressed.

1.5 Higgs boson searches

On the July 4th 2012, the CMS and ATLAS collaborations reported the observation of a new resonance at the mass of 125 GeV compatible with the SM Higgs boson [19]. This observation was based on combining the results of different searches for the Higgs boson in decay channels with gauge bosons, ZZ^* , WW^* and $\gamma\gamma$.

Currently, the significance of the Higgs boson signal exceeds 5σ for $H \rightarrow \gamma\gamma$ searches in both CMS and ATLAS [20, 21]. A similar situation is also present in the ZZ^* and WW^* channels.

The large amount of signal events in these decay channels has allowed to perform initial studies of the resonance properties. The high invariant mass resolution of $\gamma\gamma$ and ZZ channels has allowed to measure the mass of the boson with high precision: $m_H = 125.36 \pm 0.37$ (stat) ± 0.18 (syst) GeV and $m_H = 125.03^{+0.26}_{-0.27}$ (stat) $^{+0.13}_{-0.15}$ (syst) = $125.03^{+0.29}_{-0.31}$ (tot) GeV for ATLAS [22] and CMS [23], respectively.

The spin and parity (J^P) of the resonance has also been studied in the ZZ and WW channels, which are more sensitive to these physical observables. The results, shown in Figure 1.5, exclude any concurrent hypothesis at more than 99% confidence level (CL), leaving only 0^+ as an option [24].

The expected SM Higgs boson width for $m_H \approx 125$ GeV is $\Gamma_H \approx 4$ MeV, which is below

the experimental mass resolution. It has been pointed out that the ratio of on-shell and off-shell Higgs boson production cross section is proportional to the resonance width [25]; this effect is enhanced in the diboson decay channels by the interference term with the non resonant production. By simultaneously fitting the two contributions it has been possible to measure with better precision the 95% CL upper limit on the width obtaining $\Gamma_{obs}/\Gamma_H < 5.3$ for CMS [26] and $\Gamma_{obs}/\Gamma_H < 5.7$ for ATLAS [27]. This method has mild model assumptions, as it neglects the contribution of new physics in the backgrounds and in the gluon-fusion loop.

The search for Higgs boson decays to bottom quark pairs is the only feasible search to down-type quarks at hadronic colliders, as the coupling to the other two quarks are largely suppressed by their mass while the hadronic background increases dramatically. The large background contamination from heavy-flavor QCD production mandates to search for $H \rightarrow b\bar{b}$ in the experimentally clean production processes such as the associated production with a vector boson, the VBF production, or the $t\bar{t}H$ associated production.

The combination of the CDF and D0 results [28] show a 3.1σ evidence of the $H \rightarrow b\bar{b}$ decay mode. The CMS experiment reported a 2.1σ excess in the VH, $H \rightarrow b\bar{b}$ production [29], compatible with the SM predictions. The equivalent analysis for ATLAS does not show any excess [30], but is also less sensitive.

The branching fraction of the Higgs boson to muon pairs is predicted to be very low, therefore any excess might be a strong indication of new physics. Both ATLAS [31] and CMS [32] have performed a search for $H \rightarrow \mu\bar{\mu}$ decays, exploiting the extremely good dimuon mass resolution, without finding any significant excess. The analyses set an independent upper limit on the branching ratio of approximately 7 times the SM expectation.

The SM Higgs coupling to top quarks can only be observed through the $t\bar{t}H$ associated production, since the Higgs boson mass forbids the decay of the Higgs into top quark pairs. The search for $t\bar{t}H$ [33, 23] is performed in many Higgs final states. Both ATLAS and CMS exploit the $b\bar{b}$ and $\gamma\gamma$ final states; additionally CMS searches for this production mode also in multi-lepton final states, targeting the $\tau\tau$, ZZ , and WW Higgs decay modes.

The combination of these $t\bar{t}H$ searches leads in CMS to a 3.5σ excess with respect to the background-only hypothesis [23]. The combined *signal strength* (μ), i.e. the ratio $\mu = (\sigma \times BR)_{obs}/(\sigma \times BR)_{SM}$, is measured to be $\mu = 2.76^{+1.05}_{-0.92}$, which is two standard deviations away from the SM expectation ($\mu = 1$). The excess in this combination is driven by the same-sign dimuon analysis, which searches for $H \rightarrow WW$. The results from ATLAS are consistent with the background-only hypothesis [33].

The $H \rightarrow \tau\tau$ decay mode is searched by CMS in gluon fusion, VBF and associated production [34]. Events are selected in real time by dedicated triggers and according to the isolation of the final state leptons. Events are separated into categories tailored around specific production mechanisms (VBF, VH, gluon fusion) and, in the gluon fusion case,

according to the boost of the hadronic tau and Higgs candidates. The invariant mass spectrum of the Higgs final states is smeared due to the presence of neutrinos in the tau lepton decay. To recover this effect, a dedicated likelihood integration method, called SVFit [35], calculates the di-tau invariant mass taking as input the momenta of the visible Higgs products plus the information on the reconstructed missing transverse energy (E_T). The mass resolution achieved by SVFit varies between 10% and 20%, depending on the final state and the category.

Most of the backgrounds present in the direct production (VBF and gluon fusion) categories are modeled using simulated events with dedicated sideband regions to cross-check the simulation and apply correction factors where needed. The main background for these categories is given by $Z \rightarrow \tau\tau$ events which are modeled with the embedding techniques [36]: real $Z \rightarrow \mu\mu$ events are selected from collision data and the two muons are replaced with simulated tau decays.

The search for the Higgs boson in the associated production process is performed in final states including both a Z and a W boson. These searches exploit the presence of the additional leptons in the final state to suppress the backgrounds. The remaining backgrounds are estimated with MC simulation or with the misidentification rate method. Part of the work presented in this thesis has been included in the final result published by the collaboration, but additional effort has been put into estimating the uncertainties in one of the major sources of backgrounds.

A combined fit to all the channels and categories of the CMS $H \rightarrow \tau\tau$ analysis reveals a 3.2σ deviation from the background-only hypothesis and a combined $\mu = 0.78 \pm 0.27$. The best-fit mass, $m_H = 122 \pm 7$ GeV, is compatible with the Higgs mass measured in the ZZ and $\gamma\gamma$ decay channels. The equivalent ATLAS analysis [37], which lacks the associated production search, observes a 4.1σ deviation from the background-only hypothesis and measures $\mu = 1.4^{+0.5}_{-0.4}$, which is slightly in excess, but still compatible, with respect to the SM value.

The CMS $H \rightarrow b\bar{b}$ and $H \rightarrow \tau\tau$ searches have been combined leading to a 3.8σ evidence of Higgs coupling to down-type fermions [38].

The result of the single and combined signal strengths for both ATLAS [21] and CMS [23] is presented in Figure 1.6. The outcome of the various searches have been combined by the single experiments in a simultaneous fit in order to determine the couplings of the Higgs boson to the different types of particles.

Another way to interpret the results is to compute the couplings of the Higgs boson to gauge bosons (k_V) and to fermions (k_F) and to plot the one-sigma contour regions allowed by each measurement. This interpretation assumes the absence of new physics in loop processes to infer the two couplings from the measurements (e.g. assumes that the gluon fusion is almost completely mediated by a top loop). This interpretation of the results is available in Figure 1.7 for both the experiments.

All the results reported so far by both ATLAS and CMS are compatible with the inter-

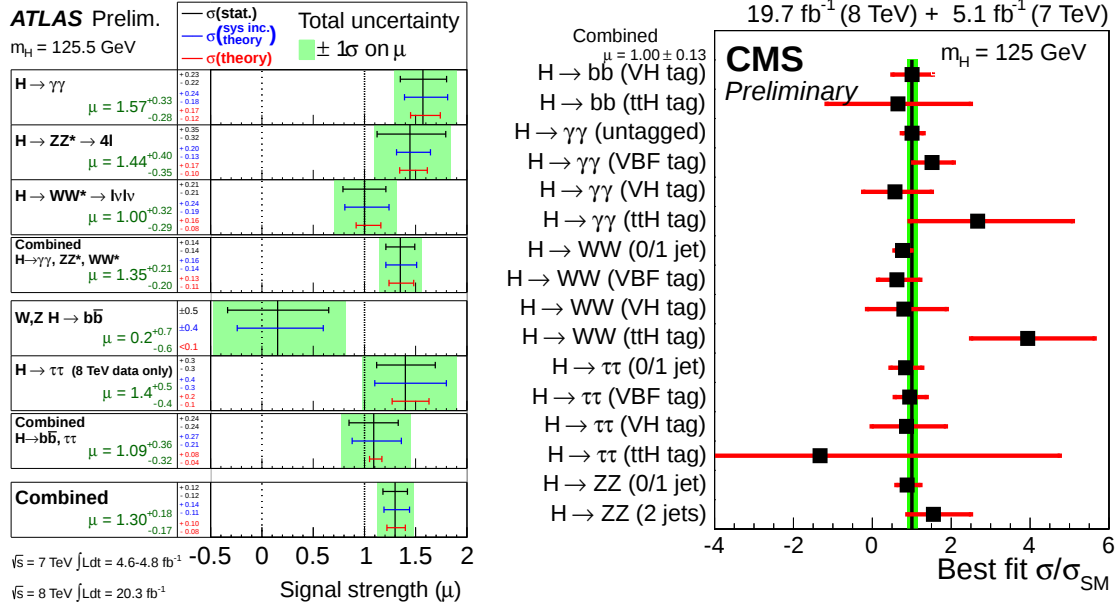


Figure 1.6: Overview of the different Higgs boson searches performed by ATLAS (left) and CMS (right) with the respective signal strength measured by each channel. The combined result is $\mu = 1.30^{+0.18}_{-0.17}$ for ATLAS and $\mu = 1.00 \pm 0.13$ for CMS.

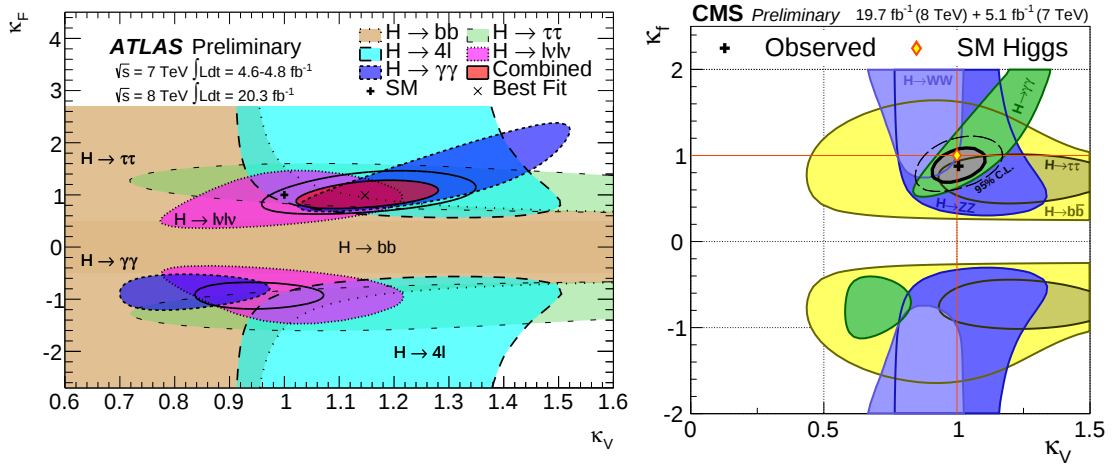


Figure 1.7: Allowed regions in the $k_V \times k_F$ space as measured by the different Higgs boson searches at ATLAS (left) and CMS (right). The colored contours represent the one standard deviation confidence intervals measured by the different searches and by their combinations. The best fit value of the couplings is marked by the symbol $\times$ (left) and by the black cross (right). The Standard Model point is located at (1, 1) and marked as a black cross on the left and as a yellow diamond on the right.

pretation of the new resonance being a purely SM Higgs boson, but the accuracy of the measurements still allows room for new physics processes.

Chapter 2

The Large Hadron Collider and the CMS experiment

The work presented in this thesis is based on data collected by the Compact Muon Solenoid (CMS) experiment, one of the four main experiments studying the collisions provided by the Large Hadron Collider (LHC). A detailed description of the LHC and of the CMS experiment can be found in [39] and [40], respectively. In this chapter we summarize the main features that are relevant for this work.

2.1 The Large Hadron Collider

The LHC and its experiments were built to explore the high-energy frontier of particle physics and to address fundamental questions of this field, such as the existence of a Higgs boson [14, 15, 41, 42, 43, 44], of extended symmetries [45], extra dimensions [46], or new elementary particles [47, 48]. All these phenomena do not have a well-defined energy range predicted by theory but they are expected to manifest at the TeV scale in case they exist in nature. A proton-proton collider was considered to be the most suitable machine for such a task, allowing higher energies with current technologies and probing a wide energy range.

The LHC is housed in the 27 km long underground tunnel where the Large Electron Positron collider (LEP) has been operating until its decommissioning in 2000. The tunnel is located under both French and Swiss territory, in proximity of the CERN research facility. Protons produced from hydrogen ionization are accelerated by a chain of older accelerators, some of them dating back to the late 1950's, before entering the LHC. A schematic view of the CERN accelerators and their connection is shown in Figure 2.1. From the last element of this chain, the Super Proton Synchrotron (SPS), protons are injected in the LHC with an energy of 450 GeV in two separate beam pipes, one containing protons running clockwise, the other with protons running in the opposite

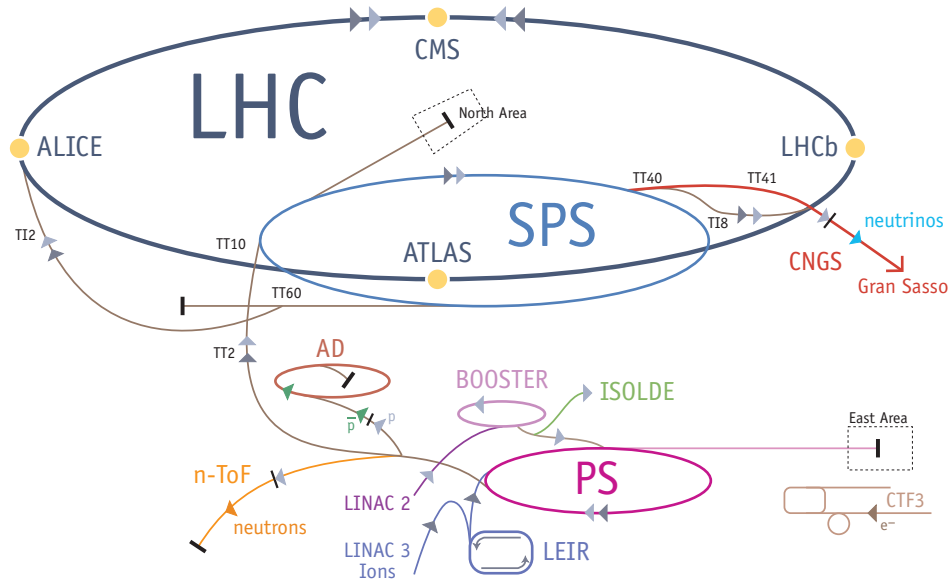


Figure 2.1: Schematic view of the CERN accelerators and their connection

direction. Protons are bent in their trajectory by 1232 superconducting dipole magnets and focused by superconducting quadrupole magnets, while 16 superconducting radio-frequency (RF) stations accelerate the two beams up to 7 TeV in steps of 0.5 MeV each turn. In order to maintain their superconducting properties the magnet coils and the cavities are cooled to a temperature of 1.9 K by a complex cryogenic system using superfluid helium as refrigerator. RF cavities can be used for accelerating the protons only if the beam is not continuous but structured in bunches. By design the LHC is built to contain 2808 bunches of 10^{11} protons each, giving a bunch time separation of 25 ns.

The LHC started its operations on September 10, 2008 and after only 9 days of operation a severe *quenching* of about 100 dipole magnets, causing the release of around two tons of helium, forced the machine to stop and to address some design failures. The main cause of the accident was found to be in some of the electrical connections between magnets. In 2009 the machine became operational again, with a reduced beam energy, in order to reduce the current flowing in the dipole magnets. The year 2010, after a careful 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, half of the design value. During 2010 and 2011 the machine commissioning continued along with the data taking. During this time the instantaneous luminosity increased continuously. The beam energy was raised to 4 TeV, corresponding to $\sqrt{s} = 8$ TeV center-of-mass energy, in 2012. The LHC design specifications and the performance achieved so far are summarized in Table 2.1. The LHC is scheduled to start a new data-taking run in 2015 at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

Table 2.1: Relevant LHC machine parameters. The design values are compared to the ones reached during the 2013 operations.

Parameter	Design value	Best value achieved
Beam energy	7 TeV	4 TeV
Number of protons per bunch	1.15×10^{11}	1.5×10^{11}
Number of bunches	2808	1368
Crossing angle	300 μm	290 μm
Beam size	17 μm	20 μm
Emittance	3.75 μm	2.4 μm
Peak luminosity	$10^{34} \text{ cm}^{-2}\text{s}^{-1}$	$7.5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$

The instantaneous luminosity, $\mathcal{L}(t)$, is the number of particles per unit area per unit time available for collisions. For any given physics process the average number of events is given by:

$$N_{ev} = \sigma \int \mathcal{L}(t) dt, \quad (2.1)$$

where σ represents the cross-section for the studied process. While the cross section depends on the process, the instantaneous luminosity is entirely derived from accelerator parameters:

$$\mathcal{L}(t) = \frac{N_p^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F \quad (2.2)$$

where N_p and n_b are the number of protons per bunch and the total number of bunches, respectively, f_{rev} is the rotation frequency, ϵ_n and β^* describe the beam focusing at the interaction point, $\gamma = E_p/m_p$ is the Lorentz factor of the protons in the beam and F accounts for the crossing angle between the two beams. Although the number of bunches has always been less than half the design value during all the running period, the peak instantaneous luminosity has been only 30% lower than the design one. This result was achieved by increasing the number of protons in each bunch and increasing the focusing of the beams. The price to pay for the high luminosity, in particular in the second half of 2011 and in 2012, was an increase of the average number of proton collisions per bunch crossing, the so-called *pileup*. This effect is displayed in Figure 2.2 where the peak number of simultaneous interactions per bunch crossing recorded by the CMS detector is shown as a function of time.

The integrated luminosity is usually quoted as $\mathcal{L}$. During its operations the LHC delivered 44.2 pb^{-1} in 2010, 6.1 fb^{-1} in 2011 and 23.3 fb^{-1} at 8 TeV in 2012. The progress of the accelerator in terms of integrated luminosity is shown in Figure 2.3.

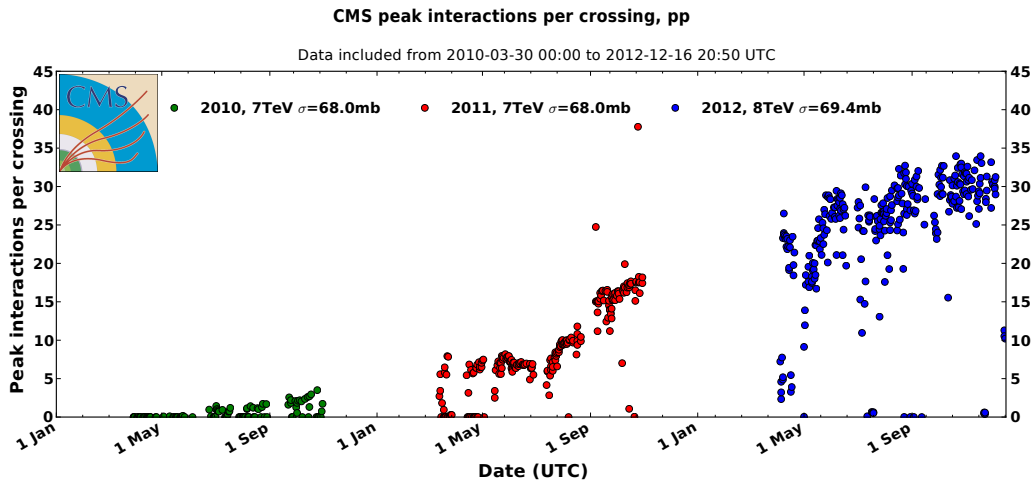


Figure 2.2: Peak Interactions per crossing versus time for p-p collisions (includes special runs), each point represents a fill. This is shown separately for the 2010 (green), the 2011 (red) and the 2012 (blue) data-taking periods.

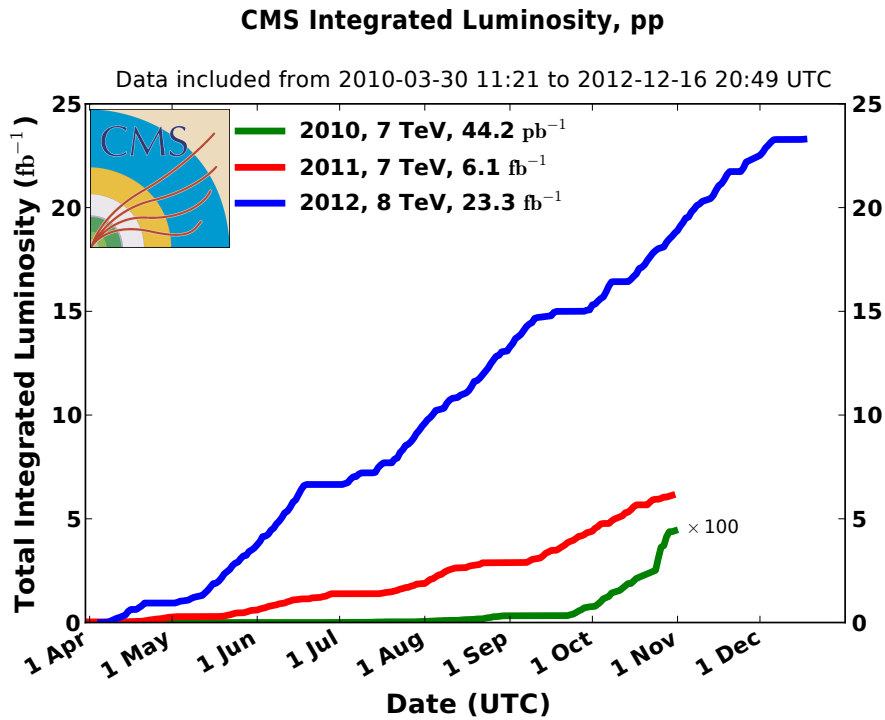


Figure 2.3: Integrated luminosity delivered to CMS during stable beams and for p-p collisions as a function of time, shown separately for the 2010 (green), the 2011 (red) and the 2012 (blue) data-taking periods.

2.2 The CMS detector

The CMS detector is located in the LHC tunnel at point 5, near the French town of Cessy, between the Jura mountains and the Geneva lake.

The CMS experiment, together with ATLAS, constitutes one of the two multi-purpose detectors operating at the LHC. Its primary task is to probe particle physics at the TeV scale searching for new phenomena as well as for the Higgs boson, the only missing piece to the SM puzzle. The experiment is also well suited to perform precision measurements of standard model processes and flavor physics studies. A heavy-ion program is carried on as well to probe QCD at very high energies and matter densities, trying to reproduce an environment similar to the conditions of the universe few instants after the Big Bang.

To carry out such ambitious research program the detector was designed to meet some baseline requirements:

- Good muon identification and momentum resolution over a wide range of momenta and angles with di-muon mass resolution of 1% at 100 GeV.
- Ability to identify the muon charge without any ambiguity for muons momenta below 1 TeV
- Good charged-particle momentum resolution, high tracking efficiency and resolution. These characteristics are especially important for objects like b -jets and tau leptons, where isolated charged hadrons and displaced vertices play a fundamental role. A high tracking resolution also plays a key role in assigning the tracks to the production vertex, mitigating the effect of pileup.
- Good electromagnetic energy resolution, with a di-photon invariant mass resolution of $\sim 1\%$ at 100 GeV and wide geometric coverage with efficient photon and lepton isolation in high pileup conditions.
- Hermetic hadronic calorimeter with fine transverse segmentation for good jet mass and missing transverse energy (E_T^{miss}) resolution.

The total proton-proton cross section at 14 TeV is expected to be roughly 100 mb, leading to an average of 20 pileup interactions with LHC design values. This value has been exceeded during the 2012 data taking period already. The number of *pileup* events together with the short 25 ns bunch spacing pose stringent requirements on the resolution, granularity and latency of the different sub-detectors. The ability to resolve overlapping vertices and assign their respective track is of primary importance. A fast trigger system is also required to reduce the event rate from the design collision rate of 40 MHz to 300Hz, the maximum rate foreseen for storing events permanently.

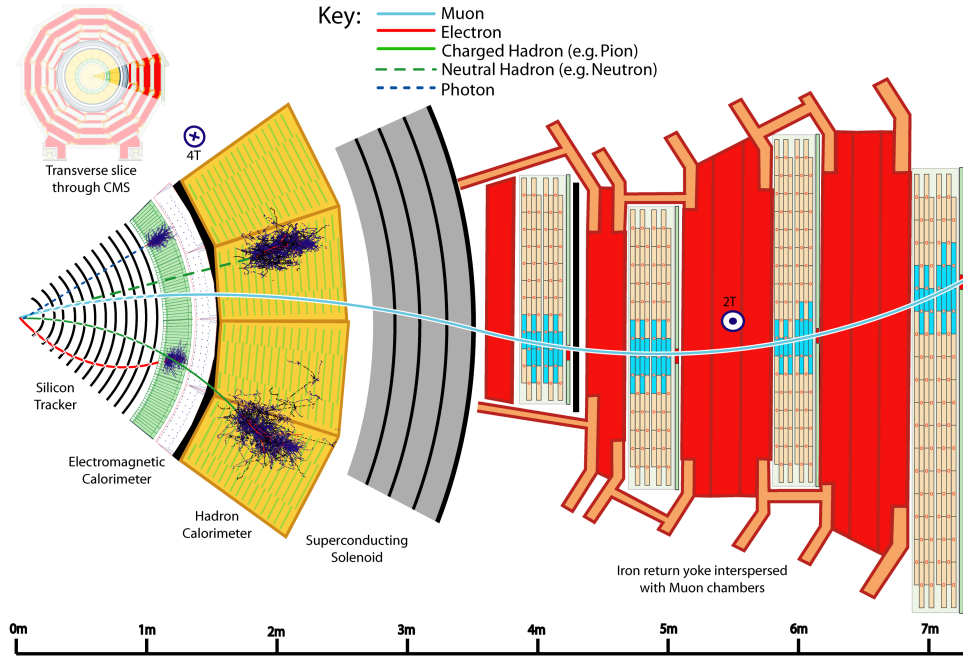


Figure 2.4: Transverse slice of the CMS barrel section, showing the trajectory of different types of particles crossing the detector and the typical signal they leave in each sub-detector.

In order to meet all these requirements CMS has been built with a 4 T NbTi superconducting solenoid magnet with an inner diameter of 6 m. Inside the magnet a large silicon tracker, the largest of its kind, is housed to track charged particles with the required resolution. Around the tracker, but still within the solenoid field, a lead tungstate electromagnetic calorimeter (ECAL) and a brass-scintillating sampling hadron calorimeter (HCAL) are installed. Inside the 1.5 m thick iron return yoke of the magnet four muon stations are installed. They consist of several layers of drift tubes or cathode strip chambers complemented by resistive plate chambers. A schematic view of a transverse slice of the CMS detector is presented in Figure 2.4

The coordinate system chosen by CMS has the y axis pointing vertically upward and the x axis pointing horizontally towards the center of LHC. The z axis is oriented along the beam line, pointing in the direction of the beam that runs anti-clockwise and points towards the Jura. A set of polar coordinates is used to describe the xy plane in the form of radius r and angle ϕ , while the angle θ is measured with respect to the positive z axis direction. Usually the polar angle θ is replaced by the *pseudorapidity*, defined as $\eta = -\ln \tan(\theta/2)$, which is Lorentz invariant for boosts along the z axis and therefore comes very handy when describing processes for which the longitudinal boost is unknown. The three-dimensional angular distance is also replaced by its Lorentz-invariant $\Delta R =$

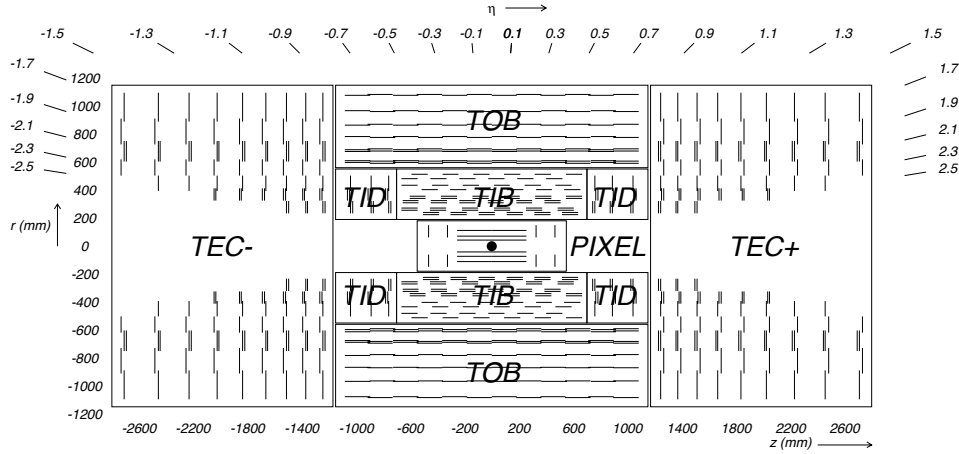


Figure 2.5: Longitudinal cross section of the CMS tracker, with pseudo rapidity coverage of the different detector elements.

$\sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\eta$ and $\Delta\phi$ are the η and ϕ coordinate differences of two points.

2.2.1 Inner Tracker

The inner tracker provides the spatial information needed to reconstruct charged tracks as well as primary and secondary vertices. A schematic view of the inner tracker is shown in Figure 2.5. To cope with the very high track density and to provide the best spatial resolution, the innermost part of the tracker is based on silicon pixel technology. The pixel detector, displayed in Figure 2.6, is composed of three cylindrical layers at distances of 4.4, 7.3 and 10.2 cm with respect to the beam axis complemented by two disks in the forward and backward regions. The total active area of the pixel detector is about of 1 m². It provides bi-dimensional measurements within its angular acceptance: $|\eta| \leq 2.5$. The pixel cell size is $100 \times 150 \mu\text{m}^2$. The solenoidal magnetic field drift spreads the charge deposited in the sensors on several neighboring pixels allowing for a better hit resolution. To enhance this effect in the forward wheels the sensors are placed in a turbine-like geometry. The combination of analogue read-out and usage of the charge sharing information allows to achieve a spatial resolution of 9-33 μm per hit [49].

The pixel detector can be extracted from the rest of the tracker to allow easy access for maintenance without interfering with the rest of the detector. This feature is particularly important due to the high radiation dose that the first layers of the pixel detector sustain, requiring additional maintenance.

The outer layers of the tracker are made of silicon strip detectors and are further divided into four sections: tracker inner and outer barrel (TIB and TOB), tracker inner disks

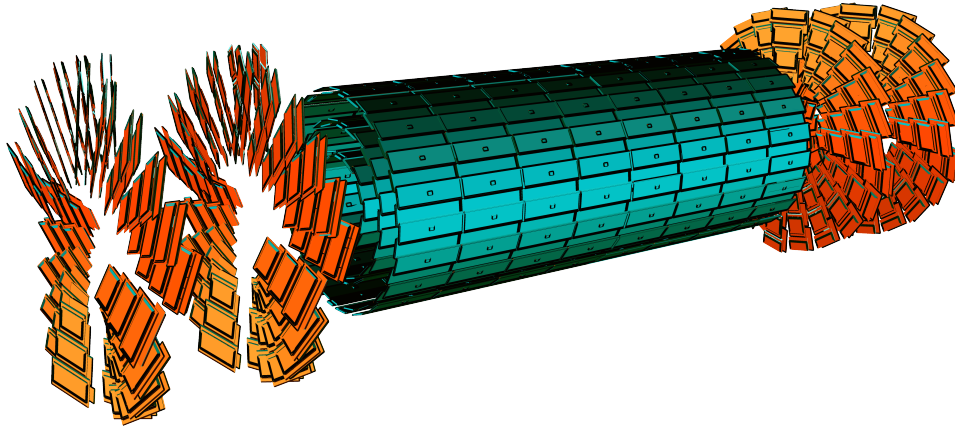


Figure 2.6: Three dimensional sketch of the CMS pixel detector, barrel modules in blue and endcap wheels in orange.

(TID) and tracker endcaps (TEC), as illustrated in Figure 2.5. TIB and TID are located at distances between 20 and 55 cm from the beam line. They consist of four barrel layers and three disks at each end. The sensors are made of $320\ \mu\text{m}$ thick silicon with a strip pitch that varies from $80\ \mu\text{m}$ in the innermost barrel layer to $140\ \mu\text{m}$ in the outermost disks. The combination of TIB and TID delivers up to four transverse measurements. The TIB-TID sections are surrounded by the TOB, which fills the remaining space up to a radius of 116 cm from the beam line. It consists of six barrel layers of $500\ \mu\text{m}$ thick sensors with a strip pitch of $183\ \mu\text{m}$ for the first four layers and $122\ \mu\text{m}$ for the remaining two. The TEC is located in the end-cap region and extends from a radius of 20 cm to 113.5 cm, and $|z| > 118\ \text{cm}$. It consists of nine disks. Each disk is composed by up to seven concentric rings of strip sensors. The thickness and the strip pitch of the sensor vary depending on the distance from the beam line.

Intrinsically strip sensors only provide a one-dimensional spatial information (ϕ). In order to measure a second coordinate (r for disks, z for barrel), an additional set of sensors is mounted back-to-back with a stereo angle of 100 mrad on the first two layers of TIB, TID and TOB and on layers 1, 2 and 5 of TEC. The strip tracker geometry is designed such that particles originating from the nominal interaction point typically cross nine or more layers of sensors, with at least four of them providing two-dimensional information.

2.2.2 Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) surrounds the inner tracker. It is divided into ECAL Barrel (EB), covering the pseudo rapidity range $|\eta| < 1.479$, and ECAL Endcaps (EE), which covers the range $1.479 < |\eta| < 3$. ECAL consist of 68 524 scintillating

lead tungstate (PbWO_4) crystals, the light of which is read out by photomultipliers.

The choice of lead tungstate was driven by its high density yielding a short radiation length (0.89 cm) and small Moliere radius (2.2 cm), together with being radiation hard and yielding fast signals, with 80% of the light yield emitted within 25 ns. To keep the calorimeter as hermetic as possible, the crystals have truncated pyramidal shape. One of the longitudinal faces is left unpolished, to moderate the non uniform light collection across the crystal length that this peculiar shape causes. Each crystal covers about 0.0174×0.0174 in the $\eta - \phi$ plane and is oriented such that its front face points towards the nominal interaction point (with a slight misalignment to mitigate the effect of the crystal surface on photon detection). The crystals are read-out by a pair of avalanche photodiodes (APD) in the barrel and by vacuum phototriodes in the endcap. Both the devices can operate in high magnetic fields with little or no efficiency degradation and exhibit a good radiation hardness.

A pre-shower detector is installed in front of the ECAL endcap, its purpose being to discriminate between single photons and photons from π^0 decays, and to enhance the angular resolution of both electrons and photons. The pre-shower detector consists of a double-layer lead-silicon calorimeter, with the lead initiating the shower and the silicon strip detector placed after each lead layer to measure the deposited energy and shower profile with high granularity.

2.2.3 Hadronic Calorimeter

The hadronic calorimeter (HCAL) serves two purposes: it measures the energy of charged and neutral hadrons from the p-p interaction while stopping them, thus allowing only muons to pass through and avoiding large quantities of energy being deposited inside the superconducting magnet. As most of the other sub-detectors, HCAL is divided in a barrel section (HB), covering the region $|\eta| < 1.3$, and an endcap section (HE), covering the region $1.3 < |\eta| < 3$. Both sections are located inside the superconducting solenoid. HCAL is a sampling calorimeter with brass passive plates, in which the hadronic shower begins and develops, interspaced with active plastic scintillator which measures the shower profile and the energy deposited.

The scintillator is segmented both in η and ϕ to provide the necessary granularity. Each scintillating tile is connected to the readout by a wavelength-shifting fiber that runs in a groove machined into the tile. The fiber guides the light emitted by the scintillator to a hybrid photodiode (HPD). The HPDs consist of a photocathode kept at high voltage. Electrons emitted by the photocathode are accelerated in the short distance (~ 3 mm) that separates the cathode from a silicon pixelated anode which amplifies the signal. These devices were chosen due to their high dynamic range, their high gain ($O(2000)$) and their capability to work in a magnetic field.

The effective thickness of the calorimeter in terms of interaction length, λ_I , ranges in the barrel from a minimum of $5.82 \lambda_I$ at $\eta = 0$ to a maximum of $10.6 \lambda_I$ at $|\eta| = 1.3$. The

ECAL crystals add about 1.1 additional interaction lengths. The total thickness of the endcap calorimeters, including the ECAL crystals, amounts to about $10 \lambda_I$.

Due to the limited stopping power of HB, especially in the central rapidity region, an additional hadronic calorimeter is placed outside of the solenoid magnet (hadron outer or HO). Its function is that of a *tail catcher*. The HO consists of one scintillating station (two for the most central region) exploiting the solenoid itself (and the first return yoke in case of the second station) as absorber. The HO extends the total thickness of the hadron calorimeter to a minimum of $11 \lambda_I$.

2.2.4 Muon System

Muon detection and triggering are of prime importance in CMS, as many new physics processes may manifest themselves via decay chains involving muons. Muon detection in the channel $H \rightarrow ZZ^*$ was also of prime importance for the discovery of the Higgs boson. For this reason a redundant system of three different kinds of detectors is used to track muons in CMS.

The muon system, which scheme is shown in Figure 2.7, is housed in the gaps between the return yoke of the solenoid magnet in the outermost region of the experiment. It consists of a Drift Tube (DT) tracking system in the barrel and multi-wire proportional chambers in the end-cap. In addition, a set of Resistive Plate Chambers (RPC) is located both in the barrel and in the endcap regions, providing a better time resolution in the regions of high particle activity, at the cost of coarser spatial resolution.

Drift Tubes

Drift tubes are employed to identify and track muons in the barrel region ($|\eta| < 1.2$), where the rate is below 1 Hz/cm^2 and the residual magnetic field less than 1 T, allowing for the usage of this technology. Four muon stations are located at increasing distance from the beam line. The first three stations are equipped with two *super-layers* (SL) that provide a measurement of the ϕ coordinate and one SL that measures the z one, as can be seen in Figure 2.8. In the last station the SL measuring the z coordinate is missing.

Each SL is made of four stacked layers of drift tubes, staggered by half a cell. This configurations eliminates blind spots and allows for an easy measurement of the muon crossing time by averaging the drift times. Each tube has a rectangular cross-section of $13 \times 42 \text{ mm}^2$ and is filled with a gas mixture of 85% Ar and 15% CO_2 , leading to a maximum drift time of 380 ns. Each SL has a spatial resolution of about $200 \mu\text{m}$ and a time jitter below 5 ns.

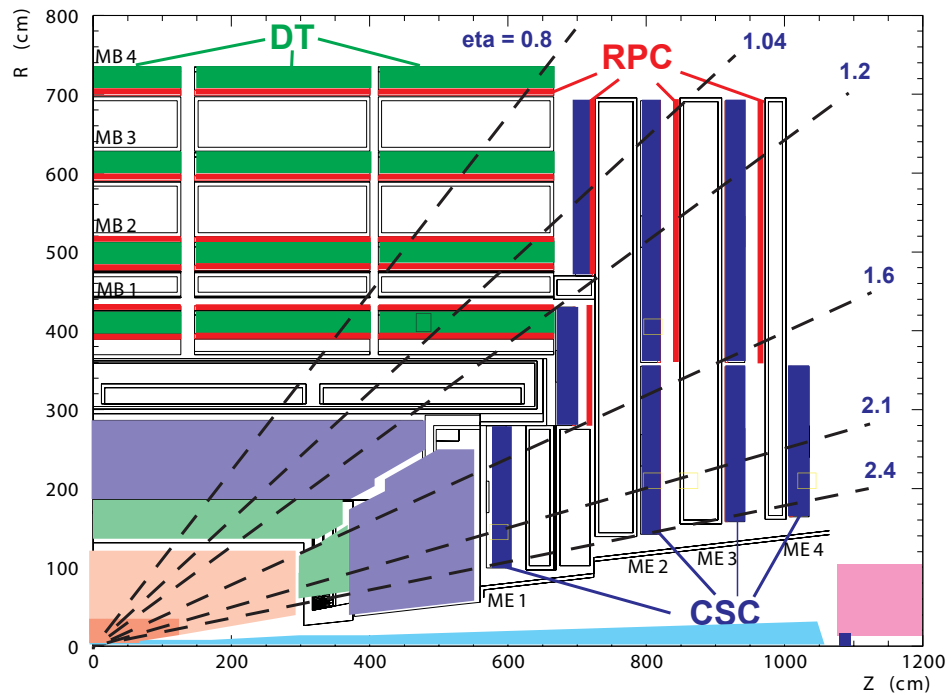


Figure 2.7: Longitudinal cross-section of the CMS detector showing the location of the muon system.

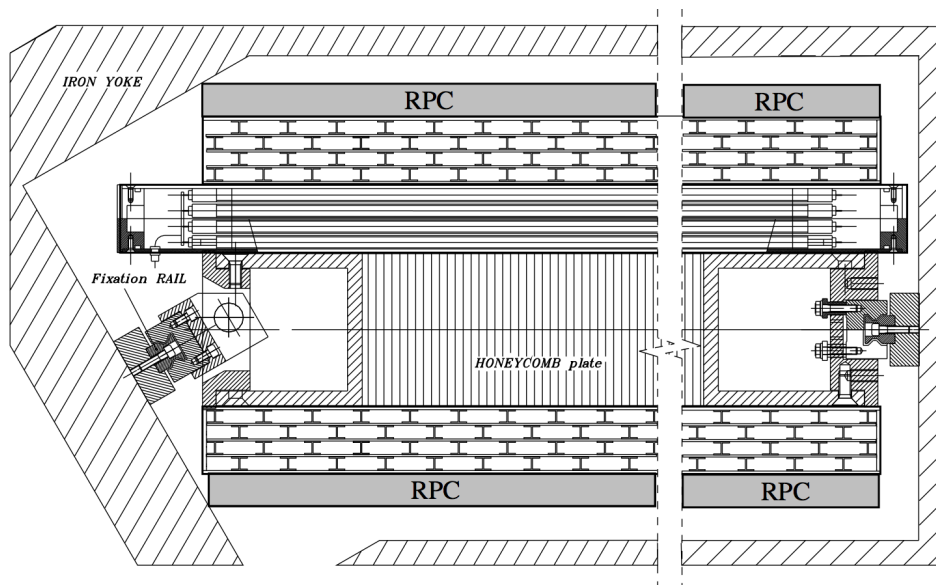


Figure 2.8: Cross-section view of a DT module.

Cathode Strip Chambers

The front wheels of the solenoid return yoke are instrumented with multi-wire proportional chambers. Each module has trapezoidal shape and covers either 10° or 20° in ϕ , forming a full disk perpendicular to the beam axis ($r - \phi$ plane). Each of these chambers has the cathode segmented radially in strips (hence the name Cathode Strip Chamber, CSC) of constant $\Delta\phi$ and wires running perpendicular to the strips, with a wire spacing of 3.2 mm.

This design allows to cope with the much higher rate with respect to the barrel and with non uniform and non zero magnetic field.

The CSC detector provides at least three measured space points for muons crossing through its acceptance ($1.2 < |\eta| < 2.4$) and provides a design resolution of approximately 2 mm at trigger level and around 200 μm after offline reconstruction.

Resistive Plate Chambers

Dual-gap RPC's provide a redundant set of spatial measurements particularly important for trigger purposes, given that their response time is much shorter than 25 ns. The layout chosen by the CMS collaboration consist of six RPC stations in the barrel and three in the endcaps. RPC stations are placed in proximity (before or after) each CSC or DT station. The first two DT stations have one RPC layer on each side of the DT station, providing at least four position measurements even for low- p_T muons, which might not reach the outer stations. Each chamber is operated in avalanche mode and consist of two gaps sharing a segmented pick-up read-out in between, which allows to operate the chambers at lower voltages (and therefore lower noise) for the same gain.

2.2.5 Trigger

As previously said, during the years 2010–2012 the LHC operated at bunch crossing period of 50 ns leading to a 20 MHz frequency, half of the design value. This frequency is well above the storage capabilities of the most modern technologies, which allow to record a few hundred events per second. The inclusive p-p cross-section is hugely dominated by low- x QCD processes that are of no or little interest for this experiment. This leads to the obvious necessity of a fast logic to isolate events of some interest for the experiment's physics program while rejecting the others. The logic to perform this selection, usually referred to as *trigger*, is divided in two stages in case of the CMS experiment: the *level one* (L1) trigger and the *high level trigger* (HLT).

The L1 trigger decision is based on a coarse reconstruction of the event, performed by custom electronics that is mostly mounted directly on the detector. The maximum processing time (called *latency*) is 3.2 μs . During this time the event is stored on the detector

electronics in pipelined memories. Due to timing constraints only the information provided by the calorimeters and by the muon system is processed. For the purpose of the trigger, the calorimeter segmentation is reduced into the so-called *trigger-towers*. The calorimeters provide to the trigger logic a set of relevant physics variables like the missing transverse energy ($\cancel{E}_T$), the scalar sum of the transverse hadronic activity (H_T), the number of jets above different energy thresholds and the locations of towers compatible with a *minimal ionizing particle* (MIP) and information on particle isolation. Electron and photon identification is also performed at this trigger level by looking at the jet hadronic over electromagnetic energy ratio (H/E) plus cluster shape information, the latter being performed with the aid of a look-up table. Muon information is mainly provided by DT and CSC sub-detectors that are complemented by timing information provided by the RPC for the purpose of determining the correct bunch crossing as well as the removal of ghost-tracks. Muon trajectories are first reconstructed coarsely by the detector dedicated electronics and then are sent to a dedicated module which merges the information of different stations and sub-detectors, to assess the transverse momentum and charge.

The final trigger decision is taken by modules that are located outside the detector volume. These modules exploit FPGA's to achieve fast response while allowing for modification of the algorithms to cope with the evolution the instantaneous luminosity or new physics demands. The maximum output frequency for L1 is 100 kHz.

Once the L1 trigger decision is made, the full event is read out from the detector buffers and is sent to the online *Data Quality Monitoring* (DQM) and to the HLT farm, consisting of over a thousand commercial processors working in parallel. During the HLT decision, a simplified version of the CMS offline event reconstruction is performed. Time consuming tasks such as tracking are performed only around the objects that caused the L1 trigger to fire and more complex algorithms like b -tagging and τ identification are performed with a simplified version of the code that is used for the final reconstruction. The algorithms running at the HLT level reduce the event rate to roughly 300Hz, within the storage and processing capabilities of the CERN facilities. Being completely software-based, the HLT is far more flexible and fast evolving than the L1 trigger, which allowed it to adjust well to the rapidly evolving machine conditions during the 2010–2012 data-taking periods. The performance achieved during data-taking have been beyond the most optimistic forecasts made at the start-up.

Once one event gets accepted by the HLT, it is transferred to the CERN computing center for storage and reconstruction.

2.3 Data storage and processing

Event information recorded by the detector is stored in RAW data format, which contains all the digitalized output of the single sub-detectors plus L1 and HLT information.

In order to be useful for physics analysis, the trajectory and the type of all particles produced in the proton-proton collision must be inferred based on the digitized hits stored in the RAW event content. This process is usually called event reconstruction and is the topic of the following chapter. The CMS detector produces about 15 TB of data for each day of operation.

The reconstruction of collected and simulated data is too CPU and storage demanding to be delegated to a single facility. In order to evenly spread the computing load, a new computing model has been created, involving different computing facilities in various parts of the world. The *world wide computing grid* [50] allows for fast data transfer plus de-localized reconstruction and analysis of data throughout the world, satisfying the huge computing demands that the LHC experiments need in order to operate successfully.

In some sense the grid can be seen as an extension of the batch queue system where the user submits jobs to a single machine, whereas, in this case, he submits them to clusters of machines, that then take care of submitting those jobs in their peculiar batch queue implementation. Computing facilities are organized in a hierarchical structure called tiers, starting from the facility that receives the data from the detector, called Tier0, scaling to smaller and smaller facilities up to Tier3's, which provide limited computing and storage for local groups. In this architecture, secure access to the resources is ensured by a system of certificates spawning children proxies of limited lifetime.

Chapter 3

Event reconstruction in CMS

While traversing the different detector layers, the particles produced in the physics collisions deposit energy in the detectors, energy that is amplified and recorded. Most sub-detectors are capable of measuring the particle position as well as the deposited energy. Each collision event results in a series of position measurements in the different layers of the tracking detectors plus energy clusters in the calorimeter. The reconstruction algorithms take care of “connecting the dots” and provide the most complete and accurate description of the event as it originated from the collision. In this task the execution time plays an important role, as the algorithms used by the reconstruction cannot be excessively resource demanding. In this chapter we will briefly describe the techniques used to reconstruct the events recorded by the CMS detector starting from the most basic objects and then moving towards the most complex ones. Finally, we describe the *Particle Flow* algorithm that gives a global description of the event. In the final section of the chapter we will describe in detail the hadronic tau reconstruction, which plays a fundamental role in this work, and the techniques used to measure its efficiency, a measurement which has been carried out as part of the doctoral work presented in this thesis.

3.1 Track reconstruction

The signals recorded by the innermost sub-detector, the inner tracker (see Section 2.2.1), are used to reconstruct the trajectory and momenta of the charged particles originating from the proton collisions as well as the location (vertex) where these interactions occurred. An accurate and detailed description of the tracking algorithms can be found in Ref. [51].

The innermost part of the tracker, made of silicon pixel detectors, records the hit position together with the charge deposited in the detector. Due to the Lorentz drift, the charge deposited by a ionizing particle is shared between neighboring pixels, improv-

ing the position resolution of the detector. The outer part of the tracker, made of silicon strip detectors, generally records only the ϕ coordinate of the hit except for few layers where an additional strip module rotated by a 100 mrad stereo angle allows for a bi-dimensional measurement of the hit positions.

Due to huge number of hits recorded in each event, reconstructing the tracks corresponding to the hits is not trivial. The strategy adopted by the CMS collaboration employs a *combinatoric track finding* (CTF) approach with very stringent requirements on the tracks, thus finding the easiest ones to reconstruct first. These requirements are progressively loosened in several iterations. After each iteration the hits used to build the tracks are removed from the pool of available hits. This procedure, called *iterative tracking*, allows to achieve a very high tracking efficiency with low fake rate and acceptable resource usage.

At the beginning of each iteration, the tracks are seeded from triplets or doublets of hits in the tracker. In case of doublets, an additional constraint on the beam-spot is used to reduce the fake rate. Only 2D hits are used at this stage to build track seeds. A first evaluation of the main track parameters (p_T and 3D impact parameter) is also performed at the seeding step. Track seeds not satisfying the minimal requirements for that iteration are discarded.

Track seeds are then propagated “inwards” (in the direction of the nominal interaction point) and “outwards” (in the direction of the calorimeter) the seeding layers in order to search for compatible hits that can be assigned to the track. The analytical propagator used at this stage assumes uniform magnetic field in between the two detector layers and neglects multiple scattering effects. New candidate hits are added to the track candidate and its parameters are updated by a *Kalman Filter* [52]. Only the track candidates satisfying a certain fit quality (evaluated with the χ^2) are retained. This process, known as *track finding*, is repeated until the innermost and outermost layers of the tracker are reached by the track propagation. During this process, the Kalman Filter retains only the latest and most accurate evaluation of the track parameters and the addition of a new hit does not require re-evaluating the full trajectory, saving a large amount of computing time. Track candidates emerging from this process are cleaned, merging those candidates sharing the majority of the hits. Mutually-exclusive hits (hits belonging to the same detector layer) are arbitrated according to the χ^2 goodness of the track fit.

Track candidates found during the track finding step are then fitted by the means of Kalman filter and smoother. This choice is as accurate as a standard fit, but more effective from the computational point of view. Trajectory parameters are evaluated both starting from the innermost hit in the tracker and from the outermost one and then averaged, avoiding any bias due to the recursive nature of the Kalman Filters. During this process, the more refined Runge-Kutta propagator is used to take into account un-homogeneities of the magnetic field and the effect of materials. Track candidates are finally selected according to the quality of the fit. Hits belonging to the tracks produced by this last step are removed from the collection of available hits and a new iteration

with different seeding conditions and track requirements begins.

A total of six iterations is performed on each event, the first ones are aimed at reconstructing most of the tracks originating from the main vertex, while the following ones focus on reconstructing tracks coming from displaced decays of long-lived particles and from vertices in periphery of the interaction region.

The solution adopted for the track reconstruction provides a tracking efficiency exceeding 90% over a wide p_T spectrum as shown in Figure 3.1.

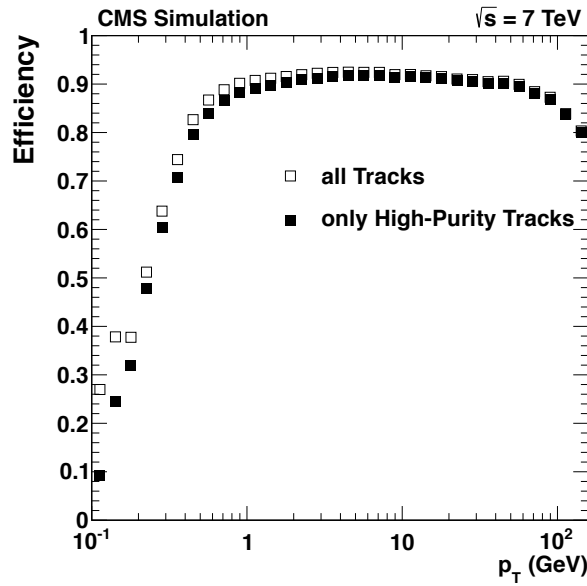


Figure 3.1: Efficiency of the track reconstruction algorithm as a function of the particle p_T . It is obtained using a sample of simulated top pair events with an average number of simultaneous p-p collisions equal to 8.

3.1.1 Reconstruction of muon tracks

Muons can be considered a special case of the tracking algorithm, given the dedicated tracking system located outside of the solenoidal magnet. This additional lever arm allows to achieve a very good momentum resolution above 200 GeV, where the inner tracker resolution begins to degrade. In CMS, the muon reconstruction is seeded both by the tracker and by the muon chambers. In the former case every track above a few GeV is considered a potential muon and a tentative match with hits in the muon system is performed. If the match succeeds, the track is refitted and it acquires the status of *tracker muon*. In the latter approach, a track is reconstructed by the CTF using only hits in the muon system. These tracks are seeded by track segments reconstructed by L1 trigger electronics. These *standalone muons* tracks are then propagated inside the inner

tracker to match with a track. As in the previous method, a successful match causes the full trajectory to be refitted and the combined object is called a *global muon*. Tracker muon reconstruction is obviously more efficient than the latter for low p_T muons as they may not leave enough hits in the muon chambers to preform a full track reconstruction, but it also suffers from a larger fake rate. The two muon collections are then merged into a single one, removing double candidates.

3.2 Vertex reconstruction

Vertices are reconstructed in CMS by a *Deterministic Annealing* (DA) [53] clustering algorithm. The tracks used during vertex reconstruction are preselected according to their transverse impact parameter with respect to the beam spot, the number of hits and the χ^2 of the track fit, to ensure that only good prompt tracks are used in the vertex reconstruction. No requirement on the track p_T is imposed in order to maximize the reconstruction efficiency in events with very little transverse momentum transfer between the colliding protons, called *minimum bias* events, which compose the majority of pile-up interactions.

Once the tracks are selected, further computations are based on the z coordinate of the points of closest approach between the tracks and the beamspot (z_i^T), and its uncertainty (σ_i^Z). In the DA framework the vertex position is evaluated by minimizing a modified version of the χ^2 with an additional term inflating σ_i^Z of each track. This parameter is called “temperature” since it behaves as the physics observable in statistical mechanics, with the χ^2 acting as free energy.

Starting from one single vertex located centrally, the temperature is iteratively lowered and the free energy term minimized for each iteration. As the temperature lowers, some tracks start being incompatible with the only available vertex and a better minimum is found by splitting the tracks in two subsets belonging to two different vertices. This splitting is done “softly”, i.e. each track can be assigned to multiple vertices with a weight between 0 and 1. With the lowering of the temperature the weights tend to assume values increasingly closer to the end-points.

The annealing process is continued down to a minimal temperature, which ensures good vertex resolution and very limited risk of splitting true vertices. After this threshold no more vertex splittings are allowed, but the temperature is still iteratively reduced and potential outliers in the vertex fits are removed. Once the final temperature ($T = 1$) is reached, tracks are assigned to the corresponding vertices if their weight is greater than 0.5 as spurious weights may still be different from zero.

Each set of tracks assigned to a vertex is then fitted with the *adaptive vertex filter* algorithm [54], consisting of an iterative weighted Kalman Filter, to evaluate the vertex position with the best possible precision.

Such refined technique allows to achieve a vertex reconstruction efficiency between

98% for cluster of two or three tracks and 100% for the rest, with a negligible fake rate of about 1%. The vertex resolution for clusters of few tracks is about $200\ \mu\text{m}$ both in the transverse and longitudinal plane, rapidly decreasing to an asymptotic value of about $20\ \mu\text{m}$ as more tracks are used to compute the vertex position.

3.3 Electron and photon reconstruction

Both electrons and photons are reconstructed starting from ECAL clusters, which are seeded from local maxima in the energy deposition. A cluster size of 5×5 crystals on average contains 95% of the energy of an unconverted photon. This average value is considerably reduced in the real environment due to photons converting into e^+e^- pairs when traversing the material in front of the calorimeter and those pairs being separated in the $r - \phi$ plane by the magnetic field. Photon conversions result in wider energy clusters, elongated in the ϕ direction. Similar arguments can be made for electrons and positrons, that emit a sequence of *bremsstrahlung* photons as they approach the surface of the calorimeter. In order to recover this energy the 5×5 clusters are grouped together in *superclusters* (SC), elongated in the ϕ direction. A more detailed description of the clustering algorithm can be found in [55].

The *bremsstrahlung* process and its intrinsic non-Gaussian energy emission makes the Kalman filter approach unsuitable. It is replaced by the Gaussian Sum Filter (GSF) [56] which handles the energy loss properly. The usage of the GSF and its larger propagation uncertainties makes a tracker-only approach unfeasible due to excessive computation time and low track purity. In order to reduce the amount of fake tracks, the seeding step starts from ECAL superclusters from which two track candidates are propagated inside the tracker under both charge assumptions looking for two compatible hits in the pixel detector. After the seeding, the tracking process continues as described in the previous section with the GSF substituting the KF. This modified version of the tracking algorithm is also used to correctly identify photon conversions within the tracker volume, looking for highly displaced vertices compatible with a massless state and matching a broad supercluster in the electromagnetic calorimeter.

Several variables are used to discriminate between fake and real electrons, such as the supercluster shape, its energy and location with respect to the impact position in the calorimeter, expected from the measured track parameters, and the presence of energy deposits in the hadronic calorimeter. All these variables are combined to perform the electron identification. Several working points, both with cut-based and MVA-based approach are available. The latter exploits TMVA's [57] Boosted Decision Tree (BDT) implementation to separate electrons from charged hadrons and photons. The BDT behavior has been trained on real electrons selected from a $Z \rightarrow ee$ sample and fake ones from a $Z \rightarrow ee + \text{Jets}$ one.

Photon identification is based on the same supercluster quantities as the electron identi-

fication but doesn't require a charged track to be matched to the ECAL energy deposit.

3.4 Particle Flow

The redundancy of signals that particles traversing the CMS detector leave in multiple sub-detectors is exploited by the *Particle Flow* (PF) algorithm in order to improve the overall description of the event. The particle flow method furthermore allows for cross-cleaning of spurious signals and achieve a global description of the event.

Building blocks for the PF algorithm are the different types of track collections plus the calorimeter energy deposits. Local maxima in the calorimeter deposits serve as seed for "topological clusters" which are grown iteratively, by adding adjacent cells with an energy deposit above a customizable threshold. After the clustering stage the algorithm links the different blocks according to their position in the $\eta - \phi$ plane. Tracks are considered linked to calorimeter clusters if the track trajectory propagated to the average shower depth falls within the cluster boundaries, within some distance that accounts for resolution effects. Calorimeter clusters are linked if the center of the inner cluster fits within the boundaries of the outer one. The radial ordering of the linking step reflects the higher spatial resolution of the inner calorimeters (pre-shower, ECAL, HCAL).

After the linking step is completed, the algorithm starts building the final objects, starting with muons. Each global muon that has a combined momentum measurement compatible with the track reconstructed in the inner detector is called a *PF muon*. The corresponding track is removed from the track collection and an estimate of the associated energy deposit in the calorimeter is also removed from the list of linked clusters. Electrons are the second category of objects that is built with the procedure detailed above. Tangents to the electron trajectory in correspondence of the tracker layers are propagated to the ECAL in search for bremsstrahlung photons. If the electron candidate passes the selection cuts a *PF Electron* is created and the track as well as the ECAL deposits linked to the track are removed. A tighter selection is performed on the remaining tracks, requiring that the tracker p_T resolution is smaller than the calorimetric energy resolution. This requirement is found to reject about 0.2% of tracks in hadronic jets, 90% of which being fake tracks. The algorithm finally exploits the redundancy in the transverse momentum measurement performed by the tracker and by the calorimeter to build *PF charged hadrons, photons and neutral hadrons*. The PF algorithm operates under the assumption that all the hadrons (charged and neutral) traversing the detector are pions. Photons and neutral hadrons are produced from the calorimetric energy in the clusters once the energy of the linked tracks is subtracted. In the rare case that a calorimetric cluster has a total energy that is incompatible and lower than the sum of the linked tracks a dedicated procedure aimed at recovering mis-reconstructed muons is performed, recovering muon identification efficiency at no expense of increasing the fake rate.

3.4.1 Muon identification

The muon identification step in the PF reconstruction is necessary in order to suppress the hadronic punch-through, i.e. very energetic charged pions passing through the hadronic calorimeter and reaching the innermost muon station. The work presented in this thesis uses the tight working point of the PF muon identification, also referred to as “PFTight” within CMS. The corresponding selection cuts are:

- The candidate muon must be reconstructed both as global and PF muon;
- The normalized χ^2 of its global track fit must be less than 10;
- There must be a least one DT or CSC hit included in the global fit;
- At least two muon stations must be matched to the candidate;
- The associated tracker track must have a transverse impact parameter $d_{xy} < 2$ mm and a longitudinal distance $d_{xy} < 5$ mm;
- It must have at least one hit in the pixel detector and more than 5 hits in the whole tracker.

3.4.2 Isolation

A large fraction of the leptons produced in a proton-proton collision are due to the decay of mesons within jets. In order to separate these leptons from the ones which originate from the decay of a heavy resonance it is important to quantify the *isolation* of the lepton, i.e. the amount of hadronic activity surrounding the lepton trajectory in a cone of a certain radius. The isolation can be computed separately with respect to charged and neutral particles.

While the effect of pileup can be easily removed from the charged isolation by using only tracks associated to the selected primary vertex, the same is not possible for the neutral component. In fact, calorimeters do not have sufficient spatial resolution to uniquely associate each cluster to a certain interaction vertex. The contribution of pileup to the neutral isolation is estimated by computing the energy deposited by charged tracks from pileup (i.e. not assigned to the hard-scatter vertex) and correcting it for a factor 1:2 to account for the amount of neutral energy with respect to charged one. This estimate is called $\Delta\beta = I_{charged}^{pileup}/2$. To limit the effect of statistical fluctuations in the charged to neutral ratio on event-by-event basis, the $\Delta\beta$ correction is capped to the neutral isolation value. The *relative isolation*, i.e. the isolation divided by the p_T of the object, can therefore be written as

$$I_{rel} = \frac{I_{abs}}{p_T} = \frac{\sum p_{Tcharged} + \max(\sum E_{Tneutral} - \Delta\beta, 0)}{p_T}. \quad (3.1)$$

3.4.3 Jet clustering

The particles reconstructed by the PF algorithm are clustered into jets using the *Anti-kT* [58] algorithm. The Anti-kT belongs to the family of sequential recombination algorithms together with the Cambridge-Aachen and the kT algorithms [59, 60]. All these algorithms require the definition of the distance between two objects (in this case PF particles) and of the distance from the beam axis. In case of the Anti-kT algorithm, the distance between two objects is defined as:

$$d_{ij} = \min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right) \frac{\Delta R_{ij}^2}{r^2}, \quad (3.2)$$

where r is a constant parameter defining the size of the jet cone and $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The distance of an object from the beam is defined as:

$$d_{iB} = \frac{1}{p_{Ti}^2}. \quad (3.3)$$

The algorithm starts with a high- p_T object used as seed and then it clusters around the seed all other objects until the minimum distance between the candidate jet and the closest object is larger than d_{iB} . One can show that this algorithm is infra-red safe, meaning that the jet shape and energy do not depend strongly on the energy of low- p_T particles clustered around a high p_T object. The Anti-kT algorithm gives rise to almost-conical jets of size r in angular aperture. Special cases are overlapping jets and jets with multiple high p_T objects within the angular acceptance.

The jets used in this work have been reconstructed with the Anti-kT algorithm with a radius parameter $r = 0.5$ and their energies have been corrected to account for non-linear response of calorimeters and other instrumental effects [61].

Pileup interactions deposit a considerable amount of energy in the detector. Furthermore, the energy deposits due to different pileup interactions may overlap and be misinterpreted by the Anti-kT algorithm as one hard jet. The CMS experiments exploits a combination of tracking and jet shape variables to discriminate between pileup jets and jets originating from the hard scattering. The information provided by these variables is interpreted by a BDT trained on simulated $Z \rightarrow \mu\mu$ events [62]. The training is divided in four different $|\eta|$ categories that reflect the resolution of tracking information and granularity of the calorimeters.

3.4.4 Missing transverse energy

Neutrinos and other hypothetical (not yet observed) weakly-interacting particles that are neutral and stable leave no trace when traversing the detector and hence can not be directly reconstructed. Their presence can be inferred by utilizing the conservation of

transverse momentum in the event. The observable quantifying an imbalance is called *missing transverse energy* ($\cancel{E}_T$). It is computed as the negative vectorial sum of all the PF objects reconstructed in the event [63]:

$$\vec{\cancel{E}}_T = - \sum^{PF\ obj.} \vec{p}_T \quad (3.4)$$

The magnitude of the $\cancel{E}_T$ is found to be underestimated for various reasons including thresholds in the calorimeter clustering and signal non-linearities. This bias is significantly reduced in case the corrections to jet energies are included in the formula [61], obtaining the “type 1”-corrected $\cancel{E}_T$. The introduction of jet energy corrections (JEC) is performed by subtracting the vectorial residuals between all corrected and uncorrected jets found in the event:

$$\vec{\cancel{E}}_T^{corr} = \vec{\cancel{E}}_T - \sum_{jets} (\vec{p}_{T,jet}^{corr} - \vec{p}_{T,jet}) \quad (3.5)$$

Additional corrections, named “type 0” and “ ϕ ”, correct for biases induced by pileup and by asymmetries in the ϕ direction [64].

3.5 Tau reconstruction and identification

Hadronic taus are one of the highest level objects reconstructed from CMS data and the last in the event reconstruction chain. The reconstruction and identification algorithm exploits all the detector capabilities and the Particle Flow algorithm to achieve a reconstruction efficiency close to 60% with a sub-percent fake rate. The tau reconstruction algorithm used in CMS is called *Hadron Plus Strip* (HPS) [65, 66].

3.5.1 Reconstruction

Tau reconstruction is seeded from PF jets clustered by the Anti-kT algorithm with $r = 0.5$. In order to recover most of the photon conversions, the tau reconstruction algorithm clusters PF electromagnetic objects (electrons and photons) with deposited energy above 0.5 GeV in topological “strips” of size 0.05×0.20 in the η and ϕ direction, respectively. Strips satisfying a minimum transverse energy of 2.5 GeV are combined with the charged hadrons to form the tau candidate. The charged hadrons included in the tau reconstruction must satisfy the following requirements:

- $p_T > 0.5$ GeV;
- track fit $\chi^2 < 100$;

- track transverse distance of closest approach to the PV $d_0 < 0.03$ cm;
- track longitudinal distance of closest approach to the PV $d_z < 0.4$ cm;
- At least three hits in the tracker.

The algorithm builds all possible combinations of hadrons plus strips matching one of the following decay modes:

1. *single hadron*: corresponding to a single hadron and no strip.
2. *hadron plus one strip*: the invariant mass of the pair has to be compatible with the ρ meson, $0.4 < M_{\tau \text{ candidate}} < 1.3 \cdot \sqrt{p_T[\text{GeV}]/200}$ GeV. The upper boundary of this window is limited at 1.3 (2.1) GeV for candidates with $p_T < 200$ (> 800) GeV. The increasing window size accounts for resolution effects at high p_T .
3. *hadron plus two strips*: the invariant mass of the triplet has to be compatible with the ρ meson, $0.4 < M_{\tau \text{ candidate}} < 1.2 \cdot \sqrt{p_T[\text{GeV}]/200}$ GeV. The upper boundary of this window is limited at 1.2 (2.0) GeV for candidates with $p_T < 200$ (> 800) GeV.
4. *three hadrons*: the invariant mass of the triplet must be compatible with the a_1 meson, $0.8 < M_{\tau \text{ candidate}} < 1.5$ GeV. The tracks are required to be compatible with originating from the same vertex and to sum to unit charge.

In addition to the previous requirements, tau constituents are required to be contained within a cone of size 0.1, $3/p_T[\text{GeV}]$ or 0.05 for tau candidates with $p_T < 30$ GeV, $30 < p_T < 60$ GeV and $p_T > 60$ GeV, respectively. If more than a tau candidate can be formed within the same jet only the candidate with the largest p_T is kept.

3.5.2 Identification

Hadronic tau decays reconstructed by the HPS algorithm are required to be isolated in order to suppress fakes due to quark and gluon jets. Only charged hadrons with p_T above 1 GeV and photons with E_T above 1.5 GeV are considered when computing the isolation. To mitigate the effect of pileup the neutral isolation is corrected with the $\Delta\beta$ method computed in a cone of $\Delta R = 0.8$ around the tau candidate. The mismatch between the size of the isolation cone (0.5) and the one in which the pileup contribution is computed (0.8) leads to a conversion factor from charged pileup to neutral pileup isolation of 0.4576 that is empirically found to make the tau identification efficiency insensitive to pileup.

Three working points are provided: loose, medium and tight, corresponding to isolation thresholds of 2.0, 1.0 and 0.8 GeV respectively.

3.5.3 Light lepton rejection

The PF charged-hadron collection may contain a significant amount of electrons and muons that do not pass the selections to enter the respective PF collections, and therefore are considered to be charged hadrons by the tau reconstruction algorithm. Electrons and muons from the decay of W and Z bosons are typically isolated. As a consequence, the background due to $e \rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ fakes may be sizable. In order to reduce these backgrounds, dedicated discriminators against electrons and muons have been developed [67].

Electron rejection: The rejection of electrons is performed with the aid of a boosted decision tree. Tau candidates are classified into different categories depending on:

- The decay mode in which the tau candidate is reconstructed. Three decay modes are considered: single hadron, hadron plus one or two strips, and three hadrons. The three hadrons decay mode always passes the electron rejection discriminator;
- Whether the hadron is associated to a track reconstructed by the GSF algorithm;
- Whether a GSF electron candidate is reconstructed in the same direction as the tau candidate within a distance $\Delta R < 0.3$;
- Whether the tau candidate is reconstructed in the ECAL barrel ($|\eta| < 1.479$) or endcap ($|\eta| > 1.479$).

All the possible combinations of the previous cases are considered, leading to a total of 16 mutually exclusive categories. Each category is trained separately using a mixture of simulated events containing both signal and background processes: $Z/\gamma^* \rightarrow \tau\tau$, $Z/\gamma^* \rightarrow ee$, $W \rightarrow \tau\nu$, $W \rightarrow e\nu$, $t\bar{t}$, $H \rightarrow \tau\tau$, $Z' \rightarrow \tau\tau$, $Z' \rightarrow ee$, $W' \rightarrow \tau\nu$, $W' \rightarrow e\nu$. Events are considered as “signal” or “background” depending on whether the tau candidate is matched to a generator-level hadronic tau decay or an electron within a cone of size $\Delta R < 0.3$, respectively.

The input variables to the BDT can be divided into four groups. The variables are:

Hadronic tau variables (available for all categories):

- Invariant mass, p_T and η of the tau candidate;
- Ratio of ECAL to the sum ECAL plus HCAL energy deposits of the tau constituents;
- Ratio of ECAL plus HCAL energy deposits of the tau constituents to the tau candidate p_T ;
- $\Delta\eta$ between the tau candidate and the closest ECAL crack;

- $\Delta\phi$ between the tau candidate and the closest ECAL crack (used for candidates in the ECAL barrel only).

Strip variables (available for taus reconstructed in the decay modes hadron plus one strip or hadron plus two strips only):

- $\sqrt{p_T^\gamma \cdot (\Delta\eta)^2}$ and $\sqrt{p_T^\gamma \cdot (\Delta\phi)^2}$, the p_T -weighted RMS of distances along η and ϕ direction between all photons included in any strip and the charged hadron;
- Fraction of τ_h candidate energy carried by photons.

GSF track variables (available only when the charged hadron is associated to a GSF track):

- $\ln(p_T)$, η and normalized χ^2 of the GSF track;
- PF electron MVA output for the PF charged hadron;
- $(N_{hits}^{GSF} - N_{hits}^{KF}) / (N_{hits}^{GSF} + N_{hits}^{KF})$, where N_{hits}^{GSF} and N_{hits}^{KF} are the number of tracker hits associated to the GSF and KF tracks, respectively;

GSF electron variables (only available in case a GSF electron is found near the hadronic tau candidate):

- The ratio between the total ECAL energy and the electron momentum measured at the IP;
- Normalized χ^2 , $\ln(p_T)$, η , p_T resolution (σ_{p_T}/p_T), and numbers of hits in the tracker of the electron GSF track;
- The ratio between the bremsstrahlung photon energy as measured by the ECAL and by the track, $\sum E_\gamma / (P_{in} - P_{out})$, where P_{in} is the momentum at the interaction vertex and P_{out} is the one on the ECAL surface;
- $F_{brem} = (P_{in} - P_{out}) / P_{in}$, the ratio between the bremsstrahlung photon energy as measured by the track and the electron momentum measured at the IP;

Four working points are defined: *loose*, *medium*, *tight* and *very tight*, with nominal efficiency ranging from 95% to 80% in steps of 5% and corresponding decreasing $e \rightarrow \tau_h$ fake rates.

Muon rejection: uses a cut-based approach, looking for segments in the muon chambers near the direction of the tau. Two working points are provided:

- **Loose:** tau candidates are rejected if track segments in at least two muon stations are found within $\Delta R < 0.5$ or if the sum of ECAL and HCAL energy deposits associated to the leading charged hadron are less than 20% of the momentum of its associated track.
- **Tight:** in addition to the loose working point selection no hits in the two outermost muon stations have to be found in a cone of $\Delta R = 0.5$ from the tau direction.

3.5.4 Performance

Identification efficiency

The hadronic tau identification efficiency has been measured in 2011 data using an integrated luminosity of 1.2 fb^{-1} collected at $\sqrt{s} = 7 \text{ TeV}$. The measurement is part of the doctoral work presented in this thesis and was based on a *tag-and-probe* method applied to $Z \rightarrow \tau_\mu \tau_h$ events, where τ_μ indicates $\tau \rightarrow \mu \nu_\mu \nu_\tau$.

Events used for the efficiency measurement have been selected as follows:

- The event is required to have fired the single-muon trigger path with $p_T^\mu > 17 \text{ GeV}$. The chosen trigger path prevents any bias on the hadronic tau leg selection, while the p_T threshold corresponds to the lowest unpre-scaled single muon trigger path for the analyzed run period;
- The event must contain at least one vertex with ≥ 4 degrees of freedom, reconstructed within 24 cm along z from the nominal interaction point and within 2 cm in the transverse plane from the beamspot position;
- The event must contain one muon with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.1$, with a relative isolation below 0.3 in a cone of $\Delta R < 0.4$ (no $\Delta\beta$ corrections to the muon isolation are applied);
- The event must contain a tau-jet candidate with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.3$. In order to reject electrons and muons being identified as tau-jet candidates it has to be separated from any segment in the muon system by $\Delta R > 0.5$ and to be identified as hadron-like by the PF electron and photon MVA identification. In addition, the *leading* (highest p_T) object in the tau-jet candidate is required to be a charged hadron with $p_T > 5 \text{ GeV}$ and the p_T sum of charged and photon PF objects in an annulus of size $0.15 < \Delta R < 0.6$ around the leading track must be below 2.5 GeV. These last two requirements are found to significantly reduce the backgrounds while having a high efficiency for genuine taus;

- A veto on any additional global muon is imposed on the event.

Events passing the previous selection are then divided into six mutually-exclusive categories (called A, B, C1p, C1f, C2, and D). C1p and C1f categories contain $Z \rightarrow \tau_\mu \tau_h$ events where the τ_h passes or fails the hadronic tau identification criteria, respectively. The other categories are dominated by the two main backgrounds to this analysis, $W + Jets$ and QCD events, and are used in a simultaneous fit to constrain the yields of these to backgrounds in the signal region. Events are assigned to the proper category as follows:

- In the C1p, C1f, C2, and D categories the muon relative isolation is required to be below 0.1;
- In the C1p, C1f, C2, and A categories the muon and the tau-jet leading track are required to be of opposite charge;
- In the C1 categories it is required that $M_T(\mu, \vec{E}_T) < 40$ GeV and that $p_\zeta - 1.5 \cdot p_\zeta^{vis} > -20$ GeV, where:

$$M_T(\mu, \vec{E}_T) = \sqrt{(E_T^\mu + \vec{E}_T)^2 - (\vec{p}_T^\mu + \vec{E}_T)^2} \quad (3.6)$$

$$p_\zeta = (\vec{p}_T^\mu + \vec{p}_T^{\tau jet} + \vec{E}_T) \cdot \vec{\zeta} \quad (3.7)$$

$$p_\zeta^{vis} = (\vec{p}_T^\mu + \vec{p}_T^{\tau jet}) \cdot \vec{\zeta} \quad (3.8)$$

where $\vec{\zeta}$ is the versor bisecting the angle between the muon and the tau-jet candidate in the transverse plane [68].

- The categories C1p and C1f contain the events where the tau-jet candidate passes or fails the tau identification working point, respectively.

A graphical illustration of the event categories is shown in Figure 3.2.

The processes considered signal and background in the tau identification efficiency measurement are: $Z \rightarrow \tau\tau$, $Z \rightarrow \mu\mu$, $W \rightarrow \ell\nu + Jets$, $t\bar{t} + Jets$ and QCD multijet production. All the kinematic distribution shapes used for this processes in the analysis are taken from MC simulation and normalized to the NLO predictions with the exception of the QCD shape in the C1p and C1f regions, where, to mitigate the lack of MC statistics, the shape is taken from the data distribution in the B region with the additional requirements of $M_T < 40$ GeV and $p_\zeta - 1.5 \cdot p_\zeta^{vis} > -20$.

The yield of each process contributing to the measurement is expressed as a product of the total yield (N^i), the probability to pass (or fail) the muon isolation ($p_{\mu Iso}^i$), the probability to pass (or fail) the leading track charge requirement (p_{ch}^i), the probability

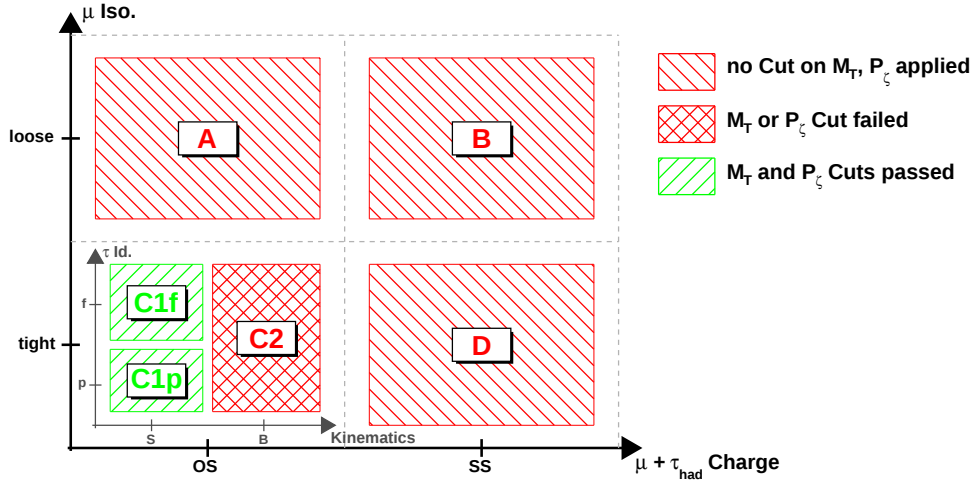


Figure 3.2: Illustration of signal and control regions used in the efficiency measurement and their requirements.

to pass (or fail) the topological cuts (p_{topo}^i) and the probability pass (or fail) the given tau-identification discriminator, $p_{\tau ID}^i$. The parameters listed above, initialized to their MC expectation, are inferred from data with a simultaneous template fit of all the regions and the different signal and background components. The physical observables used for the template fit is the invariant mass of the visible decay products, also called *visible mass*, (M_{vis}), for the C1p and C1f categories and the transverse mass of the muon and $\cancel{E}_T$ for the others. Figure 3.3 shows the results of the template fit for the loose discriminator.

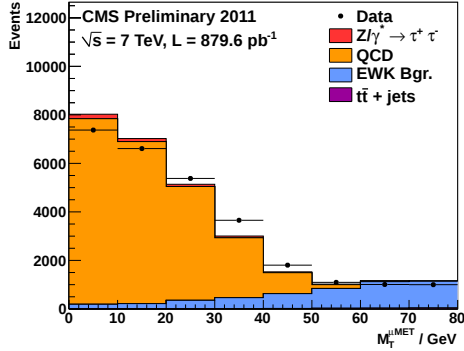
The inferred probability to pass the tau identification discriminator is related to its efficiency by:

$$\epsilon_{\tau ID} = p_{\tau ID} \cdot \epsilon_f \cdot \epsilon_p \cdot \epsilon_i \cdot k_{jet \rightarrow \tau} \cdot k_p \cdot k_i \quad (3.9)$$

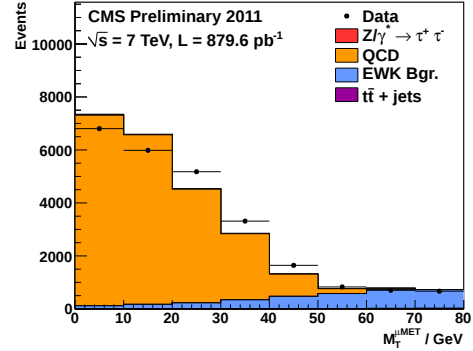
Where the symbols ϵ denote the preselection efficiencies and k the correction factors. The correction factors account for the fraction of taus that would pass the tau identification, but fail either the jet isolation (k_i), or the leading track requirement (k_p), and for the small amount of quark and gluon jets present in the $Z \rightarrow \tau\tau$ templates ($k_{jet \rightarrow \tau}$). These correction factors are extracted from MC simulation.

The measured hadronic tau identification efficiency is compared to the MC prediction. The efficiencies measured for all the tau identification discriminators are found to be in agreement with the MC prediction within the uncertainties, even though a small trend towards lower values may be seen for the tighter working points. The hadronic tau identification efficiency for the working point used in the search presented in this thesis amounts roughly to 60%, with a total uncertainty of 6%. The uncertainty represents the sum of statistical and systematic uncertainties added in quadrature. The main source of statistical error comes from the large uncertainties in the $Z \rightarrow \tau\tau$ yield in the C1f region,

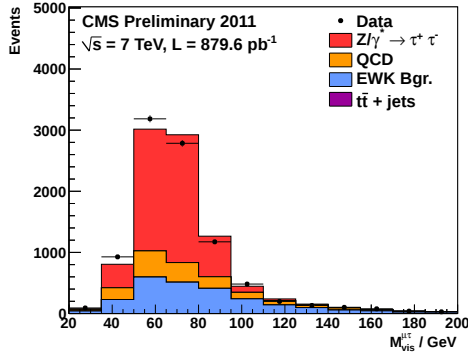
(A)



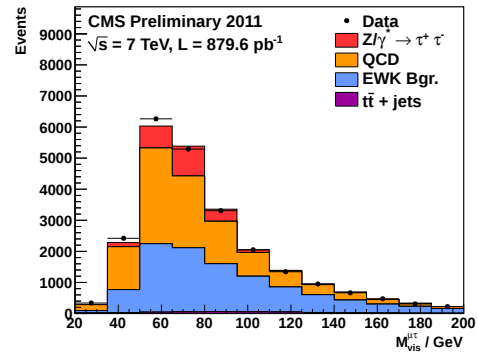
(B)



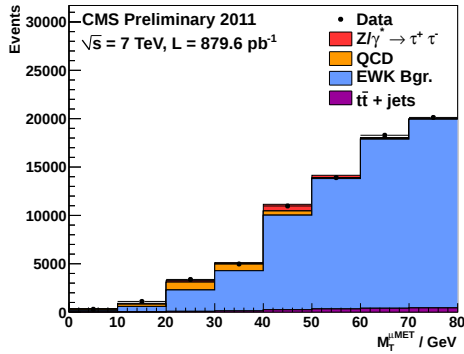
(C1p)



(C1f)



(C2)



(D)

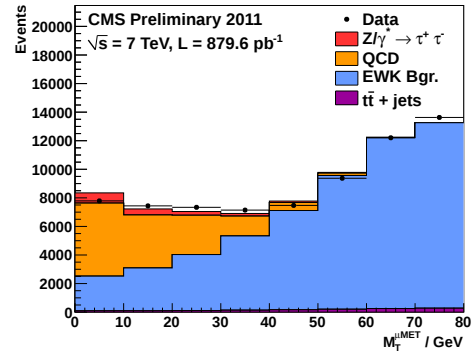


Figure 3.3: Measured distributions of M_T for the regions A , B , $C2$ and D and of M_{vis} for $C1p$ and $C1f$ compared to the sum of templates associated to the signal and background processes. “EWK Bgr.” denotes the sum of $Z \rightarrow \mu^+ \mu^-$ and $W + \text{jet}$ background contributions. The templates are scaled by normalization factors obtained by the fit for the loose working-point of the HPS combined isolation discriminator.

Table 3.1: Summary of systematic uncertainties considered in the hadronic tau identification efficiency measurement, and their effect on the final result.

systematic uncertainty	effect
Muon momentum scale	$\ll 1\%$
τ -Jet energy scale	$< 1\%$
Hadronic track reconstruction	3.9%
Track momentum scale	$< 1\%$
Loose Isolation	2.5%
Jet $\rightarrow \tau_h$ fake rate	1.2%
k_p uncertainty	1.7%
k_i uncertainty	2.1%
Statistical uncertainty	2.6%
Total	6.0%

while the leading systematic uncertainty is the reconstruction efficiency for hadronic tracks. A summary of the systematic uncertainties considered in the measurement can be found in Table 3.1.

Jet fake probability

The jet $\rightarrow \tau_h$ fake probability is measured using W+Jets and QCD events. Events are selected online by a single-muon (W events) or single-jet (QCD events) trigger. W+Jets events are required to contain a well identified and isolated muon with $M_T(\mu, \cancel{E}_T) > 50$ GeV. The jet $\rightarrow \tau_h$ fake probability is computed for all taus with $p_T > 20$ GeV, $|\eta| < 2.3$ and passing the combination of “decay mode finding” plus the isolation discriminator under study with respect to all jets of $p_T > 20$ GeV and $|\eta| < 2.3$ [69]:

$$p(jet \rightarrow \tau_h) = \frac{N(p_T^{\tau_h} > 20 \text{ GeV} \& |\eta^{\tau_h}| < 2.3 \& \text{Decay} - \text{Mode} \& \text{Discriminator WP})}{N(p_T^{jet} > 20 \text{ GeV} \& |\eta^{jet}| < 2.3)} \quad (3.10)$$

The jet $\rightarrow \tau_h$ fake probability measured in W+jets events as a function of jet p_T , η , ϕ and as a function of the number of reconstructed vertices is shown in Figure 3.4. The jet $\rightarrow \tau_h$ fake probability is found to be strongly dependent to the jet p_T . An increase of the fake probability is also observed at high η values, where the tracking efficiency decreases, affecting the discriminating power of the isolation-based discriminators. A mild dependence on the number of pileup vertices is also observed. This effect is due to the $\Delta\beta$ correction, which gradually switches off the neutral isolation discrimination as the pileup increases, leaving only the track-based isolation. Please note that the $\Delta\beta$ correction has been tuned to keep the efficiency constant as a function of the number

of pileup interactions, causing the fake-rate to increase; another possible solution could have been keeping the fake-rate stable at the cost of reducing the efficiency. The slight mismatch between data and MC prediction is due to a mismodeling in the simulation of *out-of-time pileup*, i.e. the effect that collisions from previous bunch crossings have in the current event. This effect is visible only in high-latency detectors such as the ECAL.

Electron and Muon fake probability

The rejection power of the anti-electron and anti-muon discriminators is computed in $Z \rightarrow \ell\ell$ events with a tag-and-probe technique. Well identified and isolated leptons are used to select the events both online and offline. The event is required also to contain a tau candidate passing only the decay-mode finding plus loose isolation discriminators (probe). The dimuon invariant mass distribution is fitted both in regions passing and failing the anti-lepton requirement to extract the rejection power and the ratio between data and MC.

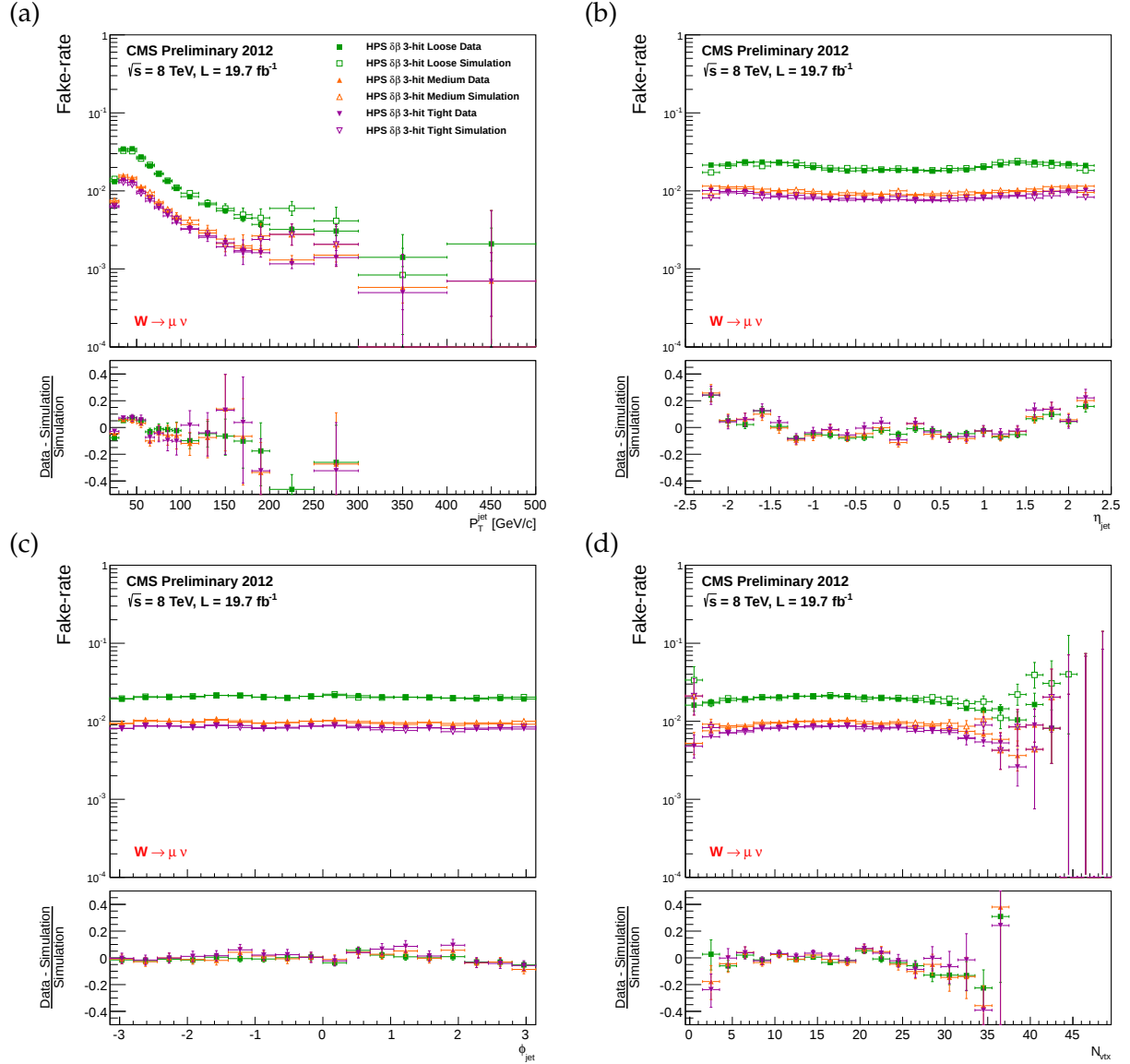


Figure 3.4: Probabilities for quark and gluon jets in W +jets events to pass the *HPS combined isolation 3-hit* discriminator, as function of jet P_T (a), η (b), ϕ (c) and as function of N_{vtx} , the number of vertices reconstructed in an event (d). The probabilities measured in data are compared to the Monte Carlo expectation.

Chapter 4

Search for the Higgs boson in tau decays in the process $pp \rightarrow WH$

As described in Chapter 1, the Higgs boson is produced at the LHC via three main processes: the gluon fusion (GF), the vector boson fusion (VBF) and the associated production (VH). The VH process has a production cross section one order of magnitude smaller than the dominant gluon fusion process. The presence of an additional high- p_T lepton enhances the signal over background ratio significantly, increasing the sensitivity of the VH channel to a level that is comparable to the sensitivity of the GF process.

In this chapter we describe the search for the associated production of a W and a Higgs boson, where the W boson decays into a light lepton (electron or muon, generically denoted as ℓ from now on) and a neutrino, and the Higgs boson decays into a tau pair, with one tau decaying leptonically and the other tau hadronically. Three final states are considered: $\mu\mu\tau_h$, $e\mu\tau_h$ and $ee\tau_h$.

4.1 Event Selection

Events are selected in real time by the double lepton triggers, which require the presence of either two muons, two electrons or a muon plus an electron. Evolving running conditions and increasing instantaneous luminosity imposed a tuning of the trigger object requirements in order to meet the constraints in trigger bandwidth.

The unrescaled triggers with the lowest p_T thresholds available for each running period have been chosen. This choice results in p_T thresholds ranging from 7 up to 17 GeV and from 7 to 8 GeV for the leading and sub-leading muons in the double muon trigger. The p_T thresholds of the muon plus electron and of the double electron triggers have been kept constant at 17 GeV for the leading object and 8 GeV for the sub-leading lepton. The electron isolation criteria have been tightened as instantaneous luminosity

increased. A more detailed description of the electron and muon triggers can be found in [70, 71].

In the offline event selection, leptons are selected according two sets of criteria called “loose” and “tight”. Tight lepton requirements are applied when selecting signal events. The loose lepton selection, which is a subset of the tight one, is used to enhance the event statistics for the purpose of modeling the backgrounds as described in the following sections.

- Loose muons are required to:
 - be reconstructed as global or tracker muons;
 - have $p_T > 10$ GeV, $|\eta| < 2.4$, $|d_Z| < 0.2$ cm;
 - have at least one hit in the pixel detector, to discriminate against in-flight decays;
 - the jet nearest to the muon in ΔR is required not to pass the loose b -tagging discriminator. This cut is applied to reject $t\bar{t}$ events.
- Loose electrons are required to:
 - have $p_T > 10$ GeV, $|\eta| < 2.5$, $|d_Z| < 0.2$ cm;
 - have no missing tracker hits;
 - have “tight” charge agreement, i.e. the curvature of the CTF, the GSF and the pixel-only tracks should agree. This requirement reduces the electron charge mis-identification;
 - the jet closest to the electron is required not to pass the loose b -tagging discriminator, to reject $t\bar{t}$ events
- Loose taus are required to:
 - have $p_T > 20$ GeV, $|\eta| < 2.3$, $|d_Z| < 0.2$ cm;
 - pass the “decay-mode finding” discriminator as described in Section 3.5;
 - not overlap with any electron passing loose MVA identification criteria plus isolation $Iso_{rel} < 0.3$, computed with $\Delta\beta$ correction applied in a cone of radius $\Delta R < 0.4$.

The offline event selection is different for the three channels and is detailed in the following sections:

$\mu\mu\tau_h$

- The event must pass the lowest-threshold unprescaled double muon trigger;
- Both muons must be matched to the corresponding trigger candidates;
- The leading muon must have $p_T > 20$ GeV, to be compliant with the online thresholds;
- Both muons must pass the PF Tight identification working point;
- The $\Delta\beta$ -corrected relative PF isolation of the leading muon, computed in a cone of $\Delta R < 0.4$ has to be less than 0.15 (0.1) for muons with $|\eta| < 1.479$ ($|\eta| > 1.479$);
- The $\Delta\beta$ -corrected relative PF isolation of the sub-leading muon, computed in a cone of $\Delta R < 0.4$ has to be less than 0.2 (0.15) for muons with $|\eta| < 1.479$ ($|\eta| > 1.479$);
- The hadronic tau candidate is required to pass the loose HPS combined isolation working point to suppress the backgrounds with jets misidentified as τ_h , the loose electron rejection working point plus the tight muon rejection working point. The latter is applied to suppress the background process $Z \rightarrow \mu\mu + \text{jet}$ where a muon is misidentified as the hadronic tau and the jet as a muon.

 $e\mu\tau_h$

- The event must pass the lowest-threshold unprescaled muon plus electron trigger, which imposes a 17 GeV threshold on the leading lepton p_T and a 8 GeV threshold on the sub-leading one, independently on the flavor of the light leptons;
- Both the electron and the muon must be matched to the corresponding trigger candidates;
- The muon is required to have $p_T > 20$ GeV if the trigger accepting the event has a 17 GeV threshold on the muon candidate;
- The electron is required to have $p_T > 20$ GeV if the trigger accepting the event has a 17 GeV threshold on the electron candidate;
- The electron is required to pass the “loose” MVA identification working point;
- The electron $\Delta\beta$ -corrected relative PF isolation in a cone of $\Delta R < 0.4$ has to be less than 0.15 (0.1) for candidates with $|\eta| < 1.479$ ($|\eta| > 1.479$);
- The muon $\Delta\beta$ -corrected relative PF isolation in a cone of $\Delta R < 0.4$ has to be less than 0.15 (0.1) for candidates with $|\eta| < 1.479$ ($|\eta| > 1.479$);

- The hadronic tau candidate is required to pass the loose HPS combined isolation working point to suppress the backgrounds with jets misidentified as τ_h .
- To suppress $Z \rightarrow e^\pm e^\mp + \text{jet}$ background with an electron misidentified as the hadronic tau and a jet as the muon, events with $|M_{e\tau} - M_Z| < 20 \text{ GeV}$, where $M_{e\tau}$ is the invariant mass of the electron and the hadronic tau and $M_Z = 91.2 \text{ GeV}$, i.e. the mass of the Z boson, are required to contain a tau passing the medium electron rejection working point, otherwise the tau is required to pass the loose one;
- To suppress $Z \rightarrow \mu^\pm \mu^\mp + \text{jet}$ background with a muon misidentified as the hadronic tau and a jet as the electron, events with $|M_{\mu\tau} - M_Z| < 20 \text{ GeV}$, where $M_{\mu\tau}$ is the invariant mass of the muon and the hadronic tau, are required to contain a tau passing the tight muon rejection working point, otherwise the tau is required to pass the loose one;

$ee\tau_h$

- The event must pass the lowest-threshold unprescaled double electron trigger;
- Both the electrons must be matched to the corresponding trigger candidates;
- The leading electron must have $p_T > 20 \text{ GeV}$, to be compliant with the online thresholds, and pass the tight MVA electron identification working point;
- The $\Delta\beta$ -corrected relative PF isolation of leading electron, computed in a cone of $\Delta R < 0.4$ is required to be less than 0.15 (0.1) for electrons with $|\eta| < 1.479$ ($|\eta| > 1.479$);
- The sub-leading electron must pass the loose MVA electron identification working point and its $\Delta\beta$ -corrected relative PF isolation in a cone of $\Delta R < 0.4$ is required to be less than 0.2 (0.15) for electrons with $|\eta| < 1.479$ ($|\eta| > 1.479$);
- The hadronic tau is required to pass the loose HPS combined isolation working point, the loose electron rejection working point and the tight muon rejection working point;
- To suppress $Z \rightarrow e^\pm e^\mp + \text{jet}$ background with an electron misidentified as the hadronic tau and a jet as electron, events with $|M_{e_{1,2}\tau} - M_Z| < 10 \text{ GeV}$, where $M_{e_{1,2}\tau}$ is the invariant mass of the hadronic tau and either electron, are required to contain a tau passing the tight electron rejection working point.
- In events with $10 \text{ GeV} < |M_{e_{1,2}\tau} - M_Z| < 20 \text{ GeV}$ the tau is required to pass the medium electron rejection working point.

- This channel suffers from an additional source of background that is due to DY events in which the charge of either electron is mis-measured. A detailed description of the studies conducted to model this background is presented in the following section of the note. In order to suppress this type of background events with $|M_{ee} - M_Z| < 10$ are rejected.

In all the channels the three leptons are required to be separated by a ΔR distance of at least 0.4, in order to avoid that their isolation cones overlap.

In addition to the selection criteria given above, the electrons and muons in the events are required to be of the same charge. This requisite greatly suppresses the background contamination from $Z/\gamma \rightarrow \ell^\pm \ell^\mp + \text{jet}$ and $t\bar{t}$ in which a jet is misidentified as hadronic tau.

To reduce the contamination from ZZ and $t\bar{t}$ backgrounds, events with additional isolated electrons, muons, hadronic taus or b -tagged jets are rejected. None of these vetoes but the b -jet one have a significant effect on the signal efficiency of the analysis. The b -jet veto was found to degrade the final expected limit by about 4%, but is nevertheless included to avoid any overlap of the signal region with searches in the $t\bar{t}H$ channel [72].

The offline event selection has been optimized for each channel separately, maximizing the expected significance to discover a 125 GeV SM Higgs boson signal. To perform this maximization, the full background estimation procedure, detailed in the following sections, has been used with a simpler systematics set-up. Electron identification and lepton isolation have been optimized simultaneously, while the $Z \rightarrow \ell\ell + \text{jet}$ background rejection has been optimized separately. The electron charge mis-identification rejection has been optimized in a separate procedure, considering only the $ee\tau_h$ channel.

4.2 Background estimation

The main background processes can be divided into two types: reducible (or fake) and irreducible backgrounds. The former category includes all those processes in which at least one of the final state leptons is mis-identified. The main process contributing to this category is the associated production of W and jets, where two jets are mis-identified as a light lepton and a hadronic tau, respectively. Further relevant background contributions arise from QCD multi-jet, $t\bar{t}$ and Z +Jets production. The Z +Jets contamination in particular needs a more detailed discussion. As the two light leptons are required to have same charge, only two configurations are possible: in the first case one of the leptons from the Z decay is mis-identified as a hadronic tau while a jet is mis-identified as light lepton; in the second case one of the leptons is mis-reconstructed with the wrong charge. This effect is only relevant for electrons and is discussed in Section 4.2.6. The irreducible background is composed only by the dominant WZ production, which has the same signature as the process under study, and the ZZ one, in case one of the light leptons fails to be reconstructed.

4.2.1 Irreducible backgrounds

The irreducible backgrounds are modeled by the Monte Carlo simulation. The hard scattering process amplitudes are generated by MADGRAPH [73], a leading order (LO) matrix element generator. The decay of long-living particles originating from the parton scattering and the hadronization process of quarks and gluons is handled by PYTHIA [74]. PYTHIA also adds additional jets to simulate the presence of the underlying event. This process, as well as the hadronization, is performed with the aid of empirical parton fragmentation functions. The tau decay, which was found to be imprecisely simulated by PYTHIA, is simulated by TAUOLA [75], a dedicated library.

The presence of pileup in MC events is simulated by adding simulated *minimum bias* events to the generated “hard” process. The distribution of pileup events in the simulated sample is different, and generally generated with higher pileup, with respect to real running conditions. This mis-match was created on purpose to ensure good coverage of the pileup distribution even for running periods posterior to the MC creation. Simulated events are therefore reweighted according to the number of pileup events in the MC and the distribution of pileup events in data. This procedure, called *pileup reweighting*, removes the differences induced by the mis-match in the pileup distribution. The distribution of pileup events in real data is computed based on the luminosity profile and the inelastic p-p scattering cross Section [76].

Residual differences concerning the trigger selection, lepton identification and isolation are accounted for with dedicated analyses yielding a set of data to MC correction factors and corresponding uncertainty. When differing from unity, these scale factors are applied to simulated samples in the form of an event weight. Finally, events yields are scaled according to the NLO theoretical cross-section prediction [77].

Measured and simulated kinematic distributions are compared in dedicated control regions with $Z \rightarrow \mu\mu, e\mu, ee$ decays, shown in Figure 4.1.

4.2.2 Reducible backgrounds

The reducible background includes a wide range of processes with quark or gluon jets misidentified as leptons. A data-driven background estimation technique is preferred over simulation. In the work presented in this thesis the reducible backgrounds are estimated by means of the “misidentification rate” (or “fake rate”) method.

The method consists of defining “loose” as well as “tight” selections for the lepton that can be faked by the jets, with the tight selection being the same as the final analysis selection used in this work. The misidentification probability for a quark or gluon jet satisfying the loose lepton selections to pass the corresponding tight selection can be parametrized by a function $f(\vec{x})$ of a set of kinematic variables, $\vec{x}$, characterizing the event. This probability is measured in a dedicated control region resembling as closely as possible the signal region, while still being disjunct from it. The misidentification

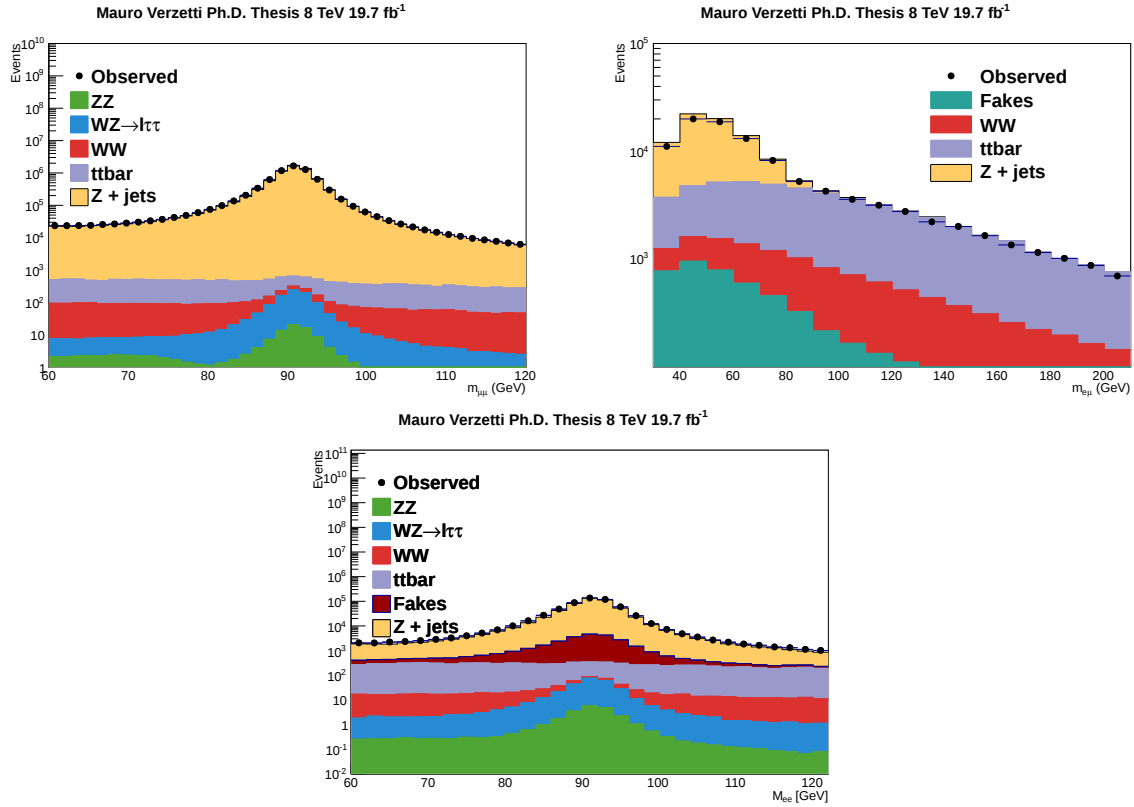


Figure 4.1: Dilepton invariant mass distribution measured with 8 TeV data in dedicated $Z \rightarrow \mu\mu$ (top left), $Z \rightarrow \tau\tau \rightarrow e\mu$ (top right), and $Z \rightarrow ee$ (bottom) control regions. Data are represented by solid points while the expectation is given by the stacked histogram. The background contributions labeled “Fakes” is estimated as described in Section 4.2.3.

probability is then applied as weight to the events passing all selection criteria of the signal region except the tight lepton requirements. The weighted distributions obtained in this way represent the background contribution expected in the signal region.

We compare the background estimates obtained by applying the method to the light leptons and to the tau leg. The former method is more inclusive as it accounts for all the sources of reducible backgrounds while the latter ignores those in which an isolated light lepton is mistakenly identified as a hadronic tau candidate. For these processes (i.e. Z +jets, WW , and part of $t\bar{t}$) the latter method relies on MC simulation.

4.2.3 Fake lepton method

Control region definition

The lepton misidentification probability is measured in a W +jets enriched control region. This region differs from the signal one by requiring the transverse mass of the leading $\ell - \cancel{E}_T$ system to exceed 55 GeV, and vetoing events with more than two identified electron or muon candidates in the event. Events containing a well identified hadronic tau are vetoed as well, to remove any overlap with the signal region. In order to mimic the presence of a hadronic tau, the event is required to contain an additional jet with p_T above 20 GeV.

Events selected in the control regions described above are used to train a k-Nearest Neighbor classifier (kNN) [57], which allows the parametrization of the misidentification probability as function of multiple variables. All the events selected in the control region are used to train the kNN. Each parameter (dimension) is inversely scaled by its variance over the training sample, so that the (scaled) parameters are approximately normally distributed. Leptons passing the tight lepton selection are marked as signal, those failing as background. To find the misidentification probability at a particular point in the parameter space, the classifier searches for the k nearest objects to the given point in the training sample. The predicted efficiency is then the number of signal events among the k neighbors and divided by k .

The input variable used for the training are the lepton p_T , the p_T of the jet associated to the lepton, and the number of jets with $p_T > 20$ GeV in the event. The first two variables, especially the jet p_T , are found to be strongly correlated to the probability for a jet to be mis-identified as a lepton. In particular, the difference of these two variables is linked to the isolation value of the lepton. In the training of the muon misidentification probabilities the logarithm of the jet p_T is used instead of the jet p_T to reduce the steep dependence of the probability with respect to this variable, allowing a larger number of neighbors to be used. The number of jets was found to improve the description of the background shape. This variable is particularly effective in separating the different processes underlying the reducible background (W , multi-jet, and $t\bar{t}$). Another peculiarity of this variable is that it is bound to be ≥ 1 in the training sample, while it can be

zero in the main analysis. The difference is resolved by adding the number of hadronic taus in the event to the number of jets, resulting in a variable with the same range as the training variable.

The contribution of genuine isolated leptons from WZ and ZZ events in the control region is removed using simulated MC events and training dedicated kNN classifiers on them. The output of the diboson kNN's is then scaled by:

$$\frac{N_{training}^{MC} \mathcal{L}^{data}}{N_{training}^{data} \mathcal{L}_{eff}^{MC}} \quad (4.1)$$

where $N_{training}^i$ is the number of events used for the training of either data or MC and $\mathcal{L}^i$ denotes the corresponding integrated luminosities. The output of the diboson kNN is finally subtracted to the output of the kNN trained on data.¹

kNN optimization and systematic uncertainties

The only adjustable parameter in the kNN training is the number of neighbors used to extract the probability. The optimal number of neighbors depends on the probability distribution of the different observables (more or less steeply varying) and on the statistics of the training sample. A large number of neighbors yields very accurate results, at the price of a reduced ability to describe abrupt changes in the probability function.

It is therefore important to define a procedure that determines the optimal number of neighbors to be used for each sample as function of the tight working point employed in the final analysis. This evaluation is performed with a bootstrapping technique. Each training sample is initially split into two sub-samples, one containing 90% of the events and the other the remaining 10%. In order to remove any bias due to the run period (different pileup conditions) the smaller dataset is evenly sampled from the full dataset. This procedure is repeated ten times to obtain a set of mutually exclusive pairs of “large” and “small” samples. Subsequently, a training of the kNN classifier is performed on each large sample and applied to the small one. This procedure ensure a total decoupling between testing and training samples. The small size difference between the large sample and the full dataset also ensures that the results inferred from the major dataset are still statistically valid for the full one. The outcome of the ten training and testing cycles is then combined.

The full process is repeated to scan over a wide range of values for the number of neighbors, k . A strong limiting factor for this method is the significant computing resource usage, as the computing complexity of training and testing of the kNN algorithm is proportional to the number of neighbors used.

¹Another equivalent option would have been to directly inject the simulated events in the training sample with negative weight proportional to the luminosity difference, but this method has been found to yield inconsistent results, most likely due to a wrong implementation in the TMVA framework.

Table 4.1: Minimization points for each kNN scan and relative uncertainties as obtained from a parabolic fit

channel	lepton	-1σ	central value	1σ
7 TeV				
$\mu\mu\tau_h$	leading muon	15	20	60
	sub-leading muon	11	20	49
$e\mu\tau_h$	muon	7	20	30
	electron	634	1000	1807
$ee\tau_h$	leading electron	35	200	686
	sub-leading electron	381	1200	1723
8 TeV				
$\mu\mu\tau_h$	leading muon	28	40	73
	sub-leading muon	30	40	63
$e\mu\tau_h$	muon	16	20	30
	electron	664	1400	2136
$ee\tau_h$	leading electron	218	400	552
	sub-leading electron	605	1000	2201

For each choice of k , the shape of important physics observables is computed based on the weighted sum of all the events in the small samples and from the passing events. The two shapes are compared using as figure of merit the χ^2 compatibility between the two shapes. The physical observable used for the χ^2 minimization is the scalar sum of the p_T of the two leptons plus the jet, similar to the L_T variable used further on in the analysis. Other variables were also considered and found to yield similar conclusions concerning the optimal choice of k .

The minimum χ^2 value obtained by the scan is taken to define k for training the final kNN algorithm used in the analysis. A parabolic fit around the minimum is performed to estimate the one sigma confidence interval as the value for which the parabolic fit yields the value of $\chi^2_{min} + 1$. This confidence interval reflects both the statistical uncertainties on the training sample and the related uncertainty on the choice of the training parameters. For each channel and each lepton, three classifiers are trained with the corresponding central values and one sigma boundaries.

A summary of the chosen minima and of the confidence interval boundaries is given in Table 4.1. A graphic representation of a sample scan is shown in Figure 4.2. The misidentification rate observed in the training sample is displayed in Figure 4.3, overlaid with the output of the respective kNN classifier.

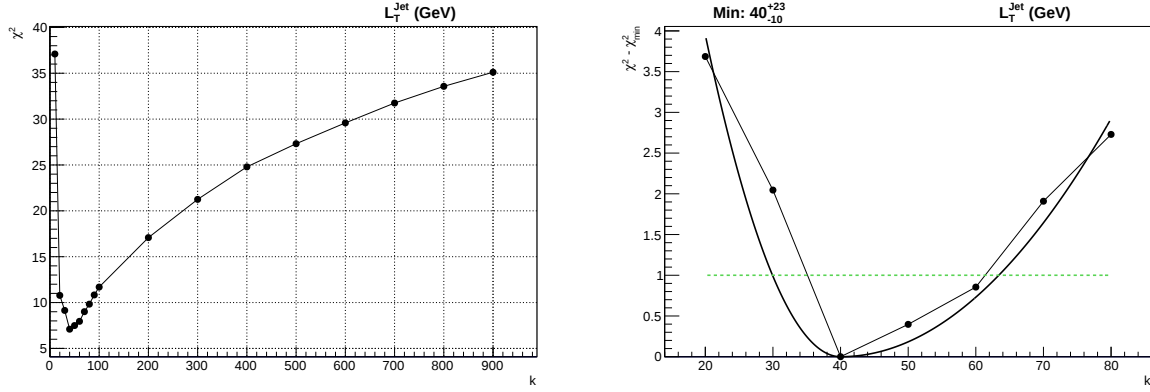


Figure 4.2: χ^2 as a function of the number of neighbors used in the kNN algorithm for background estimation (left) and corresponding fit (right) for sub-leading muons in the $\mu\mu\tau_h$ channel. The variable used for the scan is the scalar sum of the p_T of the two leptons and the jet

Background shape extraction

The contribution to the signal region from backgrounds containing a given fake object is estimated using an “anti-isolated” sideband in which the loose, but *not* the tight, selection requirement is satisfied for that object. Events can therefore be divided into three categories:

1. Events in which the first lepton fails the tight requirements are weighted by $w_1(\vec{x}) = f_1(\vec{x})/(1 - f_1(\vec{x}))$
2. Events in which the second lepton fails the tight requirements are weighted by $w_2(\vec{x}) = f_2(\vec{x})/(1 - f_2(\vec{x}))$
3. Events in which the both lepton fail the tight requirements are weighted by $w_1(\vec{x}) \cdot w_2(\vec{x})$

Each weighted category represents the contribution to the signal region of the backgrounds in which jets fake one or both the light leptons. As the events in which both reconstructed leptons are due to jet $\rightarrow \ell$ fakes are included in category 1 as well as in category 2, it is necessary to remove the double-counting by subtracting the weighted events of category three, obtaining:

$$N_{bkg} = N_1 + N_2 - N_3 \quad (4.2)$$

where N_{bkg} denotes the final estimate for the number of background events in the signal region and N_i denotes the scaled number events in category i .

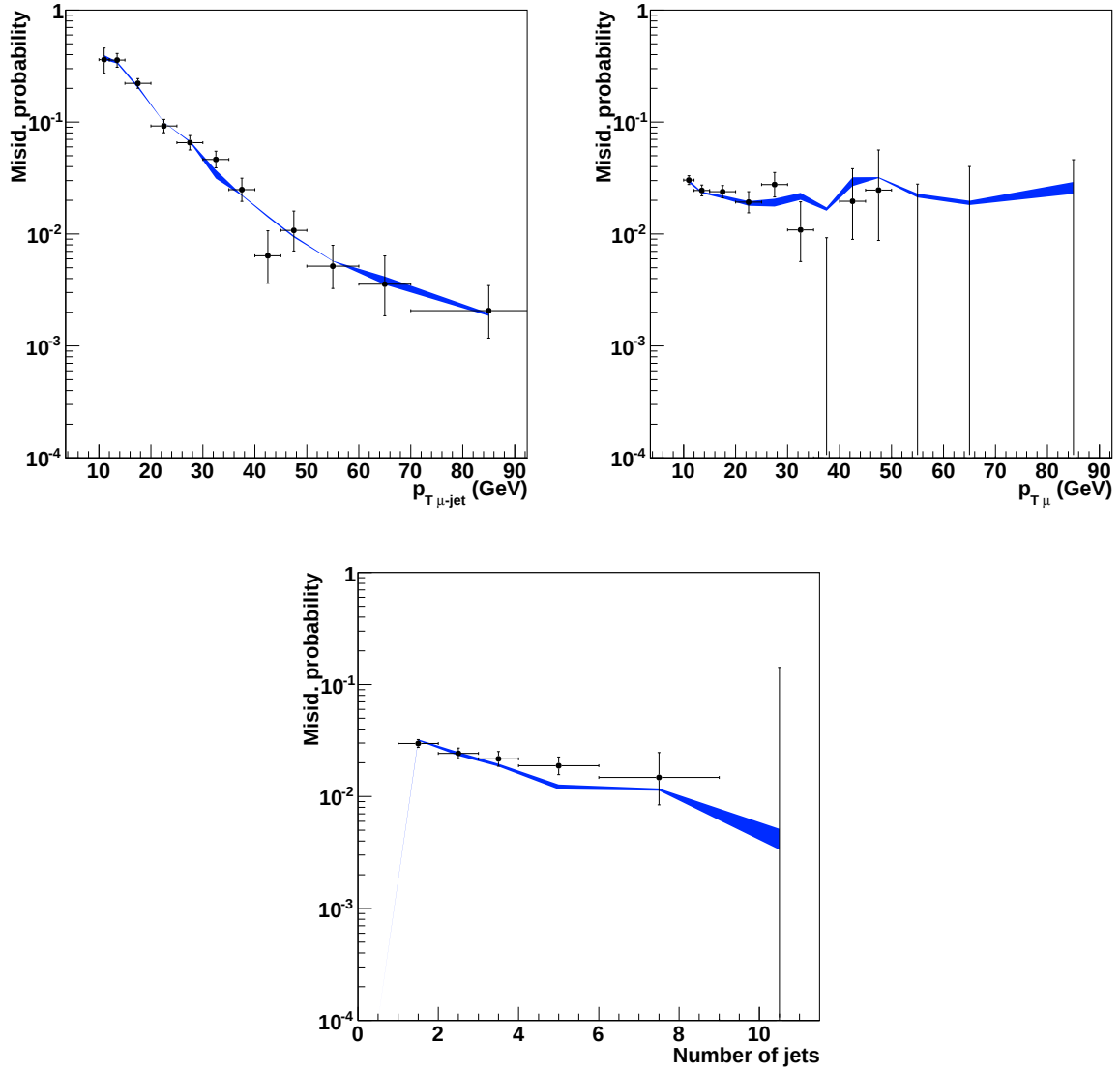


Figure 4.3: Measured misidentification probability for sub-leading muons in the training sample dedicated to the $\mu\mu\tau_h$ channel. The blue band represents the output of the kNN algorithm and its related uncertainty.

Uncertainties on the kNN training, evaluated as described above, are propagated to the background estimate for the signal region by comparing the event weights obtained with classifiers trained with the central value and with the two one-sigma boundaries. The difference is taken as systematic uncertainty.

In rare cases, especially in the tail of kinematic distributions, statistical fluctuations cause the number of weighted events in category three to exceed the sum of $N_1 + N_2$, and therefore the expected yield in the signal region becomes negative. In these cases the yield is set to the value of category three only, relying on the much larger statistics present in that sideband.

4.2.4 Validation of the fake lepton method

The method for estimating the reducible backgrounds is validated in two control regions which are dominated by fake τ_h candidates. These control regions provide an independent validation, as they are naturally exclusive with respect to the event sample used to train the kNN. Events in the control region are selected by maintaining the light-lepton same-charge requirement, but inverting the isolation requirement of the τ_h candidate and in one of the regions the charge of the tau is furthermore required to be the same as the leptons. These regions are dominated by W +Jets and $t\bar{t}$ backgrounds. Comparison between the observed and predicted kinematic distributions in the control regions is shown for the $\mu\mu\tau_h$, $e\mu\tau_h$ and $ee\tau_h$ channels in Figures 4.4–4.9 respectively. Reasonable agreement is observed, considering that the expected background in this region lacks an additional systematic uncertainty, described in the next section and not defined in this region, which has sizable effect both on the distribution and total normalization of the reducible background.

4.2.5 Fake tau method

To better estimate the solidity of the “fake lepton” method against differences in composition of the training sample and the sample to which the method is applied, a second background estimate is performed. This methodology is very similar to the previous one, but the object considered as possibly mis-reconstructed is the hadronic tau.

Region definition

The misidentification probability for τ_h candidates is measured in W +Jets events, requiring:

- Events must pass the single muon trigger;

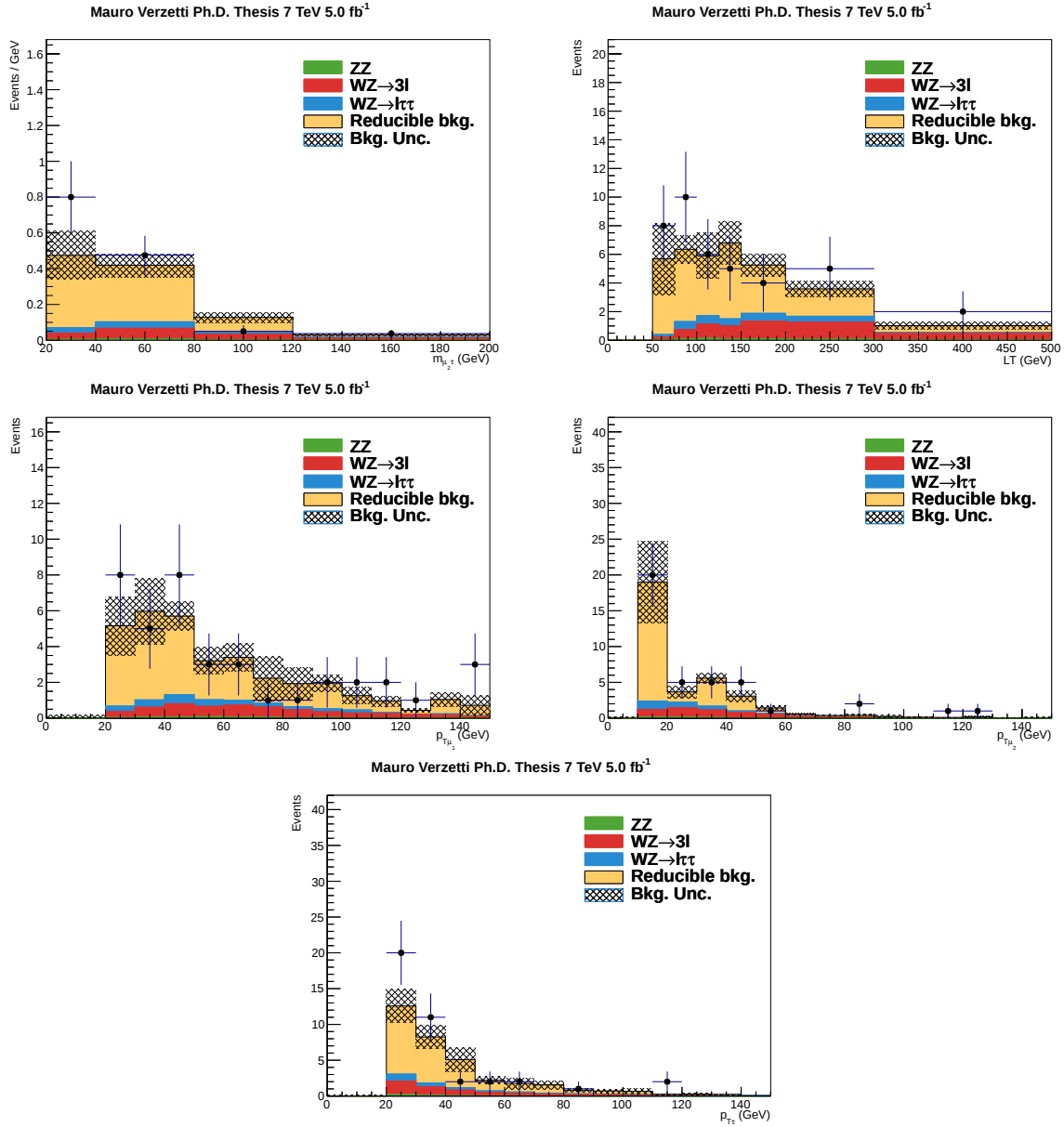


Figure 4.4: Comparison of measured and predicted backgrounds in the $\mu\mu\tau_h$ “fake tau” control region for 7 TeV data. From top left to bottom: mass of the sub-leading muon and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading muon, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

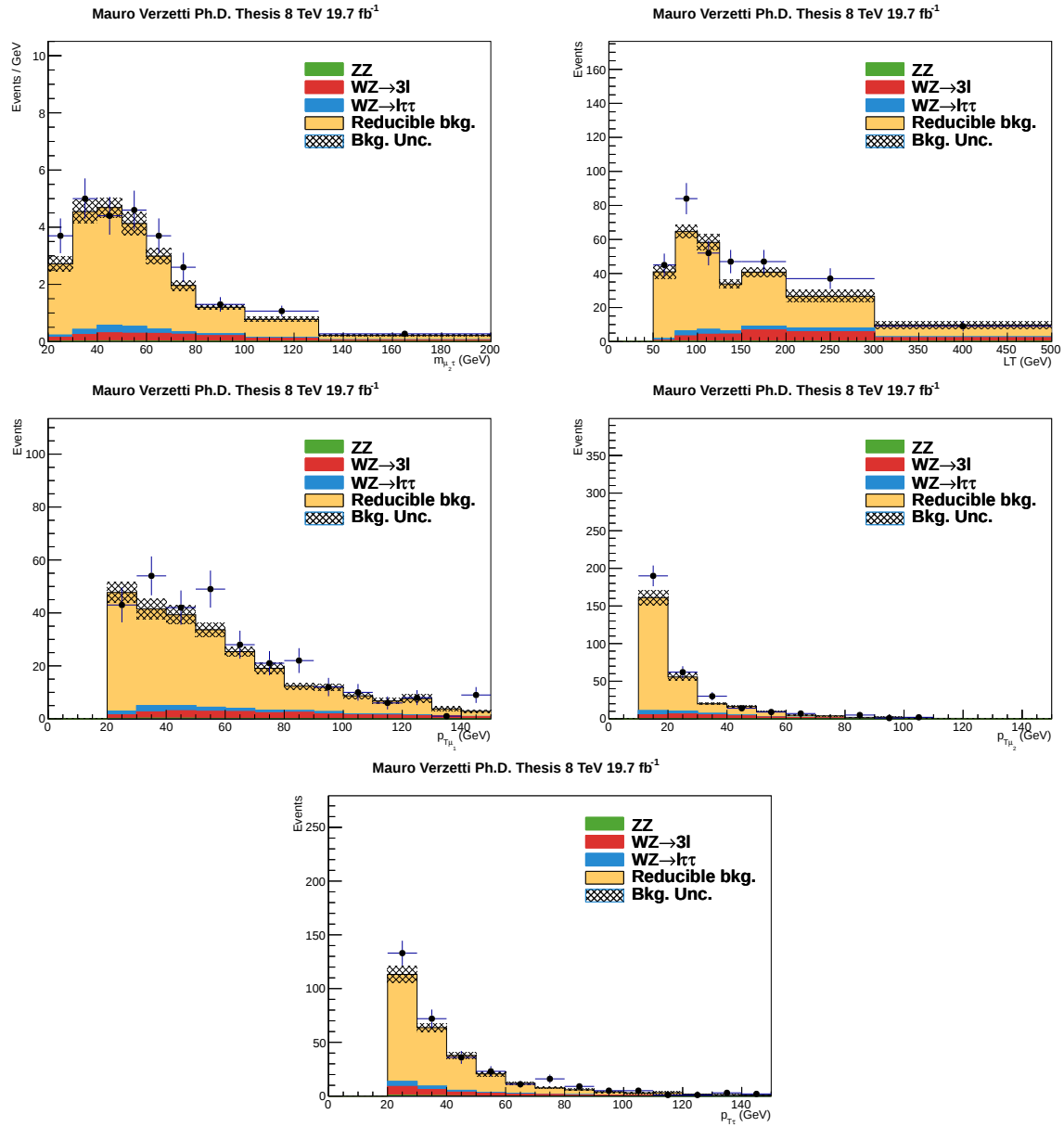


Figure 4.5: Comparison of measured and predicted backgrounds in the $\mu\mu\tau_h$ "fake tau" control region for 8 TeV data. From top left to bottom: mass of the sub-leading muon and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading muon, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

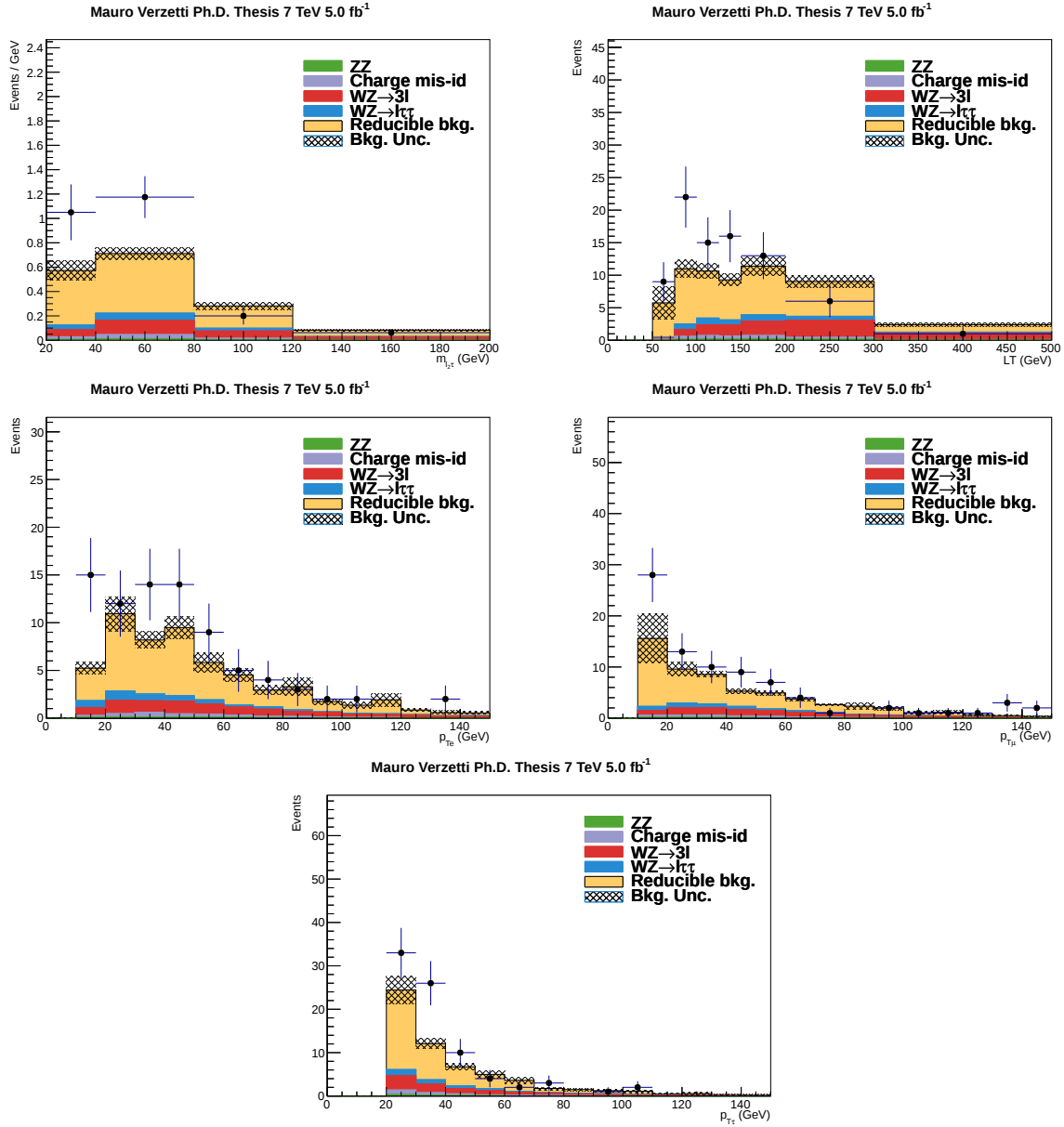


Figure 4.6: Comparison of measured and predicted backgrounds in the $e\mu\tau_h$ “fake tau” control region for 7 TeV data. From top left to bottom: mass of the sub-leading lepton and the tau system, scalar sum of the leptons p_T (L_T), p_T of the electron, muon, and hadronic tau. The reducible background contribution is estimated by the kNN method, as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

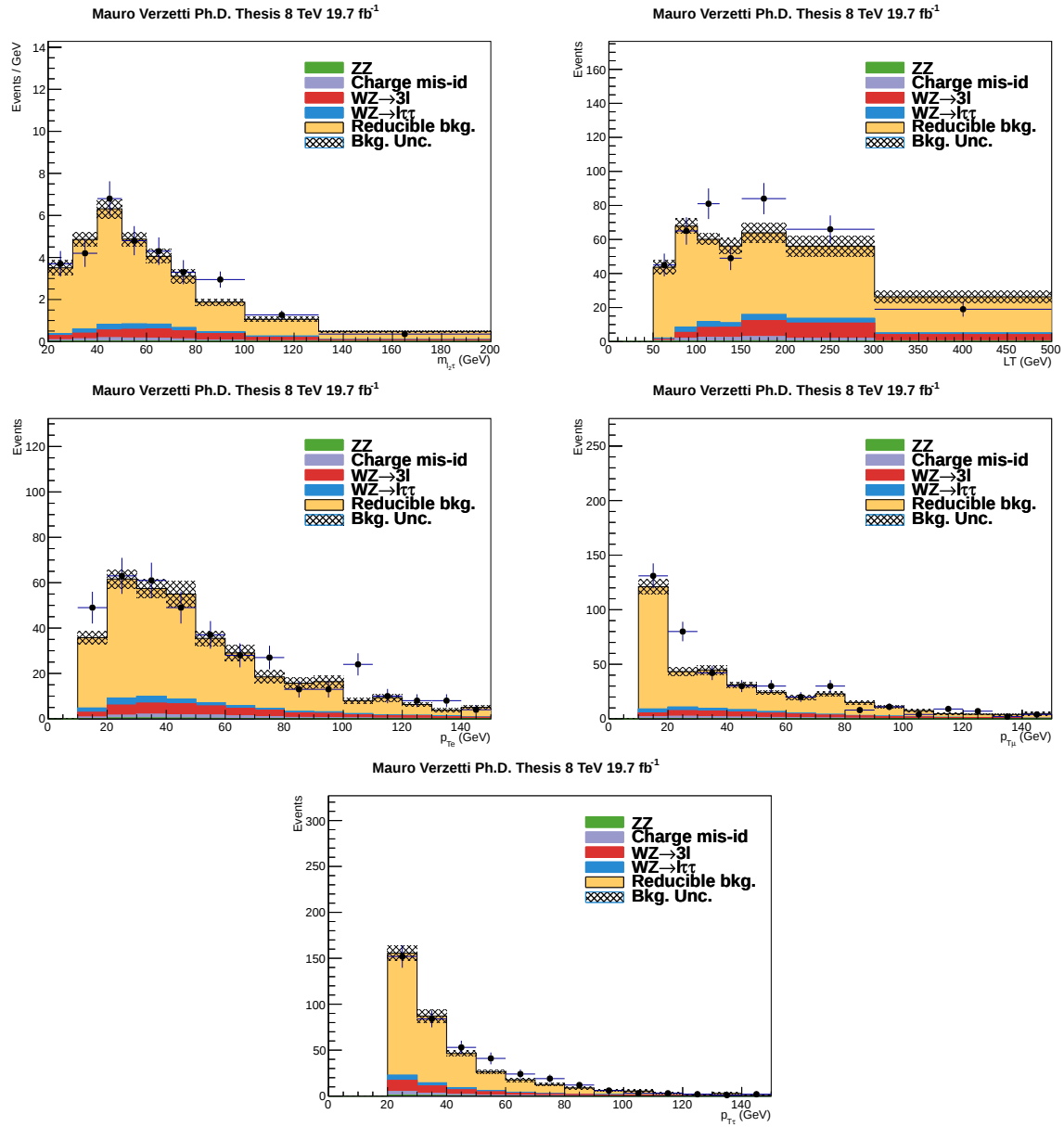


Figure 4.7: Comparison of measured and predicted backgrounds in the $e\mu\tau_h$ "fake tau" control region for 8 TeV data. From top left to bottom: mass of the sub-leading lepton and the tau system, scalar sum of the leptons p_T (L_T), p_T of the electron, muon, and hadronic tau. The reducible background contribution is estimated by the kNN method, as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

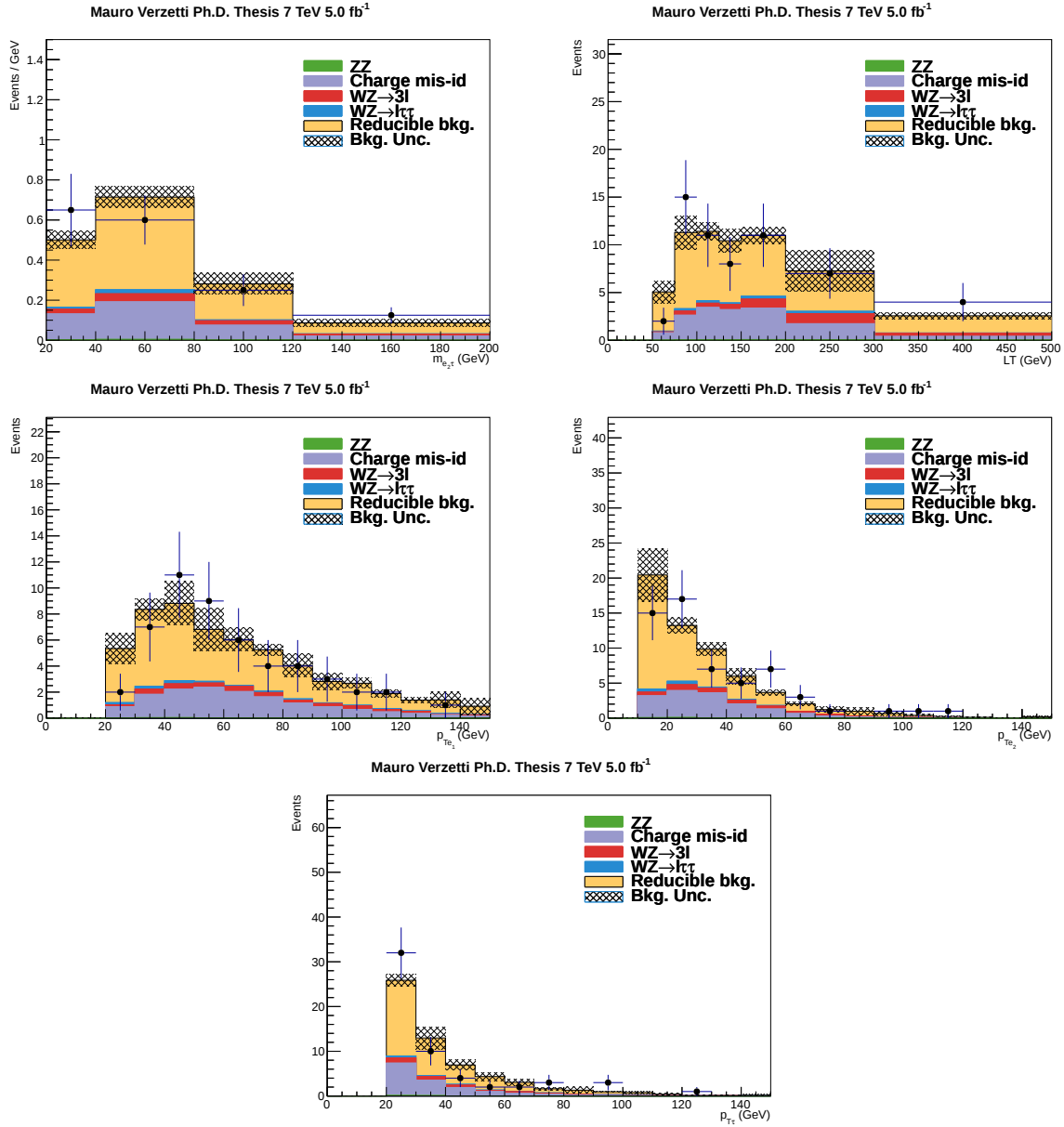


Figure 4.8: Comparison of measured and predicted backgrounds in the $ee\tau_h$ “fake tau” control region for 7 TeV data. From top left to bottom: mass of the sub-leading electron and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading electron, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

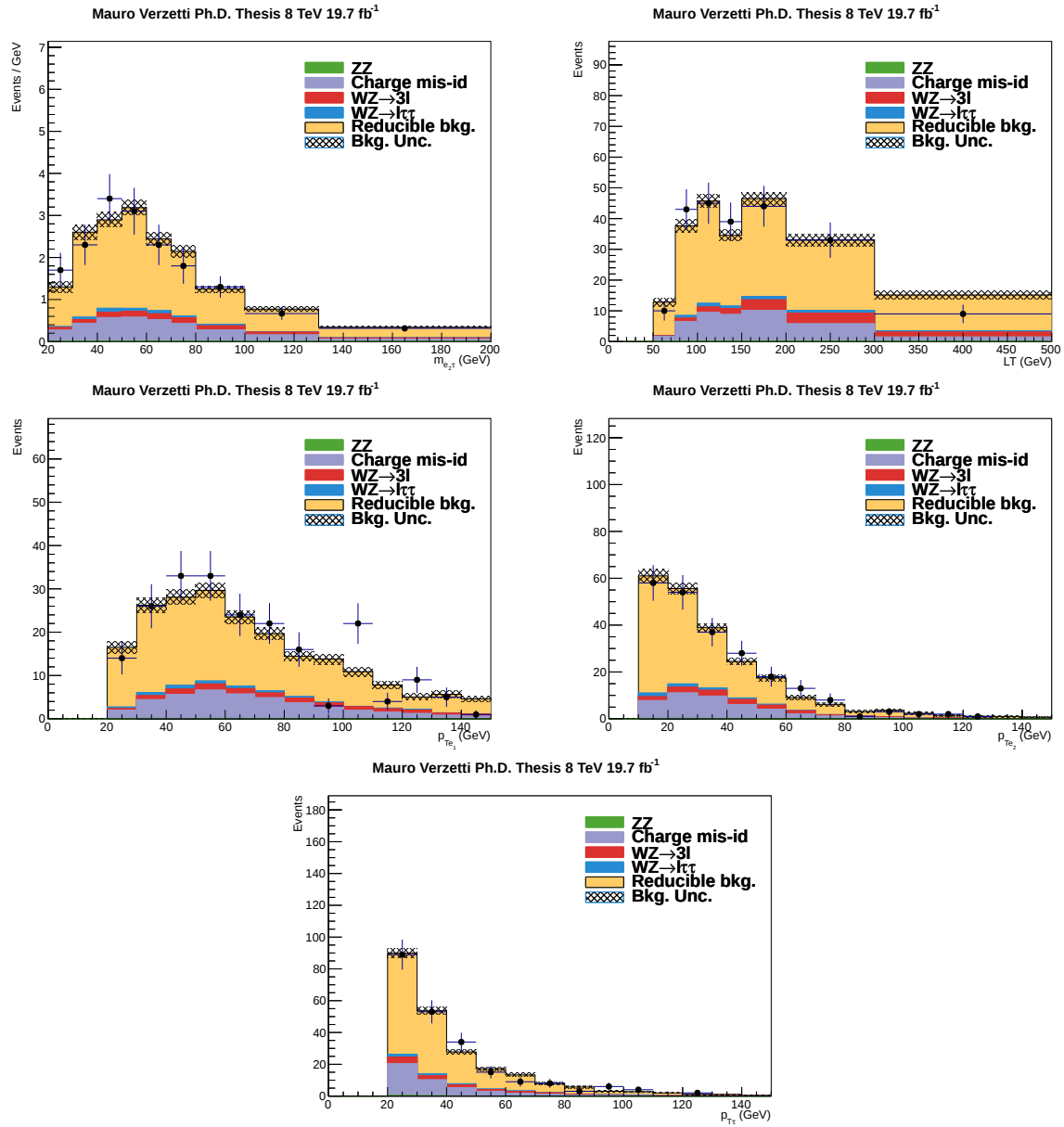


Figure 4.9: Comparison of measured and predicted backgrounds in the $ee\tau_h$ "fake tau" control region for 8 TeV data. From top left to bottom: mass of the sub-leading electron and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading electron, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

- A muon candidate with $p_T > 24$ GeV and $|\eta| < 2.1$;
- The muon must pass the tight Particle Flow identification and have a combined PF Relative Isolation $\Delta\beta$ corrected below 0.1;
- The longitudinal impact parameter of the muon track with respect to the primary vertex must be less than 0.2 cm;
- $M_T(\mu, \cancel{E}_T) > 40$ GeV;
- Two same sign tau candidates are required, both with opposite charge to the muon;
- Events are rejected if they contain another muon or electron with $p_T > 15$ GeV;
- Events are rejected if they contain a jet with $p_T > 20$ GeV and $|\eta| < 2.4$ satisfying the loose CSV b-tagging working point.

The tau fake-rate is measured individually for the 2011 and 2012 run periods in three regions of pseudo-rapidity ($|\eta| < 0.8$, $0.8 < |\eta| < 1.6$ and $1.6 < |\eta| < 2.3$) and as a function of the tau p_T . The misidentification probability is modeled by a Landau function. The width and peak parameters are adjusted to best match the measured $\text{jet} \rightarrow \tau_h$ fake-rates.

The fake rates measured in the 7 TeV and 8 TeV data are shown in Figure 4.10. The $\text{jet} \rightarrow \tau_h$ fake-rate ranges from 6% at low tau p_T in the end-cap region, to 1% at high p_T (> 80 GeV) in the central region.

Background shape extraction

The tau misidentification probability is used to determine a second and completely independent estimate of the reducible background in the signal region. In this method the tau is considered the possibly misidentified object. Events passing all the analysis selection criteria plus the loose tau selection but *not* the tight one are weighted according to the measured tau misidentification probability to estimate the reducible background.

This method only takes into account those processes in which jets are misidentified as taus. The method cannot account for the Drell-Yan, WW and $t\bar{t}$ reducible backgrounds in which the reconstructed τ_h corresponds to a genuine hadronic tau or to an isolated light lepton faking a tau. To account for these types of background the method described in Section 4.2.3 is applied to simulated events. This procedure is considered to be more reliable than taking the MC expectation in the signal region directly.

The fake tau method was found to produce results consistent with the fake-rate method applied to leptons. To account for possible residual differences, the fake lepton method is used as nominal value for the reducible background and the estimate obtained from the fake-tau method is used as systematic uncertainty on the background shape.

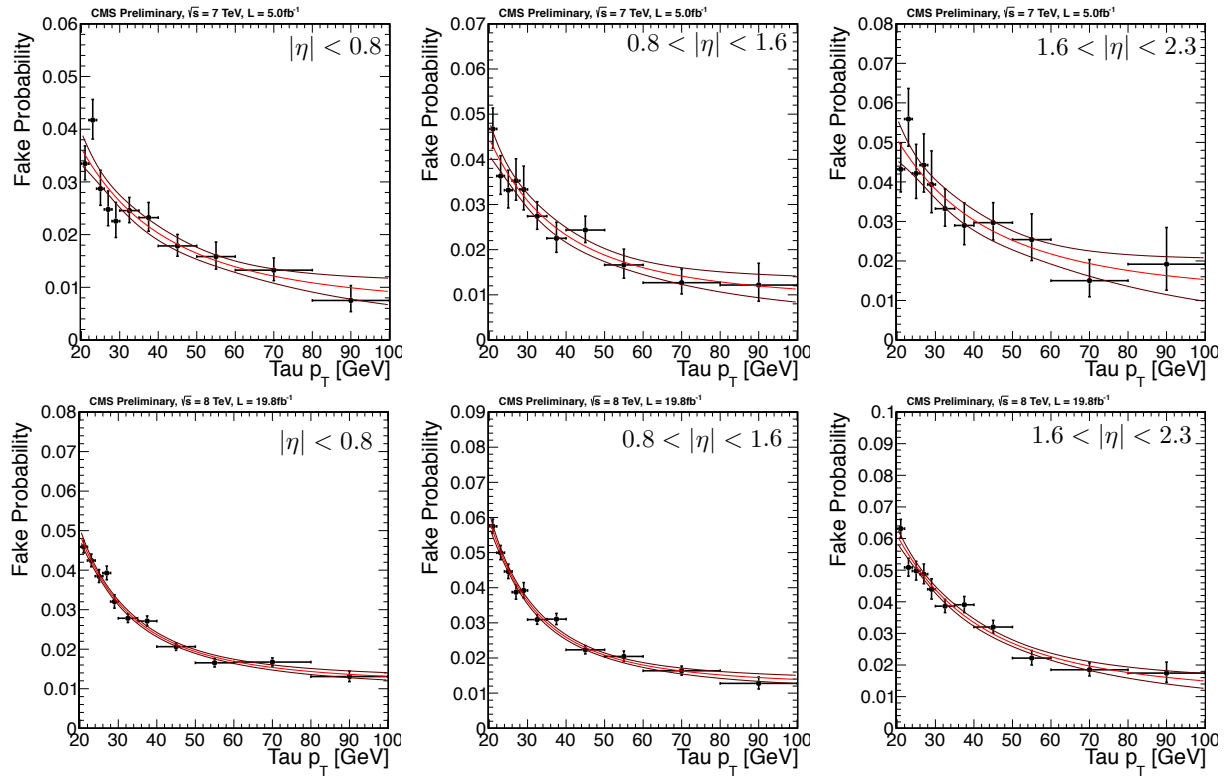


Figure 4.10: Jet $\rightarrow \tau_h$ misidentification probability in 7 TeV (top) and 8 TeV (bottom) data, in $W \rightarrow \mu + \text{jet}$ events, as function of tau p_T in barrel (left), transition (center) and endcap (right) regions. The red line shows the result fit to the data.

4.2.6 Electron charge mis-identification

The most common reason for the mis-measurement of the electron charge inside the CMS detector is the interaction of such particle with material in the tracking detectors. An electron can radiate a photon, carrying most of its momentum, which may subsequently convert into an e^+e^- pair with highly unbalanced momenta, producing an electron with charge opposite to the initial one that carries most of the initial energy. In case the bremsstrahlung and the photon conversion processes occur in proximity to the interaction point, the reconstruction algorithm cannot identify the electron as originating from a conversion. The description of this process in the simulation is handled by GEANT [78]. The probability for an electron to be reconstructed with opposite charge is measured as a function of the electron η and p_T using a $Z \rightarrow ee$ MC sample. The results are shown in Figure 4.11.

Another effect associated to the charge mis-measurement is a partial energy loss by radiation that fails to get associated to the electron candidate by the reconstruction algorithm. This effect causes a shift of the dielectron invariant mass spectrum. The precision with which this phenomenon is modeled in the simulation is not totally satisfactory. To quantify the effect, the spectra of opposite sign (OS) and same sign (SS) electron pairs were fitted with a Cruijffian function (a two-sided gaussian with exponential tails) allowing only a global shifting factor to float. The results of these fits are shown in Figure 4.12. The process is repeated for both data and simulated events. The value extracted from the former is taken as central value, while the one taken from the latter is used as systematic uncertainty.

The probability for the electron charge mis-measurement is applied as a weight on opposite sign $e^\pm e^\mp \tau_h$ events, which pass the same selection as the signal region, in order to obtain a data-driven estimate for this type of background. The invariant mass distribution obtained with this method is artificially shifted towards lower values by the amount computed in the $Z \rightarrow ee$ data.

4.2.7 Control regions

The background induced by charge misassignment is validated in a dedicated $Z \rightarrow e^\pm e^\pm$ control region. This region has the same selection as the $ee\tau_h$ signal region with the exception that events containing a tau are removed. The di-electron invariant mass spectrum in this region is shown in Figure 4.13. The red histogram represents opposite-sign dielectron events weighted by the charge misassignment probability and with the invariant mass scaled by an offset extracted from MC simulation.

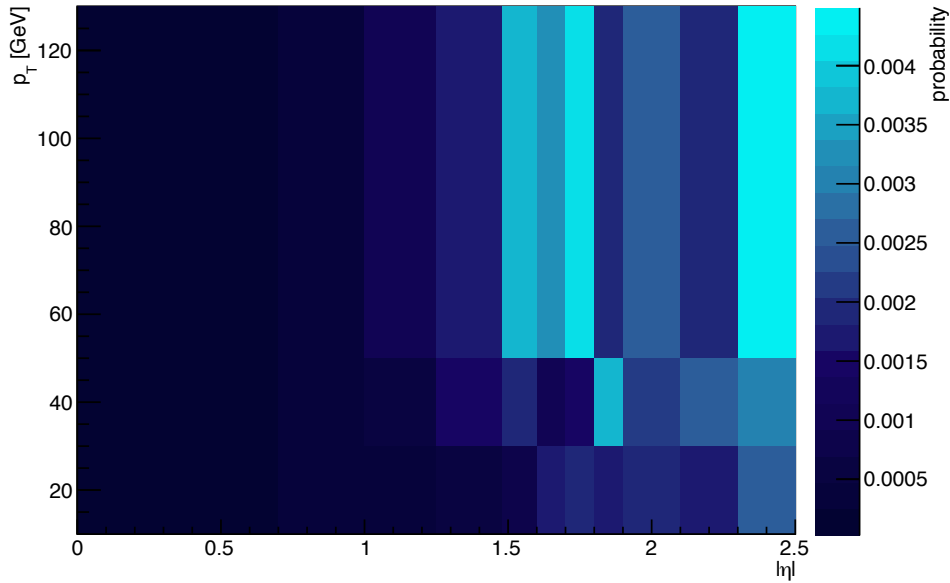


Figure 4.11: Electron charge misassignment probability (color scale) as a function of $|\eta|$ and p_T .

4.3 Systematics uncertainties and category optimization

The optimization of the analysis sensitivity and the description of the systematic uncertainties are two subjects that are tightly bound. The figure of merit for the optimization is the expected signal significance, while considering the full set of uncertainties correlated with both signal and backgrounds. An optimization purely based on the “statistical” significance, defined as $S = N_{sig} / \sqrt{N_{sig} + N_{bkg}}$, may yield an optimal configuration for which the systematic uncertainties are very large.

In this section the definitions and values necessary to have a complete picture of the category optimization process are briefly introduced. The systematic uncertainties are described after the optimization process. A detailed description of how systematic uncertainties are treated to obtain the final result is left for the sections describing the statistical interpretation of the data.

4.3.1 Category optimization

Given the strong dependence of the mis-identification rate on the p_T of the object considered, in particular the hadronic tau, it is quite straightforward to try to improve the signal purity by increasing the threshold on the objects p_T . A better but correlated so-

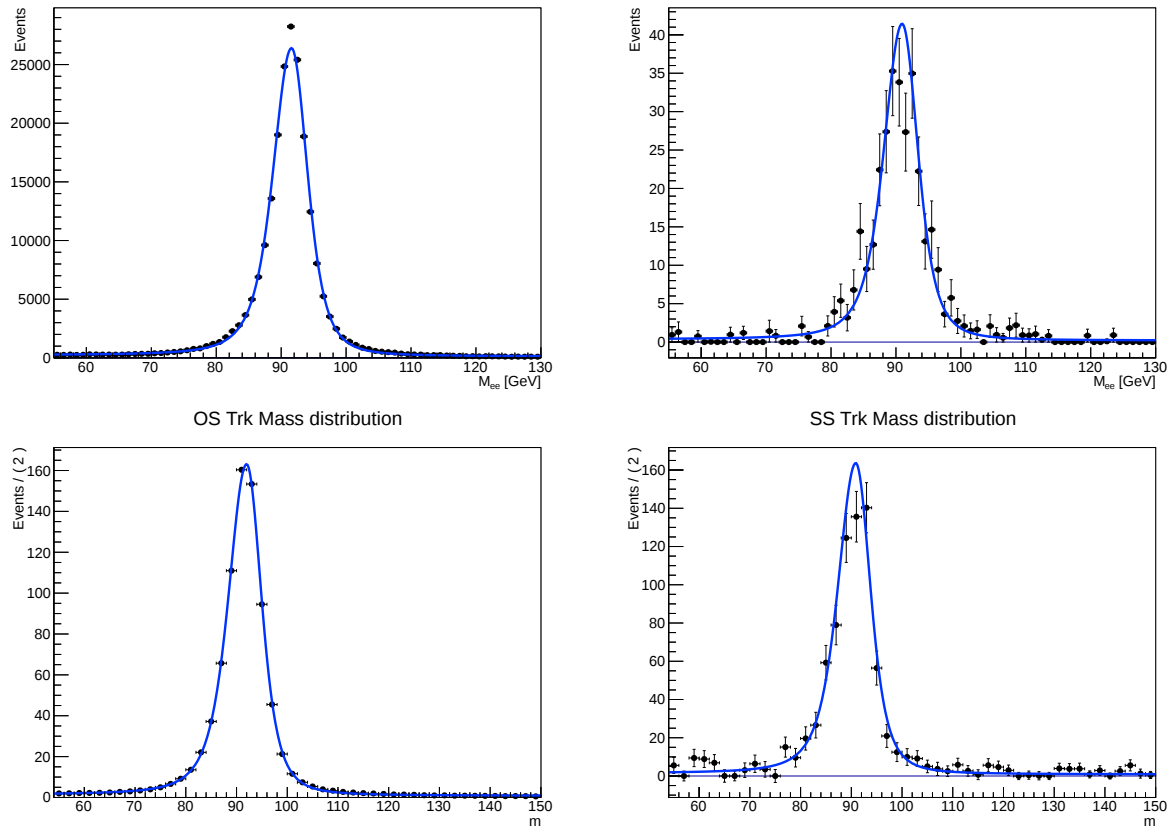


Figure 4.12: Comparison of the fit to dielectron invariant mass distributions for opposite-sign (left) and same-sign (right) dielectron candidates from Drell-Yan MC simulation (top) and $Z \rightarrow ee$ data (bottom). The measured invariant mass scale factor is 0.7% for simulated events and 1.2% for data

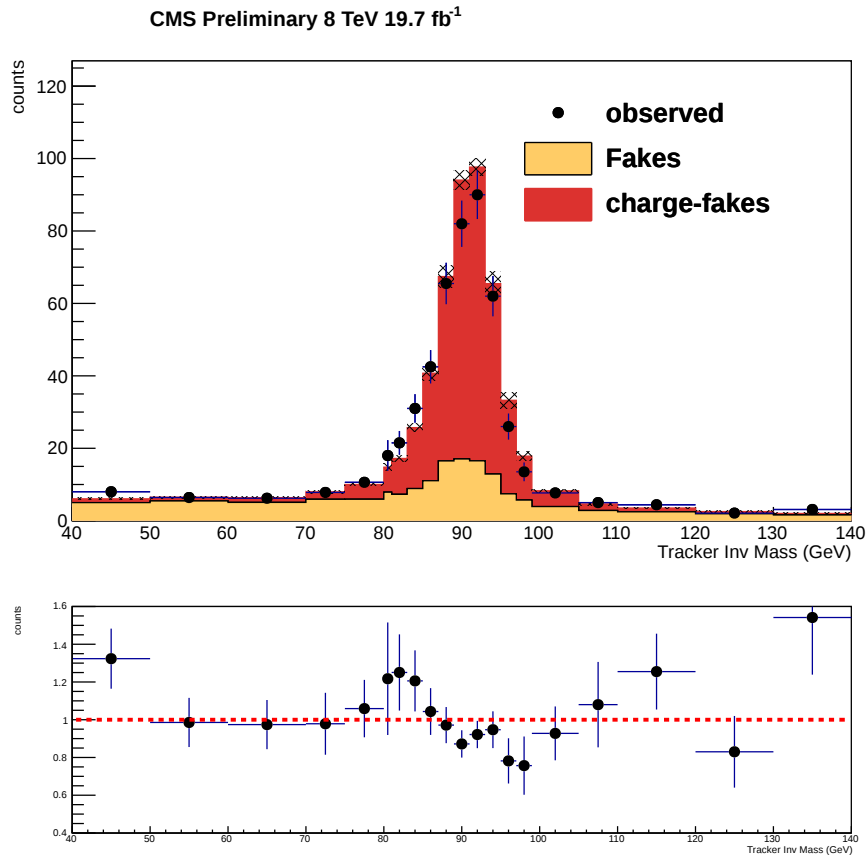


Figure 4.13: Invariant mass distribution of same-sign dielectron events: the red histogram is obtained using the information extracted from simulation (charge misassignment probability and invariant mass scale correction) and applied them to opposite-sign data. The yellow histogram is obtained by applying the fake lepton method to estimate the reducible background.

lution is to use as discriminating observable the scalar sum of the p_T of the three final state leptons, called L_T .

Two types of systematic uncertainties in the background templates are considered in this analysis: normalization uncertainties, which affect only the overall yield of a background, and shape uncertainties, which also affect distribution of the background between bins.

Normalization uncertainties are assigned to account for uncertainties in the luminosity measurement, in the correction factors for the trigger emulation, in the identification and isolation efficiencies and in the theoretical cross-sections for all the simulated processes. Normalization uncertainties are also assigned to account for the limited size of the MC samples and the very limited size of the loose-but-not-tight samples that are used to estimate the reducible backgrounds. The last two sources of background are strongly dependent on the category selection and have to be recomputed for each tested configuration.

The difference between the reducible background shape computed with the fake lepton method and the one evaluated with the fake tau method is accounted for by introducing a shape uncertainty that is allowed to vary both the shape and the normalization of the reducible background. The same procedure is applied to the reducible background shapes computed with the optimal number of neighbors and the two inferred boundaries. Single bin fluctuations in the reducible background shape due to the limited statistics in the anti-isolated regions are separately accounted for by a separate shape uncertainty for each bin (also called *bin-by-bin uncertainties*). Bin-by-bin uncertainties are not allowed to change the total yield of the background process, as this effect is collectively handled by the appropriate normalization uncertainty. It is evident that also these shape uncertainties are dependent on the category selection.

During the optimization two options were tested: either apply a cut rejecting all the events below a certain L_T threshold, or divide the events in two categories (called L_T -high and L_T -low) that get fitted simultaneously, increasing the total sensitivity of the analysis. For each L_T threshold tested, the expected frequentist significance of both options has been computed for a SM Higgs boson of 125 GeV. The L_T scan was performed by varying the cut threshold between 60 and 170 GeV in 10 GeV steps.

The categorization is expected to perform *at least* as well as the equivalent cut option, since the L_T -high category alone has the same statistical power as the corresponding cut case. The categorization choice, however, requires a significant amount of data in both the event categories in order for the optimization to be stable. For this reason the categorization option has not been investigated in case of the 7 TeV data. In the 8 TeV $ee\tau_h$ channel the difference between the two options was found to be negligible and therefore the more simple cut was chosen. Figure 4.14 show the results of the threshold scans for the three channels in the two run periods, while Table 4.2 shows the L_T threshold values that maximize the significance for each channel, the reported selection is used to extract the final result.

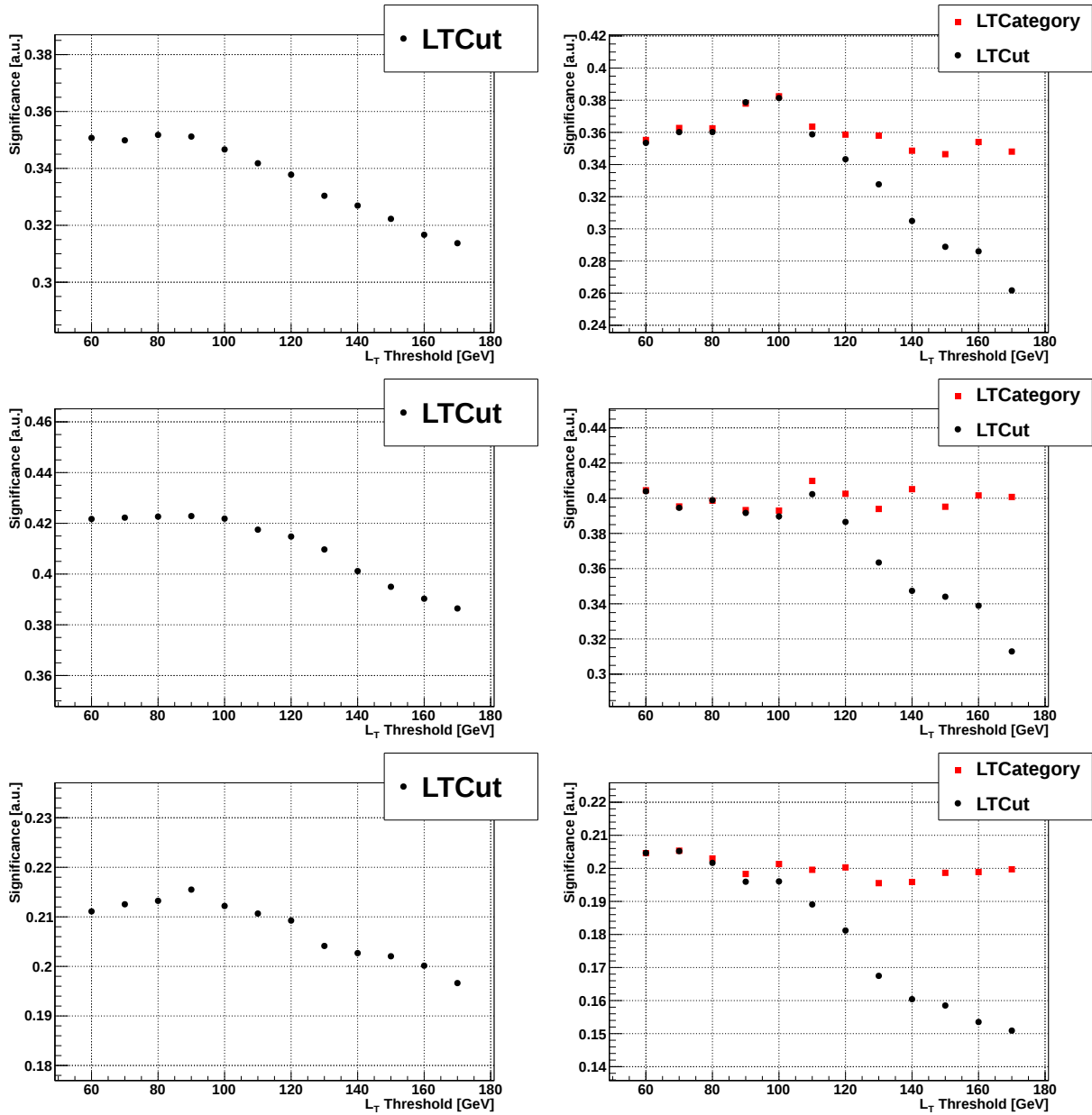


Figure 4.14: Expected significance versus L_T threshold value for the $\mu\mu\tau_h$ (top), the $e\mu\tau_h$ (center), and the $ee\tau_h$ (bottom) channels in 7 TeV (left) and 8 TeV (right) datasets for two optimization options. The L_T threshold is represented by the black dots, the separation of L_T event categories by the red boxes.

Table 4.2: Threshold values and categorization approach used in signal extraction for each channel and dataset.

Run period	Channel	Best approach	Threshold (GeV)
7 TeV	$\mu\mu\tau_h$	L_T Cut	80
	$e\mu\tau_h$	L_T Cut	90
	$ee\tau_h$	L_T Cut	90
8 TeV	$\mu\mu\tau_h$	L_T Category	100
	$e\mu\tau_h$	L_T Category	110
	$ee\tau_h$	L_T Cut	70

4.3.2 Systematic uncertainties

As mentioned previously, the sources of systematic uncertainties accounted for in this work can be divided into two groups: the ones affecting only the total normalization of a process, yielding a normalization uncertainty, and the ones also affecting the distribution of the events, yielding a shape uncertainty.

A 2.2% (2.6%) uncertainty [79] in the total integrated luminosity is applied to the 7 TeV (8 TeV) simulated samples, respectively. This uncertainty is correlated throughout all the simulated processes, categories and channels belonging to the same running period. The theoretical uncertainties on the diboson cross-sections amount to 4% due to parton distribution function (PDF) plus 4% due to the QCD renormalization scale. These uncertainties are correlated between all the diboson processes, regardless of channel, category or running period.

The statistical power of the different simulated samples are accounted for by a normalization uncertainty that varies between 3 and 10% and is uncorrelated between all the channels and categories.

The theoretical uncertainty on the signal cross section due to missing higher orders and due to parton distribution function amounts to 2.7% [80].

The uncertainties on the e and μ identification, isolation, and trigger efficiencies correspond to the uncertainties on the data-to-simulation correction factors, which have been measured via the tag-and-probe technique². The uncertainty on the τ identification efficiency is 6% (see Section 3.5.4). The effect on the simulated yields due to the

²In a similar manner to what has been described in Section 3.5.4, $Z \rightarrow \ell\ell$ events are selected with a single lepton trigger and the desired isolation and identification working points are applied on both leptons. Additionally the invariant mass of the two leptons is required to be within the Z boson mass window: $60 < m_{\ell\ell} < 120$. The trigger efficiencies are computed by applying the trigger requirement to the probe lepton, which does not fire the trigger, both on data and MC. The discrepancies between the two efficiencies, binned in p_T and η of the lepton, are applied as a scale factor on the MC and the uncertainty on such correction are taken as systematic uncertainties. A similar tag-and-probe approach on $Z \rightarrow \ell\ell$ events is taken also to correct for discrepancies in the identification and isolation efficiencies.

uncertainty on the electron and τ_h energy scale has been evaluated. The electron energy scale uncertainty (1% for $|\eta| < 1.5$, 2.5% elsewhere) does not have a measurable effect on the analysis. Varying the τ_h energy scale (within a 3% [34] uncertainty) is found to change the expected yields by 3%, but has no sizable effect on the observables shapes. A 3% normalization uncertainty is therefore included. The uncertainty in the compatibility between simulated and measured efficiency of a light quark or gluon jet to be misidentified as a b-tagged jet, causing the b-veto to fail is propagated to the analysis and found to be covered by a 1% uncertainty in the yields of the simulated samples.

Uncertainties in the reducible background estimation related to the choice of the number of neighbors, k , of the kNN algorithm and in the statistical power of the training sample are computed as described in Section 4.2.3 and propagated to the signal region as a shape uncertainty, which is also allowed to affect the total normalization of the process. The uncertainty is correlated between categories belonging to the same channel and running period.

Uncertainties due to the limited statistical power of the inverted isolation regions that are used to compute the reducible background shape are accounted for by introducing an additional normalization uncertainty that ranges between 12% and 52% plus bin-by-bin shape uncertainties. These uncertainties are uncorrelated between all categories and channels.

Finally, the possible mis-modeling caused by the differences in background processes composition between the kNN training sample and the isolation-inverted region is estimated with the “fake tau” method and applied as a shape uncertainty, uncorrelated between the different channels and categories.

A 4% (7%) uncertainty for each electron candidate in the final state is assigned to account for systematic uncertainties on the charge mis-measurement background yield in addition to the statistical power of the sample for 8 TeV and 7 TeV data, respectively. A 42% (89%) uncertainty is assigned to the invariant mass scaling factor (described in Section 4.2.6) that is applied to the shape templates for the charge mis-measured background in the 8 TeV and 7 TeV data, respectively. The uncertainty is applied as a shape uncertainty with no normalization effect.

A summary of the sources of systematic uncertainties, their values, and their treatment is given in Table 4.3.

4.4 Results

The checks and optimizations have been performed on control regions completely orthogonal to the signal region, or simply based on the expected distributions of known backgrounds. This procedure, called “blind analysis”, is aimed at reducing as much as possible any bias originating from the choices done by the analyst. Once the analysis is considered final, the distribution of observed events in the signal region of each

Table 4.3: Systematic uncertainties affecting the $\ell\ell\tau_h$ categories. Values specific to the 2011 datasets are indicated in paranthesis.

Source	Value	Type	Correlation
Luminosity	2.2%(2.6%)	norm.	correlated between channels
σ_{WZ}	4% $\oplus$ 4%	norm.	fully correlated
σ_{ZZ}	3.3% $\oplus$ 2.3%	norm.	fully correlated
σ_H (PDF)	2.7%	norm.	fully correlated
diboson MC statistics	[3% – 9%]	norm.	fully uncorrelated
signal MC statistics	[4% – 12%]	norm.	fully uncorrelated
Tau energy scale	3%	norm.	correlated between categories
Tau ID	6%	norm.	correlated between categories
Muon ID + Iso	2%	norm.	fully correlated
Electron ID + Iso	2%	norm.	fully correlated
b-tag Fake Rate	1%	norm.	fully correlated
Trigger efficiency	1%	norm.	correlated between periods
fake background statistics	[12% - 52%]	norm.	fully uncorrelated
charge mis-ID statistics	[3% - 26%]	norm.	fully uncorrelated
kNN training	–	shape	uncorrelated between channels
fake background composition	–	shape	fully uncorrelated
charge flip mass scale	–	shape	uncorrelated between channels
single bin fluctuations	–	shape	fully uncorrelated

channel are unveiled.

The physics observable used to discriminate between the signal and the different sources of background is the invariant mass of the visible products of the hadronic tau decay and the sub-leading lepton, $M_{\ell_2\tau}$. The choice of the sub-leading lepton has been found to be about 70% correlated to the true Higgs decay product, estimated using simulated data. Other observables to assign the correct light lepton as Higgs decay product such as the transverse mass of the lepton and the MET, the lepton isolation or the $\Delta\phi$ between the lepton and the $\cancel{E}_T$ have been investigated, but were found to yield inferior correlation with the true Higgs decay products.

The presence of a considerable missing energy source such as the neutrino produced in the W decay does not allow the exploitation of more refined mass reconstruction algorithms such as SVFit.

A summary of the different observed distributions, compared to the respective expectations, is shown in Figures 4.15, 4.16, 4.17 for the different channels, categories and run periods. Tables 4.4 and 4.5 summarizes the expected and observed yields in each category and run period.

4.5 Limit setting procedure and results

4.5.1 Statistical framework

The binned $M_{\ell_2\tau}$ distributions of each channel and category have been used for the statistical interpretation of the results. The choice of the binning, reflected in the plots shown in the previous section, represents a compromise between granularity and event population of each bin. The final choice for 8 TeV data is 10 GeV wide bins between 20 and 80 GeV followed by a 20 GeV wide bin up to 100 GeV and a 30 GeV wide bin up to 130 GeV; finally a 70 GeV wide bin covers the invariant mass region between $130 < M_{\ell_2\tau} < 200$ GeV. In the 7 TeV data, the small dataset size forces a much wider binning, leading to the choice of a 20 GeV wide bin between 20 and 40 GeV, followed by 40 GeV wide bins up to 120 GeV plus a single bin extending up to 200 GeV.

The expected background and signal distributions are fitted to the data to infer the value of the signal normalization, μ , by maximizing a likelihood function. The likelihood function is formulated in terms of $\mu \cdot s(\theta) + b(\theta)$, where $b(\theta)$ denotes the expected background yield and distribution, while $s(\theta)$ denotes the expected distribution of the signal process. Both $b(\theta)$ and $s(\theta)$ depend on the ensemble of *nuisance parameters* (θ), which represent the statistical and systematic uncertainties in the predictions. The multiplicative μ parameter in front of $s(\theta)$, known as *signal strength modifier*, is the parameter of interest which describe the strength of the observed signal with respect to the expected SM value (σ_{obs}/σ_{SM}). The value of the μ parameter is allowed to become negative to avoid any bias in the fit.

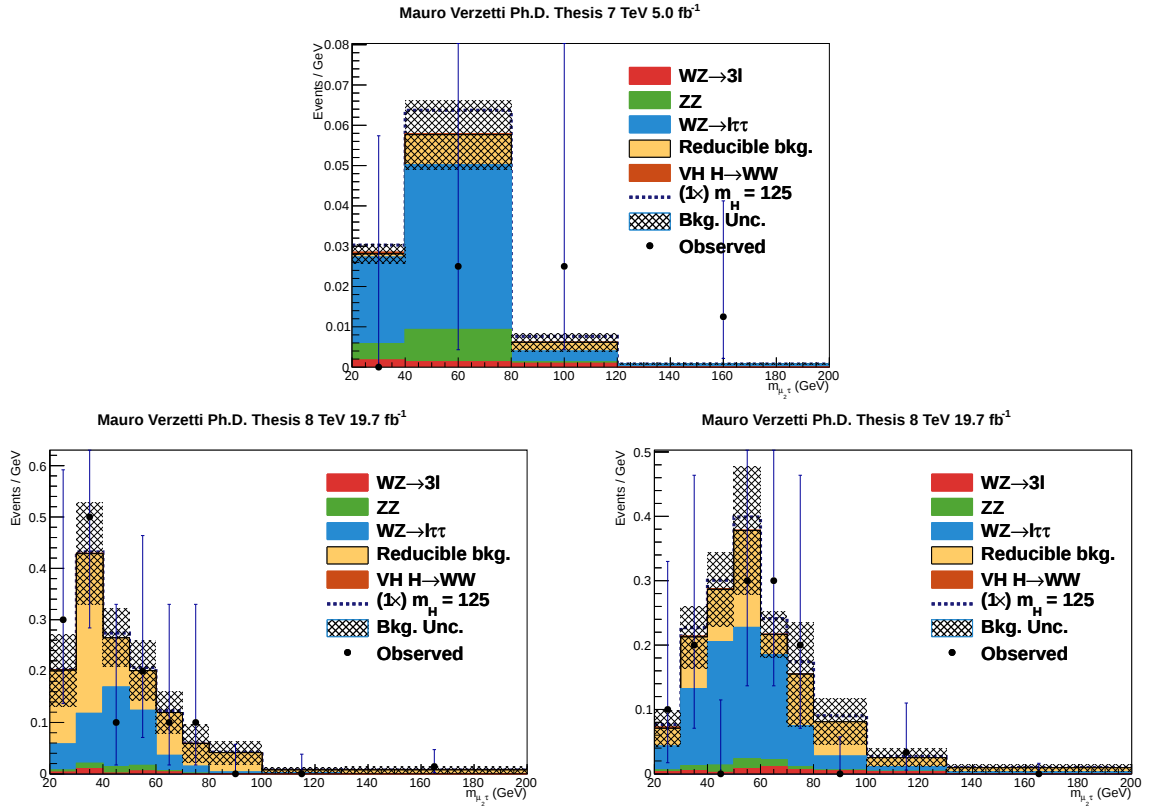


Figure 4.15: Comparison of $M_{\mu_2\tau}$ spectra observed in data and the background expectation in the $\mu\mu\tau_h$ signal region for 7 TeV data (top) and 8 TeV data (bottom). The 8 TeV data is split according the optimal categorization (LT-low on the left and LT-high on the right).

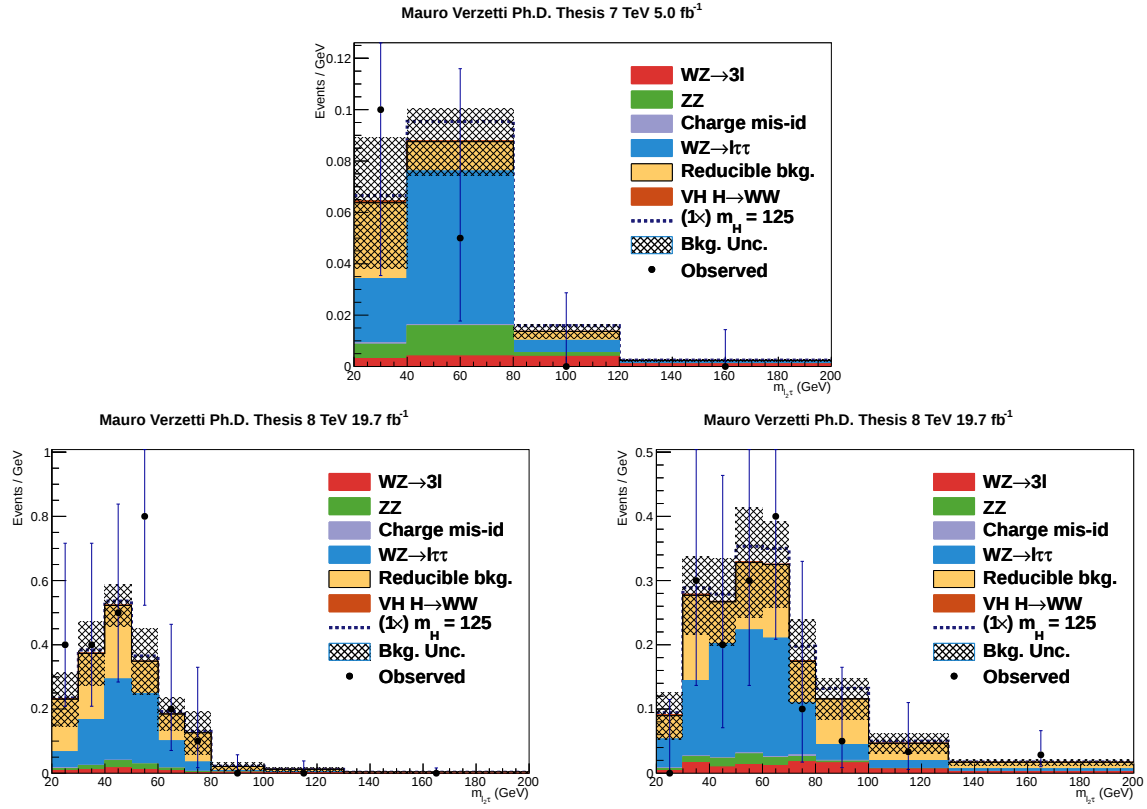


Figure 4.16: Comparison of $M_{\ell_2\tau}$ spectra observed in data and the background expectation in the $e\mu\tau_h$ signal region for 7 TeV data (top) and 8 TeV data (bottom). The 8 TeV data is split according the optimal categorization (LT-low on the left and LT-high on the right).

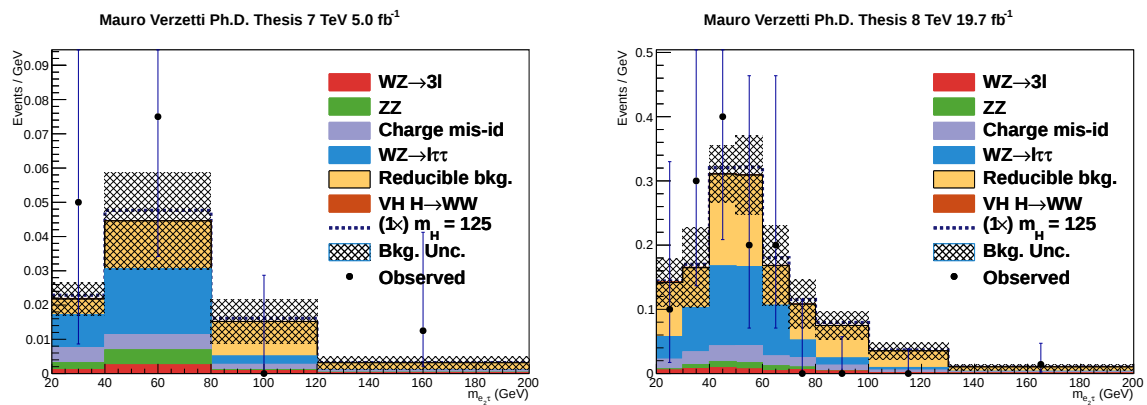


Figure 4.17: Comparison of $M_{e_2\tau}$ spectra observed in data and the background expectation in the $ee\tau_h$ signal region for 7 TeV data (left) and 8 TeV data (right).

Table 4.4: Expected event yields for the different background processes and for a 125 GeV $WH, H \rightarrow \tau\tau$ signal compared to the number of events observed in the data, split by channel and category, for 8 TeV data. The uncertainty quoted in each background process represents statistical plus systematic uncertainty.

Process	Channels and event categories					
	$\mu\mu\tau_h$		$e\mu\tau_h$		$ee\tau_h$	
	L_T Low	L_T High	L_T Low	L_T High		
WZ	4.9 ± 0.6	9.2 ± 1.0	8.7 ± 0.9	10.5 ± 1.1	5.5 ± 0.6	
ZZ	0.44 ± 0.04	0.57 ± 0.06	0.71 ± 0.07	0.72 ± 0.07	0.43 ± 0.04	
Reducible Bkg.	9.0 ± 1.4	6.5 ± 1.2	9.4 ± 1.4	8.1 ± 1.3	7.8 ± 0.9	
Charge mis-id.	–	–	0.10 ± 0.01	0.16 ± 0.02	1.6 ± 0.1	
$WH, H \rightarrow WW$	0.047 ± 0.004	0.109 ± 0.009	0.073 ± 0.006	0.16 ± 0.01	0.072 ± 0.006	
Total bkg.	14.6 ± 1.6	17.5 ± 1.6	19.4 ± 1.7	21.1 ± 1.8	16.0 ± 1.2	
$WH, H \rightarrow \tau\tau$ ($m_H = 125$ GeV)	0.25 ± 0.04	1.2 ± 0.1	0.56 ± 0.06	1.4 ± 0.1	0.62 ± 0.07	
Observed	14	12	24	17	13	

Table 4.5: Expected event yields for the different background processes and for a 125 GeV WH, $H \rightarrow \tau\tau$ signal compared to the number of events observed in the data, split by channel and category, for 7 TeV data. The uncertainty quoted in each background process represents statistical plus systematic uncertainty.

Process	Channels		
	$\mu\mu\tau_h$	$e\mu\tau_h$	$ee\tau_h$
WZ	2.3 ± 0.3	3.6 ± 0.4	1.2 ± 0.1
ZZ	0.43 ± 0.04	0.64 ± 0.07	0.24 ± 0.02
Reducible Bkg.	0.4 ± 0.2	1.2 ± 0.5	1.2 ± 0.3
Charge mis-id.	–	0.04 ± 0.01	0.38 ± 0.06
WH, $H \rightarrow WW$	0.035 ± 0.003	0.055 ± 0.005	0.020 ± 0.002
Total bkg.	3.5 ± 0.4	6.0 ± 0.7	3.3 ± 0.3
WH, $H \rightarrow \tau\tau$ ($m_H = 125$ GeV)	0.33 ± 0.03	0.45 ± 0.05	0.18 ± 0.02
Observed	3	4	5

The likelihood implemented in the fit is defined as the product of the likelihoods for each category times the likelihood function for the nuisance parameters:

$$\begin{aligned}
 \mathcal{L}(obs|\mu, \boldsymbol{\theta}) &= \prod^{cat} \mathcal{L}_c(obs|\mu, \boldsymbol{\theta}) \times \prod^{nuisance} p(\hat{\theta}, \theta) \\
 &= \prod^{cat} \text{Poisson}(obs_c, \mu, \boldsymbol{\theta}) \times \prod^{nuisance} p(\hat{\theta}, \theta),
 \end{aligned} \tag{4.3}$$

where

$$\text{Poisson}(obs_c, \mu, \boldsymbol{\theta}) = \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))}. \tag{4.4}$$

As described in Section 4.3, two kinds of systematic uncertainties were considered. Systematic uncertainties associated to the overall normalization are considered to be distributed according to a *log-normal* distribution:

$$\rho(\theta, \hat{\theta}) = \frac{1}{\theta \sqrt{2\pi \ln(k)}} e^{-\left(\frac{\ln(\theta/\hat{\theta})}{\sqrt{2 \ln k}}\right)^2} \tag{4.5}$$

with $k = 1 + \epsilon$, where ϵ is the size of the yield uncertainty.

Shape uncertainties are modeled using three distributions: the distribution for the central value plus the templates for the upward and downward shifts by one standard

deviation of the respective nuisance parameter. A family of distributions is obtained starting from these three templates using a morphing technique. The morphing function interpolates the contents of each bin as a function of the nuisance parameter. The interpolation is quadratic within the boundaries provided by the two additional shifted templates and linear beyond this point. A gaussian distribution with mean 0 and $\sigma = 1$ is assigned to the single morphing parameter. The extrapolation is built in such a way that the nuisance parameter assumes the value 0 for the central template and the values ± 1 for the shifted templates. More details on this “vertical morphing” technique can be found in [81].

In case the fit shows no significant presence of signal, which is related to a negative value of μ or compatible with zero, exclusion limits can be set.

The exclusion limit of a SM Higgs boson is statistically performed by evaluating the compatibility of the data to two hypotheses: the *null hypothesis* H_0 , which corresponds to the absence of a Higgs boson signal, and the *alternative hypothesis*, which includes the presence of a Higgs signal. The discrimination between the two hypotheses is performed by defining a test statistic under the null hypothesis and compute its value under the observed data case. The observed value of this test statistic, with respect to the theoretically expected distribution, allows to compute the probability to observe a compatibility to H_0 worse than the observed one. Such probability is called *p-value* and is used to take the decision whether to reject the hypothesis H_0 in favor of H_1 or not.

The theoretical framework under which the test statistics is built represents a modified version of the CL_s method [82] and follows the prescriptions agreed between the ATLAS and CMS statistical committees for Higgs boson searches [83].

The null hypothesis is formulated in terms of only the expected backgrounds distributions $b(\theta)$ defined previously, while the H_1 hypothesis is formulated as the sum of signal and backgrounds, $\mu \cdot s(\theta) + b(\theta)$, as initially shown.

According to the Neyman-Pearson lemma, the most powerful test statistics to discriminate between the H_0 (background only) and the H_1 (signal plus background) hypotheses is the *profile likelihood ratio* λ , defined as:

$$\lambda(\mu) = -2 \ln \frac{\mathcal{L}(obs|\mu \cdot s + b, \hat{\theta}_\mu)}{\mathcal{L}(obs|\hat{\mu} \cdot s + b, \hat{\theta})}, \quad (4.6)$$

where $\hat{\theta}$ represents the set of nuisance parameter values that maximize the likelihood (called *maximum likelihood estimator*) and $\hat{\theta}_\mu$ are the values of the nuisance parameters that maximize the likelihood for a fixed value of μ .

Equation 4.6 is complemented by the boundary condition $0 \leq \hat{\mu} \leq \mu$ that ensures that the expected signal yield is positive and that the computed confidence interval is one sided (and therefore $\mu \geq \hat{\mu}$). These conditions are imposed on the test statistic by forcing $\hat{\mu} = 0$ even when the fit yields as best result a negative value and by setting the

test statistic to 0 when $\mu < \hat{\mu}$.

$$q(\mu) = \begin{cases} -2 \ln \frac{\mathcal{L}(\text{obs}|\mu \cdot s + b, \hat{\theta}_\mu)}{\mathcal{L}(\text{obs}|\hat{\mu} \cdot s + b, \hat{\theta})} & , \hat{\mu} < 0 \\ -2 \ln \frac{\mathcal{L}(\text{obs}|\mu \cdot s + b, \hat{\theta}_\mu)}{\mathcal{L}(\text{obs}|\hat{\mu} \cdot s + b, \hat{\theta})} & , 0 \leq \hat{\mu} \leq \mu \\ 0 & , \mu < \hat{\mu} \end{cases} \quad (4.7)$$

The test statistic $q(\mu)$ is defined such that small (large) values indicate preference for the H_1 (H_0) hypothesis.

In the limit of large event statistics the distribution of the test statistic q_μ under both hypotheses collapse into well-defined analytical functions as described by the Wilkins theorem and reported in [84, 83]. This assumption allows to compute the p-value under a specific hypothesis with a numerical integration rather than toy MC methods.

A naive approach to limit setting would be computing, for each mass point, the value μ_{up} for which $p(q(\mu_{up})|H_1) = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|H_{\mu_{up}})dq_\mu = 0.05$. This approach would not be safe in case of background under-fluctuations and a small expected signal. In this case the value $p(q(\mu_{up})|H_0)$ is also a small number, potentially very close to the threshold of 0.05. This would lead to a paradox in which both H_1 and H_0 have to be rejected.

In this situation data are not sufficient to reject the H_1 hypothesis. To overcome this effect the CL_s method defines the confidence level as:

$$CL_s(\mu) = \frac{CL_{s+b}}{CL_b} = \frac{p(q(\mu)|H_1)}{p(q(\mu)|H_0)} \quad (4.8)$$

The immediate consequence of this definition is that small values of μ become harder to exclude.

4.5.2 Fit results

The reliability of the background modeling and of the systematic uncertainty model can be tested by analyzing the ancillary output of the profile maximum likelihood fit performed during the CL_s limit computation. A common practice is to compute the values:

$$\frac{\hat{\theta} - \theta}{\sigma_\theta}, \quad (4.9)$$

which are the difference between the best-fit (post-fit) value and the central (pre-fit) value of the nuisance parameter, divided by its expected uncertainty. This quantity is

Table 4.6: List of all the nuisance parameters which pull ($\Delta x/\sigma_{\text{in}}$) is either larger than ± 0.3 or the a posteriori variance ($\sigma_{\text{out}}/\sigma_{\text{in}}$) changed by more than 10% with respect to the a priori one. Δx denotes the shift in the nuisance value that best fits the data and σ_{in} , σ_{out} represent the a priori and a posteriori variance of the nuisance parameter, respectively.

Sample	Channel	name	b -only fit
			$\Delta x/\sigma_{\text{in}}, \sigma_{\text{out}}/\sigma_{\text{in}}$
7 TeV	$\mu\mu\tau_h$	Reducible bkg. Bin-by-bin bin #2	-0.20, 1.08
		Reducible bkg. Bin-by-bin bin #4	+0.14, 1.68
		kNN shape uncertainty	-0.04, 0.83
	$e\mu\tau_h$	Reducible bkg. Bin-by-bin bin #1	+0.20, 0.85
	$ee\tau_h$	Reducible bkg. Bin-by-bin bin #3	-0.31, 1.03
8 TeV	$\mu\mu\tau_h$	Reducible bkg. normalization, L_T high	-0.30, 0.95
		Reducible bkg. composition, L_T high	-0.35, 0.67
		Reducible bkg. Bin-by-bin bin #1	+0.25, 0.88
		Reducible bkg. Bin-by-bin bin #7	-0.35, 1.00
		Reducible bkg. composition, L_T low	+0.32, 0.64
		kNN shape uncertainty	+0.25, 0.82
	$\mu\mu\tau_h$	Reducible bkg. composition, L_T high	+0.52, 0.74
		Reducible bkg. Bin-by-bin bin #1	+0.32, 0.91
		Reducible bkg. composition, L_T low	-0.45, 0.62
		kNN shape uncertainty	-0.03, 0.73
	$ee\tau_h$	Reducible bkg. composition	-0.52, 0.82

commonly referred to as *pull*. The maximum likelihood fit also evaluates the a posteriori variance of each nuisance parameter. Large outliers in the pull distribution can indicate an under-estimation of the systematic uncertainty associated to the corresponding nuisance parameter. Very small a posteriori variance of a nuisance indicates that the data are constraining a certain systematic uncertainty to a better degree than the a priori knowledge. This effect is not necessarily a flaw in the modeling, but requires additional scrutiny to ensure that the correlation of such nuisance parameter among different categories is well motivated. This effect is widely exploited when creating a category with large number of selected events and low sensitivity on the sole purpose of constraining one or more systematic effects on a more sensitive, but small size, category.

Table 4.6 summarizes the largest pulls observed in the maximum likelihood (ML) fit to the data.

The agreement between the data and the background description is also tested by the goodness of fit. It is computed by randomly sampling a set of pseudo-data, called *toys*, from the expected distributions of the backgrounds, allowing for fluctuations within the variance of the nuisance parameters. These pseudo-data are then fit with the same background model to obtain a (toy) “observed” limit.

The distribution of the limits computed with the toys is then compared with the one observed in the real data. The goodness of fit, as in the case of the χ^2 method, is the right-sided integral of the distribution of toys, starting from the observed value. Figure 4.18 shows the goodness of fits for the different channels.

Figures 4.19, 4.20, 4.21 show the data and background distributions after the fitting process.

The resulting shapes after the fitting process can be combined together to have a more complete view of the results. The combination is performed in such a way that each category contributes to the total by an amount that is proportional to its signal purity evaluated as $S/(S + B)$, where S denotes the integral of the signal and B the sum of the backgrounds as from pre-fit values. The total normalization of the shapes is kept constant: i.e. the integral of the observed data is not affected by this procedure. Given the differences in the choice of binning this combination is only possible among 7 TeV and 8 TeV categories separately.

From this process, result of which is shown in Figure 4.22, one can see that the number of events observed in the mass region in which the Higgs signal is expected to appear is below the background expectation. Even though these plots have relatively limited statistical meaning, they help showcase the background under-fluctuation observed in the data.

The computation of the μ values for the different channels and their combination confirms the background under-fluctuation. Figure 4.23 shows the observed μ values for the three $\ell\ell\tau_h$ channels and their combination. Even though these values are negative, the combination is still compatible with the SM value $\mu = 1$. The right plot on Figure 4.23 shows the best fit μ values for each different $H \rightarrow \tau\tau$ channel search performed by CMS.

As described in Section 1.5 the combination of these channels, including the ones presented in this thesis, led to the first observation of the SM Higgs boson into tau lepton pairs in CMS.

4.5.3 Exclusion limits

As the channels presented in this thesis do not show any significative evidence of the searched process, exclusion limits are set at 95% confidence level and are compared to the expected limits under the hypotheses of a presence or absence of a SM Higgs boson.

The exclusion limit on the process $pp \rightarrow WH \rightarrow \ell\nu_\ell\tau\tau$ is computed for SM Higgs mass range between 90 GeV and 145 GeV in 5 GeV steps. In this computation the process $pp \rightarrow WH$ with $H \rightarrow WW \rightarrow \ell\nu_\ell\ell'\nu_{\ell'}$, which is present in our signal region, is treated as a background. The Higgs mass for this background process is fixed to 125 GeV, following the joint CMS and ATLAS mass measurement [23, 22], and the yield is fixed to the theoretical cross-section value, which agrees with current measurements. The

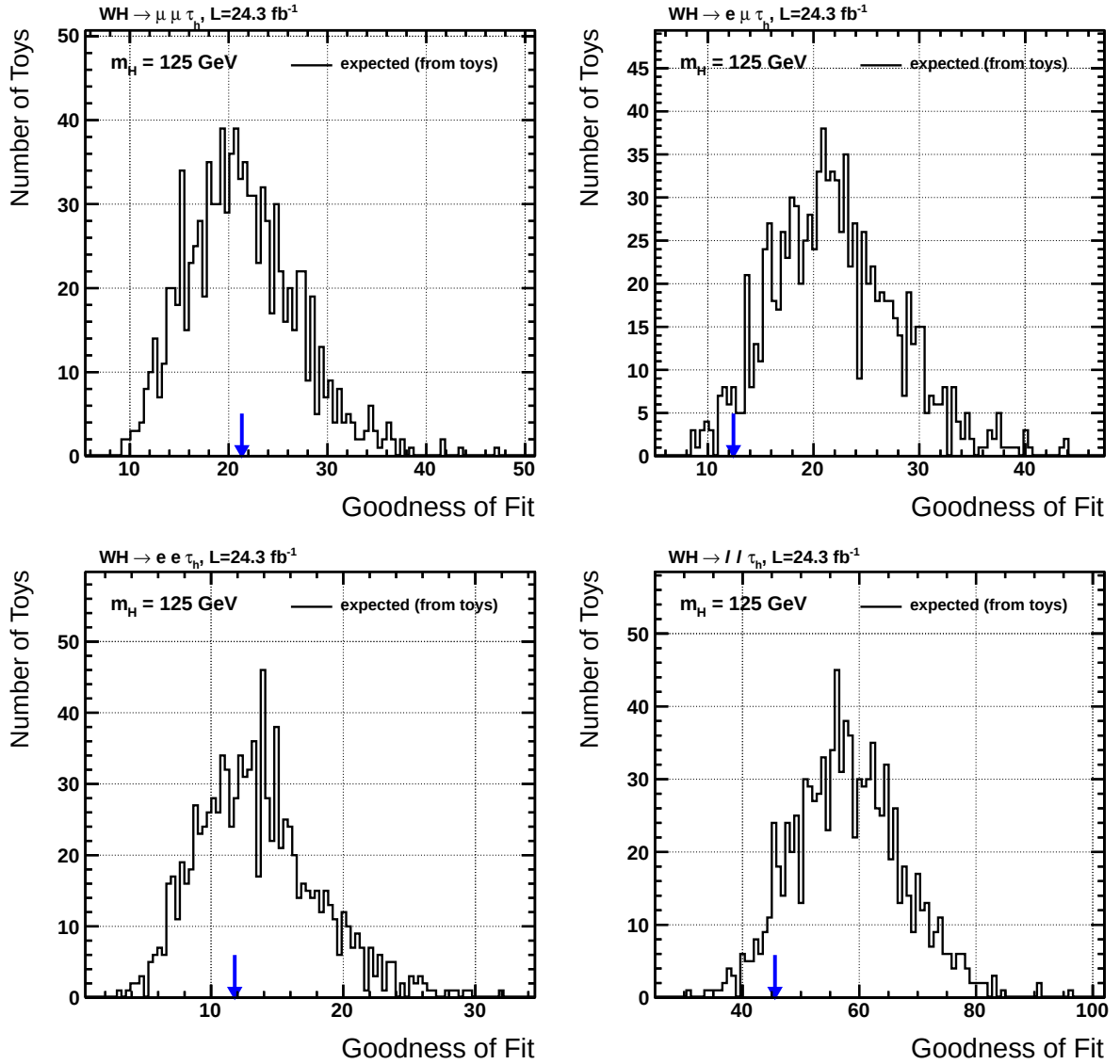


Figure 4.18: Goodness of fit as computed by toys for $\mu\mu\tau_h$ (top left), $e\mu\tau_h$ (top right), $ee\tau_h$ (bottom left), and combined (bottom right) fits. The blue arrow represents the value observed in data.

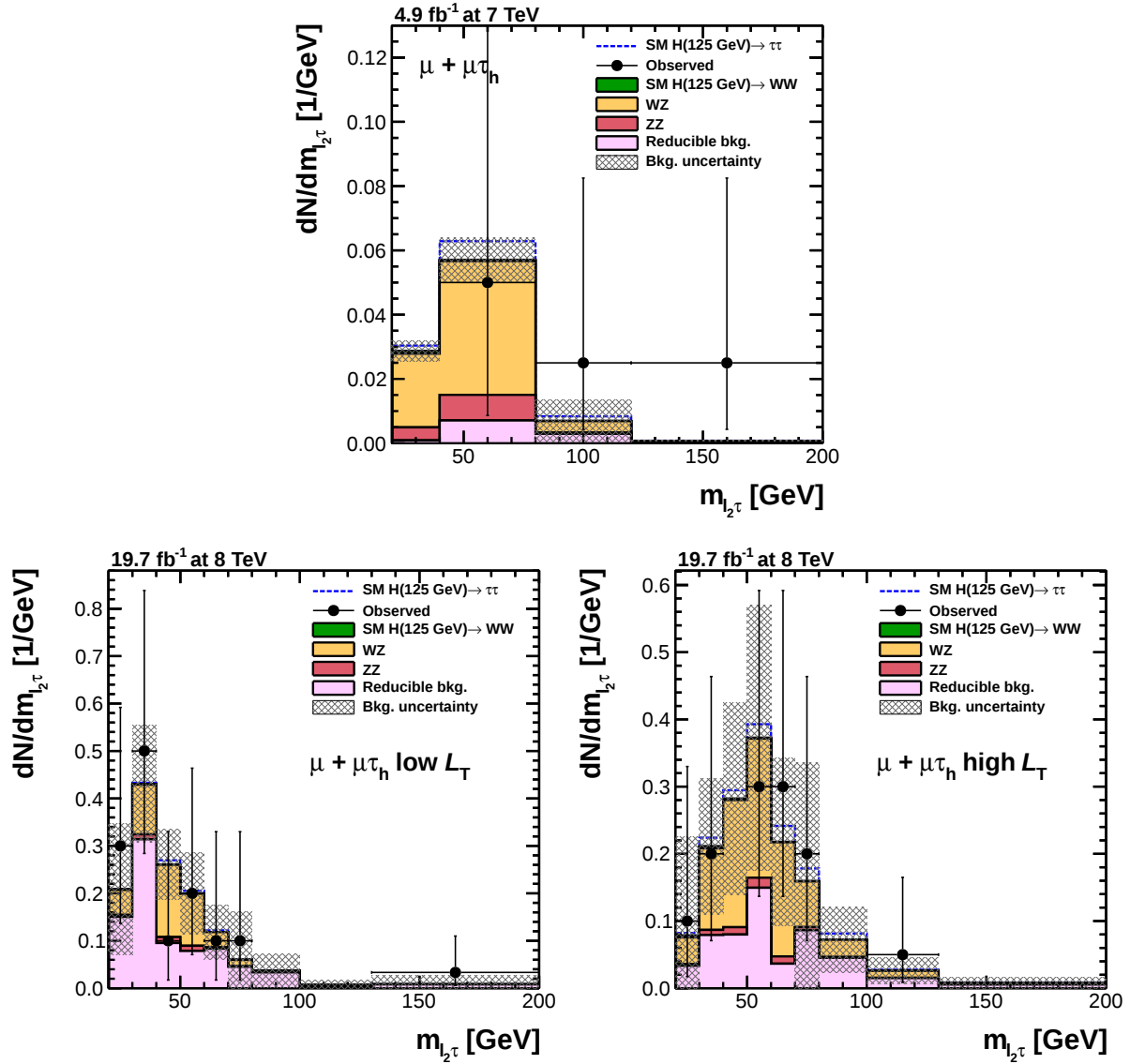


Figure 4.19: Comparison of measured and predicted $M_{\ell_2\tau}$ distributions in the $\mu\mu\tau_h$ signal region for 7 TeV data (top) and 8 TeV data (bottom). Expected backgrounds are shown for the post-fit values of the nuisance parameters. The 8 TeV data are split according the optimal categorization.

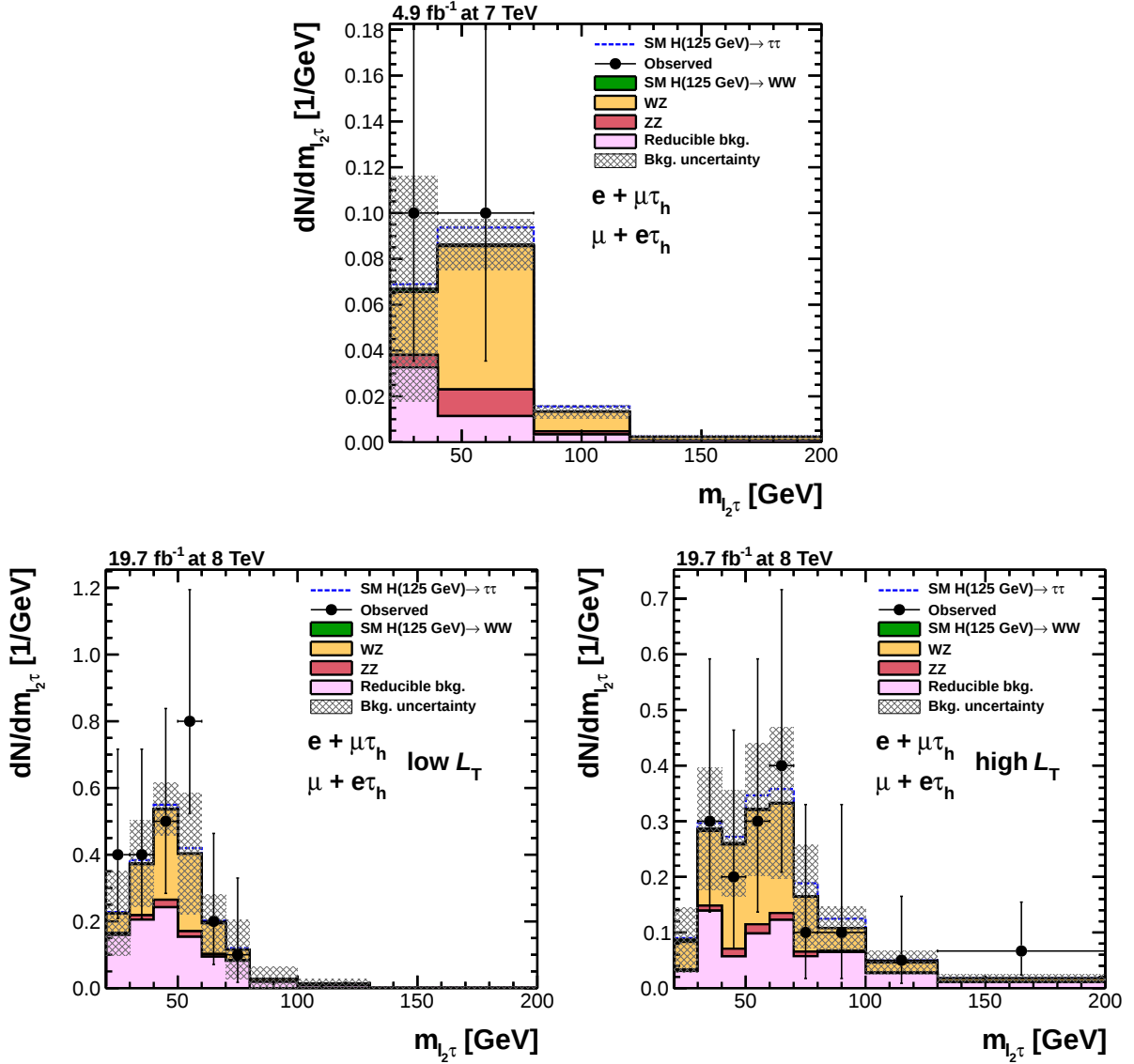


Figure 4.20: Comparison of measured and predicted $M_{\ell_2\tau}$ distributions in the $e\mu\tau_h$ signal region for 7 TeV data (top) and 8 TeV data (bottom). Expected backgrounds are shown for the post-fit values of the nuisance parameters. The 8 TeV data are split according the optimal categorization.

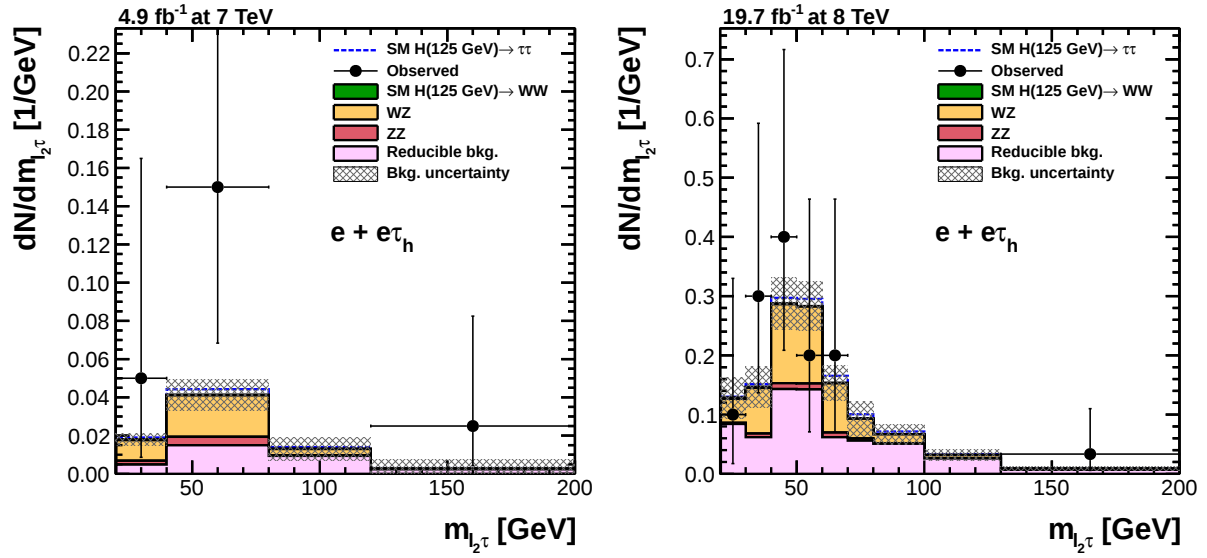


Figure 4.21: Comparison of measured and predicted $M_{\ell_2\tau}$ distributions in the $e\tau_h$ signal region for 7 TeV data (top) and 8 TeV data (bottom). Expected backgrounds are shown for the post-fit values of the nuisance parameters.

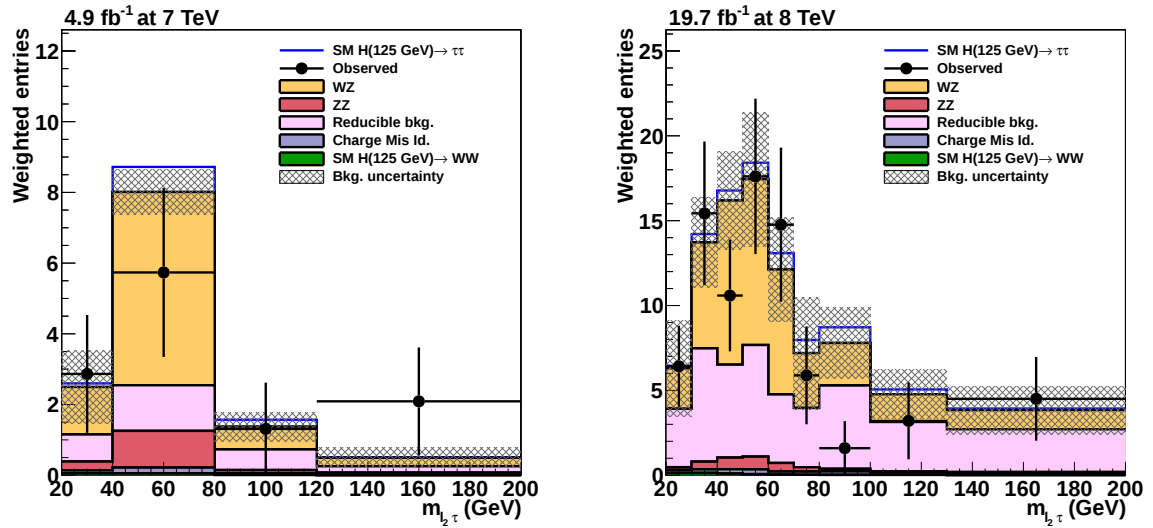


Figure 4.22: $S/(S + B)$ weighted combination of post-fit results from 7 TeV (left) and 8 TeV (right) categories. The over-all normalization of data is kept constant.

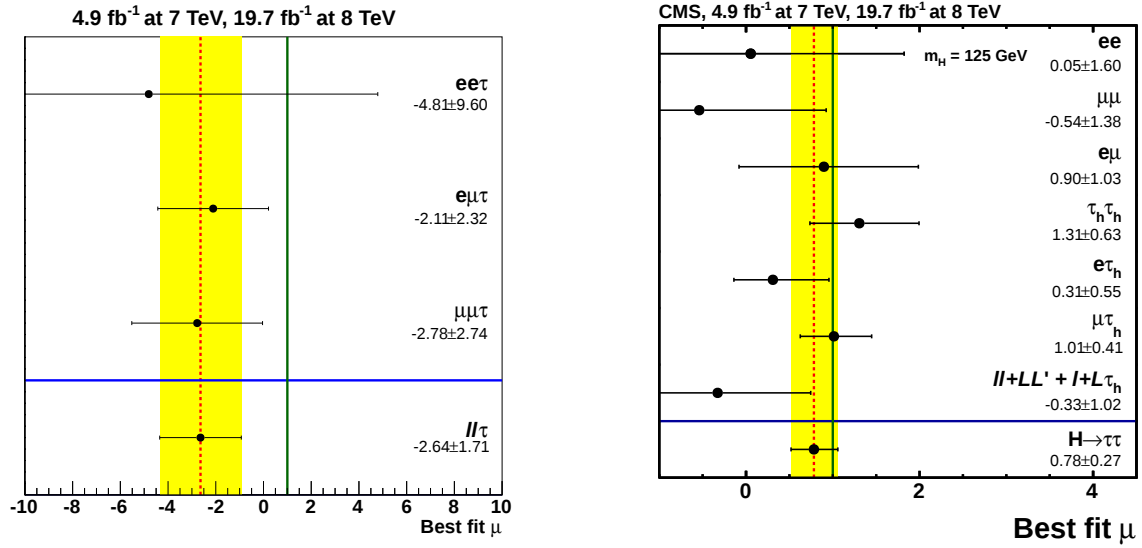


Figure 4.23: Best fit signal strength μ for the channels presented in this thesis (left) and those included in the CMS $H \rightarrow \tau\tau$ search (right) and their respective combined value.

choice of fixing the Higgs mass in the $H \rightarrow WW$ process, while still scanning for a Higgs presence in a range of masses is motivated by the possibility of having multiple Higgs bosons, close in mass, with different decay modes.

For each mass point the expected exclusion limit and its one and two sigma confidence intervals are computed by evaluating the exclusion limit (and variance) on the *Asimov dataset*³, a set of pseudo-data that reflects the average expected background.

As an evidence for the decay of the 125 GeV Higgs boson to tau pairs exists, signal-injected limits are also computed. In this case, while the observed limit remains the same, the expected limit and the respective one and two sigma confidence intervals are computed from the distribution of toy limits sampled from signal plus background distributions.

Expected and observed exclusion limits on SM Higgs boson detection in the three channels are shown in Figure 4.24, the signal-injected version of those limits is available in Figure 4.25. Figure 4.26 shows the combined limit for 7 TeV and 8 TeV datasets, while Figure 4.27 shows the combination of the whole WH analysis.

³The Asimov dataset owes its name to the famous sci-fi novelist Isaac Asimov, author of the short story *Franchise* [85], in which elections are performed by polling the opinion of a single randomly chosen representative of the population.

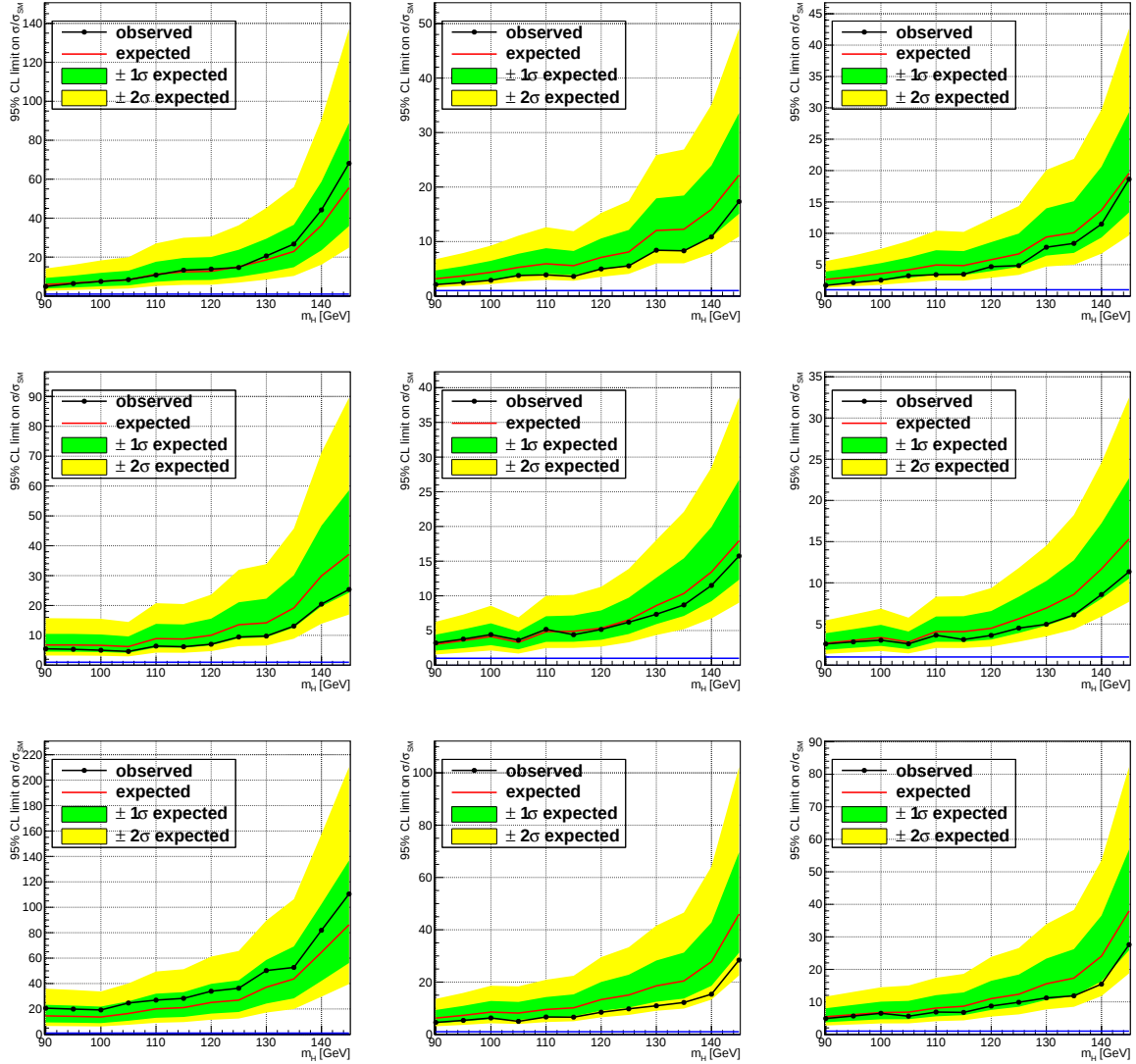


Figure 4.24: Expected and observed exclusion limits on σ/σ_{SM} for the process $WH \rightarrow \ell\ell\tau$ for 7 TeV (left), 8 TeV (middle) and combined (right) data. Top, middle and bottom plots show the limits obtained by the $\mu\mu\tau_h$, $e\mu\tau_h$ and $ee\tau_h$ channels, respectively. Values above the solid black line are excluded by the data. The benchmark points used for the computation of the exclusion limits are shown as a dot in the black line. The red line represent the expected exclusion limit in presence of background only while the green and yellow bands show the one and two standard deviation intervals on the expected limit, respectively.

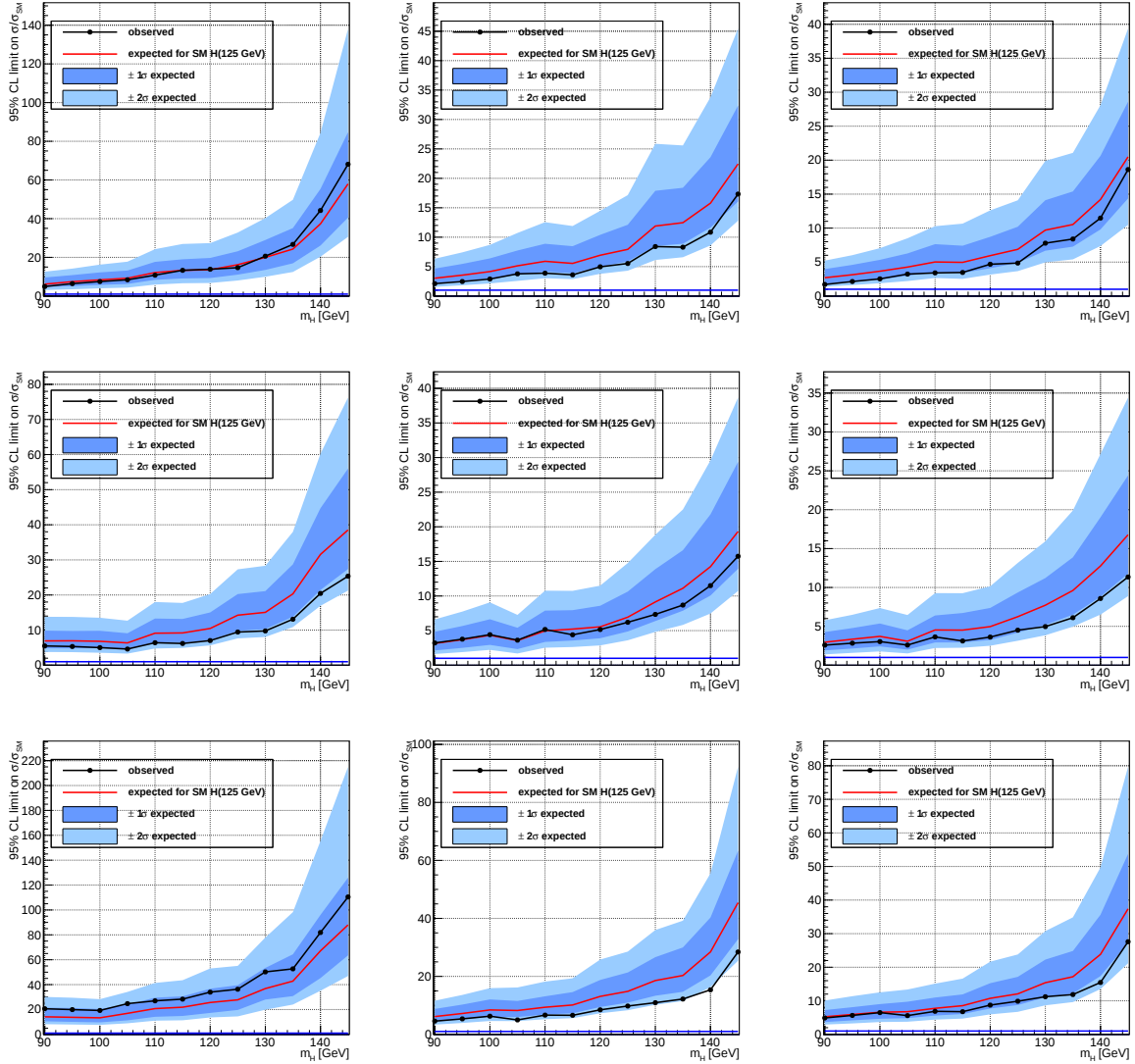


Figure 4.25: Expected and observed exclusion limits on σ/σ_{SM} for the process $WH \rightarrow \ell\ell\tau$ for 7 TeV (left), 8 TeV (middle) and combined (right) data. Top, middle and bottom plots show the limits obtained by the $\mu\mu\tau_h$, $e\mu\tau_h$ and $ee\tau_h$ channels, respectively. Values above the solid black line are excluded by the data. The benchmark points used for the computation of the exclusion limits are shown as a dot in the black line. The red line represent the expected exclusion limit under the hypothesis of the presence of a 125 GeV SM Higgs boson signal while the blue and azure bands show the one and two standard deviation intervals on the expected limit, respectively.

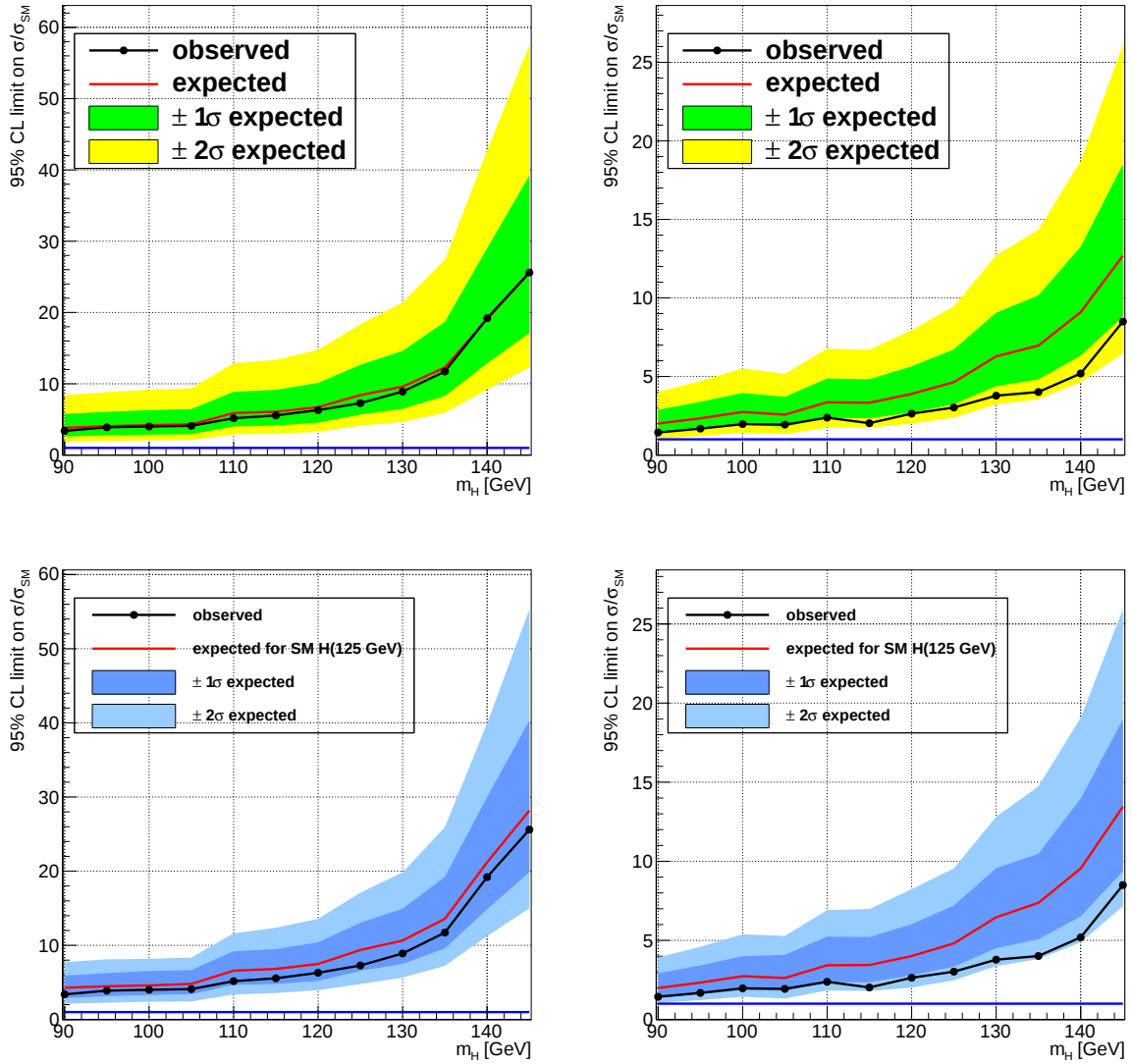


Figure 4.26: Expected and observed exclusion limits under the assumption of background only (top) and background plus SM Higgs signal (bottom) for the 7 TeV (left) and 8 TeV (right) datasets.

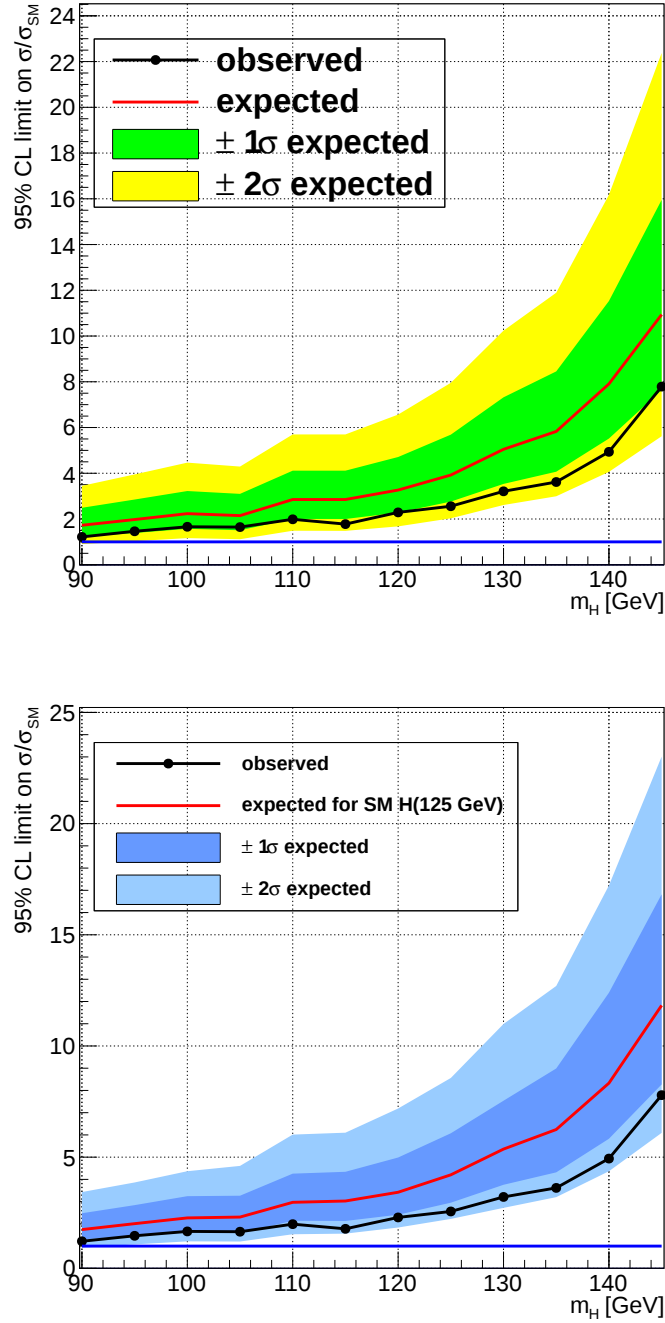


Figure 4.27: Expected and observed exclusion limits under the assumption of background only (top) and background plus SM Higgs signal (bottom) for the combination of 7 TeV and 8 TeV datasets.

Chapter 5

Conclusions and possible future developments

A search for the process $WH \rightarrow \ell\nu_\ell\tau\tau \rightarrow \ell\nu_\ell\tau_\ell\tau_h$ has been performed. This channel, combined with the other VH, $H \rightarrow \tau\tau$ channels, is sensitive to SM Higgs boson production in the mass range 90–130 GeV. The VH production mode furthermore provides a clean signature that could be exploited in future accurate measurements of Higgs coupling to taus.

The search has been performed in three distinct channels: $\mu\mu\tau_h$, $e\mu\tau_h$, and $ee\tau_h$, analyzing data from LHC Run I collected by the CMS detector. The total integrated and analyzed luminosity amounts to 4.9 fb^{-1} , recorded at the center-of-mass energy $\sqrt{s} = 7 \text{ TeV}$, plus 19.7 fb^{-1} recorded at $\sqrt{s} = 8 \text{ TeV}$. Events are selected by requiring two isolated and well identified electrons or muons plus an isolated and well identified hadronic tau. An additional separation of the selected events in two categories according to the scalar sum of the p_T of the leptons is performed to increase the sensitivity to the signal.

The main backgrounds for this channel are the WZ production, which has the same signature as the signal process, and the W+Jets production, in which two jets are misidentified as light lepton and hadronic tau, respectively. Despite the performance of the CMS detector and reconstruction algorithms to discriminate electrons, muons, and hadronic taus from jets, the large W+Jets production cross section causes this reducible, fake, background to yield a relevant contribution to the event sample selected in the signal region.

The reducible backgrounds, which include W+Jets and all the backgrounds with at least one mis-identified final state electron, muon or hadronic tau, are modeled in a data-driven way based on the sidebands with inverted lepton isolation. The probability for a jet to be mis-identified as a lepton is computed in dedicated signal-free regions, with a topology as close as possible to the signal region. The mis-identification probabilities are evaluated with a functional fit in the case of hadronic taus, and using a k-Nearest Neighbor algorithm in case of light leptons.

A dedicated study on the performance of the kNN estimator has been carried out in order to maximize the agreement between the background and its modeling and to infer the systematic uncertainties of statistical nature connected to the limited size of the training sample and the choice of the training variables.

The measurement of the hadronic tau identification efficiency presented in this thesis has been performed with 2011 data and has been used as input for previous versions of the $H \rightarrow \tau\tau$ searches performed at CMS. It shows that hadronic tau decays can be identified with approximately 60% efficiency and percent level fake probability.

The data observed show no evidence of the presence of the Higgs boson, obtaining a combined μ value of $\mu = -2.64 \pm 1.71$. As no significative evidence of signal is found, exclusion limits are set, excluding, for a SM Higgs boson of 125 GeV, a production rate above 2.5 the expected SM value.

This result has been combined with other searches for the SM Higgs boson to tau pair decays, obtaining a combined 3.2σ evidence for the $H \rightarrow \tau\tau$ decay process. The combined signal strength and best fit mass are both compatible with 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.

Future analysis improvements might come from a better assignment of the electron or muon as Higgs decay product and by a better mass reconstruction technique. In principle both tasks can be achieved with a single algorithm exploiting the topology of the event including the missing transverse energy information. The SVFit algorithm, already used in other channels of the $H \rightarrow \tau\tau$ searches is a very good candidate to perform these tasks, but a MVA approach is also a viable option. The latest developments in the tau identification within the CMS collaboration are also very promising, as they are expected to significantly reduce the $\text{jet} \rightarrow \tau_h$ fake background while retaining the same detection efficiency.

Reducible background is the second major source of background for the analysis, with a sizable increase in the 8 TeV dataset due to the larger instantaneous luminosity. The dire running environment of LHC Run II will see an increase of this source of backgrounds. More detailed studies of the misidentification probability, with the aim to bring the fake tau method to the same strength of the fake lepton one, would allow a significant reduction of the uncertainties.

Appendix A

Additional plots

This appendix contains additional plots documenting various stages of the analysis which were not fully reported in the main body of the thesis to ease the reading process.

A.1 Background estimation plots

Figures A.1, A.2, and A.3 show the χ^2 scans as a function of the number of neighbors employed in the classifier as described in Section 4.2.3 for the 8 TeV datasets of $\mu\mu\tau_h$, $e\mu\tau_h$, and $ee\tau_h$ channels, respectively.

Figures A.4, A.5, and A.6 show the measured fake rate for the corresponding datasets as a function of the kNN training variables.

A.2 Background control regions

Figures A.7–A.12 show the agreement obtained in the control regions with same selection as the signal region but inverted tau isolation and with the requirements that the three objects have the same measured electric charge.

It can be noted that the population of this region is inferior to the one shown in the Section 4.2.4 as the contribution from Z+jets and $t\bar{t}$ is absent in this region, which is also the cause for the slightly worse agreement between observed data and expected backgrounds.

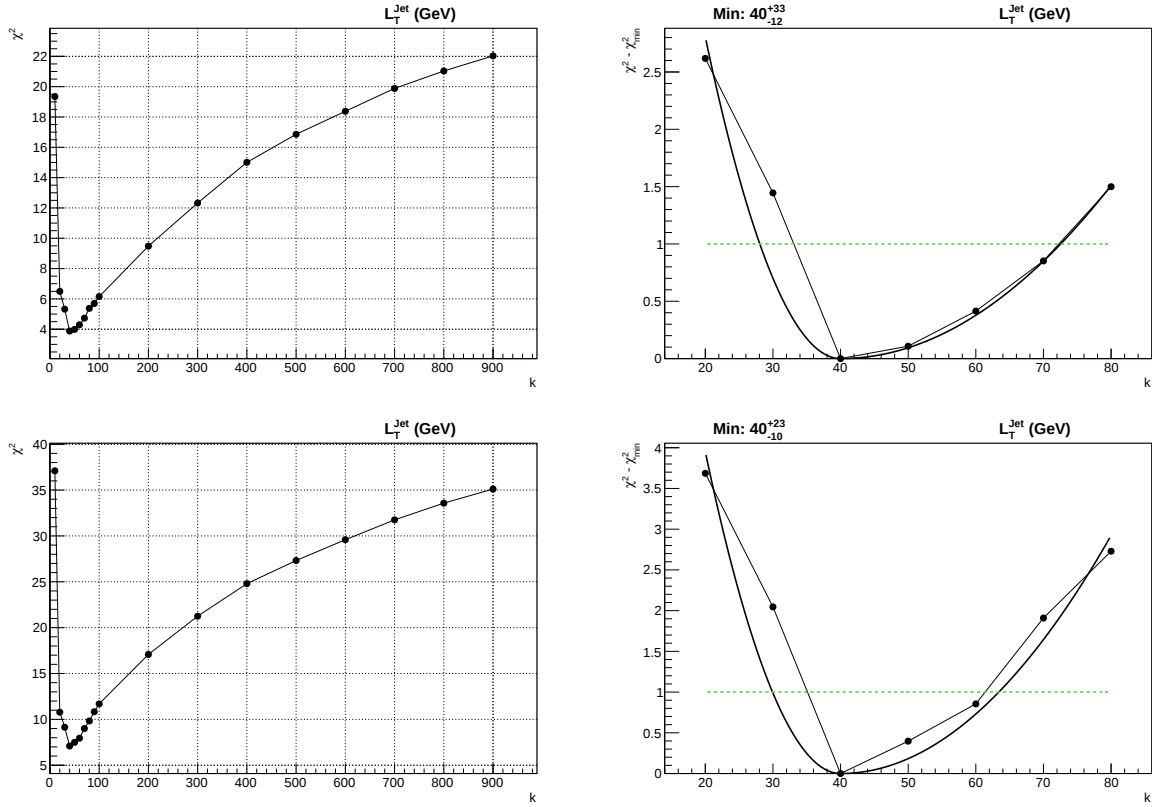


Figure A.1: χ^2 scan of different neighbors value (left) and corresponding minima fit (right) for leading (top) and sub-leading (bottom) muons in $\mu\mu\tau_h$ channel. The variable used for the scan is the scalar sum of the p_T of the two leptons and the jet

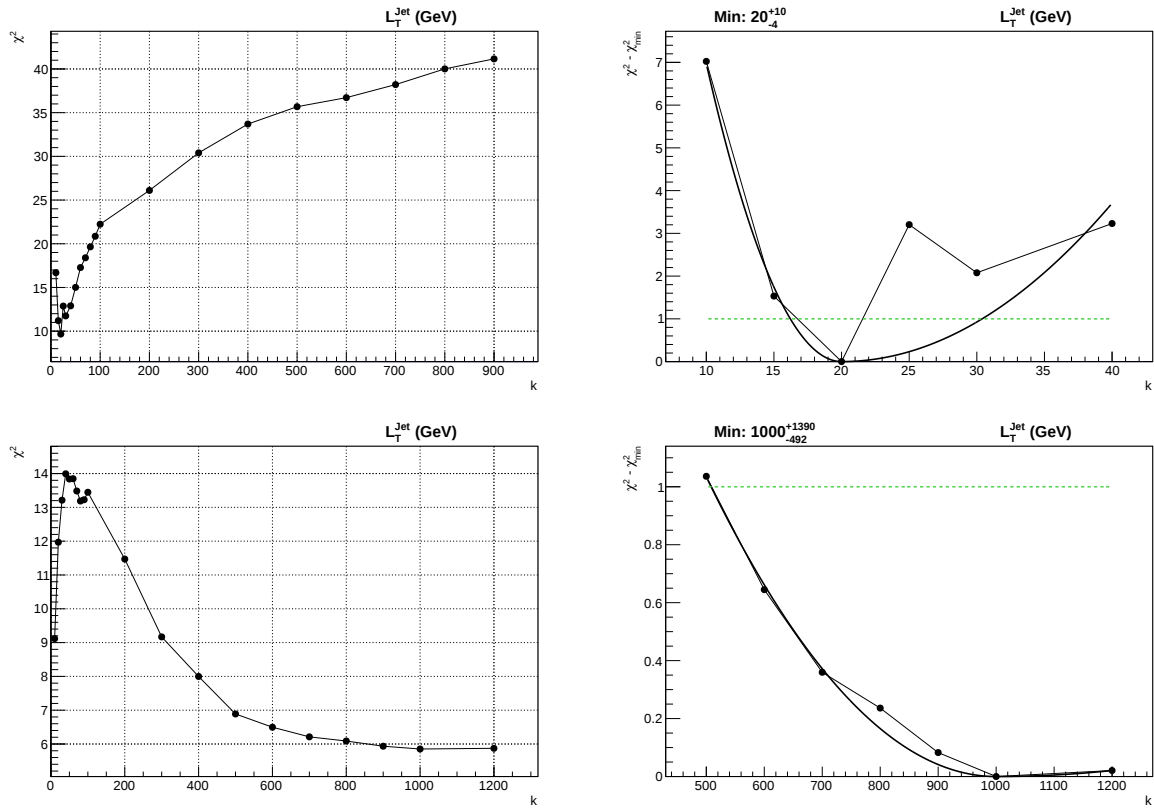


Figure A.2: χ^2 scan of different neighbors value (left) and corresponding minima fit (right) for muon (top) and electron (bottom) in $e\mu\tau_h$ channel. The variable used for the scan is the scalar sum of the p_T of the two leptons and the jet

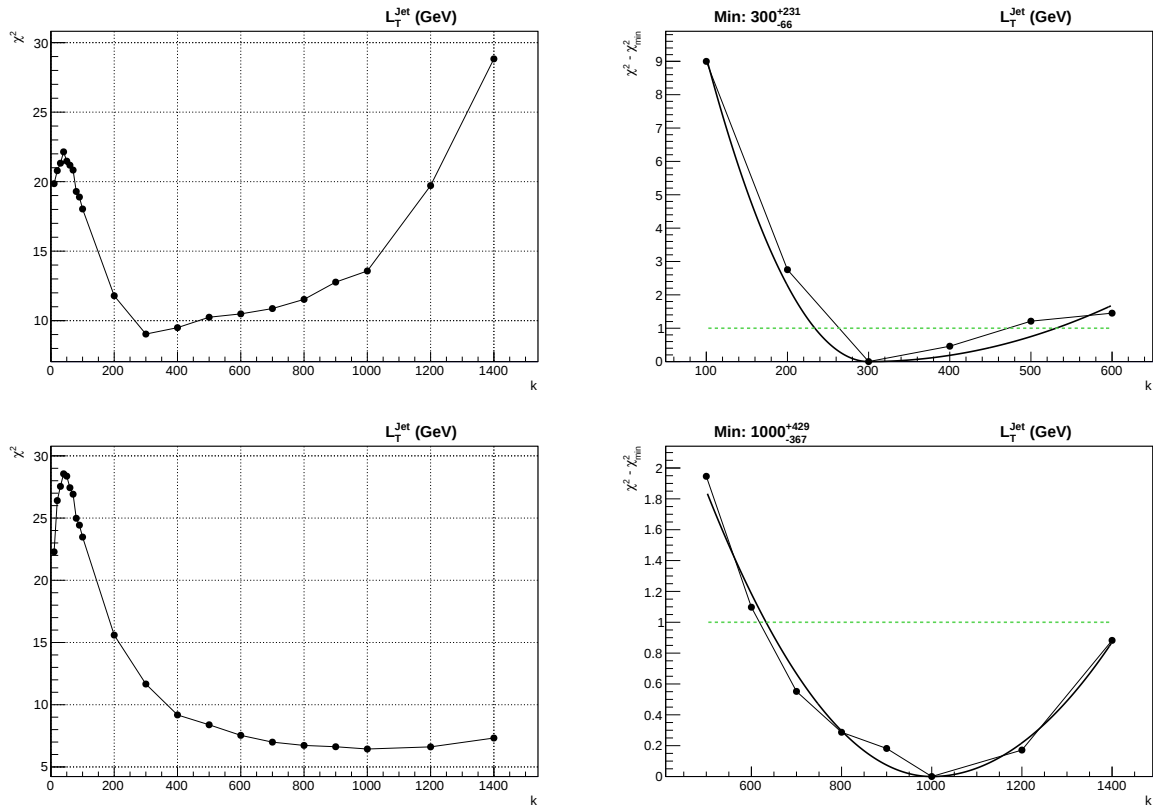


Figure A.3: χ^2 scan of different neighbors value (left) and corresponding minima fit (right) for leading (top) and sub-leading (bottom) electrons in $ee\tau_h$ channel. The variable used for the scan is the scalar sum of the p_T of the two leptons and the jet

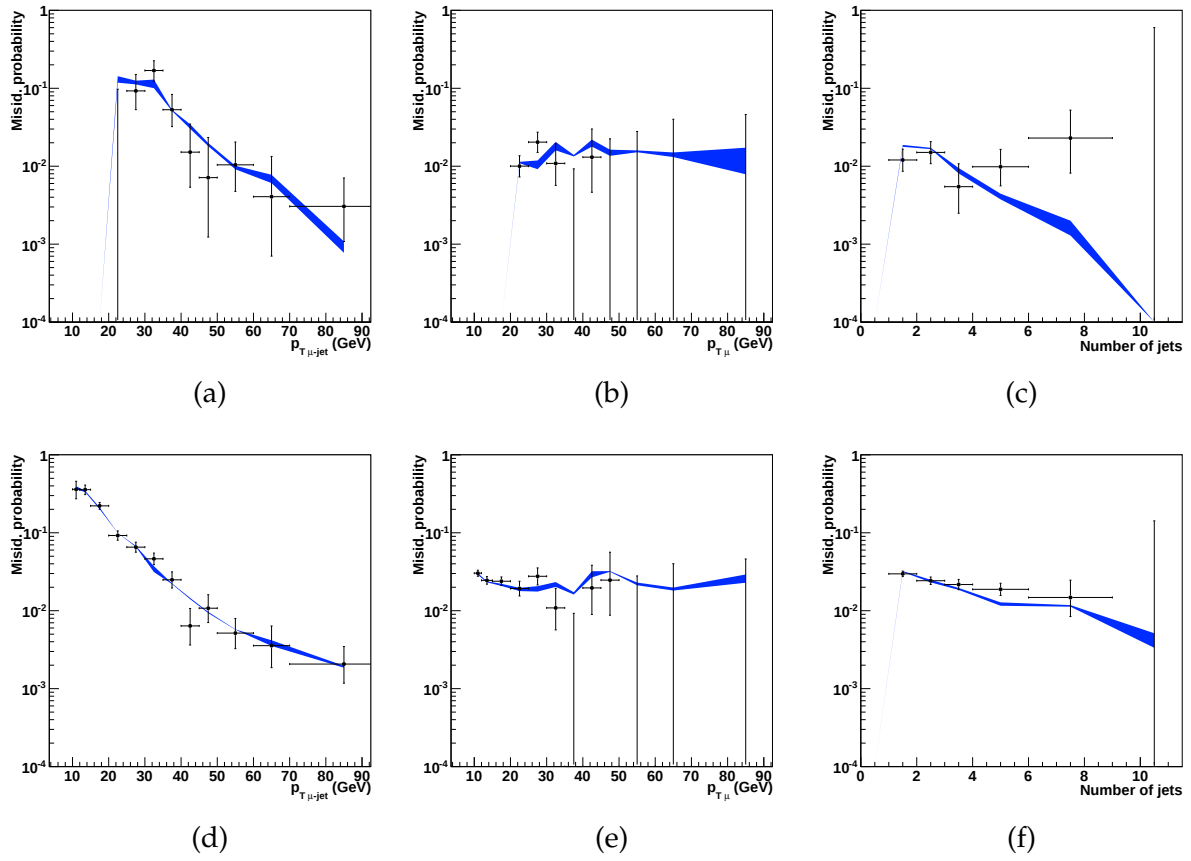


Figure A.4: Measured misidentification probability for leading (top) and sub-leading (bottom) muons in the training sample dedicated to the $\mu\mu\tau_h$ channel. The blue band represents the output of the kNN algorithm and its related uncertainty.

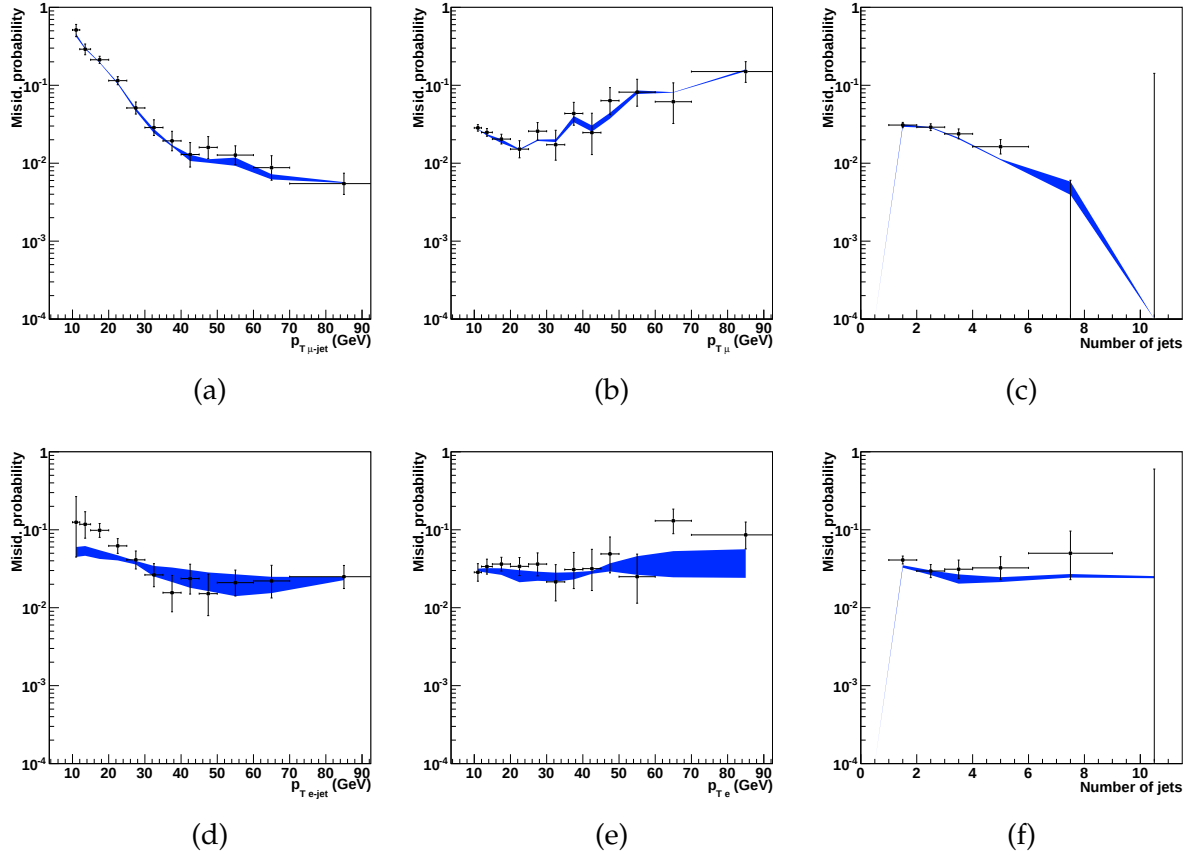


Figure A.5: Measured misidentification probability for muons (top) and electrons (bottom) in the training sample dedicated to the $e\mu\tau_h$ channel. The blue band represents the output of the kNN algorithm and its related uncertainty.

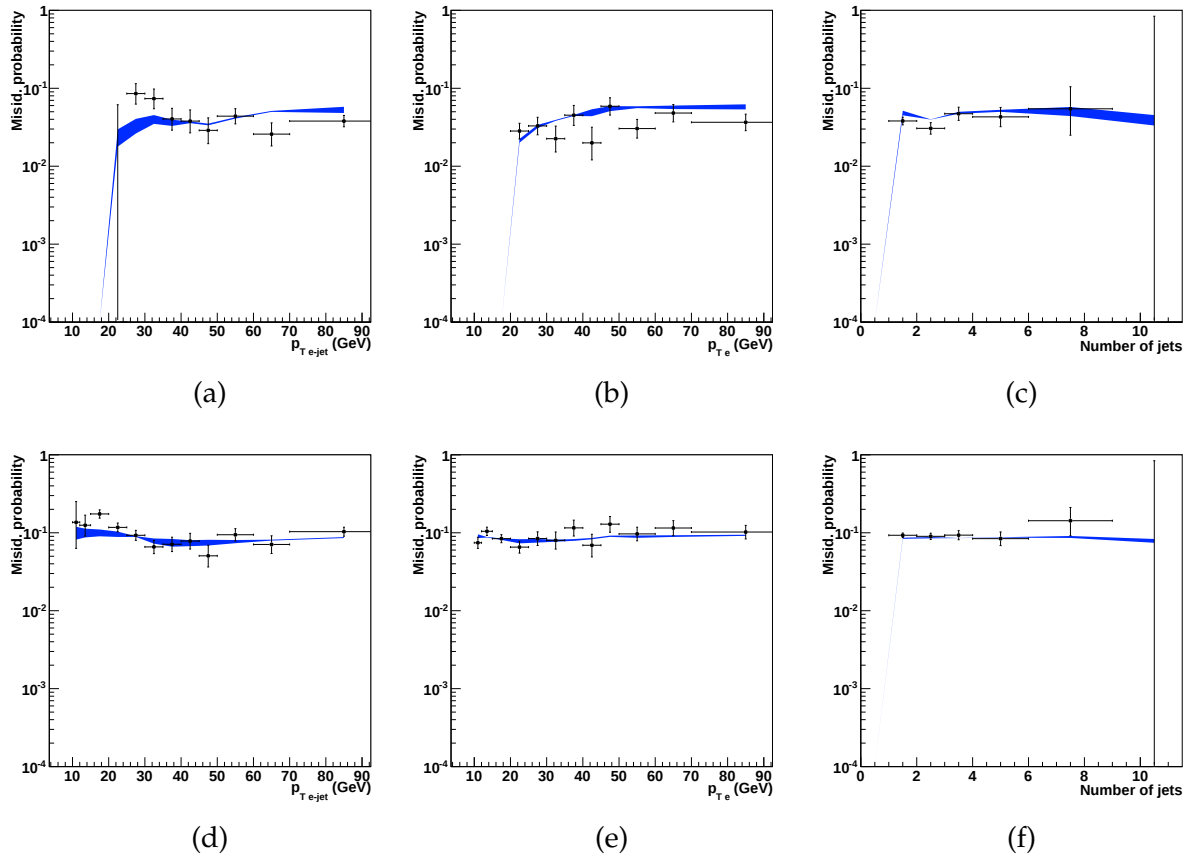


Figure A.6: Measured misidentification probability for leading (top) and sub-leading (bottom) electrons in the training sample dedicated to the $ee\tau_h$ channel. The blue band represents the output of the kNN algorithm and its related uncertainty.

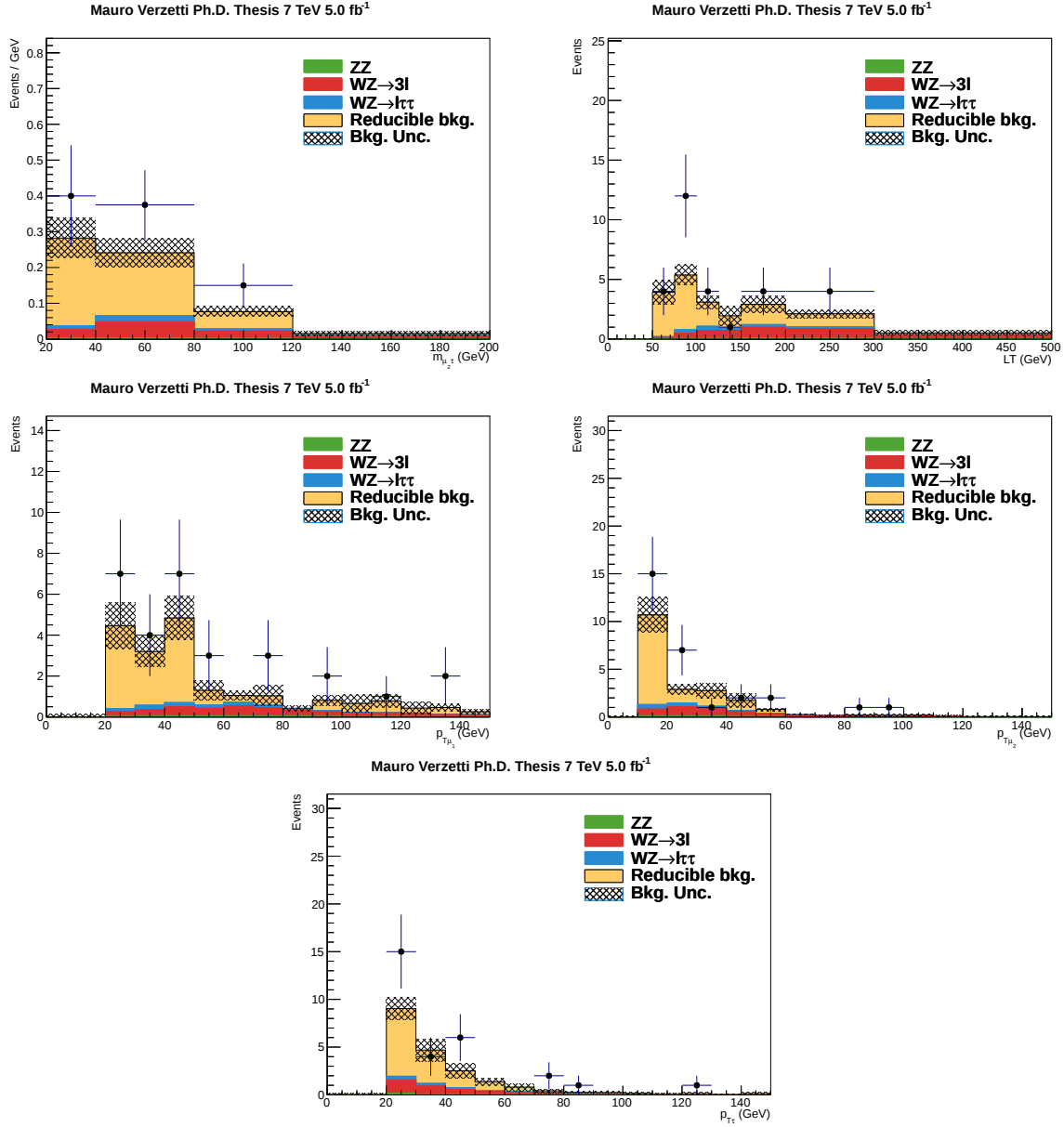


Figure A.7: Comparison of measured and predicted backgrounds in the $\mu\mu\tau_h$ "fake tau, charge ± 3 " control region for 7 TeV data. From top left to bottom: mass of the sub-leading muon and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading muon, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, in the same manner as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

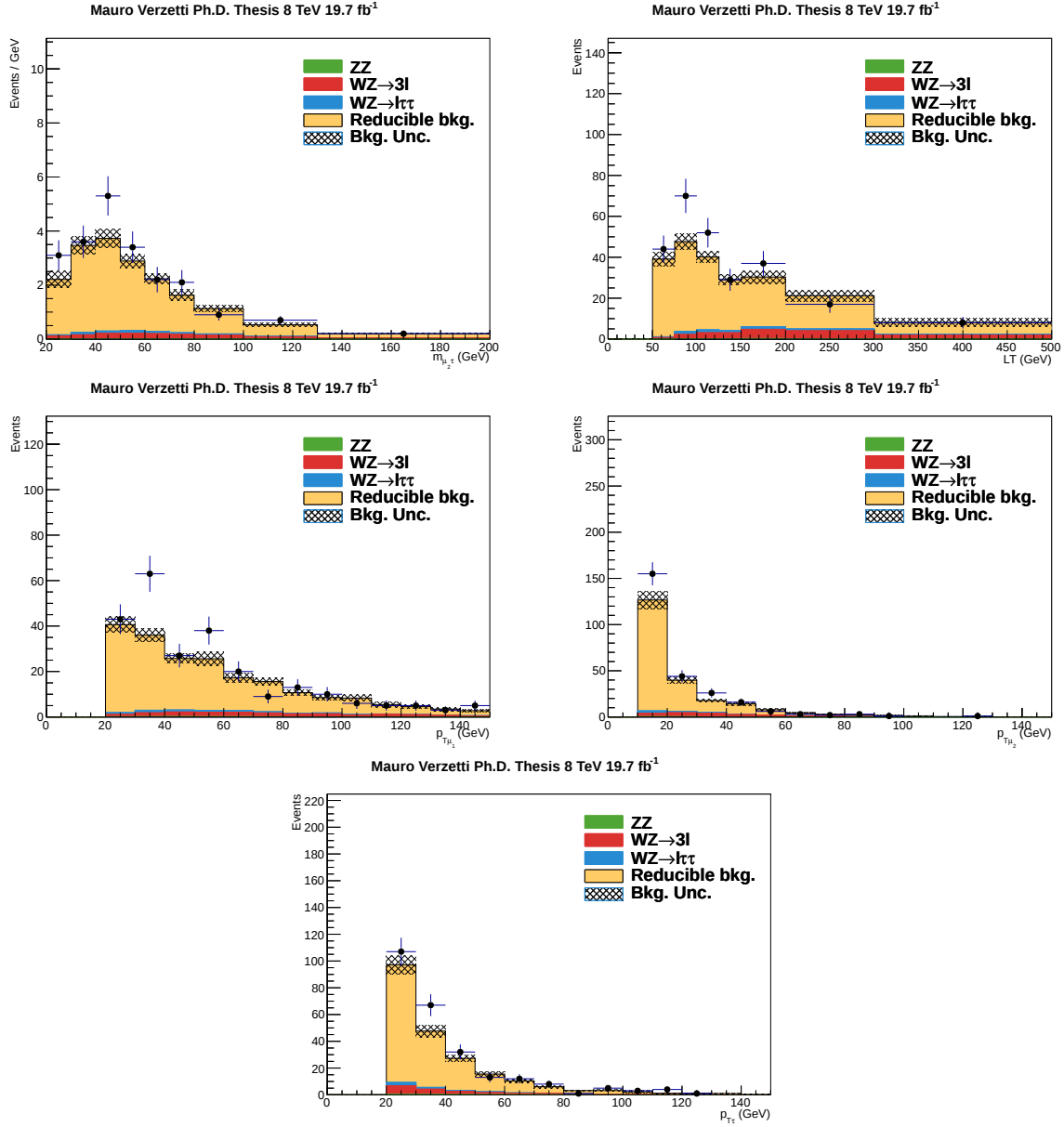


Figure A.8: Comparison of measured and predicted backgrounds in the $\mu\mu\tau_h$ "fake tau, charge ± 3 " control region for 8 TeV data. From top left to bottom: mass of the sub-leading muon and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading muon, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, in the same manner as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

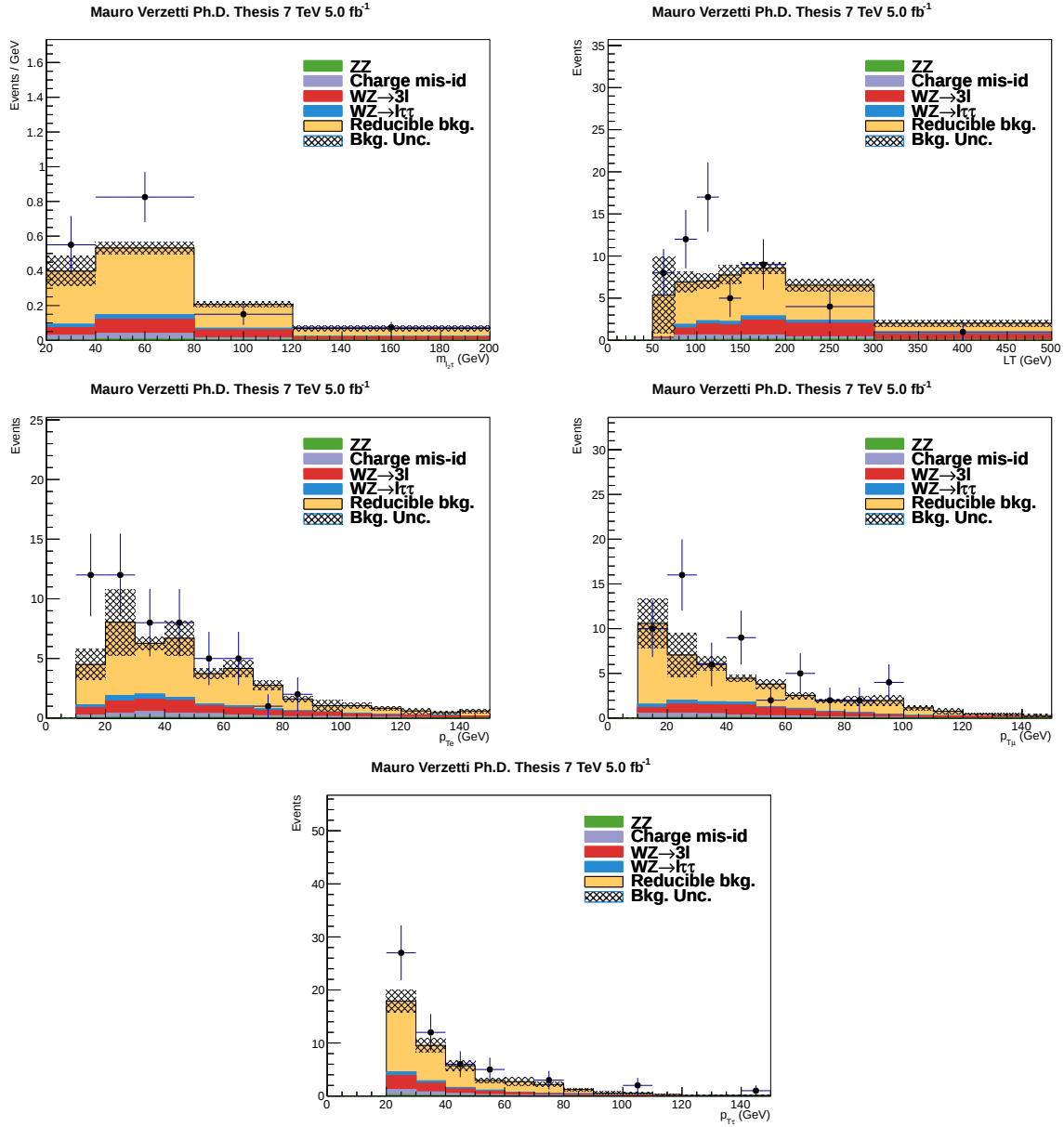


Figure A.9: Comparison of measured and predicted backgrounds in the $e\mu\tau_h$ "fake tau, charge ± 3 " control region for 7 TeV data. From top left to bottom: mass of the sub-leading lepton and the tau system, scalar sum of the leptons p_T (L_T), p_T of the electron, muon, and hadronic tau. The reducible background contribution is estimated by the kNN method, in the same manner as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

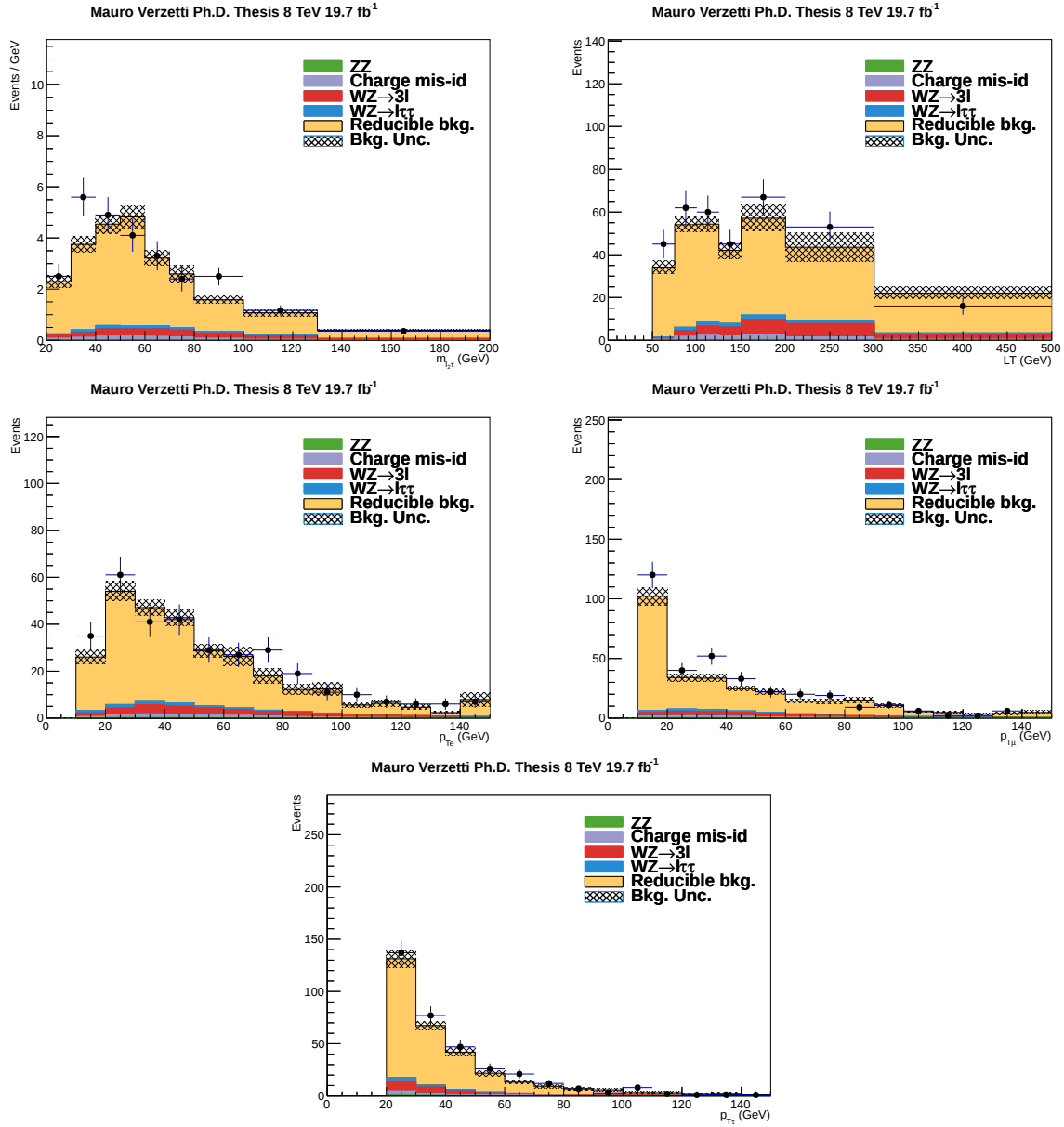


Figure A.10: Comparison of measured and predicted backgrounds in the $e\mu\tau_h$ "fake tau, charge ± 3 " control region for 8 TeV data. From top left to bottom: mass of the sub-leading lepton and the tau system, scalar sum of the leptons p_T (L_T), p_T of the electron, muon, and hadronic tau. The reducible background contribution is estimated by the kNN method, in the same manner as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

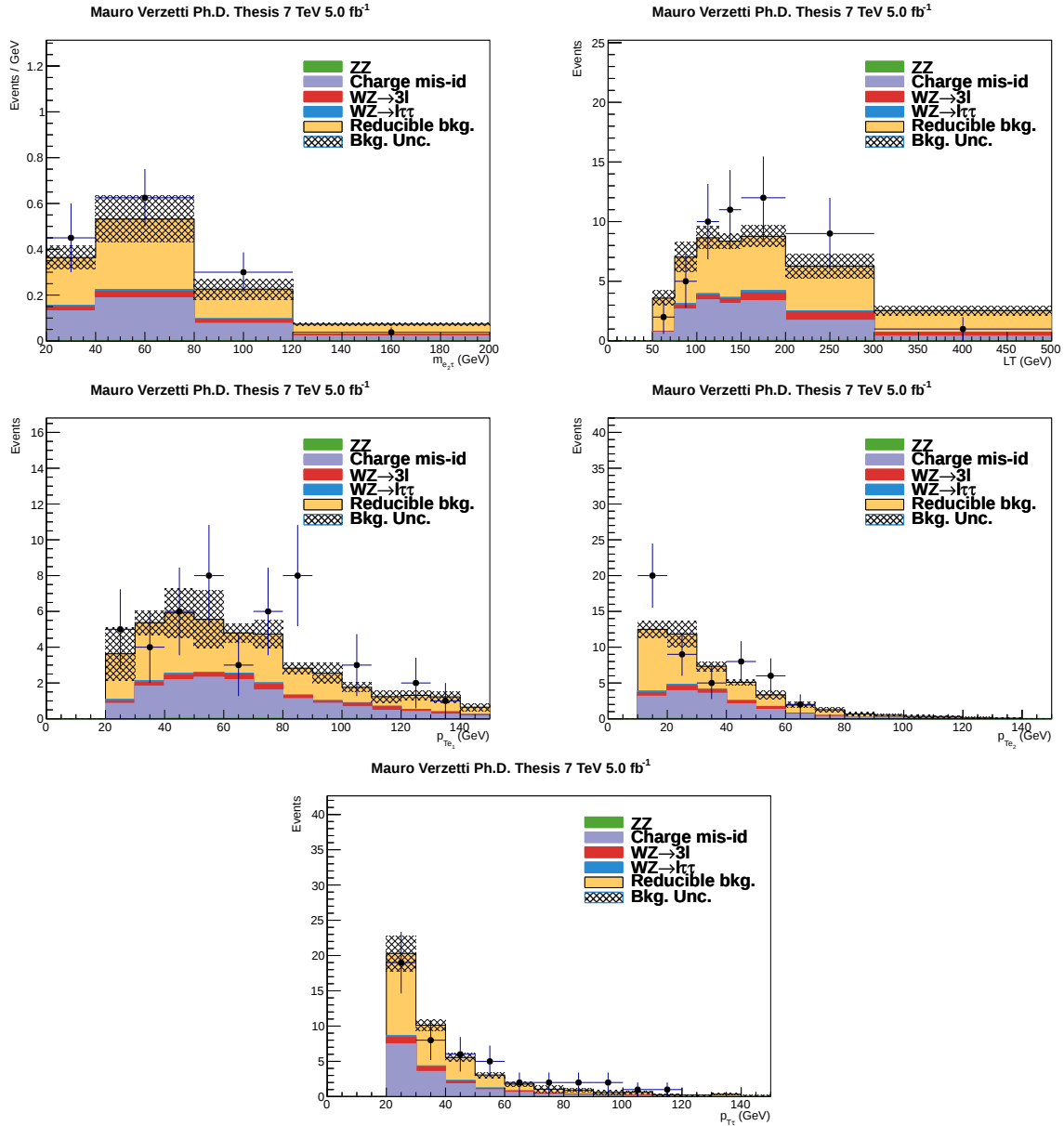


Figure A.11: Comparison of measured and predicted backgrounds in the $ee\tau_h$ "fake tau, charge ± 3 " control region for 7 TeV data. From top left to bottom: mass of the sub-leading electron and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading electron, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, in the same manner as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

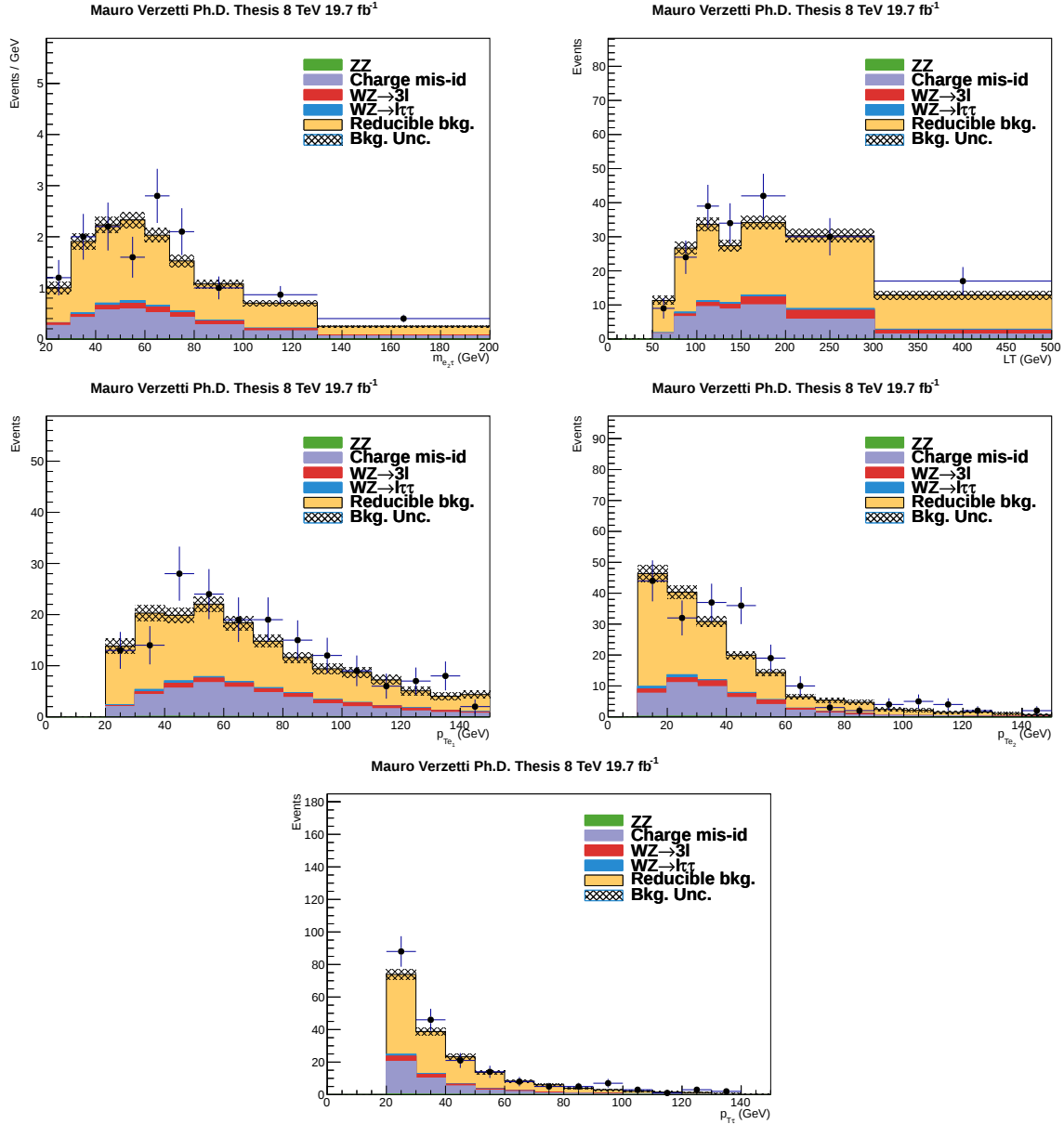


Figure A.12: Comparison of measured and predicted backgrounds in the $ee\tau_h$ “fake tau, charge ± 3 ” control region for 8 TeV data. From top left to bottom: mass of the sub-leading electron and the tau system, scalar sum of the leptons p_T (L_T), p_T of the leading and sub-leading electron, and p_T of the hadronic tau. The reducible background contribution is estimated by the kNN method, in the same manner as in the signal region. The shaded band represents the uncertainty on the sum of the background contributions.

Bibliography

- [1] OPERA Collaboration, “Observation of a first ν_τ candidate in the OPERA experiment in the CNGS beam”, *Phys.Lett.* **B691** (2010) 138–145, [arXiv:1006.1623](#). [doi:10.1016/j.physletb.2010.06.022](#). 2
- [2] Particle Data Group Collaboration, “Review of Particle Physics (RPP)”, *Phys.Rev.* **D86** (2012) 010001. [doi:10.1103/PhysRevD.86.010001](#). 2
- [3] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group 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](#). 2
- [4] BELLE Collaboration, “Observation of a Resonancelike Structure in the $\pi^\pm\psi'$ Mass Distribution in Exclusive $B \rightarrow K\pi^\pm\psi'$ Decays”, *Phys. Rev. Lett.* **100** (2008) 142001, [arXiv:0708.1790](#). [doi:10.1103/PhysRevLett.100.142001](#). 2
- [5] LHCb Collaboration, “Observation of the resonant character of the $Z(4430)^-$ state”, *Phys. Rev. Lett.* **112** (2014) 222002, [arXiv:1404.1903](#). [doi:10.1103/PhysRevLett.112.222002](#). 2
- [6] J. H. Christenson, J. W. Cronin, V. L. Fitch et al., “Evidence for the 2π Decay of the K_{20} Meson”, *Phys. Rev. Lett.* **13** (1964) 138–140. [doi:10.1103/PhysRevLett.13.138](#). 6
- [7] S. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl.Phys.* **22** (1961) 579–588. [doi:10.1016/0029-5582\(61\)90469-2](#). 7
- [8] A. Salam and J. C. Ward, “On a gauge theory of elementary interactions”, *Nuovo Cim.* **19** (1961) 165–170. [doi:10.1007/BF02812723](#). 7
- [9] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266. [doi:10.1103/PhysRevLett.19.1264](#). 7
- [10] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (1963) 531–533. [doi:10.1103/PhysRevLett.10.531](#). 8

- [11] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog.Theor.Phys.* **49** (1973) 652–657.
[doi:10.1143/PTP.49.652](#). 8
- [12] J. Goldstone, A. Salam, and S. Weinberg, “Broken Symmetries”, *Physical Review* **127** (August, 1962) 965–970. [doi:10.1103/PhysRev.127.965](#). 12
- [13] Y. Nambu, “Quasiparticles and Gauge Invariance in the Theory of Superconductivity”, *Phys.Rev.* **117** (1960) 648–663.
[doi:10.1103/PhysRev.117.648](#). 12
- [14] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321. [doi:10.1103/PhysRevLett.13.321](#). 12, 25
- [15] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964), no. 2, 132. [doi:10.1016/0031-9163\(64\)91136-9](#). 12, 25
- [16] H. Yukawa, “On the interaction of elementary particles”, *Proc.Phys.Math.Soc.Jap.* **17** (1935) 48–57. 15
- [17] J. Pumplin, D. Stump, J. Huston et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **0207** (2002) 012,
[arXiv:hep-ph/0201195](#). [doi:10.1088/1126-6708/2002/07/012](#). 17
- [18] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti et al., “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”, CERN-2013-004 (CERN, Geneva, 2013) [arXiv:1307.1347](#). 19
- [19] CMS Collaboration, “Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **1306** (2013) 081, [arXiv:1303.4571](#).
[doi:10.1007/JHEP06\(2013\)081](#). 20
- [20] CMS Collaboration, “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, *Submitted to Eur. Phys. J. C* (2014)
[arXiv:1407.0558](#). 20
- [21] 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). 20, 22
- [22] ATLAS Collaboration, “Measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels with the ATLAS detector using 25 fb⁻¹ of pp collision data”, *submitted to Phy. Rev. D* (2014) [arXiv:1406.3827](#). 20, 97
- [23] CMS Collaboration, “Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model”, *CMS-PAS-HIG-14-009* (2014). 20, 21, 22, 97

- [24] CMS Collaboration, “Constraints on anomalous HVV interactions using H to 4l decays”, *CMS-PAS-HIG-14-014* (2014). 20
- [25] F. Caola and K. Melnikov, “Constraining the Higgs boson width with ZZ production at the LHC”, *Phys.Rev.* **D88** (2013) 054024, [arXiv:1307.4935](#). [doi:10.1103/PhysRevD.88.054024](#). 21
- [26] 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](#). 21
- [27] ATLAS Collaboration, “Determination of the off-shell Higgs boson signal strength in the high-mass ZZ final state with the ATLAS detector”, *ATLAS-CONF-2014-042*, *ATLAS-COM-CONF-2014-052* (2014). 21
- [28] CDF Collaboration, 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](#). 21
- [29] 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) 012003, [arXiv:1310.3687](#). [doi:10.1103/PhysRevD.89.012003](#). 21
- [30] ATLAS Collaboration, “Search for the bb decay of the Standard Model Higgs boson in associated W/ZH production with the ATLAS detector”, *ATLAS-CONF-2013-079*, *ATLAS-COM-CONF-2013-080* (2013). 21
- [31] ATLAS Collaboration, “Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector”, *submitted to Phy. Lett. B* (2014) [arXiv:1406.7663](#). 21
- [32] CMS Collaboration, “Search for the standard model Higgs boson in the dimuon decay channel in pp collisions at $\sqrt{s}=7$ and 8 TeV”, *CMS-PAS-HIG-13-007* (2013). 21
- [33] ATLAS Collaboration, “Search for $H \rightarrow \gamma\gamma$ 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). 21
- [34] CMS Collaboration, “Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons”, *JHEP* **1405** (2014) 104, [arXiv:1401.5041](#). [doi:10.1007/JHEP05\(2014\)104](#). 21, 87

- [35] L. Bianchini, J. Conway, E. K. Friis et al., “Reconstruction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood techniques”, *J.Phys.Conf.Ser.* **513** (2014) 022035. doi:10.1088/1742-6596/513/2/022035. 22
- [36] M. Bluj, A. Burgmeier, T. Fruboes et al., “Modelling of $\tau\tau$ final states by embedding τ pairs in $Z \rightarrow \mu\mu$ events”, *AN-2011/020 CMS Internal Note*. 22
- [37] 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). 22
- [38] CMS Collaboration, “Evidence for the direct decay of the 125 GeV Higgs boson to fermions”, *Nature Phys.* **10** (2014) arXiv:1401.6527. doi:10.1038/nphys3005. 22
- [39] L. Evans, “The Large Hadron Collider (LHC)”, *New J.Phys.* **9** (2007) 335. doi:10.1088/1367-2630/9/9/335. 25
- [40] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. 25
- [41] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508. doi:10.1103/PhysRevLett.13.508. 25
- [42] G. Guralnik, C. Hagen, and T. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587. doi:10.1103/PhysRevLett.13.585. 25
- [43] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons”, *Phys.Rev.* **145** (1966) 1156–1163. doi:10.1103/PhysRev.145.1156. 25
- [44] T. Kibble, “Symmetry breaking in nonAbelian gauge theories”, *Phys.Rev.* **155** (1967) 1554–1561. doi:10.1103/PhysRev.155.1554. 25
- [45] S. P. Martin, “A Supersymmetry primer”, *Adv.Ser.Direct.High Energy Phys.* **21** (2010) 1–153, arXiv:hep-ph/9709356. 25
- [46] I. Antoniadis, K. Benakli, and M. Quiros, “Direct collider signatures of large extra dimensions”, *Phys.Lett.* **B460** (1999) 176–183, arXiv:hep-ph/9905311. doi:10.1016/S0370-2693(99)00764-9. 25
- [47] M. Beltran, D. Hooper, E. W. Kolb et al., “Maverick dark matter at colliders”, *JHEP* **1009** (2010) 037, arXiv:1002.4137. doi:10.1007/JHEP09(2010)037. 25
- [48] L. Randall and R. Sundrum, “An Alternative to compactification”, *Phys. Rev. Lett.* **83** (1999) 4690–4693, arXiv:hep-th/9906064. doi:10.1103/PhysRevLett.83.4690. 25

- [49] CMS Collaboration, “CMS Tracking Performance Results from early LHC Operation”, *Eur. Phys. J.* **C70** (2010) 1165–1192, [arXiv:1007.1988](#).
[doi:10.1140/epjc/s10052-010-1491-3](#). 31
- [50] P. Malecki, T. Szymocha, M. Turala et al., “Computing needs of LHC experiments and prospects for a world wide Grid for particle physics”, *Phys.Part.Nucl.* **36** (2005) 10–20. 38
- [51] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Submitted to JINST* (2014)
[arXiv:1405.6569](#). 39
- [52] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”,
Nucl.Instrum.Meth. **A262** (1987) 444–450. [doi:10.1016/0168-9002\(87\)90887-4](#).
40
- [53] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression and related Optimisation Problems”, *Proceedings of the IEEE* **Vol. 86, Issue 11** (1998). 42
- [54] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting”, CMS Public Note CMS-NOTE-2007/008, (2007). 42
- [55] CMS Collaboration, “CMS Physics Technical Design Report, Volume I: Detector Performance and Software”, *CERN-LHCC-2006-001 CMS-TDR-008-1* (2006). 43
- [56] W. Adam, R. Fruehwirth, A. Strandlie et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *J. Phys. G: Nuclear and Particle Physics* **31** (2005), no. 9, N9. 43
- [57] A. Hoecker, P. Speckmayer, J. Stelzer et al., “TMVA: Toolkit for Multivariate Data Analysis”, *PoS ACAT* (2007) 040, [arXiv:physics/0703039](#). 43, 66
- [58] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm”, *JHEP* **0804** (2008) 063, [arXiv:0802.1189](#).
[doi:10.1088/1126-6708/2008/04/063](#). 46
- [59] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys.Rev.* **D48** (1993) 3160–3166, [arXiv:hep-ph/9305266](#).
[doi:10.1103/PhysRevD.48.3160](#). 46
- [60] Y. L. Dokshitzer, G. Leder, S. Moretti et al., “Better jet clustering algorithms”, *JHEP* **9708** (1997) 001, [arXiv:hep-ph/9707323](#).
[doi:10.1088/1126-6708/1997/08/001](#). 46
- [61] 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](#). 46, 47

- [62] CMS Collaboration, “Pileup Jet Identification”, *CMS-PAS-JME-13-005* (2014). 46
- [63] CMS Collaboration, “Performance of the Missing Transverse Energy Reconstruction by the CMS experiment in the 8 TeV pp data”, *CMS-PAS-JME-13-003* (2014). 47
- [64] Marionneau, M. and Veelken, C., “Study of MET Performance in the full 2012 Dataset using $Z \rightarrow \mu\mu$ events”, *AN-2013/233 CMS Internal Note*. 47
- [65] CMS Collaboration, “Tau identification in CMS”, *CMS-PAS-TAU-11-001* (2011). 47
- [66] The Tau Physics Object Group, “Performance of tau reconstruction and identification in pp collisions at $\sqrt{s} = 8$ TeV”, *AN-2014/008 CMS Internal Note*. 47
- [67] L. Bianchini, I. Naranjo Fong, C. Veelken, “Development of a new tau identification discriminator against electrons”, *AN-2012/417 CMS Internal Note*. 49
- [68] C. Cuenca Almenar, “Search for the neutral MSSM Higgs bosons in the $\tau\tau$ decay channels at CDF Run II”,. 52
- [69] The Tau Physics Object Group, “Tau ID Performance Plots”, *DP-2014/015 CMS Internal Note*. 55
- [70] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, [arXiv:1206.4071](#). [doi:10.1088/1748-0221/7/10/P10002](#). 60
- [71] CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7$ TeV”, *CMS-PAS-EGM-10-004* (2010). 60
- [72] CMS Collaboration, “Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons”, *CMS-PAS-HIG-13-019* (2013). 63
- [73] J. Alwall, P. Demin, S. de Visscher et al., “MadGraph/MadEvent v4: The New Web Generation”, *JHEP* **0709** (2007) 028, [arXiv:0706.2334](#). [doi:10.1088/1126-6708/2007/09/028](#). 64
- [74] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **0605** (2006) 026, [arXiv:hep-ph/0603175](#). [doi:10.1088/1126-6708/2006/05/026](#). 64
- [75] S. Jadach, Z. Was, R. Decker et al., “The tau decay library TAUOLA: Version 2.4”, *Comput.Phys.Commun.* **76** (1993) 361–380. [doi:10.1016/0010-4655\(93\)90061-G](#). 64

- [76] TOTEM Collaboration Collaboration, "Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at $\sqrt{s} = 8\text{ TeV}$ ", *Phys.Rev.Lett.* **111** (2013), no. 1, 012001. doi:10.1103/PhysRevLett.111.012001. 64
- [77] J. M. Campbell and R. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10, arXiv:1007.3492. doi:10.1016/j.nuclphysbps.2010.08.011. 64
- [78] J. Allison, K. Amako, J. Apostolakis et al., "Geant4 developments and applications", *Nuclear Science, IEEE Transactions on* **53** (2006), no. 1, 270. doi:10.1109/TNS.2006.869826. 80
- [79] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS:Winter 2012 Update", *CMS-PAS-SMP-12-008* (2011). 86
- [80] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", *CERN-2011-002* (CERN, Geneva, 2011) arXiv:1101.0593. 86
- [81] J. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", *Proceedings of PHYSTAT 2011* (2011) arXiv:1103.0354. 94
- [82] A. L. Read, "Modified frequentist analysis of search results (The CL(s) method)", *CERN-OPEN-2000-205* (2000). 94
- [83] ATLAS and CMS Collaboration, "Procedure for the LHC Higgs boson search combination in summer 2011", *ATL-PHYS-PUB-2011-011*, *ATL-COM-PHYS-2011-818*, *CMS-NOTE-2011-005* (2011). 94, 95
- [84] 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. 95
- [85] I. Asimov, "Franchise, The complete stories". Isaac Asimov. Doubleday, 1990. 102