Measurement of the CP properties of the Higgs boson in its decays to τ leptons with the CMS experiment

Dissertation

vorgelegt von

Andrea Cardini

aus

Florenz, Italien

Hamburg

2021

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used any resources and aids other than those acknowledged. The written version submitted corresponds to the one stored electronically. The dissertation presented in this form, has not already been accepted in an earlier doctoral procedure or assessed as unsatisfactory.

Hamburg, 04. Mai 2021

Andrea Cardini

D'azzurro, a due piante di cardo d'oro, decussate

"Dubium sapientiae initium"

René Descartes

Abstract

The discovery of the Higgs boson in 2012 by the CMS and ATLAS collaborations marked the start of the exploration of the Higgs sector of particle physics. This thesis explores one of these measurements: the study of the CP properties of the Yukawa coupling between the Higgs and the τ lepton. The study discussed in this thesis is performed using data collected in proton-proton collisions at 13 TeV by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The dataset used in the measurement was collected throughout the Run 2 data-taking period at the LHC, corresponding to an integrated luminosity of 137 fb^{-1}.

In $H \to \tau\tau$ decays the amount of CP-mixing in the Yukawa coupling can be expressed in terms of the CP mixing angle $\varphi_{\tau\tau}$, defined such that $\varphi_{\tau\tau} = 0, \pi/2$ and $\pi/4$ correspond to a pure CP-even, CP-odd and maximal CP-mixing cases. This angle can be inferred by measuring the angle between the decay planes of the two τ leptons (φ_{CP}). The decay planes reconstruction was performed using dedicated methods for each decay channel. The estimation of most background processes was performed using data-driven methods, while a neural network was used to identify Higgs decays with respect to the dominant backgrounds.

The measurement was performed in this thesis in the final state where a tau lepton decays muonically and the other one hadronically ($\tau_\mu\tau_h$), and combined with one performed in the fully hadronic final state ($\tau_h\tau_h$). The measured CP admixture is $\varphi_{\tau\tau} = (4 \pm 17)°$ [$(0\pm23)°$], corresponding to a significance for the exclusion of a pure CP-odd hypothesis of 3.2σ (2.3σ). The results are in good agreement with the SM prediction of a pure CP-even coupling, and allow to exclude at 95% confidence level models that predict CP-even and CP-odd Yukawa couplings of comparable magnitude.

Zusammenfassung

Die Entdeckung des Higgs-Bosons im Jahr 2012 durch die Zusammenarbeit der CMS- und ATLAS-Experimente hat den Beginn der Erforschung des Higgs-Sektors der Teilchenphysik markiert. Die Eigenschaften des Higgs-Bosons mit einer Masse von 125 GeV, wurden im Laufe der Jahre eingehend untersucht. Zu diesen Eigenschaften gehört die Struktur unter Ladungskonjugation und Paritätssymmetrie (CP) der Higgs-Kopplungen. Diese Dissertation untersucht eine dieser Messungen: die Untersuchung der CP-Eigenschaften der Yukawa-Kopplung zwischen dem Higgs und dem τ-Lepton. Die Messung von dieser Doktorarbeit wird unter Verwendung von Daten durchgeführt, die bei Proton-Proton-Kollisionen bei einer Massenschwerpunktsenergie von 13 TeV durch das Compact Muon Solenoid (CMS) - Experiment am Large Hadron Collider (LHC) gesammelt wurden. Der für die Messung verwendete Datensatz wurde während des gesamten Datenerfassungszeitraums von Lauf 2 am LHC gesammelt, was einer integrierten Luminosität von 137 fb^{-1} entspricht.

In $H \to \tau\tau$-Zerfällen kann die Menge der CP-Mischung in der Yukawa-Kopplung als CP-Mischwinkel $\varphi_{\tau\tau}$ ausgedrückt werden, der in der Tat als folgendes definiert, dass $\varphi_{\tau\tau} = 0, \pi/2$ und $\pi/4$ einem reinen CP-geraden, CP-ungeraden und maximalen CP-Mischfall entsprechen. Dieser Winkel kann durch Messen des Winkels zwischen den Zerfallsebenen der beiden τ-Leptonen (φ_{CP}) abgeleitet werden. Die Rekonstruktion der Zerfallsebenen wurde unter Verwendung spezieller Methoden für jeden Zerfallskanal durchgeführt.

Die Abschätzung der meisten Untergrundprozesse wurde unter Verwendung von datengetriebenen Methoden durchgeführt, während ein neuronales Netzwerk verwendet wurde, um Higgs-Zerfälle in Bezug auf die dominanten Untergründe zu identifizieren.

In dieser Doktorarbeit die Messung wurde im Endzustand durchgeführt, in dem ein Tau-Lepton muonisch und das andere hadronisch ($\tau_\mu\tau_h$) zerfällt und mit einer Studie kombiniert, die im vollständig hadronischen Endzustand ($\tau_h\tau_h$) durchgeführt wurde. Die Schätzung der CP-Beimischung ist $\varphi_{\tau\tau} = (4 \pm 17)° \ [(0 \pm 23)°]$, entsprechend einer Signifikanz für die Trennung zwischen CP-geraden und CP-ungeraden Hypothesen von $3.2\sigma \ (2.3\sigma)$. Die Ergebnisse stimmen gut mit der SM-Vorhersage einer reinen CP-geraden Kopplung überein und erlauben es, Modelle mit einem Konfidenzniveau von 95% auszuschließen, die CP-gerade und CP-ungerade Yukawa-Kopplungen vergleichbarer Größe vorzuhersagen.

Contents

Introduction

"Apri la mente a quel ch'io ti paleso e fermalvi entro; ché non fa scienza, senza lo ritenere, aver inteso." (Open your mind to what I will show you, and keep that knowledge within you; for it is not science, to listen without understanding.)

"Divina Commedia, Paradiso, Canto V", Dante Alighieri

The discovery of the Higgs boson, announced in 2012 by the CMS and ATLAS collaborations [1, 2], marked the start of the exploration of the Higgs sector of particle physics. After its discovery, several Higgs decay channels have been observed, most notably the decay via vector bosons, and the ones via a pair of fermions. The $H \to \tau\tau$ and $H \to b\bar{b}$ decays were discovered in 2017 [3] and 2018 [4] respectively, while evidence for the $H \to \mu\mu$ decay was found in 2020 [5]. The properties of the Higgs boson (125 GeV) have been studied extensively throughout the years. In particular, the properties of the Higgs sector under CP symmetry have been investigated. The Standard Model (SM) of particle physics predicts a pure CP-even coupling for the Higgs boson both to bosons and fermions, and any anomalous CP-odd component in the couplings would be a clear sign of new physics. The CP properties of the Higgs couplings to vector bosons have been investigated extensively throughout the years, focusing both on the production via Vector Boson Fusion (VBF) and the $H \to VV$ decays. This thesis work explores the CP properties of a different interaction: the Yukawa coupling between the Higgs and the τ lepton.

To properly discuss the methods and the results of this thesis work, the theoretical framework has to be clarified. Chapter 1 presents an overview of the Standard Model of particle physics with greater focus being dedicated to the τ lepton and its decays which played a crucial role in the study being presented. The Electroweak Symmetry Breaking (EWSM) or Brout-Englert-Higgs mechanism is also described, together with the Yukawa interaction, as an introduction to the Higgs boson physics.

Chapter 2 provides a description of the CP structure of the Higgs coupling both to vector-bosons and to fermions. The latter interaction is discussed in greater detail as the focus for the main analysis shown in this thesis work. A greater focus is given to the experimental signatures which allow to investigate the CP structure of the Yukawa coupling of the Higgs to τ leptons.

The study discussed in this thesis is performed using data collected in proton-proton collisions at 13 TeV by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The dataset used in the measurement was collected throughout

the Run 2 data-taking period at the LHC, corresponding to an integrated luminosity of 137 fb^{-1}. To better understand how the proton-proton collisions are reconstructed, Chapter 3 provides an overview of the CMS detector. The focus of the chapter is kept on the main subsystems active during the data-taking period. The description of the object reconstruction in CMS is provided in Chapter 4, which is dedicated to the physics objects reconstruction techniques, and in particular to the ones targeting τ leptons, used in CMS. An overview of the reconstruction algorithm is provided alongside the decay mode identification for τ candidates. Due to the presence of several sources of misidentification for hadronically decaying τ leptons, part of the chapter is dedicated to DeepTau, a Neural Network (NN) based τ identification algorithm recently introduced to reduce the misidentification rate. The DeepTau identification provided a noticeable improvement in the rejection of jets and leptons faking hadronically decaying τ leptons (τ_h). This is shown in Chapter 5 where the measurement of the electron misidentification rate as τ_h is presented. This measurement was performed by myself to improve the modelling of the electron misidentification as τ_h in simulated $Z \to ee$ events.

Chapter 6 presents the core measurement which I performed in this thesis: the study of the properties of the $H \to \tau\tau$ Yukawa coupling under CP symmetry in the $\tau_\mu\tau_h$ channel. After a brief overview of the analysis, the event reconstruction and selection is presented (Sections 6.3 and 6.6). This section outlines methods developed specifically for this analysis: a multivariate analysis based decay mode identification for τ_h, and the impact parameter and primary vertex reconstruction. The estimation of most background processes was performed using data-driven techniques, developed in the CMS collaboration, and is presented in Section 6.5. The event classification, applied to increase the separation between signal and background processes, was performed using a NN, as shown in Section 6.7. The measurement of the Yukawa coupling CP structure was performed by comparing simulated signal templates, parametrized to account for different levels of CP-admixture in the couplings, to the recorded data. A negative log-likelihood minimization was performed using simulated templates of Higgs decays and background processes as detailed in Section 6.8. Finally, the results of the measurement are detailed in Section 6.9 using the $\tau_\mu\tau_h$ final state and the data recorded by the CMS experiment in Run 2. In Section 6.9.2, a study is presented to show how the choice of the primary vertex affected the expected significance.

This measurement was combined with the corresponding study in the $\tau_h\tau_h$ channel. A brief description of this complementary measurement is shown in Chapter 7, with the focus being placed on the differences with the $\tau_\mu\tau_h$ channel. The combined results are interpreted as a measurement of CP-even and CP-odd contributions to the Higgs Yukawa coupling to tau leptons. Section 7.3 shows how the measured couplings can be interpreted in the context of some theories beyond the SM.

The Standard Model of particle physics

Contents

This thesis work focuses on the study of the CP properties of the Yukawa interactions between the Higgs boson and the τ leptons. The study was performed by reconstructing and analyzing events where a Higgs boson decays into a pair of τ leptons. To properly understand this process and how its properties were studied it is crucial to set a clear theoretical framework in which this analysis is performed.

This framework is the Standard Model (SM) of particle physics [6–8]. This theory aims at describing the nature and behavior of elementary particles and their interactions. It relies on Quantum Field Theory (QFT) [9], which describes particles as excited states of quantum fields, and their interaction as mediated by the exchange of force carriers, namely the gauge bosons. This description is applied to the Electromagnetic, Weak and Strong interactions, which are three out of the four fundamental forces of Nature. The fourth force, Gravity, is not described within the SM. This is because a self-consistent theory for a quantized gravitational field has not been proven to exist yet. This thesis work uses the SM as the base theoretical framework, therefore Gravity will not be included among the interactions between particles, unless explicitly addressed as such.

Elementary particles in the SM are defined as point-like objects which cannot be decomposed into, or are not bound states of, other particles. Experimentally this definition means that for a particle to be considered elementary in the SM, its internal structure, if present, must not have been observed experimentally at any energy. For each elementary particle, an antiparticle also exists and is characterized by having the same mass and spin and opposite additive quantum numbers. The elementary constituents of matter in the SM are *fermions*, i.e. particles with half-integer spin. These particles are classified into three generations, in ascending order of mass, and two families: *quarks* and *leptons*. Members of both families can interact via electroweak interaction, while only the quarks experience the strong force. The particle content of the SM is shown in Fig. 1.1, while a more detailed description of their properties is postponed to Section 1.5.
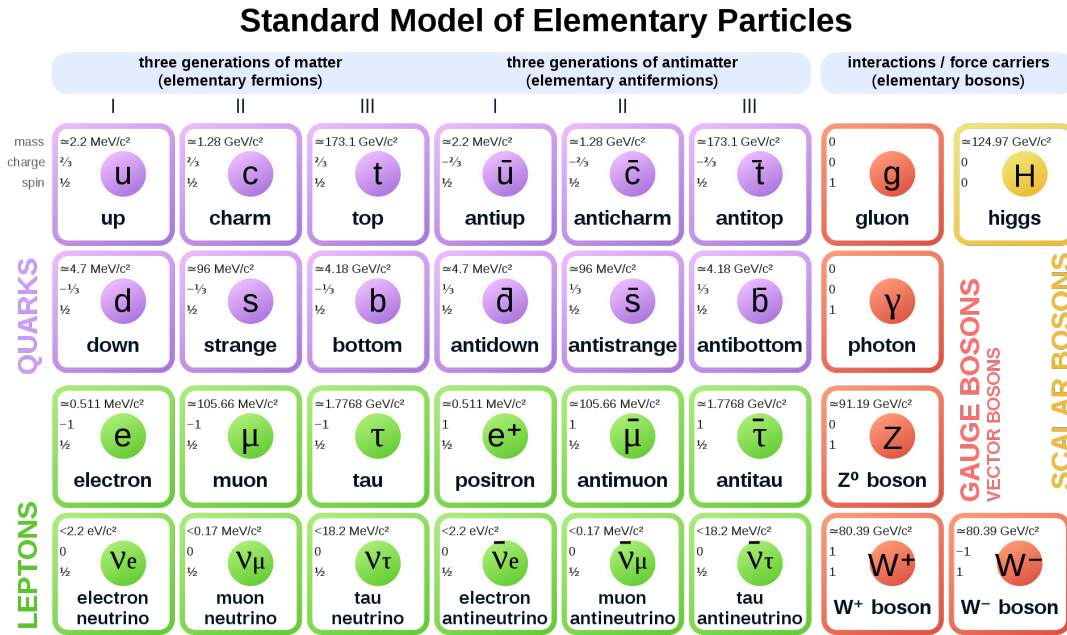
Figure 1.1: Elementary particles in the SM [10].

The Standard Model was developed between the 60s and the 70s and its Lagrangian has a symmetry of non-Abelian gauge group of the type $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, corresponding to the two symmetries that make up the model: $SU(3)_C$ for *Quantum Chromo-Dynamics* (QCD) [11], $SU(2)_L \otimes U(1)_Y$ for *Electroweak* (EW) theory [9]. Of the two symmetries, the *color symmetry* ($SU(3)_C$) is exact while the electroweak symmetry ($SU(2)_L \otimes U(1)_Y$) is broken because it would impose on both the mediator bosons and the fermionic particles to be massless, which is experimentally disproved. The symmetry is broken by introducing a scalar field, called the *Higgs field*, whose Lagrangian directly assigns mass to the W and Z mediator bosons of the weak interaction. The electroweak

symmetry breaking (EWSB) [12, 13] is further described in Section 1.3, while Section 1.2 provides greater details on the electroweak and strong interactions and their corresponding mediator bosons.

1.1 Symmetries of the SM Lagrangian

The symmetries of the SM Lagrangian [8, 14] which are relevant for this thesis work are:

- **Poincaré transformations**: translation of the spatial coordinates and Lorentz boosts when moving from one frame of reference to another;

- **gauge symmetries**: a group of transformation of the fields which keeps all physical observables unchanged;

- **discrete symmetries**: parity, charge-conjugation and time-reversal.

In general, a symmetry is defined as a group G of transformations for which an operation $\times$ is defined, such that $G \times G \longrightarrow G$ (if $x, y \in G$ then $xy \in G$), with the following properties:

- ***associativity***: $(xy)z = x(yz) \;\; \forall \, x, y, z \in G$;

- existence of the **identity** element: $\exists \, I \, \in \, G \, : \;\; Ix = xI = x \, \forall x \, \in G$;

- existence of the **inverse** element: $\forall x \, \exists y \, : \;\; xy = yx = I$.

In this section, I is used to represent both the identity operator of a group and the identity matrix[1] of an appropriate size, associated with a representation of the group.

The SM is a relativistic QFT, its laws should therefore be invariant under the space-time symmetries of nature. These transformations form the Poincaré group, whose generators are the translation operators P^μ, $\mu \in \{0, 1, 2, 3\}$ and the generators of the Lorentz group $J^{\mu\nu}$, $\mu, \nu \in \{0, 1, 2, 3\}$. This group forms a Lie algebra with the following commutation rules:

$$\left[P^\mu, P^\nu \right] = 0 \, , \tag{1.1}$$

$$\left[P^\mu, J^{\nu,\lambda} \right] = i \left(\eta^{\mu\nu} P^\lambda - \eta^{\mu\lambda} P^\nu \right) \, , \tag{1.2}$$

$$\left[J^{\mu\nu}, J^{\lambda\kappa} \right] = i \left(\eta^{\mu\kappa} J^{\nu\lambda} + \eta^{\nu\lambda} J^{\mu\kappa} - \eta^{\nu\kappa} J^{\mu\lambda} - \eta^{\mu\lambda} J^{\nu\kappa} \right) \, , \tag{1.3}$$

---

[1]Diagonal matrix with elements $I_{i,j} \equiv \delta_{i,j}$, with $\delta_{i,j}$ Kronecker delta.

with $\eta^{\mu\nu}$ metric of the Minkowski space. The translations form an Abelian sub-group of the Poincaré group, which means that its operators can commute with each other. The Lorentz transformations are the combination of spatial rotation and Lorentz boosts, and form a non-Abelian group. When discussing the fields appearing in the SM Lagrangian it is common to refer to scalar and vectorial objects. Such names refer to the properties of a field under Poincaré transformation. In particular, a field is called scalar if it is invariant under Poincaré transformations, while vectors transform like $x'^{\mu} = T^{\mu}_{\nu}\, x^{\nu}$ and tensors like $M'^{\mu}_{\nu} = T^{\mu}_{\lambda}\, M^{\lambda}_{\kappa}\, T^{\kappa}_{\nu}$, with T^{μ}_{ν} generic transformation of the Poincaré group.

The SM Lagrangian is invariant under the gauge symmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, which is the direct product of 3 symmetry groups. Following Noether's theorem [15], a conserved charge is associated with each one of these symmetries. The $U(1)_Y$ is the group of unitary matrices[2] of dimension 1. The conserved quantity associated to this symmetry is the *hypercharge*[3]. The $SU(2)_L$ and $SU(3)_C$ are the non-Abelian groups of special unitary matrices[4] of dimension 2 and 3 respectively. The *weak isospin* and the *color charge* are the conserved quantities associated to $SU(2)_L$ and $SU(3)_C$ respectively. The dimension[5] of each group corresponds to the number of gauge fields that mediate the associated interaction. The $SU(3)_C$ group has dimension 8, leading to the 8 gluons present in the SM. $SU(2)$ has dimension 3 corresponding to the 3 fields A^{μ}, $\mu \in \{1,2,3\}$, while the field B is associated to the $U(1)$ group. Section 1.3 is dedicated to show how the $SU(2)_L \otimes U(1)_Y$ symmetry is broken in the electromagnetic $U(1)_{em}$ and how a linear combination of A^{μ} and B fields can be used to define the fields associated to a photon or the W and Z bosons.

Another important transformation to address is the *chirality*, a symmetry based on the chiral projectors[6], χ_L and χ_R. It is a discrete symmetry, meaning that it does not change a system in a continuous way, and cannot be associated to an infinitesimal transformation. It is also finite, since its operators allow only two possible eigenvalues: 0 and 1. The two operators are orthogonal to each other and $\chi_L + \chi_R = I$, we can therefore separate the state of a particle φ in its *left* (φ_L) and *right* (φ_R) components, which are their projections under the chiral operators χ_L and χ_R respectively. More precisely the $\chi_{L(R)}$ projector selects the left(right)-handed component of a particle and the right(left)-handed component of an antiparticle. Left-handed particles and right-handed antiparticles transform as a doublet under $SU(2)_L$, while right-handed particles and left-handed antiparticles transform as a singlet. Fermions with left-handed chirality have weak isospin T=1/2, while those with right-handed chirality have T=0.

---

[2]A unitary matrix U is characterized by having an inverse equal to its transpose conjugate: $U^{\dagger} \times U = I$.

[3]Sometimes referred to as weak hypercharge.

[4]A unitary matrix with determinant equal to 1.

[5]The dimension of a group corresponds to the number of generators its fundamental representation possesses. For unitary groups $U(n)$ it is n^2, while $SU(n)$ groups have dimension $n^2 - 1$.

[6]A projector P is an operator with only 1 or 0 as possible eigenvalues, and with the property that $P^2 \equiv P$.

The SM Lagrangian is invariant under CPT symmetry, the combination of 3 distinct transformations: *parity* (P), *charge-conjugation* (C) and *time-reversal* (T). The invariance under CPT is a requirement for any self-consistent Lorentz invariant quantum field theory. The 3 individual symmetries are instead not necessarily conserved. Before approaching which parts of the Lagrangian are invariant under these 3 symmetries it is useful to describe them in more detail. They are all discrete symmetries, whose operators have only two eigenvalues: +1 and -1.

The parity is the inversion of the spatial coordinates: a Lorentz vector representing the coordinates of an object $v = (t, x, y, z)$ is transformed under parity in $Pv = (t, -x, -y, -z)$. It is a unitary and Hermitian[7] operator therefore it has the property $P \cdot P \equiv I$. Under the parity operator, physical quantities such as energy and spin remain unaltered while helicity and momenta change sign.

The charge-conjugation (also known as C-parity) transforms a particle in its corresponding antiparticle. More generally it changes the sign of all additive quantum numbers of a particle or field. Like the parity operator, it is also Hermitian and unitary, therefore $C \cdot C \equiv I$. Momenta, energy, spin, and helicity are all unaltered by C-parity, while electric and magnetic field change sign.

The time-reversal is an anti-unitary operator: $T^{-1}(-i)T \equiv i$. It inverts momenta, energy and spin, while position and helicity remain unchanged. Due to the experimental difficulties of studying the properties of time-reversal it has been historically preferred the study of the CP properties of interactions. Due to CPT being a symmetry satisfied by the SM Lagrangian, a violation under CP corresponds to a violation of time-reversal.

The weak interaction violates both parity and charge-conjugation [16]: a left-handed particle is transformed into a right-handed particle under parity and into a left-handed antiparticle under charge-conjugation, while the W boson can couple only to a left-handed particle or a right-handed antiparticle. However electromagnetic and strong interaction both conserve parity and charge-conjugation. CP symmetry is minimally violated by the weak interaction, as observed in the Cronin and Fitch experiment [17].

1.2   Strong and electroweak interactions

The SM Lagrangian can be written as a sum of three terms, one related to the QCD, one to the EW interaction, and one to the Higgs field:

$$\mathscr{L}_{SM} = \mathscr{L}_{QCD} + \mathscr{L}_{EW} + \mathscr{L}_{Higgs} . \tag{1.4}$$

Quantum chromodynamics (QCD) is a non-Abelian field theory formulated in 1973 by Fritzsch [18], Politzer [19], Gross, and Wilczek [20], to describe the phenomenology of hadronic interactions and spectroscopy. It is derived from the quark model of baryons

---

[7]A Hermitian or self-adjointed operator is defined by being equal to its adjoint conjugate: $H^{\dagger} \equiv H$.

and mesons proposed by Gell-Mann in the 60s [21]. When this model was first suggested there was no experimental evidence of quarks, and it also seemed to require a violation of the Pauli exclusion principle in order to describe some states of hadronic spectroscopy, e.g. the Δ^{++} baryon. To account for that, a new charge of "color" was introduced, with three different values for particles (red, blue, and green) and antiparticles (anti-red, anti-blue, and anti-green). The number of colors is derived by the gauge symmetry of the strong interaction, $SU(3)_C$. The force carriers of the strong interaction are named **gluons**. Since no evidence of gluons being massive particles has been found [22, 23], $SU(3)_C$ is considered to be an exact symmetry of Nature. To visualize how the gluons interact with the quarks let us consider a quark with a red color charge. This quark can lose its red color charge and acquire a green color by emitting a gluon. In this example the emitted gluon would possess a color charge in the form of a red/anti-green pair, since the color charge is conserved by the strong interaction. More generally, gluons possess color charge as well, although in a form different from the one possessed by quarks and antiquarks. As a result, there are 8 possible values for the gluon color charge, and gluons can interact both with each other via strong interaction and with themselves.

The consequence of this self-interaction is that the coupling strength of the strong interaction (denoted as α_S) grows with the distance between two colored objects. At very short distances the coupling strength approaches zero, leading to a phenomenon known as *asymptotic freedom*. This behavior of α_S has two major consequences:

1. In the proton scattering at very high energies, the strong force is virtually absent between proton constituents and the quarks and gluons can be considered free within the protons.

2. It is possible to apply a perturbative approach to the study of the strong interaction (pQCD, perturbative QCD) in case of interactions at very short distances.

As the distance between two colored objects is increased, the coupling strength grows and the perturbative approach does not hold. This property is referred to as *confinement* and explains why gluons and quarks, collectively referred to as *partons*, have never been observed as isolated particles. Two colored objects cannot be separated from each other and appear only in composite, neutral-colored[8] states, named *hadrons*. Experimentally, following a hadron collision the produced partons fly away from the interaction point. Because of the running nature of the strong coupling constant, the interaction strength between the traveling partons grows as they move apart leading to the creation of quark-antiquark pairs. Furthermore, the partons thus created recombine with each other to form neutral colored objects, creating a group of hadrons moving in a collimated direction. This process is called *hadronization* and these "cones" of color neutral particles are called *jets*. The hadrons can be classified depending on the number of quarks and

---

[8]Sometimes also called *white-colored*.

anti-quarks which serve as their constituents. The most common hadrons found in jets are the *mesons*, quark-antiquark pairs, and the *baryons*, bound states of 3 quarks or 3 anti-quarks. Evidence for the existence of more complex bound states, like *tetraquarks*, bound states made of two quarks and two antiquarks, and *pentaquarks*, bound states of 4 quarks and an antiquark or 4 antiquarks and a quark, have been found by the Belle experiment [24, 25], and at the LHC by the LHCb [26–28] experiment.

The theory of the electroweak interaction was formulated by Glashow [29], Salam [30], and Weinberg [31] and is a local gauge theory based on the symmetry $SU(2)_L \otimes U(1)_Y$. As stated in the previous section, the conserved charges of this group are the hypercharge and the isospin. The invariance under the local gauge symmetry of the group $SU(2)_L \otimes U(1)_Y$ requires all the particles to be massless. This is also true for the mediator bosons of the EW interaction: γ (*photons*), $W^\pm$ and Z. Because of the structure of the EW interaction, these particles must be *vector bosons*, i.e. they must have spin equal to 1.

Photons are the mediators of the electromagnetic force and interact with all electrically charged particles. Being massless and electrically neutral they do not possess a self-interaction property like the gluons. Among the vector bosons, photons have been the first to be theorized [32], in the Quantum Electrodynamics (QED) theory [33–36].

The **W$^\pm$ bosons** mediate the weak interaction via *charged current* and can couple only to left-handed particles and right-handed antiparticles. The **Z boson** mediates the weak interaction via *neutral current*, coupling particles with antiparticles having the same chirality. Table 1.1 summarizes the mediator bosons and their masses.

Table 1.1: Experimental values for the mediator bosons masses [23].

Interaction	Particle name	Mass (GeV)
Electroweak	*photons* (γ)	0
	$W^\pm$	80.385 ± 0.015
	Z	91.1876 ± 0.0012
Strong	*gluons* (g)	0

1.3 Electroweak symmetry breaking

Experimentally, both the fermions and the W and Z bosons have mass, as can be seen in Table 1.2 for the fermions and Table 1.1 for the mediator bosons. This means that the EW symmetry must be broken. At low energies it has to fall back to the charge symmetry of electromagnetism.

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{em} \ . \tag{1.5}$$

Table 1.2: Experimental values for fermionic masses [23].

Elementary particle	mass
e	0.511 MeV
μ	105.6 MeV
τ	1776 MeV
ν_l	~ 0
u	2.3 MeV
d	4.8 MeV
s	95 MeV
c	1.275 GeV
b	4.66 GeV
t	173 GeV

To account for this, Higgs [13], Englert, and Brout [12] proposed the *spontaneous symmetry breaking*, also known as BEH model or Higgs mechanism. This model introduces a new complex scalar[9] field represented by an isospin doublet:

$$\varphi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix} , \tag{1.6}$$

with an associated potential

$$V(\varphi) = \mu^2 \varphi^\dagger \varphi + \lambda(\varphi^\dagger \varphi)^2 . \tag{1.7}$$

By requiring $\lambda > 0$ and $\mu^2 < 0$ the potential presents a minimum which is degenerate, as shown in Fig. 1.2.

The vacuum ground-state of the field has infinite degenerate solutions corresponding to:

$$\varphi\varphi^\dagger = \frac{-\mu^2}{2\lambda} . \tag{1.8}$$

By operating a $SU(2)_L \otimes U(1)_Y$ gauge transformation of the scalar field φ a particular minimum can be selected. By choosing $\varphi_1 = \varphi_2 = \varphi_4 = 0$ the vacuum state of the Higgs field can be written as:

$$< \varphi >=< \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \varphi_3 \end{pmatrix} >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} , \tag{1.9}$$

---

[9]With spin equal to 0 and positive parity.
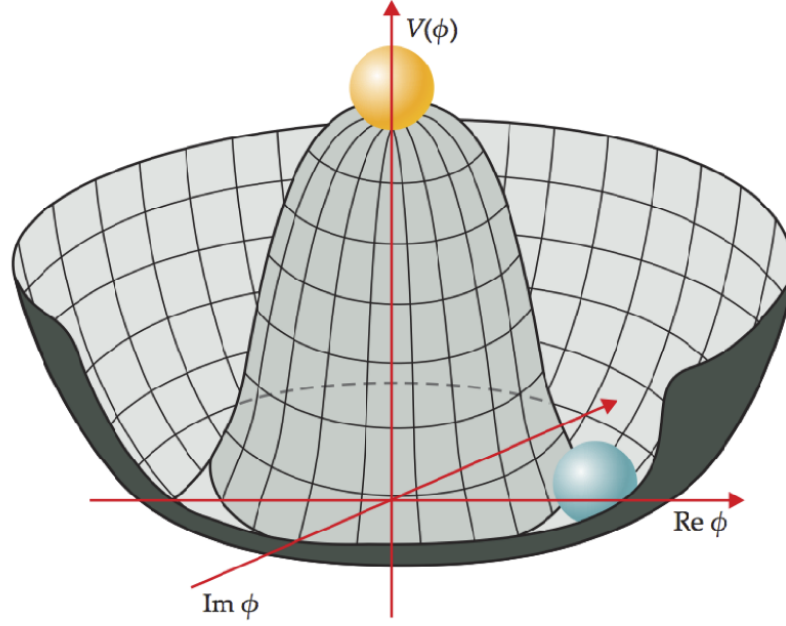
Figure 1.2: Higgs potential [37]: The potential presents a local maximum for $\varphi \equiv 0$ resulting in a more symmetrical states corresponding to a massless particle; the lowest energy state is degenerate and is defined by $\varphi\varphi^\dagger = \frac{-\mu^2}{2\lambda}$.

with $v = -\mu^2/\lambda$, vacuum expectation value of the field. This choice spontaneously breaks EWK symmetry. The most general $SU(2)_L \otimes U(1)_Y$ gauge invariant, Lorentz invariant and renormalizable Lagrangian that can be written for the scalar field is:

$$L_\varphi = |\underbrace{(\partial_\mu - ig\vec{A}_\mu \cdot \vec{T}_L^{(\varphi)} - ig'B_\mu Y^{(\varphi)})\varphi}_{\text{Gauge covariant derivative of } \varphi: \ (D_\mu\varphi)}|^2 - \underbrace{(\mu^2\varphi^\dagger\varphi + \lambda(\varphi^\dagger\varphi)^2)}_{\text{Higgs potential, } V(\varphi)} \ , \tag{1.10}$$

where $\vec{A}_\mu$ and B_μ are the fields assoctiated withthe $SU(2)_L$ and $U(1)_Y$ groups respectively and $\vec{T}_L^{(\varphi)}$ and $Y^{(\varphi)}$ are the generators of $SU(2)_L \otimes U(1)_Y$:

$$\vec{T}_L^{(\varphi)} = \frac{1}{2}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right] = \frac{1}{2}\vec{T} \ , \tag{1.11}$$

$$Y^{(\varphi)} = -\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = -\frac{1}{2}I \ . \tag{1.12}$$

At the ground-state the gauge covariant derivative term can be written as:

$$L_{mWZ} = \left|(g\vec{A}_\mu \cdot \vec{T}_L^{(\varphi)} + g'B_\mu Y^{(\varphi)}) < \varphi >\right|^2$$

$$= \frac{1}{2}\left|(\frac{g}{2}\vec{A}_\mu \cdot \vec{T} - \frac{g'}{2}B_\mu \cdot I)\cdot \begin{pmatrix} 0 \\ v \end{pmatrix}\right|^2$$

$$= \frac{1}{2}v^2\left|\begin{pmatrix} \frac{1}{2}g(A^1_\mu - iA^2_\mu) \\ -\frac{1}{2}(gA^3_m u + g'B_\mu) \end{pmatrix}\right|^2 \tag{1.13}$$

$$= \left(\frac{v^2 g^2}{4}W^\dagger_\mu W^\mu + \frac{v^2(g^2 + g'^2)}{8}Z_\mu Z^\mu\right) \ ,$$

where in the last equation a linear combination of the $\vec{A}_\mu$ and B_μ fields has been defined:

$$W^{\pm\mu} = \frac{1}{\sqrt{2}}(A^\mu_1 \pm iA^\mu_2) \ , \tag{1.14}$$

$$Z^\mu = -\frac{gA^\mu_3 + g'B^\mu}{\sqrt{g^2 + g'^2}} \ . \tag{1.15}$$

This linear combination of fields correspond to the W and Z boson fields which gain mass through the Higgs mechanism. The masses of the two bosons can be written as a function of the Higgs vev:

$$m_W = \frac{v|g|}{2} \ , \tag{1.16}$$

$$m_Z = \frac{v\sqrt{g^2 + g'^2}}{2} \ . \tag{1.17}$$

It should be noted that another linear combination of the $\vec{A}_\mu$ and B_μ fields can be defined as orthogonal to W^μ, $W^{\mu*}$, and Z^μ:

$$A^\mu = \frac{g'A^\mu_3 - gB^\mu}{\sqrt{g^2 + g'^2}} \ . \tag{1.18}$$

This linear combination of the fields is not associated with a mass term in the Lagrangian and represent the photon field. Therefore the Higgs mechanism gives mass to the W and Z bosons while the photon remains massless.

1.4 The Higgs boson

In the previous section, the ground-state of the Higgs field has been shown to give mass to the W and Z bosons. It is possible to define an excitation of the field by adding a scalar term H to the ground-state field:

$$\varphi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} . \tag{1.19}$$

The gauge covariant derivative of φ can be rewritten as:

$$
\begin{aligned}
(D^\mu \varphi)^\dagger (D_\mu \varphi) =& \frac{1}{2} \partial^\mu H \partial_\mu H + m_W^2 W_\mu^\dagger W^\mu + \frac{m_Z^2}{2} Z_\mu Z^\mu \\
&+ \frac{2m_W^2}{v} W_\mu^\dagger W^\mu H + \frac{m_Z^2}{v} Z_\mu Z^\mu H \\
&+ \frac{m_W^2}{v^2} W_\mu^\dagger W^\mu H H + \frac{m_Z^2}{2v^2} Z_\mu Z^\mu H H .
\end{aligned}
\tag{1.20}
$$

The term $\partial^\mu H \partial_\mu H$ is the kinetic energy of the Higgs boson particle, while the terms which are linear in H are associated to the couplings between the Higgs boson and the vector boson (HVV couplings). The presence of these terms has a consequence for the CP nature of the Higgs boson: they predict the possibility of a Higgs directly coupling to a pair of Z or W bosons. The $H \to ZZ^*$ and $H \to WW^*$ decays would be suppressed if the Higgs was CP-odd, since a particle with spin-parity $J^P = 0^-$ cannot decay into two particles with $J^P = 1^-$. Because of this, as shown in greater detail in Chapter 2, the search for CP-odd couplings in HVV interactions is performed looking at higher order terms. Quartic couplings of the form WWHH and ZZHH are also predicted, and suppressed by a factor 2 with respect to the HVV couplings.

A mass term for the Higgs boson can be retrieved by looking at its potential ($V(\varphi)$) which can be rewritten as:

$$
\begin{aligned}
V(H) =& \frac{\mu^2}{2}(v+H)^2 + \frac{\lambda}{4}(v+H)^4 \\
=& -\frac{\lambda v^4}{4} + \lambda v^2 HH + \lambda v HHH + \frac{\lambda}{4} HHHH \\
=& -\frac{m_H^2 v^2}{8} + \underbrace{\frac{m_H^2}{2} HH}_{\text{mass term}} + \underbrace{\frac{m_H^2}{2v} HHH}_{\text{trilinear coupling}} + \underbrace{\frac{m_H^2}{8v^2} HHHH}_{\text{quartic self-coupling}} ,
\end{aligned}
\tag{1.21}
$$

where $m_H = \sqrt{2\lambda} v = \sqrt{2|\mu^2|}$ is the Higgs mass. It should be noted that the trilinear and quadrilinear terms in H predict that the Higgs boson can couple with itself.

Finally, the Higgs field can generate the mass terms for the fermions via Yukawa interaction, which for a generic fermion f can be written as:

$$L_Y^f = - \frac{g_f}{\sqrt{2}} \bar{\psi}_L (v + H) \psi_R - \frac{g_f}{\sqrt{2}} \bar{\psi}_R (v + H) \psi_L$$

$$= - \underbrace{m_f (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)}_{\text{fermionic mass term}} - \underbrace{\frac{m_v}{v} (\bar{\psi}_L H \psi_R + \bar{\psi}_R H \psi_L)}_{\text{coupling of the Higgs to fermions}} \; , \tag{1.22}$$

where ψ_L and ψ_R are the left and right-handed components of the fermion field and $m_f = \frac{g_f v}{\sqrt{2}}$ is the fermionic mass. An interesting feature can be observed by looking at the terms which are linear in H in the above contributions to the Higgs Lagrangian:

$$L_H^{Iorder} = \frac{2m_W^2}{v} W_\mu^\dagger W^\mu H + \frac{m_Z^2}{v} Z_\mu Z^\mu H - \frac{m_f}{v} (\bar{\psi}_L H \psi_R + \bar{\psi}_R H \psi_L) \; . \tag{1.23}$$

The Higgs couplings to fermions scale as the fermionic mass, while the couplings to bosons are proportional to the square of the bosonic mass.

1.4.1 Higgs production at the LHC

During the Run 2 data-taking period at the LHC, protons collided with a center of mass energy of $\sqrt{s} = 13$ TeV. As referenced in Section 1.2, the protons are baryons, bound states of quarks. The main production mechanisms of the Higgs at the LHC require therefore quarks and gluons in the initial state. Such methods are:

- gluon-gluon fusion (ggH, Fig.1.3.a): the Higgs boson is produced via a quark loop initiated by two gluons from the initial state. In absence of partons coming from the initial state radiation, the Higgs will be produced with $p_T \sim 0$ and decay in the central region of the detector.

- vector boson fusion (VBF, Fig.1.3.c): the Higgs is produced as the fusion of two vector bosons[10] as the name implies. The quarks from the initial state propagate to the final state after radiating a Z or W boson. This leads to events where two high-momentum jets can be observed in the forward or backward region of the detector.

- Higgs-strahlung or production associated with a vector boson (VH, Fig.1.3.d): here the Higgs is produced from the decay of a virtual W or Z boson. In this production mechanism, the Higgs is expected to be boosted and recoiling against high momentum leptons and jets coming from the vector bosons decay.

---

[10]When referring to the Higgs couplings it is common to use the term *vector bosons* to refer exclusively to the W and Z bosons since, as shown in Section 1.3, the Higgs boson does not couple directly to photons.

- top-associated production (ttH, Fig.1.3.b): this production mechanism is similar to the ggH, with the main difference being the presence of a top and an anti-top quarks in the final state.

- b-associated production (bbH): this production mechanism is similar to the ttH one, with the major difference being that instead of being produced in association with a top anti-top pair, the Higgs is produced in association with a bottom anti-bottom one.

(a) ggH

(c) VBF

(b) ttH

(d) VH
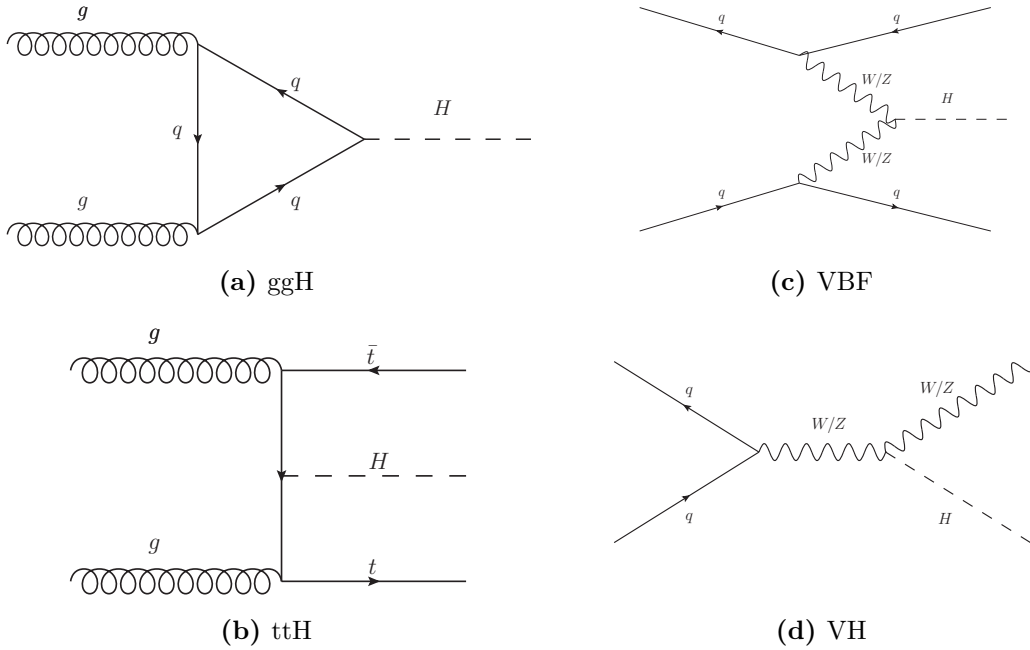
Figure 1.3: Higgs production processes: Feynman diagrams of the 4 main production mechanism of the Higgs at the LHC: gluon-gluon fusion (**top-left**), vector boson fusion (**top-right**), Higgs-strahlung (**bottom-right**) and top associated production (**bottom-left**).

Fig. 1.4 shows the Higgs production cross-sections as a function of the center of mass energy for a proton-proton collision. At an energy of $\sqrt{s} = 13$ TeV the main production mechanism is the ggH, shown in figure with the label $pp \to H$. Gluons dominate the proton parton distribution function (PDF) at low momentum fraction and energy scales of interest (100 GeV and above) [8] therefore their fusion is one of the most common hard scattering processes produced at the LHC. Being the force carriers of the strong interaction, they cannot couple directly with the Higgs and instead produce the Higgs through a quark loop, dominated by the top quark as it has the largest direct coupling to the Higgs boson. The VBF production mechanism, labeled as $pp \to qqH$ in Fig. 1.4, has the second largest cross-section at the LHC: the Higgs boson is produced as the fusion

of two vector boson radiated by quarks from the initial state. The quarks propagate to the final state, producing two high-momentum jets which can be used to identify this production mechanism. Other processes where the Higgs boson is produced via coupling with vector bosons are the W- and Z-strahlung production modes, labeled as $pp \to WH$ and $pp \to ZH$ in Fig. 1.4. The heavy-quark associated productions, namely with top and bottom quarks, have competitive cross-sections, but of the two only the top-associated production has been observed [38].
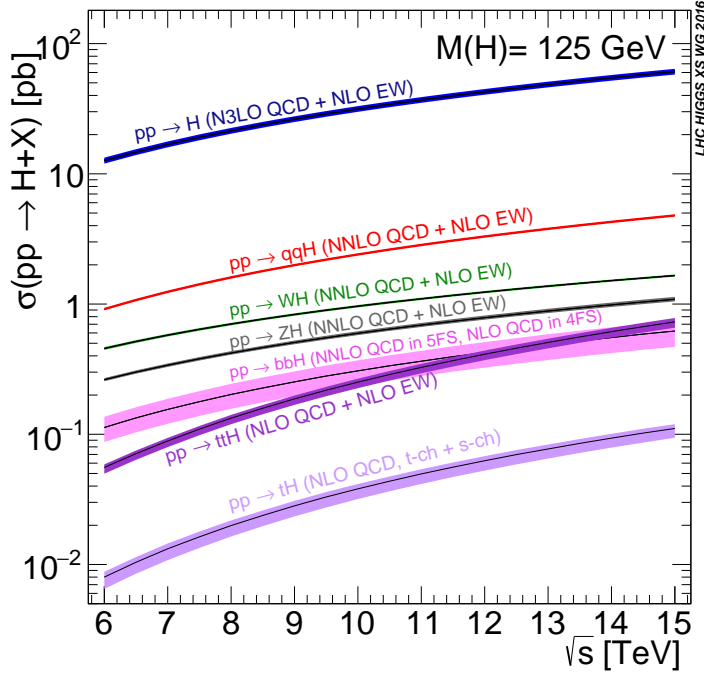
Figure 1.4: Cross-sections for Higgs production at the LHC [39]: The production cross-sections for the Higgs processes are shown as a function of the center of mass energy of the proton-proton collision, for a Higgs boson of mass 125 GeV.

1.4.2   Higgs decays

In 2012, a particle with properties compatible with that of the Higgs boson was observed by both the CMS [2] and ATLAS [1] experiments at the Large Hadron Collider (LHC). The discovery was performed by combining several decay channels: $H \to ZZ^* \to 4l$, $H \to \gamma\gamma$, $H \to WW^* \to e\mu2\nu$, $H \to \tau\tau$ and $H \to b\bar{b}$. The di-photon and 4 lepton decay channels were the main contributors to the Higgs discovery and, as shown in Fig. 1.5, have been used in later analyses [40] to measure with high precision the Higgs mass:

$$m_H = 125.38 \pm 0.14 \text{ GeV} . \tag{1.24}$$
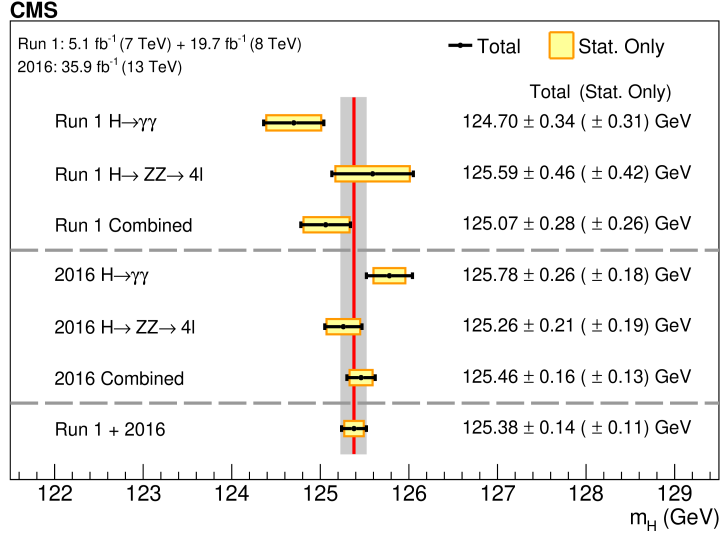
Figure 1.5: Combination of different measurements of the Higgs mass [40].

Fig. 1.6 shows the branching ratios (BR) for the Higgs decays. The ones to vector bosons and fermions are a result of the direct couplings shown in Section 1.4, while the $H \to gg$ proceeds via a quark loop like the one shown in Fig. 1.3.a and the $H \to \gamma\gamma$ decay via a fermion or W boson loop. As can be seen in the figure, the ZZ^* and $\gamma\gamma$ channels have a BR much smaller than other decay channels, but have more distinct experimental signatures which allow them to be more easily identified in a proton-proton collision.

The $H \to \tau\tau$ was the first direct decay to fermions to be discovered as an isolated process [3], using data collected during Run 1 and in 2016 at the LHC. Using Run 2 data also the decays into a pair of b-quarks have been observed [4] and, more recently, evidence for the Higgs decay into a pair of muons has been found [5]. The measurements have found properties of the Higgs decays consistent with SM predictions. In particular, Fig. 1.7 shows how the couplings of Higgs to fermion and vector-bosons align with the proportionality rule mentioned in Section 1.4 ($g_f \propto m_f$ and $g_V \propto m_V^2$).

Since this thesis work focuses on the properties under CP-symmetry of the Higgs boson coupling to fermions a more detailed discussion of the process is shown in Chapter 2.

1.5 Fermionic sector

The previous sections focused on the interactions present in the SM and the associated elementary particles. One element of the SM which still needs to be discussed are the
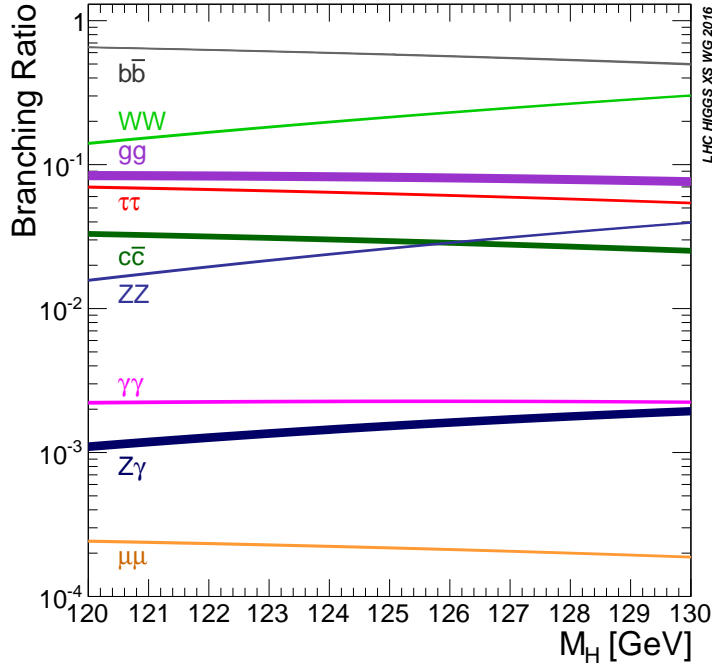
Figure 1.6: Branching ratios of the Higgs decays: The decay branching ratios for the Higgs boson are shown as a function of the Higgs mass [39].

constituents of matter, which are fermions with spin 1/2 and can be classified into two families:

- leptons, particles which interact only via electroweak interactions;

- quarks, particles which interact both via electroweak and strong interactions.

Each particle can be characterized using its mass and its quantum numbers under the gauge symmetries of the SM, those being:

- hypercharge (Y)

- isospin (T)

- color charge (C_C)

The mass of the particle depends on its coupling with the Higgs field as shown in Section 1.4 according to the formula:

$$m_f = \frac{g_f v}{\sqrt{2}} \ , \tag{1.25}$$
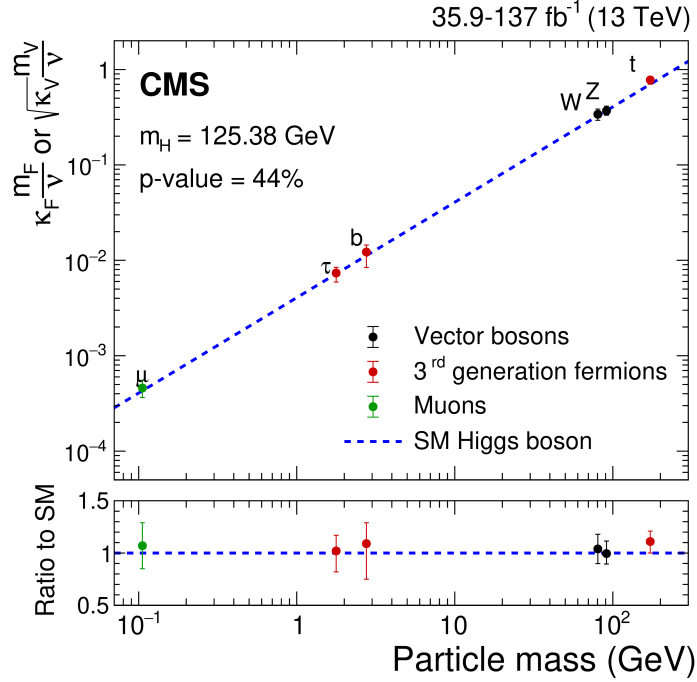
Figure 1.7: Higgs couplings with fermions and vector bosons [41]: Measured reduced coupling modifiers for fermions ($g_f = \sqrt{2}m_f/v$) and weak bosons ($g_v \propto m_V^2/v$) compared to their SM predictions. The error bars represent 68% CL intervals for the measured parameters. The lower panel shows the ratios of the measured coupling modifiers values to their SM predictions.

while the electric charge Q (in units of the electron charge e) is given by the expression of Gell-Mann-Nishijima [42]:

$$Q = T_3 + \frac{Y}{2} \, , \tag{1.26}$$

where T_3 is the eigenvalue of the third component of the isospin operator. Based on the $SU(2)_L$ symmetry, the left-handed component of fermions transforms as a doublet, while the right-handed one transforms as a singlet. Quantum numbers of fermions are therefore written separately for each chiral component. As shown in Table 1.3, the particles can be divided into groups according to their quantum numbers. Two of these groups are the left-handed charged leptons and neutrinos, which can be written as $SU(2)_L$ doublets as follows:

$$\begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L = \left\{ \underbrace{\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L}_{\text{I gen. lepton}} \; ; \; \underbrace{\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L}_{\text{II gen. lepton}} \; ; \; \underbrace{\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L}_{\text{III gen. lepton}} \right\} \, . \tag{1.27}$$

Table 1.3: Quantum numbers for leptons and quarks.

Elementary particle	Q	T	T_3	Y	C_C
l_L	-1	1/2	-1/2	-1	0
ν_{lL}	0	1/2	1/2	-1	0
l_R	-1	0	0	-2	0
u_L	2/3	1/2	1/2	1/3	r,g,b
d_L	-1/3	1/2	-1/2	1/3	r,g,b
u_R	2/3	0	0	4/3	r,g,b
d_R	-1/3	0	0	-2/3	r,g,b

In the quark sector the left-handed doublets are:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L = \left\{ \underbrace{\begin{pmatrix} u \\ d \end{pmatrix}_L}_{\text{I gen. quark}} \;\; ; \;\; \underbrace{\begin{pmatrix} c \\ s \end{pmatrix}_L}_{\text{II gen. quark}} \;\; ; \;\; \underbrace{\begin{pmatrix} t \\ b \end{pmatrix}_L}_{\text{III gen. quark}} \right\} , \tag{1.28}$$

while the right-handed singlets for leptons and quarks are: $l_R = e_R$; μ_R ; τ_R, $u_R = u_R$; c_R ; t_R and $d_R = d_R$; s_R ; b_R.

It should be noted that the right-handed neutrinos are absent in the table since a ν_{lR} would have all quantum numbers equal to 0 and be *sterile*, i.e. it would interact neither via electroweak nor strong interactions. As a consequence, the neutrinos are massless in the SM, since they do not possess a Yukawa coupling with the Higgs boson. The discovery of neutrino oscillation [43] (Nobel prize in 2015 [44]) implies, however, that neutrinos have a mass, albeit noticeably smaller than the one of other fermions. To account for this inconsistency, some extensions to the SM are required. Two of the main models used to explain massive neutrinos are: the see-saw mechanism [45], and assume that the neutrinos are Majorana particles[11] [7]. Due to the effects of neutrino mass being negligible in the calculation of the cross-section for the Higgs decays into τ leptons and their consecutive decays, in this thesis neutrinos are assumed to be massless.

Particles which share the same quantum numbers are said to belong to different *generations*. The fermions in the SM are thus divided into 3 generations, in order of increasing mass, as shown in Eq. 1.27 and 1.28. An additive quantum number of *flavor* is introduced to separate particles belonging to different generations. Starting from the leptons, we can introduce 3 additive quantum numbers named *lepton flavors*: λ_e, λ_μ and λ_τ. Particles in the same doublet share the same lepton flavor, e.g. e and ν_e have $\lambda_e = 1$ and all other lepton flavors equal to 0. Lepton flavor is an accidental symmetry of the SM, meaning that all

---

[11]A Majorana particle is defined by coinciding with its antiparticle.

renormalizable terms of the SM Lagrangian preserve the $U(1)_e \otimes U(1)_\mu \otimes U(1)_\tau$ symmetry. Although some experimental results, related to neutrino oscillation [43], beauty-quark decays [46], and the muon gyromagnetic moment [47], may point to a violation of lepton flavor universality, this is still assumed correct throughout this thesis. This is due to the possible violation being considered small for the physical processes relevant for this dissertation. A corresponding quark flavor symmetry is manifestly broken by the electroweak interaction. The mass eigenstates of down-type quarks do not coincide with the electroweak eigenstates. They are related to each other via the Cabibbo-Kobayashi-Maskawa matrix V_{CKM}:

$$\mathrm{d}_L^{\mathrm{m}i} = V_{\mathrm{CKM}}^{ij}\mathrm{d}_L^{\mathrm{w}j} \; , \tag{1.29}$$

with $\mathrm{d}_L^{\mathrm{m}}$ and $\mathrm{d}_L^{\mathrm{w}}$ eigenstates of down-type quarks with respect to the free particle and the electroweak Hamiltonians. In the next section, the τ lepton is described in greater detail and the mixing of flavor eigenstates is briefly mentioned when discussing the hadronic decays of τ leptons.

1.5.1   The tau lepton

The τ is the third generation charged lepton. The existence of a heavy charged lepton was first hypothesized after the discovery of the muon. The expected signatures of a heavy lepton which could decay leptonically and hadronically had been predicted by Yung-Su Tsai [48] in 1971. The first evidence for the existence of a heavy charged lepton was found in 1975 at the Stanford Linear Accelerator Center (SLAC) [49]. The discovery was made by finding 64 events identified as the process $e^+e^- \to e^\pm + \mu^\mp + \geq 2$ undetected particles. The need for 2 neutral or charged particles which escaped the detector without interacting came from the energy conservation. Fig. 1.8 shows the observed cross-section for the aforementioned process, from which the authors of the paper [49] concluded *"that the signature $e - \mu$, events cannot be explained either by the production and decay of any presently known particles or as coming from any of the well-understood interactions which can conventionally lead to an e and a p. in the final state. A possible explanation for these events is the production and decay of a pair of new particles, each having a mass in the range of 1.6 to 2.0 GeV/c".* Other experiments later confirmed the existence of a third charged lepton and measured its mass and spin.

Having a mass of $m_\tau = (1776.86 \pm 0.12)$ MeV [23], it is the heaviest lepton and the only one which can decay both hadronically and leptonically. Its average lifetime is $\mathscr{T}_\tau \simeq (2.903 \pm 0.005) \times 10^{-13}$ s [23], which for a τ lepton of 30 GeV corresponds to a decay length of:

$$\lambda_\tau = c \cdot \mathscr{T}_\tau \cdot \beta\gamma \simeq (9 \times 10^{-2} \text{ mm}) \cdot (30 \text{ GeV}/1.78 \text{ GeV}) \simeq 1.5 \text{ mm} \; , \tag{1.30}$$

with c speed of light and $\beta\gamma = p/m$, with momentum (p) and mass (m) written in natural units. The innermost layer of the silicon tracker in CMS is at a distance of $\sim$3 cm from
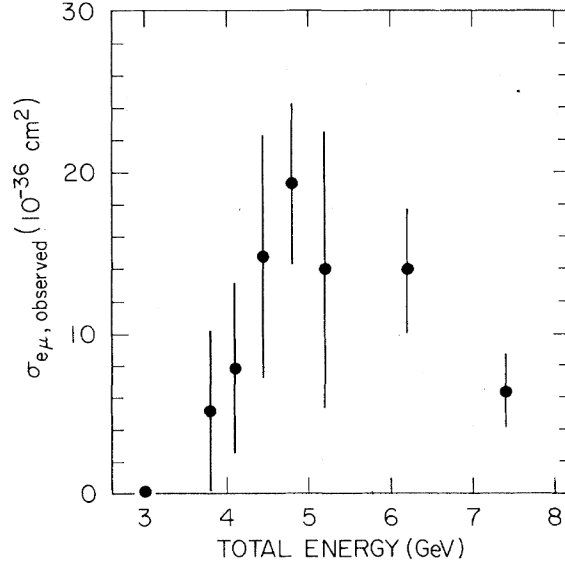
Figure 1.8: Observed cross-section for the $e^+e^- \rightarrow e^\pm + \mu^\mp + \geq 2$ undetected particles process as a function of center-of-mass energy [49].

the beamline. This means that the fraction of τ leptons produced close to the beam line (also referred to as *prompt*) that decay after reaching the innermost layer of the detector is negligible. The τ leptons are therefore reconstructed based on their decay products, so it is useful to discuss these processes in greater detail.

As previously stated, leptons do not have color charge but they possess isospin and hypercharge. They can only interact via electroweak interaction. When looking at τ lepton decays the only interaction to be considered is the weak interaction via charged current. The Feyman diagram for a general τ decay is shown in Fig. 1.9, the τ lepton decays into a ν_τ by emitting a W boson. The decay is completed by W boson coupling either with a $l - \nu_l$ (with $l = e, \mu$) or a quark-antiquark pair.

The leptonic decay assumes the form $\tau^- \rightarrow \nu_\tau \bar{\nu}_l l^-$ and its corresponding decay width can be written as [48]:

$$
\Gamma \begin{pmatrix} \tau^- \rightarrow \nu_\tau \bar{\nu}_l l^- \\ \tau^+ \rightarrow \bar{\nu}_\tau \nu_l l^+ \end{pmatrix} = \frac{G_F^2 m_\tau^5}{3 \times 2^7 \pi^4} \frac{8}{m_\tau^4} \int_0^{p_{max}} p^2 dp \int d\Omega \left[3m_\tau - 4E - \frac{m_l^2}{E} + \frac{3m_l^2}{m_\tau} \right.
$$
$$
\left. \mp \left(\vec{w} \cdot \frac{\vec{p}}{E} \right) \left(4E - m_\tau - \frac{3m_l^2}{m_\tau} \right) \right],
$$

(1.31)

with G_F Fermi constant, $\vec{p} = p\hat{p}$ lepton momentum of module p and unit vector $\hat{p}$, m_l lepton mass, $E = \sqrt{p^2 + m_l^2}$, $\vec{w}$ τ polarization vector and $p_{max} = (m_\tau^2 - m_l^2)/2m_l^2$. An interesting property that can be inferred is that the decay width for $\tau^+ \rightarrow \bar{\nu}_\tau \nu_l l^+$ can be
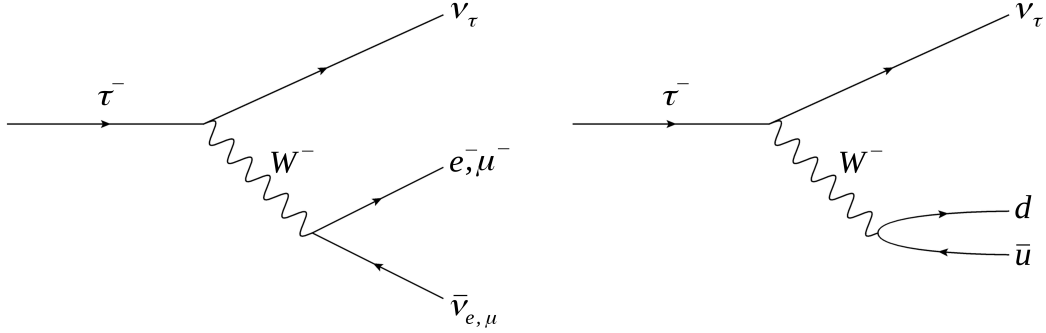
Figure 1.9: Feynman diagrams for a leptonic (**left**) and hadronic (**right**) τ^- decay.

obtained by the corresponding decay for τ^- by changing the sign of $\vec{w}$. The term $\vec{w} \cdot \vec{p}$ violates parity but preserves symmetry under CP transformation. By integrating 1.31, averaging over the τ spin and assuming $m_l = 0$, the decay width becomes

$$\Gamma = \Gamma_0 = \frac{G_F^2 m_\tau^5}{3 \times 2^6 \pi^3} \; , \tag{1.32}$$

while for decays to a generic massive charged lepton the decay width can be written as

$$\Gamma \begin{pmatrix} \tau^- \to \nu_\tau \bar{\nu}_l l^- \\ \tau^+ \to \bar{\nu}_\tau \nu_l l^+ \end{pmatrix} = \Gamma_0 (1 - 8y + 8y^3 - y^4 - 12y^2 \ln y) \; , \tag{1.33}$$

with $y = (m_l/m_\tau)^2$.

This allows to calculate the BR of τ leptonic decays:

$$\mathscr{B} = \frac{\Gamma(\tau^- \to \nu_\tau \bar{\nu}_l l^-)}{\Gamma_{tot}} = \mathscr{T}_\tau \Gamma(\tau^- \to \nu_\tau \bar{\nu}_l l^-) \; , \tag{1.34}$$

which after substituting the average lifetime ($\mathscr{T}_\tau$) with the measured value leads to:

$$\mathscr{B}(\tau^- \to \nu_\tau \bar{\nu}_e e^-) = (17.85 \pm 0.03)\% \; , \tag{1.35}$$

$$\mathscr{B}(\tau^- \to \nu_\tau \bar{\nu}_\mu \mu^-) = (17.35 \pm 0.03)\% \; . \tag{1.36}$$

Eq. 1.31 also allows to infer some properties of the τ lepton by considering the angular distribution of the charged lepton among the decay products. By substituting $m_l = 0$, the equation simplifies to:

$$\Gamma \begin{pmatrix} \tau^- \to \nu_\tau \bar{\nu}_l l^- \\ \tau^+ \to \bar{\nu}_\tau \nu_l l^+ \end{pmatrix} = \frac{\gamma_0}{2\pi} \int d\Omega \int_0^1 dx \; x^2 [3 - 2x \mp (\vec{w} \cdot \hat{p})(2x - 1)] \; , \tag{1.37}$$

with $x = E_l/E_\tau$ fraction of total energy carried by the charged lepton. If the massless lepton carries most of the τ energy ($x \simeq 1$) then the neutrinos must be emitted in the opposite direction with respect to the charged lepton momentum. Since neutrinos and antineutrinos have opposite helicities their spins must sum up to 0, resulting in the lepton being emitted in the direction opposite to the spin τ, while the antilepton would be emitted in the direction of the τ spin (Fig. 1.10).
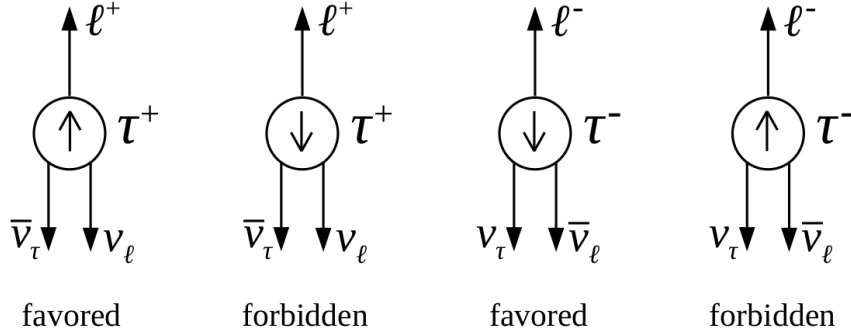
Figure 1.10: Schematic representation of a τ lepton decay into a charged lepton ($l^{\pm}$) and two neutrinos (image done by the thesis author taking inspiration from [48]). In the case where the charged lepton l^+ (l^-) carries most of the τ energy ($x = 1$ in Eq. 1.37) its direction of flight tends to be (anti-)parallel with the spin of the decaying τ^+ (τ^-).

Assuming $x << 1$ the neutrinos are now emitted back to back and their spin must sum up to 1. In this configuration, the lepton must be emitted in the direction of the τ spin, while the antilepton is emitted in the direction opposite to the τ spin. This means that the spin correlation between the τ and the charged lepton in the decay is inverted moving from $x = 0$ to $x = 1$.

As previously mentioned the τ lepton can decay hadronically in a quark-antiquark pair:

$$\tau^- \to \nu_\tau W^- \to \nu_\tau \bar{u} d \ , \tag{1.38}$$

where u and d represent up-type and down-type quarks using the notation shown in Eq. 1.28.

Given that quarks are bound together via the process of hadronization, it is expected that the τ lepton would decay into mesons. In the case of a two-body decay, the τ lepton would decay into a neutrino and a charged meson with spin 0 or 1, taking into account charge conservation and spin summing rules. The lightest charged charmed meson is the $D^{\pm}$ with a mass of (1869.62 ± 0.20) MeV, therefore a τ can decay only into mesons formed by up, down or strange quarks (and their corresponding antiquarks). In general

the matrix element for τ hadronic decays can be written as:

$$\mathscr{M} = \frac{G_F}{\sqrt{2}}|V_{CKM}|L^\mu H_\mu \; , \tag{1.39}$$

with $|V_{CKM}|$ element of the Cabibbo-Kobayashi-Maskawa matrix corresponding to the mixing of down and strange quarks, L^μ and H^μ respectively leptonic and hadronic charged currents. The leptonic current can be easily written as:

$$L^\mu = \bar{\nu}_\tau \gamma^\mu \frac{1-\gamma^5}{2}\tau \; , \tag{1.40}$$

while the hadronic current depends on the final state created. In general, among the τ decay products there is at least one charged meson with spin 0 or 1, and the remaining particles form a neutral system. Eq. 1.37 can be therefore generalized to include also hadronic tau decays:

$$\Gamma \begin{pmatrix} \tau^- \to a^-(q) + X \\ \tau^+ \to a^+(q) + \bar{X} \end{pmatrix} = \int_0^1 dx \int \frac{d\Omega}{4\pi} A(x) \pm B(x)(\vec{w}\cdot\hat{q}) \; , \tag{1.41}$$

with $a^\pm$ representing a charged meson with momenta $\vec{q} = q\hat{q}$ and $x = \frac{2m_\tau E_a}{m_\tau^2 + m_a^2 - m_X^2}$ fraction of the total energy carried by the charged meson. The $A(x)$ and $B(x)$ functions depend only on the fraction of energy carried by the charged meson: $A(x)$ determines the decay rate of $\tau \to a + X$, while $B(x)$ is related to the spin-analyzing power of the charged particle. These functions are called *spectral functions* and their theoretical predictions can be found in [48].

Let us consider a case where the τ decays only in a meson and a neutrino:

$$\tau^- \to h_{0,1}^- \nu_\tau \; , \tag{1.42}$$

with the subscript 0,1 representing the spin of the charged meson. If the meson has spin 0, it is emitted preferably in the direction of the τ^- spin. Correspondingly for a τ^+ decay the h_0^+ meson would be emitted preferably in the direction opposite to the τ^+ spin. This is true based only on the consideration that the neutrinos have negative helicity and the antineutrinos have positive helicity, therefore it should not depend on the fraction of energy carried by the charged meson. Therefore, the function $B(x)$ must be constant for these decays. Furthermore, since it is a two-body decay the energy is divided between the decay products based only on their masses. Assuming the neutrinos to be massless we have:

$$E_h = \frac{m_\tau^2 + m_h^2}{2m_\tau} \; . \tag{1.43}$$

Indeed for the $\tau^- \to \pi^- \nu_\tau$ and $\tau^- \to K^- \nu_\tau$ decays the spectral functions assume the values:

$$A(E_{\pi,K}) = f_{\pi,K}\, \delta\left(E_{\pi,K} - \frac{m_\tau^2 + m_{\pi,K}^2}{2m_\tau}\right) , \tag{1.44}$$

$$B(E_{\pi,K}) = 1 , \tag{1.45}$$

where $f_{\pi,K}$ is a factor constant with respect to the energy.

A similar consideration can be done for the behavior of $A(x)$, in the case of τ decays to neutrino and spin 1 meson. In these decays, $B(x)$ becomes more complicated as it depends on the projection of the meson spin along its momenta. There are two configurations for the decay of a spin 1/2 particle into a spin 1/2 and a spin 1 particles: $s_z(h_1, \nu_\tau) = (+1, -1/2)$ or $s_z(h_1, \nu_\tau) = (0, +1/2)$, with s_z representing the projection of the spin onto the flight direction of the decay products. As shown in Fig. 1.11 for the a_1 meson, if the meson is transversely polarized it is emitted preferably in the direction of the τ^- spin,
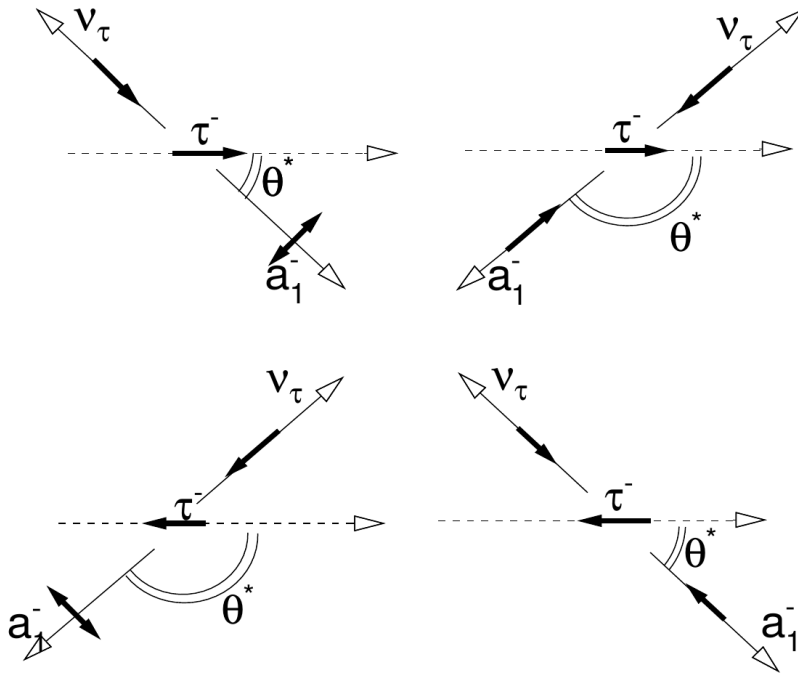
Figure 1.11: Schematic representation of a τ lepton decay to a spin 1 meson (the a_1 meson in this particular example) [50]. The dark colored arrows represent the lepton polarizations, the thin arrows represent the momenta of the decay products and the arrow with a dashed line is the τ momentum. The dark colored arrow is placed aligned with (orthogonal to) a thin arrow to indicate a longitudinal (transverse) polarization.

while if it is longitudinally polarized it is emitted preferably in the opposite direction. These considerations can be still applied to multi-body decays provided that there are intermediate resonances which allow to represent the decay as a cascade of two-body decays.

In order to determine the correlation between the τ spin and the momenta of its decay products precise measurement of the energy dependence of $B(x)$ is crucial. As previously mentioned for leptonic decays, this spectral function can change sign based on the energy of the τ decay products resulting in an inversion of the spin correlation. The same is verified for mesonic resonances [48], which are generally broad [23] allowing for the spin correlation to vary depending on how the τ energy is divided among its decay products. The modeling of the τ spectral functions was performed by several experiments measuring τ hadronic decays, e.g. the CLEO [51], BaBar [52] and Belle [53] collaborations.

Table 1.4 summarizes the main tau decay channels and their corresponding measured branching ratios. Tau leptons decay hadronically almost two thirds of the times. This is due to the quarks having 3 possible values for the color charge, and therefore being preferred in the decay.

The main mesonic resonances found in the τ decays are: pion ($\pi^\pm$), rho ($\rho^\pm$) and $a_1^\pm$. Their properties are listed in Table 1.5 along with the ones of the neutral mesons with similar properties. For the charged pion no decay channel is shown as in a collider experiment such as CMS the charged pion is unlikely to decay before reaching the hadronic calorimeter and be detected as a charged hadron. All values listed in Tables 1.4 and 1.5 are taken from [23].

Table 1.4: Main tau decay channels.

Decay channel	Main resonance	BR (%)
Leptonic decays		35.2
$\tau^- \to \nu_\tau e^- \bar{\nu}_e$		17.8
$\tau^- \to \nu_\tau \mu^- \bar{\nu}_\mu$		17.4
Hadronic decays		64.8
$\tau^- \to \nu_\tau \pi^-$	$\pi(140)$	10.8
$\tau^- \to \nu_\tau \pi^- \pi^0$	$\rho(770)$	25.5
$\tau^- \to \nu_\tau \pi^- \pi^0 \pi^0$	$a_1(1260)$	9.3
$\tau^- \to \nu_\tau \pi^- \pi^0 \pi^0 \pi^0$		1.0
$\tau^- \to \nu_\tau \pi^- \pi^- \pi^+$	$a_1(1260)$	9.3
$\tau^- \to \nu_\tau \pi^- \pi^- \pi^+ \pi^0$		4.6
$\tau^- \to other$		3.9

Sub-dominant decay channels, like the ones involving kaons ($K^\pm$, K^0 and $\bar{K}^0$ mesons) or the omega meson ($\omega(782)$), are grouped together in the $\tau \to other$ decay in Table 1.4. Charged kaons have identical spin-parity and similar average lifetime to the π mesons, becoming indistinguishable from the latter in collider experiments such as CMS. As shown in Eq. 1.45, the spin-analyzing power in $\tau^- \to \nu_\tau K^-$ decays is analogous to the one found for $\tau^- \to \nu_\tau \pi^-$, resulting in an identical treatment for the two decays. Important differences arise instead for τ decays into systems comprising both charged (neutral) kaons and neutral (charged) pions. These decays do not involve intermediate mesonic resonances, meaning they cannot be described as a cascade of two-body decays resulting in a diluted spin correlation.

The $\omega(782)$ meson has properties similar to the ρ^0 meson, but is characterised by having weak isospin equal to 0. It can be produced in tau decays via the process $\tau^- \to \nu_\tau \pi^- \omega$, which has a branching fraction of $\sim 2\%$, and mostly decays into a $\pi^+ \pi^- \pi^0$ system before interacting with the CMS detector. It is therefore reconstructed as a $\tau^- \to \nu_\tau \pi^- \pi^- \pi^+ \pi^0$ decay.

Table 1.5: Main mesons appearing in τ decays.

Meson	Mass (MeV)	J^P	Main decay channels	BR (%)
$\pi^\pm$	139.57039 ± 0.00018	0^-	$\sim$ stable[12]	
π^0	134.9768 ± 0.0005	0^-	$\pi^0 \to \gamma\gamma$	98.8
			$\pi^0 \to \gamma e^+ e^-$	1.2
$\rho^\pm$	775.11 ± 0.34	1^-	$\rho^\pm \to \pi^\pm \pi^0$	~ 100
ρ^0	775.26 ± 0.25	1^-	$\rho^0 \to \pi^\pm \pi^\mp$	~ 100
$a_1^\pm$	1230 ± 40	1^+	$a_1^\pm \to \rho^{\pm,0} \pi^{0,\pm}$	~ 100

---

[12]This particle is considered stable in the context of a proton proton collision as it usually reaches the calorimeters before decaying.

CP properties in the Higgs sector

Contents

The properties under CP symmetry of the Higgs boson are an important test for the SM. As shown in Section 1.4, the SM predicts the existence of one Higgs boson, and its coupling to fermions and vector bosons are expected to be invariant under CP symmetry. The presence of CP violation in the Higgs sector, if found, would provide a strong indication for physical phenomena not predicted by the SM and be used to constrain theories beyond the SM.

The main analysis described in this thesis work, presented in Chapter 6, focuses on the measurement of the Yukawa coupling of the Higgs boson to tau leptons. It assumes the presence of both CP-even and CP-odd components, in order to estimate the amount of CP violation in the interaction. This is done by taking advantage of the spin correlation between the τ leptons originating from the Higgs boson decay, and between the former and their decay products. Section 2.2.2 is dedicated to describe the theory and experimental techniques to study the spin correlations and to measure the CP properties of the Yukawa coupling between the Higgs boson and the tau leptons.

Before approaching the theoretical framework of that measurement it is useful to provide a brief description of other measurements targeting the CP properties of the Higgs sector of particle physics. While these complementary measurements are not the focus of this thesis, they can still provide a useful comparison with the measurement performed in order to highlight some of its features.

The first interaction to be discussed is the linear coupling at leading order between the Higgs boson and two vector bosons (see Eq. 1.20). Taking into account that the Higgs is a spin 0 particle, while the vector bosons have spin-parity $J_V^P = 1^-$, this implies

that the Higgs boson must have spin-parity $J_H^P = 0^+$. This can easily be proven for $H \to ZZ^*, WW^*, \gamma\gamma^*$ decays. For example, the final state for $H \to ZZ^*$ decays comprises two spin 1 particles which, taking into account the rules of spin summation and that the decaying Higgs boson has spin 0, can have an orbital angular momentum equal to 0 or 2. The wave function for the di-boson system has to be symmetric under parity given that they are identical in nature, of integer spin, and that the orbital angular momenta of the system is even. This implies that, unless the decay itself violates parity conservation, the parity of the initial state must also even. The tree-level HVV coupling of $H \to ZZ^*$ is:

$$\mathscr{L}_{HZZ} = \frac{m_Z^2}{v} Z_\mu Z^\mu H \ , \tag{2.1}$$

where m_Z is the mass of the Z boson and v the Higgs vev which are both real parameters. Such an interaction term is invariant under parity operator. This leads to the conclusion that in the SM, the spin-parity of the Higgs boson must be $J_H^P = 0^+$. The presented treatment takes the SM as a base and elaborates on the spin-parity of the Higgs boson, a more general treatment for the spin-parity properties of the HVV coupling can be found in [54–57].

During Run 1 at the LHC the CMS and ATLAS experiments have set limits on the properties of the Higgs boson, and in particular have found its spin-parity to be consistent with $J^{PC} = 0^{++}$ [58,59]. All the constraints placed on the spin-parity of the Higgs boson during Run 1 have been derived in the context of anomalous HVV couplings. During Run 2 the CP properties of the Higgs boson have been investigated in the following scenarios:

- anomalous HVV couplings;

- effective ggH coupling;

- Yukawa coupling in $H \to \tau\tau$ decays and ttH production.

From the experimental point of view the HVV and ggH couplings can be treated similarly, on account of both being couplings between a spin 0 and two spin 1 particles. The main differences between the two cases descends from the theory interpretation of the constraint on the spin-parity. Section 2.1 will provide more insight on the theory framework for the HVV anomalous couplings, while Section 2.2 will focus on the CP properties of the Higgs in its Yukawa couplings.

The ggH effective coupling is discussed in the latter section, as the ggH coupling is mediated by a quark loop. Anomalous ggH couplings are therefore interpreted as a modification of the Higgs boson Yukawa coupling to quarks and, assuming that the quark loop is dominated by bottom and top quarks, combined with the studies performed in the ttH production mechanism. Finally, Section 2.2.2 describes the measurement of the CP properties of the Higgs boson in its decays to tau leptons.

2.1 Anomalous CP-odd couplings in bosonic interactions

The scattering amplitude describing the interaction between a spin-zero H boson and two spin-one gauge bosons VV can be written as [60]:

$$\mathscr{A}(HVV) \simeq \left[a_1^{VV} + \frac{k_1^{VV} q_1^2 + k_2^{VV} q_2^2}{(\Lambda_1^{VV})^2} + \frac{k_3^{VV}(q_1 + q_2)^2}{(\Lambda_Q^{VV})^2} \right] m_V^2 \varepsilon_{V1}^* \varepsilon_{V2}^*$$

$$+ a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \bar{f}^{*(2)\mu\nu} \,, \tag{2.2}$$

where $f^{(i)\mu\nu} = \varepsilon_{Vi}^\mu q_{Vi}^\nu - \varepsilon_{Vi}^\nu q_{Vi}^\mu$ is the field strength tensor of a gauge boson with momentum q_{Vi} and polarization ε_{Vi}, $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma} f^{(i)\rho\sigma}$ is the dual field strength tensor with m_{Vi} pole mass of a gauge boson, Λ_1^{VV} and Λ_Q^{VV} are the scales of BSM physics.

These couplings have been investigated by the ATLAS and CMS experiments already during Run 1 [58, 61] in search for a CP-violating coupling. As an example, the search for CP violation in $H \to ZZ \to 4$ leptons[1] allowed to strongly constrain the relative contribution of the a_3^{ZZ} coupling. This is shown in Fig. 2.1, in the form of the agreement of a test statistic based on the pure CP-even and CP-odd hypotheses with the observed data. The test statistic is defined based on the treatment described in [62], as the negative logarithmic ratio of the likelihood functions for the two pure CP hypotheses. These likelihood functions have been constructed respectively by setting all couplings but a_1^{ZZ} equal to 0 for the CP-even hypothesis, while for the CP-odd hypothesis is obtained by setting all couplings besides a_3^{ZZ} to 0. The observed value of the test statistic (red arrow) is compared to distributions of *pseudo experiments* obtained for the CP-even and CP-odd hypotheses with toy Monte Carlo simulations. Data shows a clear preference for the CP-even hypothesis while a pure CP-odd state is ruled out at 99% confidence level (CL). In order to interpret such result as a limit in terms of the CP-odd coupling, the relative fraction of the a_3^{ZZ} coupling contribution to the overall cross section is used. This *effective fractional cross section* is defined as:

$$f_{a_3}^{ZZ} = \frac{\left| a_3^{ZZ} \right|^2 \sigma_3}{|a_1^{ZZ}|^2 \sigma_1 + |a_2^{ZZ}|^2 \sigma_2 + |a_3^{ZZ}|^2 \sigma_3 + \tilde{\sigma}_{\Lambda_1}/(\Lambda_1)^4 + \dots} \,, \tag{2.3}$$

with σ_i representing the cross section of the $H \to ZZ$ process computed assuming $a_j^{ZZ} = 0 \ \forall j \neq i$, $\tilde{\sigma}_{\Lambda_1}$ is the effective cross section corresponding to $\Lambda_1 = 1$ TeV, and the denominator includes contribution of all anomalous couplings. By writing the likelihood function as depending on the f_{a_3} effective fraction this analysis allowed to place a limit on the a_3^{ZZ} coupling of

$$\left| \frac{a_3^{ZZ}}{a_1^{ZZ}} \right| < 2.3 \,, \tag{2.4}$$

---

[1]Leptons is used here to indicate charged light leptons, namely muons and electrons.
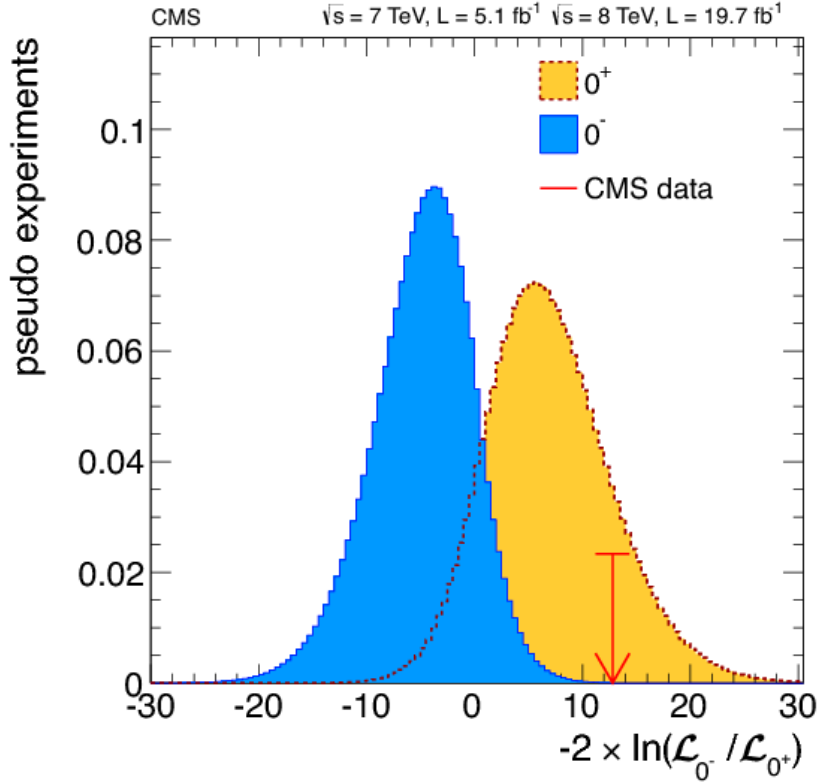
Figure 2.1: Distribution of the test statistic $q = -2\ln\left(\mathcal{L}^{0-}/\mathcal{L}^{0+}\right)$ of the pseudoscalar boson hypothesis (blue) tested against the SM Higgs boson hypothesis (yellow). The arrow indicates the observed value obtained with Run 1 data from the CMS experiment [58].

and $f_{a_3}^{ZZ} < 0.51$ at 95% CL.

Subsequent analyses followed a more general approach, extending the measurement to other anomalous couplings and accounting for the interference between different couplings at the level of cross section calculation. Such mixing is represented as the argument function for the ratio between higher order couplings and the SM one:

$$\phi_{a_i} = \arg\left(\frac{a_i}{a_1}\right) \ . \tag{2.5}$$

The latest CMS studies investigated the HVV couplings in the 4 lepton final state [63] and in the vector boson fusion production of a Higgs decaying into a pair of tau leptons [64]. The main features common to these analyses, which distinguish them from the study of the CP properties in the Yukawa coupling, is the use of an effective field theory approach: the anomalous couplings are introduced in the Lagrangian as additional terms, without focusing on the physical processes which could produce these couplings. Limits on the anomalous couplings are then estimated based on their effect on the overall

cross section, represented by the effective fractional cross sections, and the kinematics of the Higgs boson and its accompanying jets. The combination of these analyses allowed to place strong constraints on the effective fractions and phases associated to anomalous HVV couplings. Fig. 2.2 shows the limits placed on the CP-odd coupling written in the form the form $f_{a_3} \cos(\phi_{a_3})$, with a value of 0 representing the SM prediction, i.e. with all anomalous couplings equal to 0. The values +1 and -1 represent instead a model where the SM coupling is negligible with respect to the CP-odd one. As shown in the figure, a pure CP-odd coupling is excluded with a significance higher than 5 standard deviations, i.e. at a level better than 99.9999% CL. More details on the methods used in the aforementioned analyses and on the limits calculated on other anomalous couplings can be found in Appendix A.
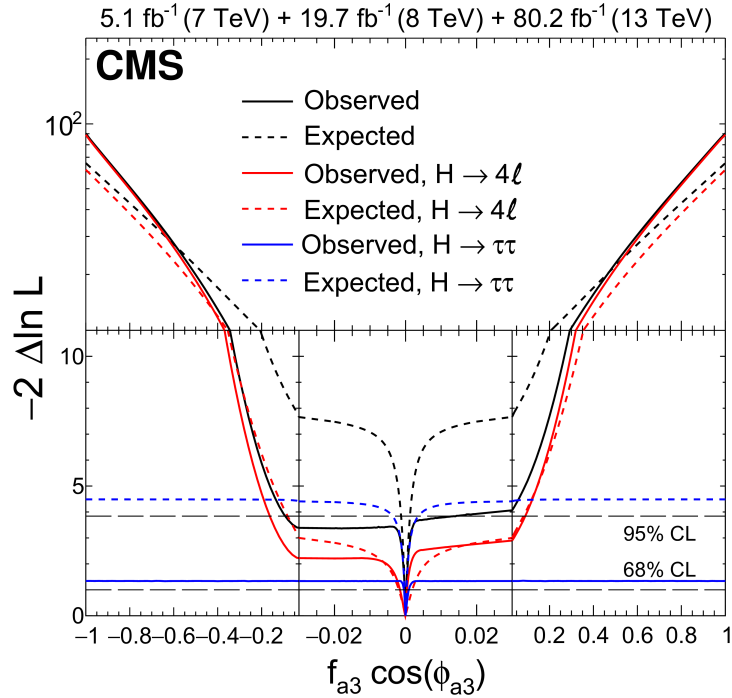
Figure 2.2: Combination of limits on the anomalous CP-odd HVV coupling obtained by studying the $H \to \tau\tau$ process [64] and $H \to 4$ leptons one [63]. The observed (solid) and expected (dashed) likelihood scans of $f_{a_3} \cos(\phi_{a_3})$ is shown. The scales of the x and y axes are varied: the x axis is always in linear scale, but a zoom is applied in the range between -0.03 and +0.03, while y is shown linear scale in the lower part of the plot and in logarithmic scale for higher values of $-2\Delta \ln L$ [64].

2.2   CP structure of the Yukawa interaction

The Yukawa Lagrangian can be modified at leading order to account for a CP-odd component:

$$\mathscr{L}_{Y,f} = -\bar{\psi}_f(y_f + i\gamma^5\tilde{y}_f)H\psi_f \ , \tag{2.6}$$

where ψ_f is the fermionic field, H is the excitation of the Higgs field with respect to the vacuum state, γ^5 is the fifth Dirac matrix, and y_f ($\tilde{y}_f$) is the CP-even (CP-odd) coupling of the Higgs boson to the fermion of interest. Compared to the expression in Eq. 1.22, in this notation the projections of the fermionic field under chiral operators are not written explicitly, and the adjoint conjugate of the Lagrangian is also dropped to simplify the notation. As shown in Eq. 1.22, the Yukawa coupling between the Higgs boson and the fermions has module m_f/v, with m_f being the fermion mass and v the Higgs vacuum expectation value (vev). It is possible to factorize this term in Eq. 2.6 in order to define the reduced Yukawa couplings:

$$\kappa_f = \frac{v}{m_f} \cdot y_f \ , \tag{2.7}$$

$$\tilde{\kappa}_f = \frac{v}{m_f} \cdot \tilde{y}_f \ , \tag{2.8}$$

and rewrite Eq. 2.6 as:

$$\mathscr{L}_{Y,f} = -\frac{m_f}{v}\bar{\psi}_f(\kappa_f + i\gamma^5\tilde{\kappa}_f)H\psi_f \ . \tag{2.9}$$

In this notation the SM prediction corresponds to $\kappa_f = 1$ and $\tilde{\kappa}_f = 0$, while a pure CP-odd coupling would be represented by $\kappa_f = 0$ and $\tilde{\kappa}_f = 1$. The Lagrangian depends linearly on the fermionic mass, as such the best candidates to study the properties of the Yukawa interaction under CP symmetry are the processes in which the Higgs boson couples with the heaviest quarks and leptons. When considering the Higgs production mechanisms at the LHC (see Section 1.4.1), the ones that could be used to study the CP properties of the Yukawa interaction are the ggH and ttH. Section 2.2.1 provides a brief overview of the studies performed on these processes. The Yukawa coupling to tau leptons can be instead investigated in $H \to \tau\tau$ decays, as this coupling does not contribute to the main Higgs production mechanisms at the LHC, at least when considering these processes at leading order. Other fermionic decays of the Higgs boson, i.e. to bottom quarks and muons, offer less sensitivity to the CP properties of the respective Yukawa couplings. The bottom quarks produced in the decay initiate a cascade of colorless particles as a result of the hadronization process, offering low sensitivity to the initial bottom quark spin. Muons instead travel trough the particle detectors without decaying, and there is currently no established technique to measure their spin in the experiments at the LHC.

2.2.1 CP violation in ggH and ttH production mechanisms

Gluons are the force carriers of the strong interaction, they can interact directly only with particles characterized by having a color charge, and as such cannot couple directly to the Higgs boson. However, as mentioned in Section 1.4.1, the main production mechanism of the Higgs boson at the LHC is the gluon gluon fusion (ggH). The coupling between the Higgs and the gluons is operated via a quark loop, dominated by the contributions from the heaviest quarks, namely the top and bottom. The Higgs couples directly to the heavy quarks in the loop via Yukawa interaction, and as such the ggH process can be used to study this coupling and its properties under CP symmetry.

Eq. 2.2 describes the amplitude of the Higgs boson coupling to spin 1 particles, and can therefore be used to study the effective ggH coupling. The techniques used to study this coupling are therefore similar to the ones used to study HVV anomalous couplings. A dedicated effective fractional cross section to study the Higgs effective coupling to gluons is defined as:

$$f_{a_3}^{ggH} = \frac{|a_3^{gg}|^2}{|a_2^{gg}|^2 + |a_3^{gg}|^2} \, \text{sign}\left(\frac{a_3^{gg}}{a_2^{gg}}\right) \, , \qquad (2.10)$$

where the cross sections σ_3^{gg} and σ_2^{gg} have been dropped due to the observation that the cross sections for pure CP-even or CP-odd couplings in ggH are equal [64]. In a similar way an effective fraction is defined for the Hff coupling:

$$f_{CP}^{Hff} = \frac{|\tilde{\kappa}_f|^2}{|\tilde{\kappa}_f|^2 + |\kappa_f|^2} \, \text{sign}\left(\frac{\tilde{\kappa}_f}{\kappa_f}\right) \, , \qquad (2.11)$$

where the same observation is made regarding the cross section obtained for pure CP-even or CP-odd couplings.

These two expressions can be used to estimate the CP admixture respectively in the ggH effective coupling and in the Hff direct one. Under the hypothesis that the quark loop in ggH production is dominated by the top and bottom quark, it is also possible to combine these two expressions as shown in [65]:

$$\left|f_{CP}^{Hff}\right| = \left(1 + 2.38\left[\frac{1}{\left|f_{a_3}^{ggH}\right|} - 1\right]\right)^{-1} \, , \qquad (2.12)$$

where the two effective fractions are taken to have the same sign.

Like in the study of HVV anomalous couplings, the initial state kinematics are used to estimate limits on the effective fractions. This was done in [65] by investigating the ggH and ttH Higgs production mechanisms, and combining their results using the expression in Eq. 2.12. The resulting limits on the anomalous Yukawa couplings between the Higgs boson and top quarks are shown in Fig. 2.3. The observed (expected) effective fraction was found to be

$$f_{CP}^{Htt} = -0.13^{+0.56}_{-0.48} \quad (0.00 \pm 0.63) \, , \qquad (2.13)$$

at 95% CL. The limits on the corresponding couplings on κ_t and $\tilde{\kappa}_t$ are shown in the right part of Fig. 2.3. As shown, a pure CP-odd coupling for the Higgs Yukawa coupling to top quarks has been excluded at 99.7% CL.
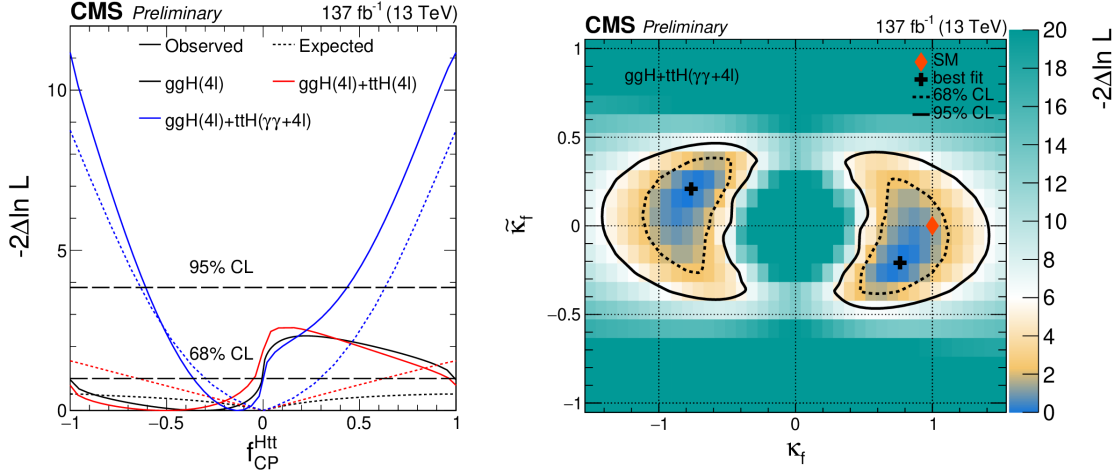
Figure 2.3: Constraints on the effective fraction f_{CP}^{Htt} (**left**) and the Higgs Yukawa couplings to top quarks, κ_t and $\tilde{\kappa}_t$ (**right**) [65]. The limits have been obtained by combining the respective measurements in the ggH and ttH production mechanisms under the assumption of top dominance in the ggH quark loop as described in Eq. 2.12.

2.2.2   CP violation in $H \rightarrow \tau\tau$ decays

The study of CP violation in ggH and ttH production mechanisms and in the HVV couplings all use multivariate discriminators determined using several variables associated to the initial or final state in order to distinguish different hypotheses on the coupling structure. This is one of the main differences between these studies and the one which is discussed in this section: the study of the structure of the Higgs Yukawa coupling to tau leptons under CP symmetry. As will be shown in this section, the spin correlation between the tau leptons coming from the Higgs boson, and between the former and their decay products, allows to construct analytically an observable which is sensitive to the CP structure of the Yukawa couplings.

Eq. 2.9 describes the Yukawa Lagrangian for a generic fermion including both CP-even and CP-odd couplings. For the interaction to tau leptons it takes the form:

$$\mathcal{L}_{Y,\tau} = -\frac{m_\tau}{v}\,\bar{\tau}(\kappa_\tau + i\gamma^5\tilde{\kappa}_\tau)H\tau\;. \tag{2.14}$$

As previously stated, the real parameters κ_τ and $\tilde{\kappa}_\tau$ represent the reduced CP-even and CP-odd Yukawa couplings of the Higgs to tau leptons. In order to study the properties

under CP symmetry of the Yukawa coupling between the Higgs boson and tau leptons it is necessary to compute the $H \to \tau\tau$ cross section as dependent on variables which are sensitive to the CP properties of the coupling structure. Since the Higgs coupling to tau leptons does not appear in the Higgs production mechanisms, at least in the interactions at leading order (LO), it is easier to discuss the Higgs decay width first and ignore the production mechanisms. As shown in [66], the $H \to \tau\tau$ decay width can be written for pure CP-even ($\kappa_\tau = 1$ and $\tilde{\kappa}_\tau = 0$) or CP-odd ($\kappa_\tau = 0$ and $\tilde{\kappa}_\tau = 1$) couplings as a function of the tau leptons polarization $s^\pm$:

$$\Gamma\left(H\left(\frac{CP-even}{CP-odd}\right) \to \tau\tau\right) = \Gamma^{unpol}\left(1 - s^-_\parallel s^+_\parallel \pm s^-_\perp s^+_\perp\right) , \tag{2.15}$$

where Γ^{unpol} represents the decay width calculated integrating over the polarization of the two tau leptons, and the projections of the $\tau^\pm$ polarization in the direction parallel (orthogonal) to the tau direction of flight is represented by $s^\pm_\parallel$ ($s^\pm_\perp$). The term $s^-_\perp s^+_\perp$ appears in the equation with a $+$ sign for a pure CP-even coupling, and with a $-$ sign for a pure CP-odd one. Eq. 2.15 can be then generalized as shown in [67] to account for a more general admixture between CP-even and CP-odd coupling:

$$\Gamma(H_{mix} \to \tau\tau) = \Gamma^{unpol}(1 - s^-_\parallel s^+_\parallel + s^-_\perp R(\varphi_{\tau\tau})s^+_\perp) , \tag{2.16}$$

where

$$\varphi_{\tau\tau} = \arctan\left(\frac{\tilde{\kappa}_\tau}{\kappa_\tau}\right) \tag{2.17}$$

is the "CP mixing angle" and is used to parametrize the mixing between the Yukawa couplings. The dependence of the Higgs decay width on the CP mixing angle is encoded in the matrix $R(\varphi_{\tau\tau})$ [2], which alters the spin correlation between the tau transverse polarizations depending on the Higgs CP nature. Eq. 2.16 is used to establish the spin

---

[2]The analytical form of this matrix can be written for a specific coordinate system. In the Higgs rest frame the two tau lepton momenta sum to 0, and their direction of flight can be used to define the z axis of a coordinate system. In this coordinate system the $s^\pm_\perp$ vectors are co-planar, with non-zero components only along the x and y axis, and the R matrix is a 3×3 matrix of the form [67]:

$$R(\varphi_{\tau\tau}) = \begin{bmatrix} R_{11} & R_{12} & 0 \\ R_{21} & R_{22} & 0 \\ 0 & 0 & -1 \end{bmatrix} , \tag{2.18}$$

where

$$R_{11} = R_{22} = \frac{\cos^2(\varphi_{\tau\tau})\beta - \sin^2(\varphi_{\tau\tau})}{\cos^2(\varphi_{\tau\tau})\beta + \sin^2(\varphi_{\tau\tau})} ,$$

$$R_{12} = -R_{21} = \frac{2\sin(\varphi_{\tau\tau})\cos(\varphi_{\tau\tau})\beta}{\cos^2(\varphi_{\tau\tau})\beta + \sin^2(\varphi_{\tau\tau})} . \tag{2.19}$$

correlation between the Higgs boson and the tau leptons it decays to. The spin correlation depends on the transverse components of the tau polarization, which can be accessed via the tau decay products. As shown in Eq. 1.41, taken from [48], the decay width of tau leptons includes a term of the form $\vec{s} \cdot \vec{q}$, scalar product between the tau spin and the momenta of a charged decay product. This term encodes the spin correlation between the tau and its decay products. Eq. 1.41 was obtained taking into account decays of the form $\tau^{\pm} \to p^{\pm} + X$, with $p^{\pm}$ one of the charged decay products, and X a neutral system of particles which includes the tauonic neutrino and all other decay products[3]. The formula can be rewritten in a Lorentz invariant frame of reference for decays of the form $\tau^{\pm} \to X^{\pm} + \nu_{\tau}$, with $X^{\pm}$ representing the charged system of tau decay products besides the tauonic neutrino [68, 69]:

$$\mathrm{d}\Gamma_{\tau \to X + \nu_{\tau}} = \frac{1}{2m_{\tau}} |\mathscr{M}|^{2} (1 + h_{\mu} s^{\mu}) \mathrm{d}Lips \; . \tag{2.22}$$

This differential expression of the decay width with respect to a Lorentz invariant phase space ($Lips$) element contains a term analogous to the scalar product $\vec{s} \cdot \vec{q}$ found in Eq. 1.41: $h_{\mu} s^{\mu}$. This term represents the dependence of the tau decay width on the tau polarization s, here written as a 4-vector. The 4-vector h^{μ} is used to encode how the tau decay products propagate depending on the tau spin, and is named *polarimetric vector*[4].

This parametrization allows to encode the spin-correlation of the tau lepton with its decay products inside a vector, and define variables sensitive to the CP-mixing in a way which is independent of the decay channel. Fig. 2.4 shows how the polarimetric vectors ($\vec{h}^{\pm}$) and the tau leptons ($\tau^{\pm}$) momenta are directed in the Higgs rest frame. By defining $\theta^{\pm}$ and $\phi^{\pm}$ as the polar and azymuthal angles defining the direction of each polarimetric vector with respect to the corresponding tau lepton direction of flight, the $H \to \tau\tau$ cross

---

The β factor depends on the τ lepton and Higgs boson masses m_{τ}, m_{H}:

$$\beta = \sqrt{1 - 4\frac{m_{\tau}^{2}}{m_{H}^{2}}} \; . \tag{2.20}$$

By approximating this factor to 1 the R matrix can be rewritten as a rotation matrix around the z axis by an angle $2\varphi_{\tau\tau}$:

$$R(2\varphi_{\tau\tau}) = \begin{bmatrix} \cos(2\varphi_{\tau\tau}) & \sin(2\varphi_{\tau\tau}) & 0 \\ -\sin(2\varphi_{\tau\tau}) & \cos(2\varphi_{\tau\tau}) & 0 \\ 0 & 0 & -1 \end{bmatrix} \; . \tag{2.21}$$

The matrix is diagonal for $\varphi_{\tau\tau} = 0$ or $\pi/2$ and the term $s_{\perp}^{-} R(\varphi_{\tau\tau}) s_{\perp}^{+}$ simplifies to a scalar products, as shown in Eq. 2.15.

[3]For example, in $\tau^{\pm} \to \nu_{\tau} \pi^{\pm} \pi^{\pm} \pi^{\mp}$ decays $p^{\pm} \equiv \pi^{\pm}$, while X corresponds to the $\pi^{\pm} \pi^{\mp} \nu_{\tau}$ system.

[4]In the oldest references found, this quantity is named *tau polarimeter*, meaning "the quantity used to measure the tau polarization". The terms *polarimeter vector* and *polarimetric vector* have later been introduced in literature and are treated as synonyms to *tau polarimeter*.

section can be written, in the limit of ultrarelativistic tau leptons, as [70, 71]:

$$d\sigma_{H \to \tau\tau}/d\cos\left(\theta^+\right) d\cos\left(\theta^-\right) d\cos\left(\phi^+\right) d\cos\left(\phi^-\right) \propto$$
$$\left(1 + \cos\left(\theta^+\right)\cos\left(\theta^-\right) - \sin\left(\theta^+\right)\sin\left(\theta^-\right)\cos\left(\phi^+ - \phi^- - 2\varphi_{\tau\tau}\right)\right) \; . \tag{2.23}$$
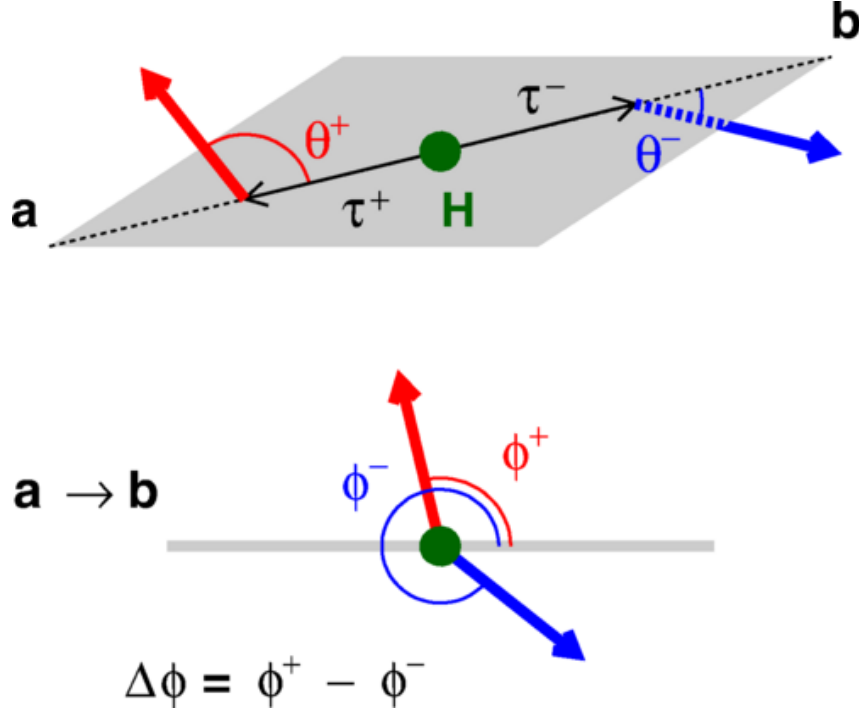
Figure 2.4: Schematic representation of the Higgs decay to polarized tau leptons in the Higgs rest frame [71]. The thick blue (red) arrow represents the polarimetric vectors for the τ^- (τ^+), while the thin black lines represent the tau leptons momenta. The upper part of the figure represents the decay in a plane on which lie the tau lepton momenta, while the lower part shows the same decay but in a plane orthogonal to the tau direction of flight. The label $\mathbf{a} \to \mathbf{b}$ is used to indicate that in the lower part of the figure the τ^+ is *coming out* of the page, while the τ^- is *entering* the page

This allows to rewrite the term $s_\perp^- R(\varphi_{\tau\tau})s_\perp^+$ in Eq. 2.16 as a function of the angle $\varphi_{\tau\tau}$ and the difference of the azymuthal coordinates of the two polarimetric vectors ($\Delta\phi = \phi^+ - \phi^-$). The scalar product of the two polarimetric vectors ($\vec{h}^- \cdot \vec{h}^+$) depends on the azymuthal and polar coordinates of the polarimetric vectors and can therefore be used to indirectly access the $\Delta\phi = \phi^+ - \phi^-$ angle. To access the $\Delta\phi$ directly, the vectorial product of the polarimetric vector with its corresponding tau momentum ($\vec{h}^\pm \times \vec{p}_\tau^\pm$) must instead be used. These two combinations of the polarimetric vectors lead to construct two CP-sensitive observables [72]:

- the *acollinearity* angle,

$$\delta = \arccos\left(\vec{h}^- \cdot \vec{h}^+\right) , \tag{2.24}$$

- and the *acoplanarity* angle,

$$\varphi^* = \Delta\phi = \arccos\left(\vec{k}^- \cdot \vec{k}^+\right) , \tag{2.25}$$

where:

$$\vec{k}^{\pm} = \frac{\vec{h}^{\pm} \times \vec{p}_\tau^{\pm}}{\left|\vec{h}^{\pm} \times \vec{p}_\tau^{\pm}\right|} . \tag{2.26}$$

The δ variable is named *acollinearity* as it represents the degree to which the two polarimetric vectors are aligned. It is defined with respect to the plane containing the two polarimetric vectors, and as such is defined between 0 and π.

The φ^* variable is named *acoplanarity* instead and is defined in the plane orthogonal to the $\tau^{\pm}$ direction of flight in the Higgs rest frame. As such, its definition can be altered in order to admit values between 0 and 2π by defining a direction of reference in this plane. A convention which can be used is the right-hand rule with respect to the τ^- direction of flight. In this coordinate system $\phi^- = 0$ and $\phi^+ = \Delta\phi = \varphi_{CP}$:

$$\varphi_{CP} = \begin{cases} \varphi^*, & \text{if } O \geq 0 \\ 2\pi - \varphi^*, & \text{if } O < 0 \end{cases} , \tag{2.27}$$

where the variable O represents the right hand rule convention: it selects cases where the cross product of the two vectors is either parallel or antiparallel with the negatively charged tau direction of flight. Analytically it is defined as:

$$O = \frac{(\vec{h}^+ \times \vec{h}^-) \cdot \vec{p}_{\tau^-}}{\left|(\vec{h}^+ \times \vec{h}^-) \cdot \vec{p}_{\tau^-}\right|} . \tag{2.28}$$

As previously stated, the polarimetric vectors allow to construct quantities which are sensitive to the spin-correlation between the two tau leptons, independently of the tau decay channel. However when actually measuring the acollinearity and acoplanarity angles the dependence of the polarimetric vector on the tau decay products must be made explicit. Based on how the Eq. 2.22 descends from Eq. 1.41, it is apparent that for $\tau \to \pi\nu_\tau$ decays the polarimetric vector coincides with the 4-momentum of the charged pion.

For leptonically decaying tau leptons the polarimetric vector cannot be reconstructed experimentally, due to the presence of two neutrinos in the decay. However, by comparing

Eq. 1.31 with Eq. 2.22 it is possible to construct a vector which serves the same purpose:

$$\vec{h} = \mathcal{N}\frac{\vec{p}}{E}\left(4E - m_\tau - \frac{3m_l^2}{m_\tau}\right) \ , \tag{2.29}$$

with $\mathcal{N}$ normalization factor and E and $\vec{p}$ energy and momentum of the charged lepton. In both cases, the vector that carries the information about the tau lepton spin is directed alongside the momentum of the charged decay product. However, in leptonic decays this vector changes sign depending on the fraction of the tau energy carried by the charged lepton.

This allows to better understand the definition of the acoplanarity angle: each cross product $\vec{h}^\pm \times \vec{p}_\tau^\pm$ can be identified as the normal with respect to the decay plane of the $\tau^\pm$, since it is defined as orthogonal to both the charged decay product, and the tau direction of flight. This holds true for the other hadronic decay channels, albeit less evident as the form taken by the polarimetric vector becomes more complicated [73]. In $\tau^\pm \to \pi^\pm\pi^0\nu_\tau$ decays, for example, the charged and neutral pions are obtained via a ρ meson decay. Their system inherits the quantum numbers of intermediate resonance, namely $J^P = 1^-$, and the polarimetric vector takes the form [73, 74]:

$$\vec{h} = \mathcal{N}(2(q \cdot N)\vec{q} - q^2\vec{N}) \ , \tag{2.30}$$

with $\mathcal{N}$ normalization factor, $q = (q_0, \vec{q})$ difference between the charged and neutral pion 4-momenta and $N = (N_0, \vec{N})$ neutrino 4-momentum, all defined in the τ rest frame. For the $\tau^\pm \to \pi^\pm\pi^\pm\pi^\mp\nu_\tau$ decays, the form taken is noticeably longer, as it is obtained via a chain of two-body decays: $\tau^\pm \to a_1^\pm\nu_\tau \to \rho^0\pi^\pm\nu_\tau \to \pi^\pm\pi^\pm\pi^\mp\nu_\tau$. The polarimetric vector depends on the 4-momenta of the 3 charged pions [69, 75], and is generally treated by using MC methods [51, 73, 76, 77]. The use of MC simulation to determine the τ polarimeter is justified by the decay modelling in simulation bringing negligible ambiguity to the measurable properties of the decay [76].

An important thing to mention is how the φ_{CP} distribution looks for the $Z \to \tau\tau$ process. This is an irreducible background having the same final state as Higgs decays to tau leptons. The Z boson has spin 1, resulting in the dominant spin correlation being the one with the tau longitudinal polarizations instead of their transverse ones. To better understand this, it is useful to write the cross section for the $Z \to \tau\tau$ process. The same variables and approximations defined for Eq. 2.23 can be used, however it is useful to generalize it to account for the interference with a virtual photon decaying to tau leptons.

The Drell-Yan production of a tau lepton pair can be written as [70]:

$$
\mathrm{d}\sigma_{DY}/\mathrm{d}\cos\left(\theta^+\right)\mathrm{d}\cos\left(\theta^-\right)\mathrm{d}\cos\left(\phi^+\right)\mathrm{d}\cos\left(\phi^-\right)\mathrm{d}E^+\mathrm{d}E^- \propto \sum_{B_1,B_2=Z,\gamma} a(B_1,B_2)
$$

$$
\times\left\{V_\tau^{B_1}V_\tau^{B_2}\left[1-\left(\cos\left(\theta^+\right)\cos\left(\theta^-\right)+\frac{1}{2}\sin\left(\theta^+\right)\sin\left(\theta^-\right)\cos\left(\phi^+ + \phi^-\right)\right)\right]\right.
$$

$$
+A_\tau^{B_1}A_\tau^{B_2}\left[1-\left(\cos\left(\theta^+\right)\cos\left(\theta^-\right)-\frac{1}{2}\sin\left(\theta^+\right)\sin\left(\theta^-\right)\cos\left(\phi^+ + \phi^-\right)\right)\right]
$$

$$
\left.+(a^{B_1}V_\tau^{B_2}+V_\tau^{B_1}A_\tau^{B_2})\left(\cos\left(\theta^+\right)-\cos\left(\theta^-\right)\right)\right\} ,
$$

(2.31)

where $V_\tau^{B_i}$ and $A_\tau^{B_i}$ are the vector and axial couplings of the photon and Z boson to tau leptons, and $a(B_1,B_2)$ depends on the coupling of the vector bosons to the quarks from the initial state and the center of mass energy [70,78]. By substituting $\varphi^* = \phi^+ - \phi^-$ into Eq. 2.31 the dependence of the cross section on the acoplanarity angle is carried in terms of the form $\cos(2\phi^+ + \varphi^*)$. By integrating the equation in ϕ^+ between 0 and 2π, those contributions are averaged to 0. This means that the differential cross section $\mathrm{d}\sigma_{DY}/\mathrm{d}\cos(\theta^+)\mathrm{d}\cos(\theta^-)\mathrm{d}\varphi^*$ does not depend on φ^* and is therefore constant with respect to that variable. This allows an efficient separation between Higgs and Z decays to tau leptons.

2.2.3   The acoplanarity angle reconstruction

The previous section focused on discussing how the CP mixing angle can be accessed by using the tau polarimetric vectors in the Higgs rest frame. The main limitation coming from the experimental point of view is the impossibility of directly reconstructing neutrinos and consequently:

- the Higgs rest frame;

- the total tau lepton momentum;

- the tau lepton polarimetric vector.

This means that in most cases the acoplanarity angle cannot be measured using Eq. 2.25, here reported for convenience of the reader without the extended definition between 0 and 2π:

$$
\varphi^* = \arccos\left(\frac{\vec{h}^- \times \vec{p}_\tau^-}{\left|\vec{h}^- \times \vec{p}_\tau^-\right|} \cdot \frac{\vec{h}^+ \times \vec{p}_\tau^+}{\left|\vec{h}^+ \times \vec{p}_\tau^+\right|}\right) .
$$

(2.32)

To reconstruct the acoplanarity angle in an experiment at the LHC some approximated methods are required. Such methods can be constructed depending on the tau decay channel studied, in order to take advantage of the characteristics of each distinct decay mode. Before describing each method in detail, it is useful to address which are the common properties they should possess.

The vector representing the normal to the tau decay plane in Eq. 2.32 is defined based on a cross product between the tau polarimetric vector and momentum: $\vec{h}^{\pm}$ and $\vec{p}_{\tau}^{\pm}$. The order in which the vectors appear in the cross product is relevant to determine the sign of the product itself, meaning that the normal to each decay plane must be defined with a well determined sign. This can be accomplished more generally by defining each plane using two vectors, both lying on the plane. Furthermore, all the aforementioned vectors are defined in the Higgs rest frame, which cannot be directly reconstructed experimentally. An approximation for the Higgs rest frame is therefore required, in order to reconstruct the acoplanarity angle experimentally.

Fig. 2.5 represents the Higgs rest frame and the vectors used to define the acoplanarity angle: each plane is defined in the Higgs rest frame by the tau momentum (represented in the figure as τ) and the corresponding polarimetric vector (represented with the letter h). Since in the Higgs rest frame $\vec{p}_{\tau+} + \vec{p}_{\tau-} = 0$, the two momenta are back-to-back and form an axis which lies on both decay planes. The normal to each plane is therefore orthogonal to both tau momenta:

$$(\vec{h}^{\pm} \times \vec{p}_{\tau}^{\pm}) \cdot \vec{p}_{\tau}^{\mp} = 0 \ . \tag{2.33}$$

Consequently, the acoplanarity can be measured in two equivalent ways:

1. by using the normal vectors to each decay plane;

2. by using vectors lying on the plane themselves and orthogonal to the tau momentum.

This can be understood by analyzing the argument of the arccosine function in Eq. 2.32: the polarimetric vectors in the calculation can be swapped with their components orthogonal to the tau momenta $\vec{h}_{\perp}^{\pm}$ without changing the resulting angle:

$$\vec{h}^{\pm} \times \vec{p}_{\tau}^{\pm} = (\vec{h}_{\perp}^{\pm} + \vec{h}_{/\!/}^{\pm}) \times \vec{p}_{\tau}^{\pm} = \vec{h}_{\perp}^{\pm} \times \vec{p}_{\tau}^{\pm} \ , \tag{2.34}$$

where $\vec{h}_{/\!/}^{\pm}$ is the component of the polarimetric vector parallel to the tau momentum, and the last equivalence holds because the cross product between two parallel vectors is 0. Furthermore, the argument of the arccosine function is a scalar product of two vectorial products between the tau momentum and its polarimetric vector. On such a product the following identity can be applied [79]:

$$(\vec{a} \times \vec{b}) \cdot (\vec{c} \times \vec{d}) = (\vec{a} \cdot \vec{c})(\vec{b} \cdot \vec{d}) - (\vec{a} \cdot \vec{d})(\vec{b} \cdot \vec{c}) \ . \tag{2.35}$$
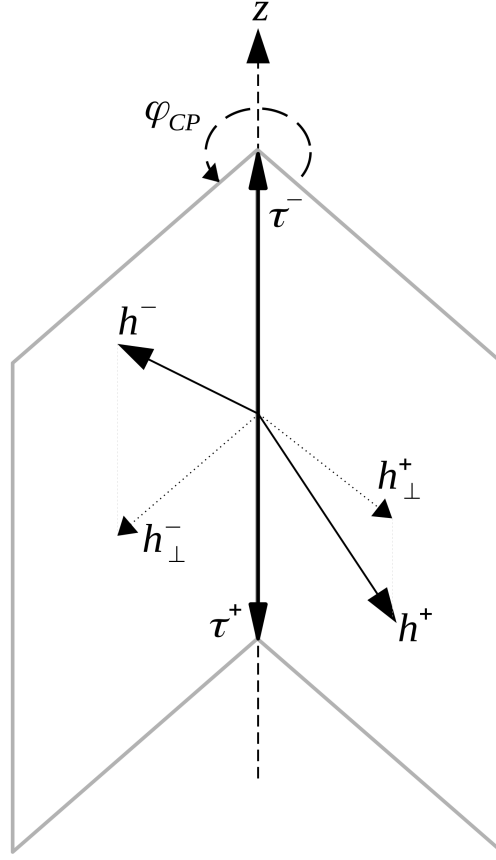
Figure 2.5: Schematic representation of the acoplanarity angle reconstruction in the Higgs rest frame using polarimetric vectors and tau momenta.

Applying that identity to the products in Eq. 2.32, it assumes the following form:

$$
(\vec{h}_\perp^- \times \vec{p}_\tau^-) \cdot (\vec{h}_\perp^+ \times \vec{p}_\tau^+) = (\vec{h}_\perp^- \cdot \vec{h}_\perp^+)(\vec{p}_\tau^- \cdot \vec{p}_\tau^+) - \underbrace{(\vec{h}_\perp^- \cdot \vec{p}_\tau^+)}_{=0}\underbrace{(\vec{h}_\perp^+ \cdot \vec{p}_\tau^-)}_{=0} \tag{2.36}
$$
$$
= -|\vec{p}_\tau|^2 (\vec{h}_\perp^- \cdot \vec{h}_\perp^+) \; .
$$

The minus sign produces a phase shift of π with respect to the angle computed with the scalar product of the $h_\perp^\pm$ vectors.

Measuring the acoplanarity angle thus requires:

1. two vectors per decay plane ($\vec{P}_\pm$ and $\vec{R}_\pm$), of which at least one must be the momentum of a charged particle ($\vec{P}_\pm$);

2. defining a frame of reference such that the vectors $\vec{P}_\pm$, boosted in that reference frame, sum to 0: $\vec{P}_+' + \vec{P}_-' = 0$.

As a function of these vectors, the acoplanarity angle can me calculated as:

$$\varphi^* = \arccos\left(\frac{\vec{R}_\perp^-}{\left|\vec{R}_\perp^-\right|} \cdot \frac{\vec{R}_\perp^+}{\left|\vec{R}_\perp^+\right|}\right) \ . \tag{2.37}$$

with $\vec{R}_\perp^\pm$ being the components of the vectors $\vec{R}^\pm$ orthogonal to the momenta $\vec{P}^\pm$. This method must now be applied to the different decay channels according to the available particle momenta and the degree of approximation of the Higgs rest frame achievable.

The tau branching fractions to the various decay modes are listed in Tab. 1.4 and are shown here in Fig. 2.6. The decay channels which can be used to reconstruct the CP properties of the Higgs boson are:

- the leptonic decays, as the spin correlation is preserved due to the neutrinos having definite helicity:

  - $\tau^- \to \nu_\tau \mu^- \bar{\nu}_\mu$, BR $\sim 17.4\%$,
  - $\tau^- \to \nu_\tau e^- \bar{\nu}_e$, BR $\sim 17.8\%$,

- the hadronic decays which involve mesonic resonances:

  - $\tau^- \to \nu_\tau \pi^-$, BR $\sim 10.8\%$,
  - $\tau^- \to \nu_\tau \pi^- \pi^0$, BR $\sim 25.5\%$,
  - $\tau^- \to \nu_\tau \pi^- \pi^0 \pi^0$, BR $\sim 9.3\%$,
  - $\tau^- \to \nu_\tau \pi^- \pi^- \pi^+$, BR $\sim 9.3\%$

Other decay channels are not considered[5], as the spin correlation is diluted over multiple decay products in absence of intermediate resonances with definite spin.

As the methods discussed are targeting the measurement of the CP properties in $H \to \tau\tau$ events reconstructed with the CMS experiment, the features and limitations of the reconstruction of tau leptons performed in this experiment are taken into account when describing the methods. A thorough description of the tau lepton identification and reconstruction in CMS is presented in Section 4.2.1, here only the features which are relevant for defining the methods to measure φ_{CP} are mentioned.

The process with the highest BR is the $\tau^- \to \nu_\tau \pi^- \pi^0$ decay which, together with the $\tau^- \to \nu_\tau \pi^- \pi^0 \pi^0$ one, presents as a signature a reconstructed charged hadron, accompanied by one or more π^0s. These decays can be grouped together in the **1 prong + π^0s** channel using the same notation of Section 4.2.1, on account of their similar signatures in

---

[5]As discussed in Section 1.5.1, $\tau \to \nu_\tau K^-$ decays have identical spin correlation properties compared to the $\tau^- \to \nu_\tau \pi^-$ ones. The two decay channels can therefore been treated with a common approach. Other decay channels involving kaons cannot be represented as a cascade of two-body decays, resulting in a diluted spin-analyzing power. They are therefore ignored in the following treatment.
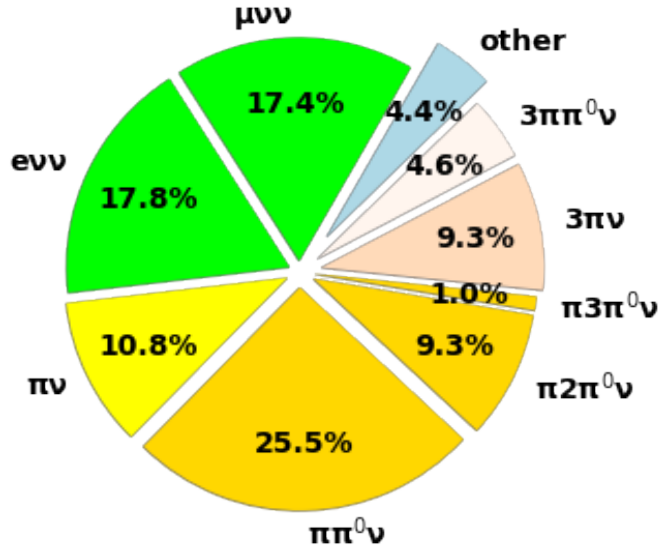
Figure 2.6: Pie chart representing the tau decay channels with relative branching fractions. The values of the branching fractions are taken from [23].

the particle detector. Both decay channels involve an intermediate resonance with the $\rho^\pm$ meson, whose decay products can be used to reconstruct the tau decay plane, and consequently φ_{CP}. To better understand this, it is useful to first analyze the $\tau^- \to \nu_\tau \pi^- \pi^0$ decay: let $\vec{p}_{\pi^\pm}$ and $\vec{p}_{\pi^0}$ be the charged and neutral pion momenta and $\vec{N}$ the neutrino momentum. While the neutrino momentum cannot be reconstructed, it lays in the same plane as the momenta of the two pions, in the tau rest frame. As previously shown in Eq. 2.30 [73, 74], the polarimetric vector for $\tau \to \rho \nu_\tau$ is:

$$\vec{h} = \mathcal{N}(2(q \cdot N)\vec{q} - q^2 \vec{N}) , \tag{2.38}$$

with $\mathcal{N}$ normalization factor, $q = (q_0, \vec{q})$ difference between the charged and neutral pion 4-momenta and $N = (N_0, \vec{N})$ neutrino 4-momentum. The polarimetric vector in the tau rest frame lays in the same plane as the pions and neutrino. Therefore a direction orthogonal to the momenta of the two pions, is also orthogonal to the tau polarimetric vector and momentum:

$$(\vec{p}_{\pi^\pm} \times \vec{p}_{\pi^0}) \not\parallel (\vec{h} \times \vec{p}_\tau) , \tag{2.39}$$

and the momenta of the charged and neutral pion can be used instead of the polarimetric vector and tau momentum to reconstruct φ_{CP}. While the two cross products lead to the definition of vectors parallel to each other, the sign of the vectors is also relevant in order to reconstruct the $H \to \tau\tau$ differential cross section. The polarimetric vector changes sign depending on the difference between the charged and neutral pion momenta. A similar change of sign must be applied to the cross product of the pion momenta in order

accurately reconstruct φ_{CP}. The first part of the polarimetric vector can be approximated in the tau rest frame as [74]:

$$q \cdot N \simeq (E_{\pi^\pm} - E_{\pi^0})m_\tau \ . \tag{2.40}$$

The normal to the ρ decay plane ($\vec{k}_\rho$) is therefore set to have the same sign as the energy difference between the two pions ($E_{\pi^\pm} - E_{\pi^0}$). With this condition in place, the pion momenta can be substituted to the tau polarimetric vector and momentum to measure φ_{CP}:

$$\frac{\vec{p}_{\pi^\pm} \times \vec{p}_{\pi^0}}{|\vec{p}_{\pi^\pm} \times \vec{p}_{\pi^0}|} \times \text{sign}(E_{\pi^\pm} - E_{\pi^0}) \simeq \frac{\vec{h} \times \vec{p}_\tau}{\left|\vec{h} \times \vec{p}_\tau\right|} \ . \tag{2.41}$$

The categorization based on the energy differences can also be applied directly to the measured acoplanarity angle. Considering a tau lepton pair decaying into $\tau_\rho \times \tau_X$ channels, with X representing an unspecified tau deacy, the acoplanarity angle can be first measured as the scalar product between the normals of each decay plane:

$$\varphi^* = \arccos\left(\frac{\vec{p}_{\pi^\pm} \times \vec{p}_{\pi^0}}{|\vec{p}_{\pi^\pm} \times \vec{p}_{\pi^0}|} \cdot \vec{k}_X\right) , \tag{2.42}$$

where $\vec{k}_X$ is the normal to the $\tau \to \tau_X$ decay plane. A phase shift of π is then applied to the angle if $y_\pm{}^6 < 0$ with

$$y_\pm = \frac{E_{\pi^\pm} - E_{\pi^0}}{E_{\pi^\pm} + E_{\pi^0}} \ , \tag{2.43}$$

implicitly extending the angle definition between 0 and 2π. The right hand rule with respect to the negative prong direction of flight can then be applied, as shown in Eq. 2.27, to measure φ_{CP}. The order in which these last two operations on the angle are performed is irrelevant due to the property $x + 2\pi = x$, $\forall x \in \{0, 2\pi\}$:

$$x \overset{X+\pi}{\longrightarrow} x + \pi \overset{2\pi-X}{\longrightarrow} \pi - x \ , \tag{2.44}$$

$$x \overset{2\pi-X}{\longrightarrow} 2\pi - x \overset{X+\pi}{\longrightarrow} 3\pi - x = \pi - x \ . \tag{2.45}$$

In cases where both tau leptons decay via $\rho^\pm$ resonance, there is the possibility of two phase shifts of π being applied, depending on the energy differences between pions, which would cancel each other. In $\tau_\rho \times \tau_\rho$ final states this results in the phase shift being applied whenever the following condition is satisfied:

$$y_+ y_- < 0 \ , \tag{2.46}$$

---

[6]This variable can also be derived analytically considering the decay of a spin-1 particle (ρ) into two spin-0 ones (π). As a p-wave decay, the angle between the $\pi^\pm$ direction of flight in the laboratory frame and the ρ spin in its rest frame is sensitive to the energy difference between the decay products.

with $y_\pm$ defined as in Eq. 2.43. The description of how a charged pion energy and momentum is reconstructed is presented in Section 4.1.5, while the typical signature of a π^0 in the CMS detector is illustrated in Section 4.2.1: it is a cluster of electrons and photons reconstructed in the proximity of the charged hadron associated to the tau decay. The module of the π^0 momentum is reconstructed as the sum of the photon and electron momenta $(\vec{p}^{\,e/\gamma})$ associated to this cluster:

$$|\vec{p}_{\pi^0}| = \left| \sum^{e/\gamma} \vec{p}^{\,e/\gamma} \right| , \qquad (2.47)$$

while its direction is taken to coincide with the one of the highest p_T photon.

As discussed in Section 4.2.1 the cluster size is reconstructed dynamically, resulting in a large overlap between decay modes identified as including one or two π^0 mesons. Due to this large overlap, the reconstruction of the decay plane for the **one prong $+$ π^0s** channel is done in the same way as for the ρ meson resonance. The neutral pions decay products are combined together as per Eq. 2.47 in order to define the total momentum of the neutral pions. This momentum is then treated as if it was of a single π^0 in the previously described method.

To discuss the frame of reference in which the vectors are boosted to compute the acoplanarity angle it is useful to consider pairs of decaying tau leptons. If both tau leptons decay in **one prong $+$ π^0s** channels, the frame of reference is defined as the zero momentum frame of the system formed by the two charged pions involved in the decay. More generally for this decay mode, the charged pion is the object used in the definition of the Higgs rest frame approximation.

This method is then extended from the **one prong $+$ π^0s** channel to the **3 prong** one. This decay channel involves an intermediate resonance with the a_1 meson, which further decays via an intermediate ρ meson resonance. In this decay channel, two oppositely charged pions should have an invariant mass close to $m_\rho = 770$ MeV. Given that all three pions momenta lie on the same plane in the tau rest frame, it is possible to define the plane as in the τ_ρ case:

1. the oppositely charged pions coming from the a_1 decay which have invariant mass closest to the ρ resonance are selected;

2. the charged pion having the same (opposite) sign as the decaying tau is then substituted to the charged (neutral) pion used for the ρ method[7].

---

[7]This includes using the energy of the charged pions to construct a $y_\pm$ variable analogous to the one shown in Eq. 2.43:

$$y_\pm = \frac{E_{\pi^\pm} - E_{\pi^\mp}}{E_{\pi^\pm} + E_{\pi^\mp}} . \qquad (2.48)$$

Similar variables can be constructed in general using different pion pairs coming from the a_1 decay. While

In these decay channels the charged pion momentum is then used in the definition of the frame of reference where the acoplanarity angle is measured.

Decay channels involving only one charged particle cannot be reconstructed with the aforementioned method, due to the lack of either a neutral system or another charged particle to be used in the plane definition. Tau leptonic decays fall into this category, together with the ones involving only one charged pion, and are referred to as **one prong** decays. While the charged particle momentum can be used to define the decay plane, a second vector is also needed. Such vector is required to lay in the tau decay plane and not be proportional (or more specifically, parallel) to the prong momentum. The impact parameter (IP) of the hadron or lepton trajectory satisfies these properties [81]. As shown in Fig. 2.7, it is orthogonal to the prong momentum in the laboratory reference frame, and lies in the same plane as the tau lepton direction of flight. The decay plane is then defined by the IP and the momentum of the prong itself, the latter being used to define the reference frame where the φ_{CP} angle is measured.
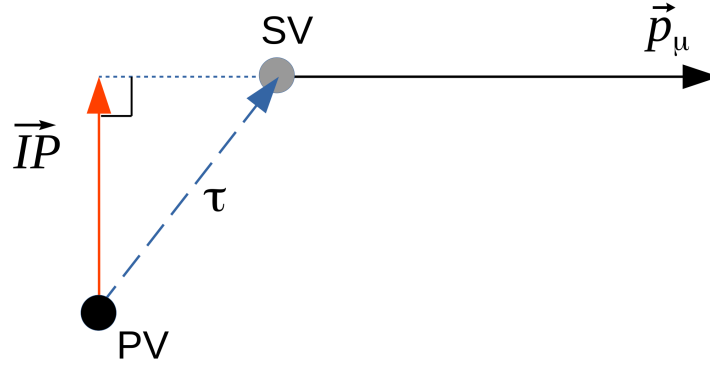

Figure 2.7: Schematic representation of the IP reconstruction in a two-dimensional plane.

In all the aforementioned cases, the frame of reference is defined as the zero momentum frame for a system formed by a charged pion, and another charged pion or lepton[8]. This is a reasonable substitution for the Higgs rest frame since neutrinos are not reconstructed, and the rest frame of the tau charged decay products can be reconstructed with relatively high precision.

A more precise approximation of the Higgs rest frame can be performed in the **3 prong + 3 prong** channel ($\tau_{a_1^{3Pr}} \times \tau_{a_1^{3Pr}}$), where each tau lepton decays into three charged hadrons, all reconstructed in the CMS detector, and a single neutrino, which is

---

such combinations were not used in the analysis described in Chapter 6, they still carry information on the a_1 meson spin and could be used to enhance the precision of the decay plane reconstruction using machine learning techniques as described in [80].

[8]The $\tau_e\tau_\mu$, $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$ channels are ignored in the presented treatment.

not reconstructed. Having three charged particles whose trajectory can be reconstructed allows to reconstruct the tau decay vertex, or secondary vertex (SV), and therefore to estimate the tau direction of flight as the vector connecting the primary interaction vertex (PV) with the SV: $\overrightarrow{SV} - \overrightarrow{PV}$. The tau momentum can also be reconstructed, by using its direction of flight and the total momentum of the three pion system, i.e. the a_1 momentum. The tau 4-momentum ($P_\tau = (E_\tau, \vec{p}_\tau)$) must be equal to the sum of the 4-momenta of neutrino ($P_\nu = (E_\nu, \vec{p}_\nu)$) and a_1 meson ($P_{a_1} = (E_{a_1}, \vec{p}_{a_1})$):

$$P_\tau = P_{a_1} + P_\nu . \tag{2.49}$$

Assuming a massless neutrino ($|P_\nu|^2 = 0$), the tau and a_1 4-momenta must satisfy the following condition:

$$|P_\tau - P_{a_1}|^2 = 0 . \tag{2.50}$$

This results in a quadratic equation [82] in the tau momentum having as parameters $p_{a_1} = |\vec{p}_{a_1}|$ and the angle between the tau and a_1 momenta: $\theta_{GJ} = \arccos(\hat{p}_\tau \cdot \hat{p}_{a_1})$, called Gottfried-Jackson angle [69]. Using the momentum of the tau lepton and its visible decay products it is possible to use the modelled 3 prong decays in TAUOLA [51, 73, 83] to infer the polarimetric vector and compute the acoplanarity angle as per Eq. 2.32.

Table 2.1: Summary of the vectors used to reconstruct the acoplanarity angle in the various decay channels.

	Vectors			
Channel	P1	R1	P2	R2
$\tau_{\mu,\pi} \times \tau_{\mu,\pi}$	$\vec{p}_{\mu,\pi}$	$\overrightarrow{IP}_{\mu,\pi}$	$\vec{p}_{\mu,\pi}$	$\overrightarrow{IP}_{\mu,\pi}$
$\tau_{\mu,\pi} \times \tau_{\rho,a_1^{1Pr}}$	$\vec{p}_{\mu,\pi}$	$\overrightarrow{IP}_{\mu,\pi}$	$\vec{p}_{\pi^\pm}$	$\vec{p}_{\pi^0}$
$\tau_{\mu,\pi} \times \tau_{a_1^{3Pr}}^\pm$	$\vec{p}_{\mu,\pi}$	$\overrightarrow{IP}_{\mu,\pi}$	$\vec{p}_{\pi^\pm}$	$\vec{p}_{\pi^\mp}$
$\tau_{\rho,a_1^{1Pr}} \times \tau_{a_1^{3Pr}}^\pm$	$\vec{p}_{\pi^\pm}$	$\vec{p}_{\pi^0}$	$\vec{p}_{\pi^\pm}$	$\vec{p}_{\pi^\mp}$
$\tau_{a_1^{3Pr}} \times \tau_{a_1^{3Pr}}^\pm$	$\vec{p}_\tau$	$\vec{h}$	$\vec{p}_\tau$	$\vec{h}$

In summary, each decay plane is defined by the momentum of a charged particle ($\vec{P}_\pm$) and another vector lying on the decay plane ($\vec{R}_\pm$) . All vectors are then boosted in the zero momentum frame of the charged particles ($\vec{P}_\pm + \vec{P}_\mp = 0$). The acoplanarity angle can then be measured by replacing $\vec{P}_\pm$ and $\vec{R}_\pm$ into Eq. 2.32 or 2.37 keeping in mind the phase shift of π between the two definitions and the one associated to **one prong + π^0s** channels. Table 2.1 summarizes the vectors chosen for each channel in the analysis presented in Chapters 6 and 7. Fig. 2.8 shows the distribution of the acoplanarity angle
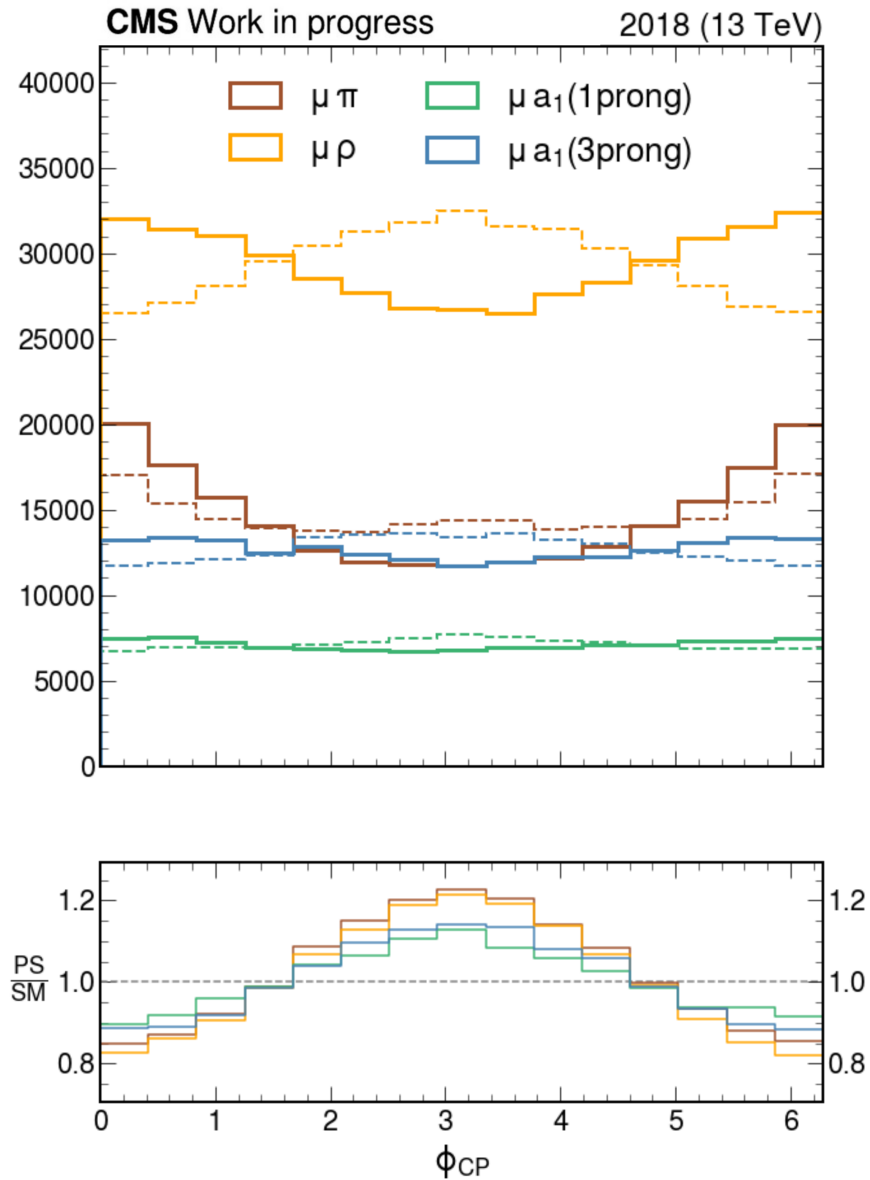
Figure 2.8: Reconstruction level distributions of the acoplanarity angle for all different decay channels in the $\tau_\mu\tau_h$ final state. The full line represents the CP-even hypothesis, while the dashed ones represent the CP-odd one. In the lower part of the plot the ratio between CP-odd (PS) and CP-even (SM) templates is shown.

for the final states involving a tau lepton decaying into a muon and another decaying hadronically. The $\tau_\mu\tau_\rho$ channel has both the highest number of collected events and the

strongest discrimination power between SM and pure pseudoscalar (PS) hypotheses. The $\tau_\mu\tau_{a_1^{1Pr}}$ and $\tau_\mu\tau_{a_1^{3Pr}}$ channels present lower sensitivity to the CP properties of the Yukawa coupling but their distributions in φ_{CP} retain a sinusoidal shape. This is not the case for the $\tau_\mu\tau_\pi$ channel: instead of having a clear sinusoidal shape, it is biased towards 0 and 2π. This is due to the uncertainty in the PV reconstruction, which is correlated between the IP used for the muon and charged pion. While distorted, the distribution can still be used effectively to separate different CP hypotheses. Sections 6.3.2 and 6.3.3 present the methods used to enhance the sensitivity of this channel.

A test was also performed to determine whether the acoplanarity angle reconstructed in data is correctly modeled in simulation. This test is presented in Appendix B.

CHAPTER 3

The Compact Muon Solenoid experiment

Contents

The analysis presented in this thesis work was performed on proton-proton (pp) collisions recorded by the Compact Muon Solenoid (CMS) [84] experiment situated at the Large Hadron Collider (LHC) [85]. The data was collected during the *Run 2* data-taking period, which comprises the years 2016, 2017 and 2018, during which the LHC provided pp collisions at a center of mass energy of $\sqrt{s} = 13$ TeV. This chapter is dedicated to describe the general conditions at which the LHC operates and the experimental setup of the CMS detector. The description of the tau lepton reconstruction in CMS is postponed to a dedicated chapter given the relevance it has for this analysis.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [85] (Fig. 3.1) is a circular accelerator with a circumference of 26,659 m built by CERN, the European Organization for Nuclear Physics.

It is installed underground in the tunnel which previously hosted the Large Electron Positron Collider (LEP) [86]. The tunnel is situated in French and Swiss territory, with a depth ranging between 50 and 175 m. The LHC has been designed to accelerate, in opposite directions, two proton beams up to a maximum energy of 7 TeV, which corresponds to a center of mass energy of 14 TeV. It can also accelerate two heavy ion bunches with a maximum energy in the center of mass of 2.76 TeV/nucleon. Given that ion collisions are outside the scope of this thesis work it is useful to concentrate on the proton-proton collisions.

Figure 3.1: LHC aerial view [87].

Proton and ion acceleration occurs in successive steps, as shown in Fig. 3.2. In particular the protons are obtained by ionizing gaseous hydrogen and accelerated by the following accelerators:

1. LINAC2 [88]: a linear accelerator, 36 m long, which accelerates the protons up to 50 MeV;

2. Proton Synchrotron Booster (PSB) [89]: circular accelerator with a circumference of 160 m, which raises the proton energy up to 1.4 GeV;

3. Proton Synchrotron (PS) [90]: 600 m long ring, which increases the beam energy up to 25 GeV

4. Super Proton Synchrotron (SPS) [91]: 7 km long circular accelerator, where proton bunches reach the energy of 450 GeV.

Protons, now grouped into *bunches*, are then injected into the LHC as two beams circulating in opposite direction. They are then accelerated up to the energy of interest.
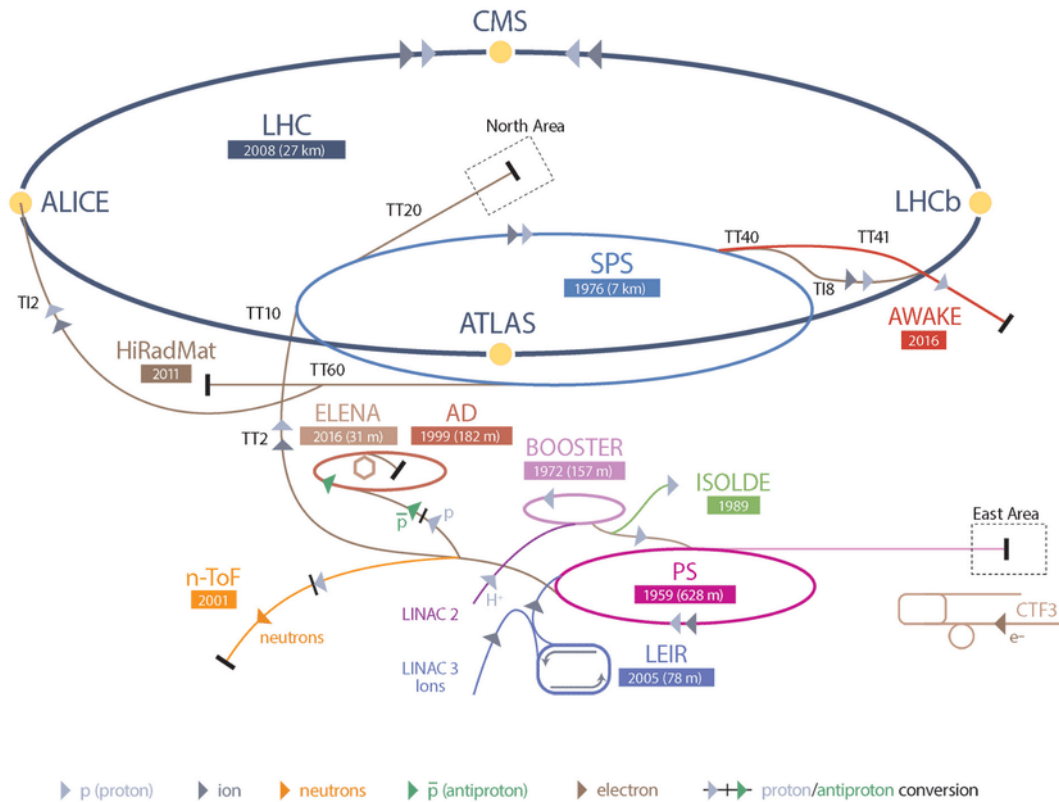

Figure 3.2: Schematic view of the LHC accelerator complex [92].

The proton bunches travel in two beams, each circulating in their own vacuum cavity. Along the LHC ring there are four points where the vacuum cavities cross each other allowing for the collision of the proton bunches. After being accelerated to the energy of interest, bunches are collided with a frequency of 40 MHz, and the products of these *bunch crossings* are recorded by the experiments hosted at each collision point:

- **CMS** (*Compact Muon Solenoid*) [84] and **ATLAS** (*A Toroidal LHC ApparatuS*) [93], respectively placed at the *collision points* 5 (P5) and 1 (P1), are two multi-purpose detectors. They are designed to study a broad range of physical processes.
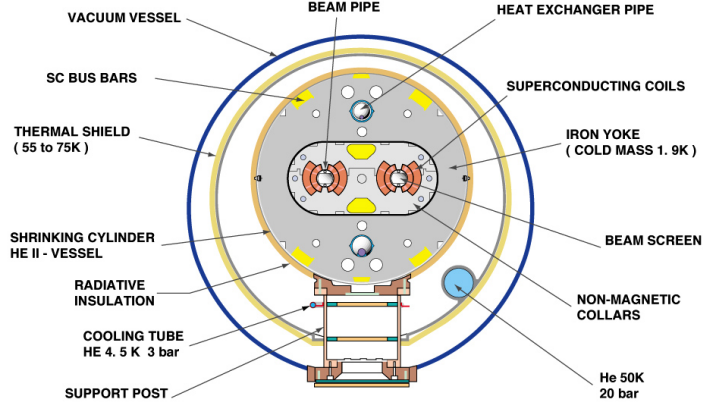
The search for the Higgs boson, culminated with its discovery in 2012, has been one of their main goals, together with the search for signals of new physics. These experiments have been designed independently, and thus can confirm or disprove the respective results.

- **LHCb** (*Large Hadron Collider beauty*) [94], located at P8, is an experiment designed for the study of flavor physics and optimized for the reconstruction of hadrons with large decay length, i.e. hadrons containing the b-quark. Unlike CMS and ATLAS it is asymmetric and covers a smaller polar angle with respect to the colliding beams. The LHCb experiment strongly contributed to the study of exotic hadronic bound states, like tetraquarks and pentaquarks.

- **ALICE** (*A Large Ion Collider Experiment*) [95], located at P2, is a detector optimized for studying heavy ions collisions. It focuses on high multiplicity events, and in particular on the study of the quark-gluon plasma, the state of matter obtained when heavy nuclei collide at high energy.

Three more experiments, are installed at the LHC: TOTEM [96], LHCf [97] and MoEDAL [98]. TOTEM stands for *TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC* and refers to a group of detectors positioned along the beam-lines spread across a length of 440 m around P5. As the name implies, it measures the total, elastic and diffractive cross-sections of proton-proton interaction. Since 2015 such measurements are carried on in combination with the CMS detector. LHCf (*LHC forward*) consists of two detectors each positioned at a distance of 140 m from P1 along the beam-lines. It studies the particles produced in proton-proton collisions and emitted very close to the beam-lines. These particles have similar kinematic conditions to the ones found in atmospheric showers initiated by cosmic rays and are used to verify the validity of the Monte Carlo (MC) simulations for cosmic rays at high energies (up to 10^{17} eV in the laboratory frame of reference). The *Monopole and Exotics Detector at the LHC* (MoEDAL) is located at P8 and focuses on the search for a magnetic monopole, i.e. a particle possessing magnetic charge, and highly ionizing Stable Massive Particles (SMPs) predicted by theories beyond the SM.

Protons are maintained in circulation through the beam pipes of the synchrotron with a magnetic system comprising over 1200 magnetic dipoles. The magnetic dipoles are able to generate a magnetic field of ∼8.33 T which is needed to bend the trajectories of the protons with momenta of the order of the TeV. To generate this magnetic field the dipoles electric coils are required to work in a state of *superconductivity*, achieved by keeping the magnets cooled to a temperature of ∼2 K. As shown in Fig. 3.3 the superconductive coils of the magnets are placed on two sides of each beam pipe obtaining a magnetic field orthogonal to the LHC ring plane. The magnetic field is directed in opposite directions in each beam pipe allowing the proton bunches to circulate in opposite directions.

CROSS SECTION OF LHC DIPOLE

VACUUM VESSEL

SC BUS BARS

THERMAL SHIELD
(55 to 75K)

SHRINKING CYLINDER
HE II - VESSEL

RADIATIVE
INSULATION

COOLING TUBE
HE 4. 5 K 3 bar

SUPPORT POST

BEAM PIPE

HEAT EXCHANGER PIPE

SUPERCONDUCTING COILS

IRON YOKE
(COLD MASS 1. 9K)

BEAM SCREEN

NON-MAGNETIC
COLLARS

He 50K
20 bar

CERN AC _HE107A_ V02/02/98

Figure 3.3: Cross section of a LHC dipole magnet [99].

Alongside the dipoles, over 8000 other magnets (quadrupoles, octupoles, etc.) are placed along the LHC ring to fine tune the beam orbit and its focusing. The beam focusing, especially in the region of the interaction points, is crucial for the collider performance. The frequency of the interactions produced by the LHC can be written as:

$$\frac{dN}{dt} = \mathscr{L}\sigma \ , \tag{3.1}$$

with σ cross section of the physical process of interest and $\mathscr{L}$ instantaneous luminosity. The latter parameter depends only on the design of the particle collider:

$$\mathscr{L} = \frac{N_p^2 N_b f_{rev} \gamma}{4\pi\beta^*\varepsilon} \times F(\varphi, d) \ , \tag{3.2}$$

with N_p number of proton per bunch, N_b number of bunches, f_{rev} frequency of revolution, γ relativistic gamma-factor for protons, β^* betatron function for the beam, which is used to estimate how much the beam is squeezed at the interaction point, ε beam emittance and $F(\varphi, d)$ geometric factor depending on the crossing angle φ and the transverse offset d between the two beams at the interaction point.

Fig. 3.4 shows the schematic view of the collision of two beams at the LHC. To maximize the number of interactions per bunch crossing the transverse size of the beams, and consequentially the β^* parameter, is reduced when approaching the interaction point. At the same time the crossing angle and transverse offset of the beam crossing are minimized in order to have proton bunches meet almost head-on. Another key parameter of the collider is the acceleration of the proton bunches. This is done by radio frequency cavities
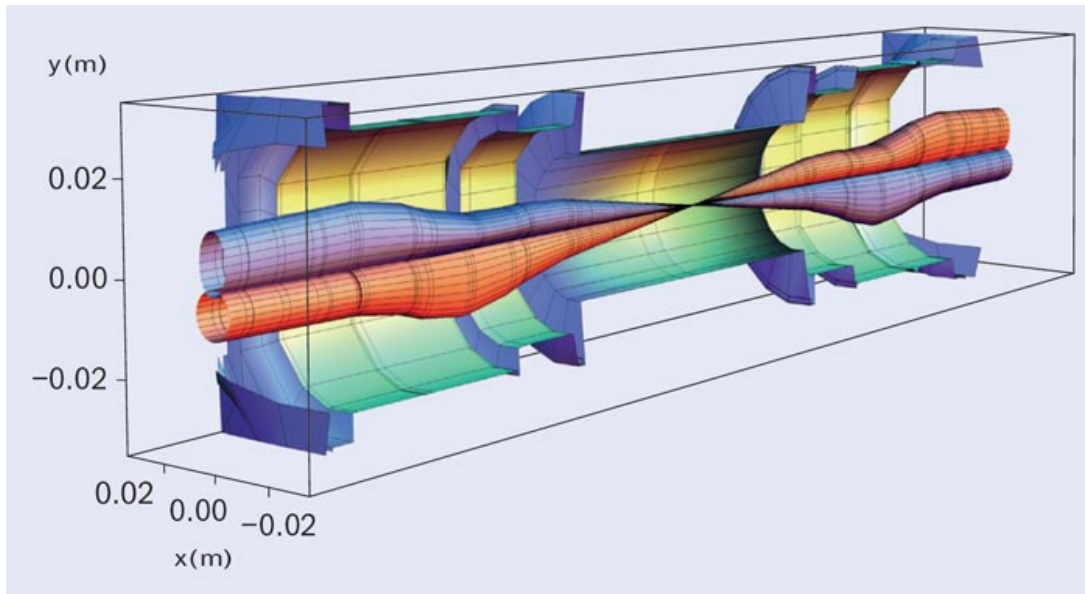
Figure 3.4: Schematic illustrating the collision of two beams at the LHC [100].

placed in straight sections of the LHC synchrotron together with the magnetic systems optimized for beam focusing. This results in the LHC being formed by alternating linear and curved sections. The interaction points, together with the beam injection and dumping systems are all located in the straight sections of the LHC ring.

3.1.1   LHC operation schedule

On March 30th 2010 the first proton-proton collision was recorded at the LHC with a center of mass energy of 7 TeV [101, 102]. This marked the start of the LHC operation schedule as a proton collider summarized in Fig. 3.5. After being operated with a steadily increasing luminosity throughout 2011, the beam energy was raised to 4 TeV, reaching a center of mass energy of 8 TeV as of April 5th 2012. Data-taking was operated continuously through the years, being interrupted only for the end of the year technical stops (EOYTS). The first data-taking period, usually referred to as *Run 1*, ended on the 16th of February 2013 and the LHC delivered a total integrated luminosity of 30 fb^{-1} to the ATLAS and CMS experiments. This first data-taking period allowed the ATLAS and CMS experiments to independently find evidence for a particle with properties consistent with the Higgs boson. During the following two years the collider was shut down to allow for the machine maintenance and repairs, this corresponds to the first *long shut-down* (LS).

Proton-proton collisions started once more on June 3rd 2015, marking the start of *Run 2*. The year 2015 also marked a milestone for the LHC, pp collisions now reached the
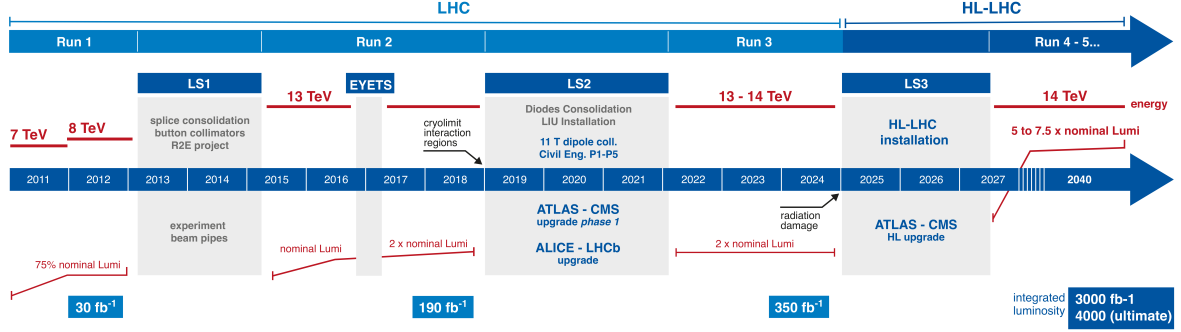
Figure 3.5: Timeline for the LHC proton-proton collision [103].

record energy of 13 TeV in the center of mass frame for the proton collisions. The beam intensity was also increased, and each beam was now made of 2240 proton bunches. Run 2 was interrupted between 2016 and 2017 for an *extended EOYTS* (EYETS) to allow the maintenance and upgrade of the LHC complex before increasing the peak instantaneous luminosity of the LHC to its nominal value: $\mathscr{L} = 10^{34}$ cm^{-2} s^{-1}. Data-taking continued till December 3rd 2018, marking the end of Run 2 and the start of the second LS (LS2).

This thesis work is based on the analysis of data collected by the CMS experiment during Run 2 [104]. During this data-taking period a total integrated luminosity of ~ 163 fb^{-1} was delivered by the LHC and the peak instantaneous luminosity reached twice the aforementioned nominal value (top plot in Fig. 3.6). The number of bunches circulated in the accelerator ring was also progressively increased throughout Run 2 as shown in the bottom plot of Fig. 3.6. In 2018 it reached the record value of 2556, approaching the LHC design value of 2800. The number of protons per bunch was instead kept roughly constant between 1.1 and 1.25 times 10^{11}. Due to the large number of protons per bunch, it is not uncommon for more than one interaction to take place in each bunch crossing. During the event reconstruction, the presence of additional interaction vertices is usually referred to as *pile-up* (PU), and is strongly affected by the beam squeezing. The latter was optimized during Run 2, as shown by the middle plot of Fig. 3.6 via the reduction of the β* parameter. In particular the major improvements in the beam squeezing were operated during the EOYTS between 2015 and 2016, and in the late part of 2017. This resulted in an increase in the number of PU interactions during Run 2, which can be observed in Fig. 3.7.

The increase in the number of PU interactions causes a *pollution* of the recorded events, as more objects are reconstructed in each bunch crossing. Several features of the CMS detector are aimed at reducing the effect of high PU and will be discussed in a later section. One of them is based on studying the high pile-up fill, a specific LHC proton injection and circulation optimized to record high PU events. Fig. 3.8 shows an event reconstructed in such a fill, with ~ 100 primary interactions being reconstructed.
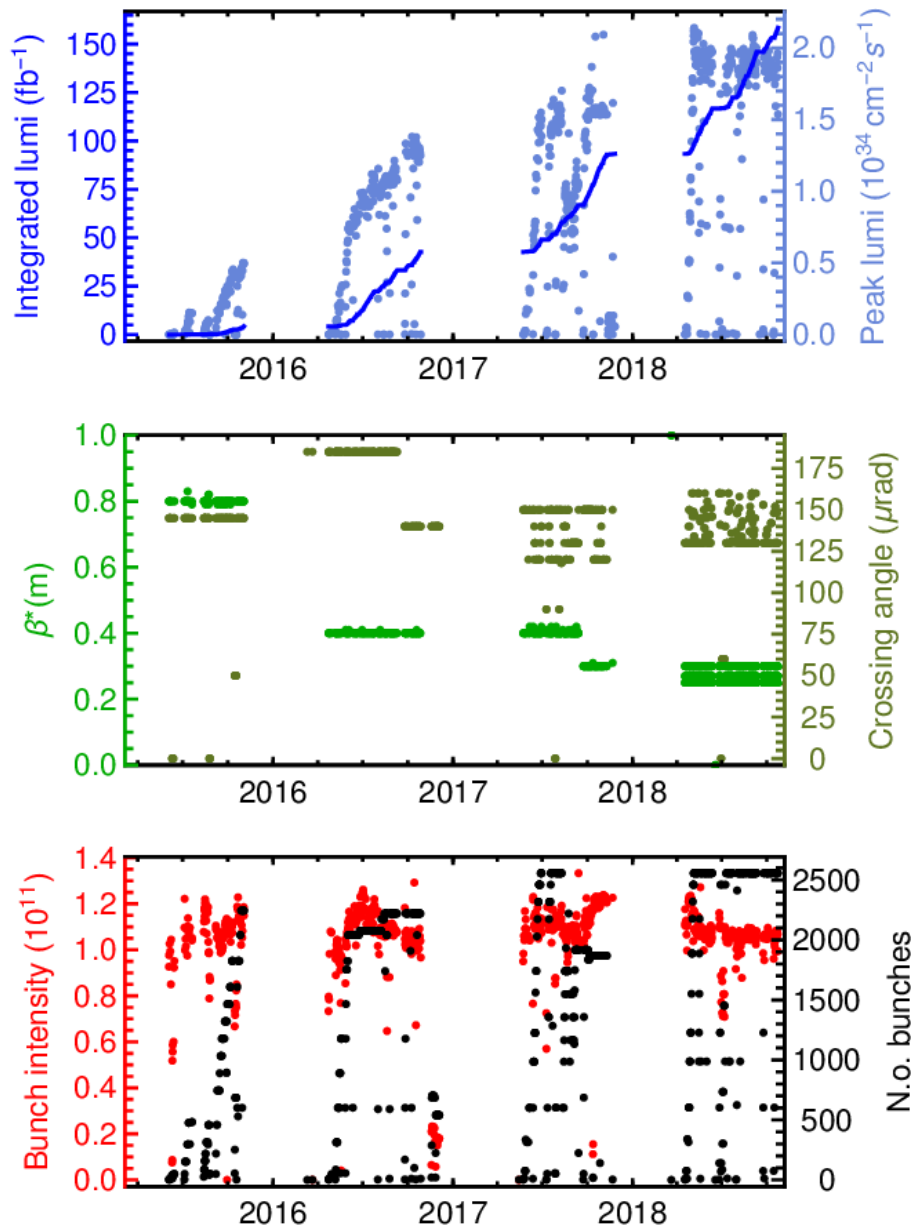
Figure 3.6: LHC operation parameters through Run 2 [104]. **Top:** Distribution of the integrated (dark blue line) and peak instantaneous (light blue dots) luminosity. **Middle:** β^* (light green) and crossing angle (dark green) distributions, defined in 3.2. **Bottom:** bunch intensity (red), i.e. number of protons per bunch, and number of bunches per beam (black).
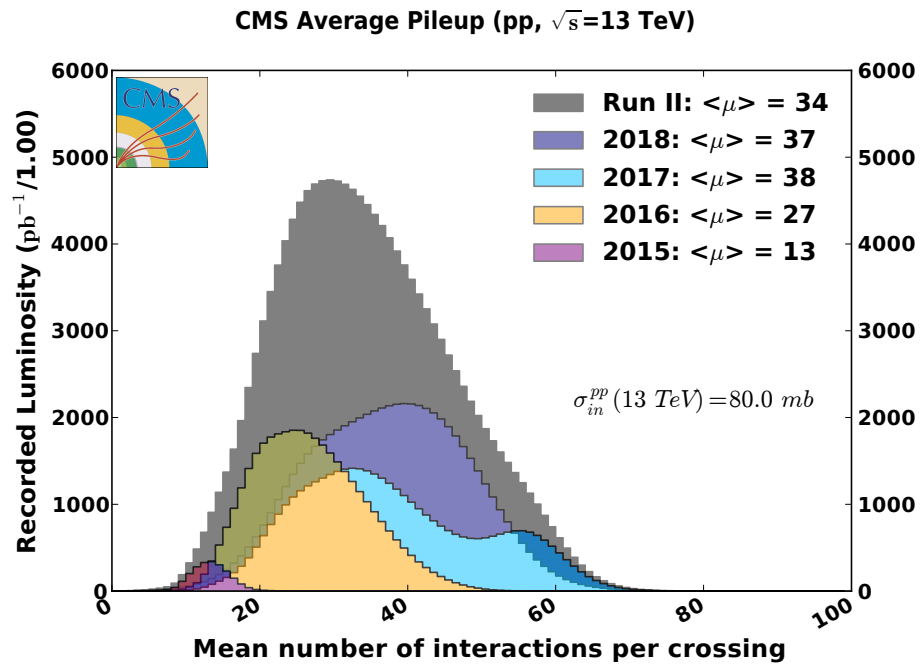
Figure 3.7: Pile-up distributions for proton-proton collisions recorded during Run 2 by the CMS experiment [105].
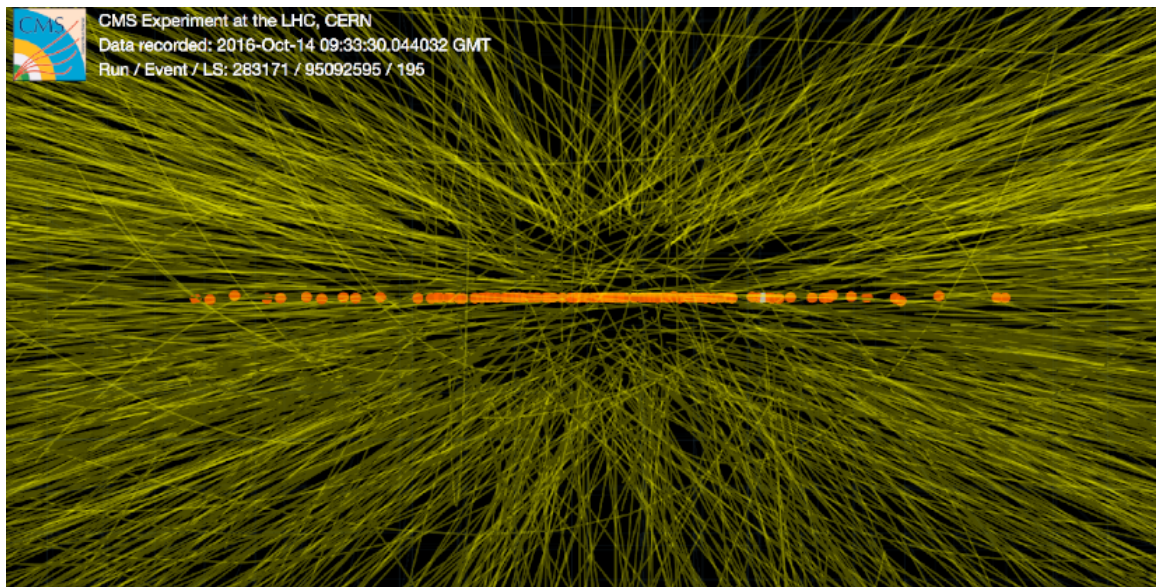

Figure 3.8: Collisions recorded on October 14th 2016 by the CMS detector during the high pile-up fill [106].

The integrated luminosity delivered by LHC and recorded by CMS is shown day-by-day in the left plot of Fig. 3.9. During Run 2 LHC delivered a total integrated luminosity of 163 fb^{-1}, of which 150 fb^{-1} were recorded by the CMS experiment. This thesis work focuses on measurements performed with tau leptons on the data collected between 2016 and 2018. The total integrated luminosity analyzed is 137 fb^{-1}, corresponding to the pp collisions where hard scattering processes have been identified.
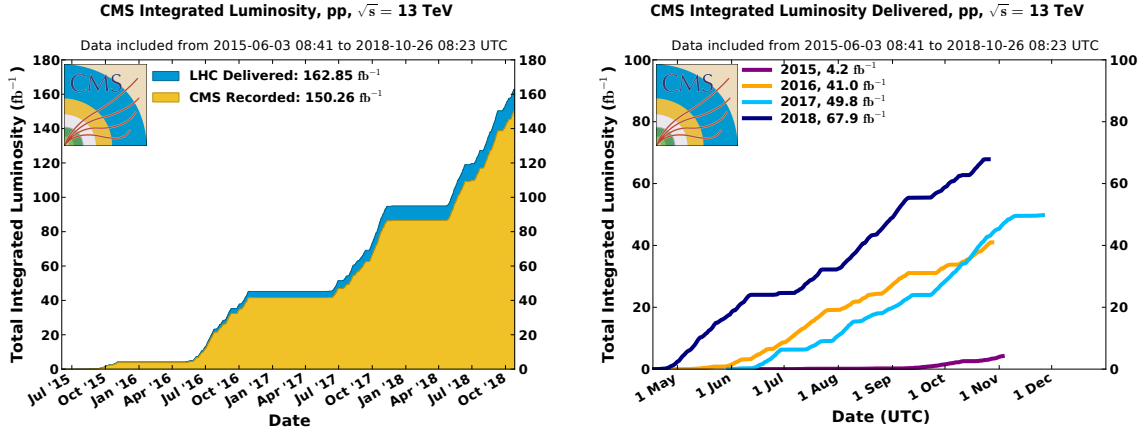

Figure 3.9: Day-by-day integrated luminosity recorded by the CMS experiment during Run 2 [105]. **Left:** the integrated luminosity is shown as a cumulative distribution across all years of Run 2 and compared to the luminosity delivered by the LHC. **Right:** integrated luminosity shown separately for each year.

After the end of 2018, the LHC entered its second long shutdown (LS2) which should end in 2021. At the moment of writing, the Run 3 data-taking period is foreseen to start in 2022. Compared to the LHC parameters of 2018 the beam energy should be raised from 6.5 to 7 TeV, while the other parameters should remain unchanged. Run 3 is foreseen to last for three years and aims to collect an integrated luminosity of over 150 fb^{-1}.

At the end of 2024 a new LS should start, leading to a high-luminosity phase for the LHC, named HL-LHC. During this phase several upgrades are planned for the accelerator complex, which should allow to increase the peak instantaneous luminosity to at least 5 times the nominal value. Several upgrades are foreseen also for the detectors [107] to account for the increase in the number of PU interactions and in the collision rate. The plan for the HL-LHC is to collect between 3000 and 4000 fb^{-1} by 2040 [103], increasing the total recorded luminosity by at least 10 times.

3.2   The CMS experiment

The Compact Muon Solenoid (CMS) is one of the two multipurpose detectors at the LHC, designed to study a wide class of physical processes. It is located at P5, near the town of Cessy, in France, and its name is based on three notable features it possesses:

- it is a *compact* detector, weighting 14,000 tonnes while being 28.7 m long and with an overall diameter of 15 m;

- it is designed to achieve an optimal *muon* detection;

- its magnetic field of 3.8 T (at the interaction point) is provided by a *solenoid* magnet, and is one of the largest superconducting solenoids ever built.

Since conception, one of its main goals has been the search for the Higgs boson. After its discovery, CMS physics has focused on the precise measurements of its characteristics as well as its couplings with other Standard Model particles. The experiment has also
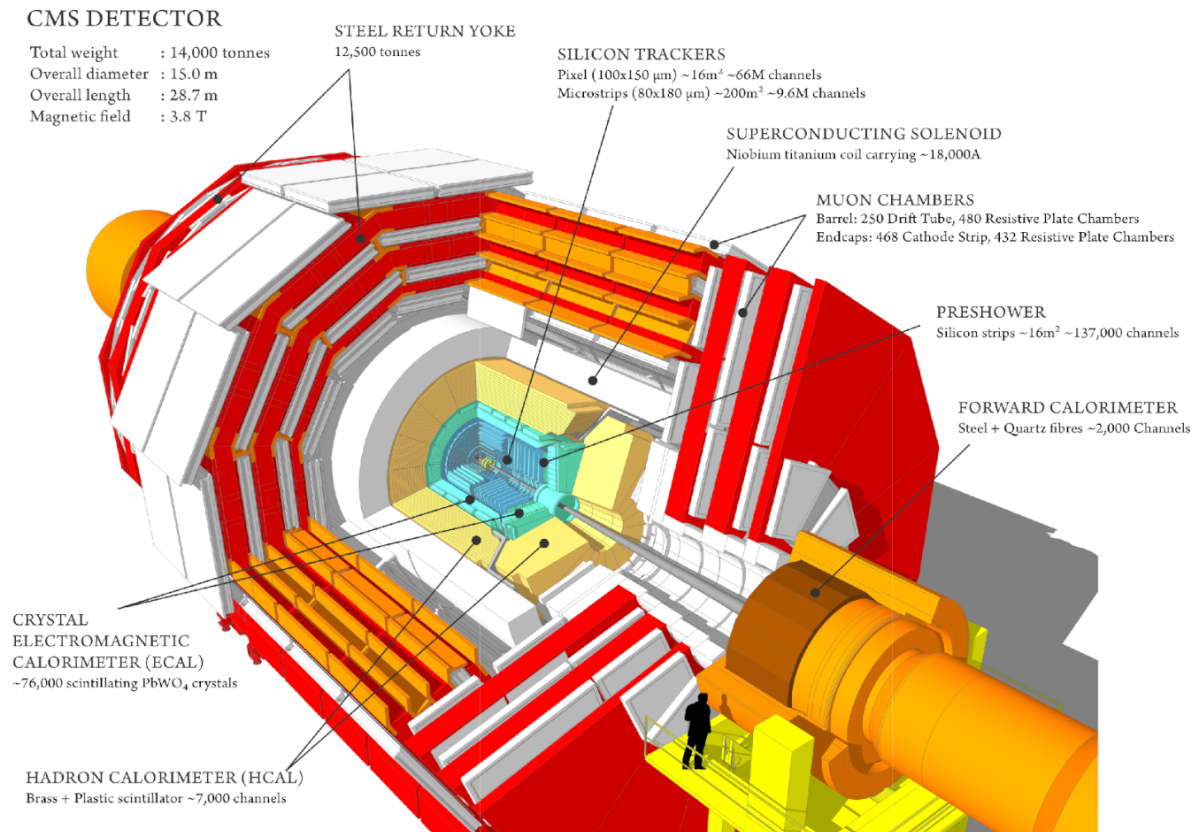

Figure 3.10: Overview of the CMS detectors [108].

measured the properties of several other processes predicted in the SM and continues its efforts in the search for signals of new physics at the TeV scale.

The compactness of the detector (shown in Fig. 3.10) emerges from the need to host its various subdetectors, aimed at particle identification and reconstruction, in an artificial cavern built at a depth of 100 m. At the core of the design philosophy is also the modularity of the detector, which allows easier access at its subdetectors for maintenance purposes, and to be able to install new components efficiently. This latter element is relevant, in particular, for what concerns the upgrades planned for the HL-LHC which will take place starting in 2024. Most subdetectors are built as part of a *wheel* structure, which can be moved inside the cavern to access its various modules. The photo in Fig. 3.11 was taken on January 30th 2019, during the LS2, by the thesis author, and shows two wheels of the detector separated from each other.

As can be seen both in the schematics in Fig. 3.10 and in the wheels shown in

Figure 3.11: View of the CMS detector during the LS2, two wheels of the detector can be observed: the one on the right is one of the endcap wheels and encased in its center are part of the barrel detectors of the tracker and calorimetric system, the wheel on the left instead is part of the barrel region and shows the muon chambers and the return yoke of the magnet.

Fig. 3.11, the detector is build with a cylindrical symmetry around the beam line. It is also symmetrical with respect to the nominal beam intersection point. Taking these symmetries into account the CMS subdetectors can be divided in two main groups:

- *barrel*: the central region of the detector and the closest to the primary interaction points;

- two *endcaps*: the regions which are placed further from the interaction point, and are used to study objects emitted at a low angle with respect to the beamline.

In order to describe the coverage and placing of the subdetectors, it is convenient to define a frame of reference. The most natural choice is to use Catesian (x,y,z) or cylindrical coordinates (r,θ,φ), due to the cylindrical symmetry of CMS. Both frames of reference are defined as centered in the nominal beam crossing point, with a z-axis directed along the beamline. The x-axis is directed radially towards the center of the LHC ring while the y-axis is directed vertically and pointing upwards. The orientation of the z-axis is then defined so as to form a right-handed tern. Given a point of Cartesian coordinates $\vec{p} = (x, y, z)$, its position in cylindrical coordinates (r,θ,φ) is then defined as:

$$r = \sqrt{x^2 + y^2}$$

$$\theta = \arccos\left(\frac{z}{\sqrt{x^2 + y^2 + z^2}}\right) , \tag{3.3}$$

$$\varphi = \arctan\left(\frac{y}{x}\right) .$$

For convenience, it is common to use the *pseudorapidity* (η) instead of the polar angle θ. The relation linking these two quantities is:

$$\eta = -\ln\tan\frac{\theta}{2} . \tag{3.4}$$

In this coordinate system objects produced at $|\eta| >> 1$ will be reconstructed in the very forward, or backward region of the detector, while objects produced in the barrel will have $|\eta| \lesssim 1$.

As a general purpose detector, CMS is required to be capable of identifying and reconstructing a wide range of particles. This is done by achieving the following objectives:

- **Identifying charged particles:** This is done by bending their trajectories using the magnetic field produced by the superconducting solenoid magnet and reconstructing the trajectories via the tracking system. This also allows to measure the momenta of charged particles.

- **Measuring the particle energy:** This is achieved by the calorimetric system which is divided in the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). These subdetectors are also used to identify neutral particles which do not interact with the tracking system.

- **Identify weakly interacting particles:** Muons, neutrinos and other weakly interacting particle are not stopped inside the calorimeters. Reconstruction of neutrinos will be discussed in a later section, while muons these are reconstructed by the muon chambers, which are placed alternated with the magnet return yoke.

3.2.1   Solenoid magnet

The superconducting solenoid magnet is a key component of the CMS detector. In order to bend the trajectory of charged particles, and allow the measurement of their momenta and charge, a strong magnetic must be generated. The magnet occupies the 2.9 m $< r <$ 3.8 m, $|\eta| < 1.5$ region, making it the largest magnet of its type ever constructed. With the exception of the muon chambers, all other detectors are hosted inside the solenoid coil. This allows the tracking and energy measurements to be performed without the particles interacting with the bulk of the magnet material. The magnet generates a magnetic field of 3.8 T in the central region of the detector, obtained by making a current of ~ 18 kA flow through the aluminum-coated NbTi superconducting cables which make up the solenoid winding. This in turn requires to maintain the whole solenoid at a temperature of 4 K by means of a liquid helium cooling system. The magnetic field lines are closed through a *yoke* of steel with a diameter of 14 m and a length of 21.6 m. The yoke weights over 10,000 tonnes. The magnetic field in the return yoke has an intensity of 1.8 T and is directed in the opposite direction with respect to the magnetic field in the central region. The layers of the return yoke are interspaced with the muon chambers, allowing a more precise reconstruction of the muon trajectory.

3.2.2   Silicon tracker

The tracking system is located in the region with $r < 1.2$ m and $|\eta| < 2.5$, making it the detector closest to the beam intersection point. It consists of concentric layers of Silicon based detectors, which can be grouped in two distinct subdetectors:

- the **pixel** detector: it is placed in the region $r < 160$ mm, allowing the reconstruction of decay vertices for particles with decay lengths of the order of centimeter;

- the **microstrip** (or more simply strip) detector: it surrounds the pixel detector and covers the space between the beam pipe and the ECAL preshower.

As a whole, the purpose of the silicon tracker is to reconstruct the trajectories of charged particles, which are bent due to the magnetic field. This allows also to identify the primary

interaction vertices by extrapolating tracks towards the nominal interaction point and determining the points from which they most likely originated. It can also reconstruct *secondary vertices*, decay vertices for particles with average decay length of the order of the centimeter, like hadrons that contain b quarks.

During the EYETS between 2016 and 2017, the pixel detector was upgraded [109] from *Phase-0*, corresponding to its design for Run 1 and 2016 data-taking, to *Phase-1*, which will be kept unchanged till the end of Run 3. Since this thesis work uses data collected across Run 2, the description of the silicon tracker is presented for the Phase-1 tracker, with a brief mention to its Phase-0 design.

The track reconstruction efficiency is driven by several effects:

- the multiple scattering of a particle with the tracker layers;

- the wrong assignment of a hit to a track;

- the alignment of the tracker modules.

The multiple scattering process causes the charged particles to lose energy in the detector layers and alters their trajectory, causing inefficiency in the track reconstruction. The amount of material traversed by particles has to be kept to a minimum to reduce this effect. Fig. 3.12 shows the material budget for the tracker system as a function of pseudorapidity. In order to reduce the ambiguity when assigning a point to a track, the tracker is required to have high granularity and redundancy of measured points per track. As shown in Fig. 3.13, a particle travelling through the detector will encounter up to 8 tracker layers, allowing for a precise track reconstruction.

The pixel vertex detector has an essential role in identifying the primary vertices of the interactions and, if present, the secondary vertices as well. It also provides the *seed*, i.e. the starting point, for the algorithm which fits a track onto the measured tracker hits. Its barrel section (BPIX) is made, after the Phase-1 upgrade, by 4 concentric cylinders at an average distance $r = 30, 68, 102, 160$ mm. Each cylindrical structure is divided in two semi-cylinders each equipped with a support structure and cooling system. The two endcap regions of the pixel detector take the name of *forward pixel detector* (FPIX) and comprise a total of 3 disk structures. The FPIX modules are assembled on half-disk support structures, each one covering the radius range between 45 and 161 mm. Each half-disk is made up of an outer ring containing 34 pixel modules, and an inner ring with 22 modules. These modules are identical to the ones installed in BPIX and each one is made up of 66,560 pixel sensors. These sensors have a size of 100×150 μm^2 and made with a substrate of type n on which $n+$ electrodes are mounted. 1184 pixel modules are installed in the BPIX while the FPIX hosts 672 modules, for a total of over 100 million pixel sensors.

During 2016 and Run 1 there was one less layer in the BPIX and one less disk in the FPIX, furthermore with the Phase-1 upgrade the distance of the innermost layer of
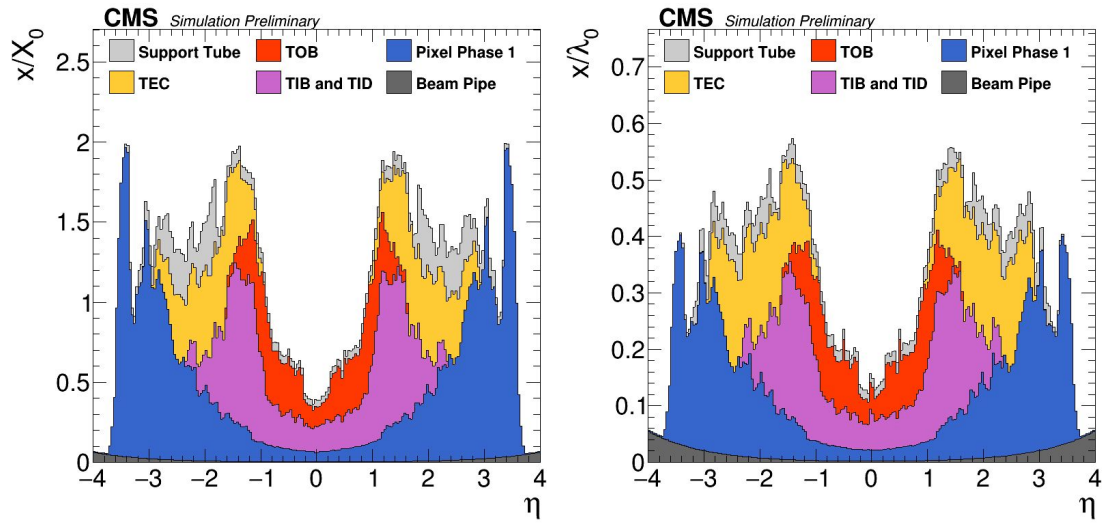
Figure 3.12: Tracker Material Budget: material budget as a function of pseudorapidity for different subdetectors in units of radiation length (**left**) and hadronic interaction length (**right**) [110].
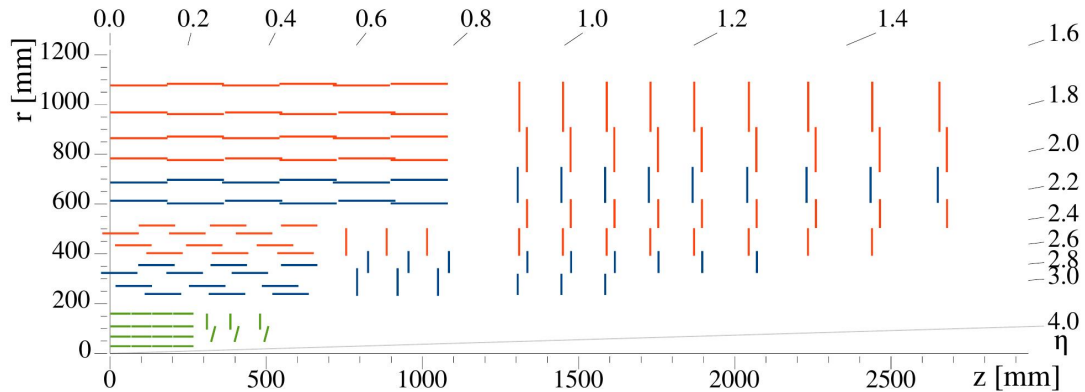

Figure 3.13: Layout of the CMS tracking system section in the r-z plane: the pixel detector is shown in green, while single-sided and double-sided strip modules are shown in red and blue respectively [111].

BPIX from the beamline has been reduced from 44 to 30 mm. A comparison between the current design of the pixel tracker and the Phase-0 design is presented in Fig. 3.14.

The resolution for the single track point identification is 10 µm in the barrel and 15 µm in the endcaps [113].

The microstrip detector is approximately 5.6 m long and has a diameter of 2.4 m. Like the pixel detector, the strip detector is formed by cylindrical elements in the barrel and disks in the endcap. It is placed around the pixel detector and is divided in four sections
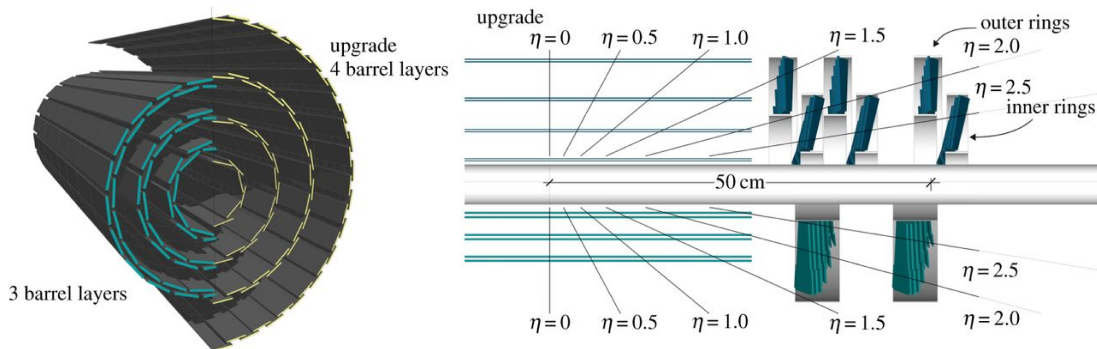
Figure 3.14: Schematic view of the Pixel detector upgrade during the EYETS between 2016 and 2017. The sections labeled "upgrade" (yellow area of figure on the left and top area of figure on the right) correspond to the Phase-1 design, and are compared to Phase-0 [112].

(Fig. 3.13):

- Tracker Inner Barrel (TIB): made of 4 cylinders and placed in the inner part of the barrel section;

- Tracker Inner Discs (TID): 3 disks placed in the inner part of the endcaps;

- Tracker Outer Barrel (TOB): formed by 6 cylinders in the outer part of the barrel;

- Tracker EndCaps (TEC): 9 external disks in the endcap sections.

The modules contained in each section include silicon sensors with a substrate of type n on which $p+$ strips are implanted. Both the pitch between the strips and the thickness of the modules vary according to their position in the detector. Modules in the inner layers are 320 μm thick and contain sensors with a strip pitched every 80 μm, while in the outer layers thickness and strip pitch increase to 500 μm and 205 μm respectively.

Another important element to consider regarding the tracker system is the alignment of its modules. Compared to the detector design the actual position, orientation and curvature of the modules is slightly shifted. These shifts are caused by several factors, from simple misalignment during module construction and installation, to detector aging effects. Track reconstruction is based on the assignment of successive hits in the tracker layers to a particle trajectory. The ideal case is shown in the left part of Fig. 3.15, the layers are positioned like the detector design, and the tracks are reconstructed by fitting a trajectory thorough the various hits. A more realistic case is shown in Fig. 3.15 on the right, where the modules position, orientation and curvature distortions are enhanced for illustration purposes.

A random misalignment of a module would lead to the degradation of the track reconstruction efficiency, but a systematic one might lead to a bias in the track reconstruction
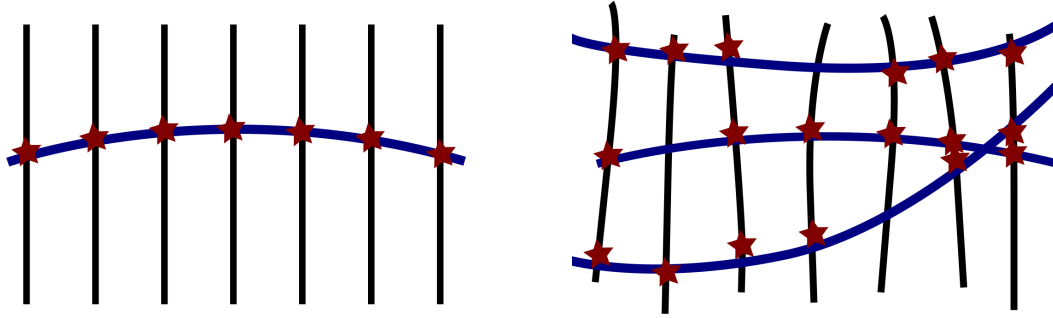
Figure 3.15: Track reconstruction in the context of alignment. The black straight lines represent the silicon modules, seen transversally; the dark blue curved line represents tracks; the red stars represent the track hits [114]. **Left:** ideal representation of the silicon modules. **Right:** more realistic case, with modules being distorted or misaligned.

and therefore in the physical measurements. In CMS a track based alignment is performed, using tracks collected from cosmic rays and pp-collisions. The alignment process is executed via a minimization of the following χ^2:

$$\chi^2(\mathbf{p}, \mathbf{q}) = \sum_{j}^{\text{tracks}} \sum_{i}^{\text{hits}} \left(\frac{m_{ij} - f_{ij}(\mathbf{p}, \mathbf{q})}{\sigma_{ij}} \right)^2 \, , \tag{3.5}$$

where $\mathbf{p}$ stands for the alignment parameters and $\mathbf{q}$ for the track parameters, m labels the measurements and f the predictions, and σ represents the uncertainties of the measurement. The minimization is operated with MILLEPEDE-II [115] and HIPPY [116] algorithms. This allows to recover a hit measurement precision of the order of the design one, showed in Table 3.1.

Tracker sub structure	Hit efficiency (μm)
pixel	9-13
strip	20-60

Table 3.1: Hit resolution for pixel and strip modules [114].

3.2.3   Calorimetry

The main function of a calorimeter is the measurement of the incoming particles energy. This is achieved by stopping the neutral and charged particle in high density material optimized for the type of particle of interest. There are two types of calorimeters in CMS:

- the **electromagnetic** calorimeter (ECAL): optimized for reconstruction and energy measurement of electrons, positrons and photons;

- the **hadronic** calorimeter (HCAL): optimized to stop and measure the energy of heavier, strongly interacting particles, like hadrons.

Muons and neutrinos do not interact within the calorimeters and their energy reconstruction is discussed in Sections 4.1.3 and 4.1.6. Particles interacting in the calorimeter are converted into a cascade of lower energy particles, which in turn lose energy by interacting with the medium of the calorimeter. Each calorimeter is characterized by an *active material*, which converts the energy lost by the incoming particles into photons, usually of ultraviolet frequency, which are in turn collected by photomultipliers (PM). The electric current produced by the photon cascade is then used to estimate the energy carried by the incoming particle. Some calorimeters are characterized by the presence of a material which causes the incoming particles to lose energy, without that being converted into scintillating light, called *absorber*. These calorimeters are classified as *sampling calorimeters*, while if they only comprise of an active material they are called *homogeneous calorimeters*.

Electromagnetic calorimeter

ECAL is a homogeneous calorimeter with a cylindrical geometry and formed by $PbWO_4$ scintillating crystals. The crystals have the shape of a truncated pyramid and are grouped into 5×5 matrices called *towers*.

Property	value
Density	8.3 g/cm^3
Radiation length	0.89 cm
Molière radius	2.2 cm
Scintillating time	15 ns

Table 3.2: Properties of $PbWO_4$ scintillating crystals [117].

The characteristics of scintillating $PbWO_4$ crystals are listed in Table 3.2. Particle showers in these crystal are contained in the transverse dimension, due to the small Molière radius, allowing for a measurement of the incoming particle direction of flight. The low scintillating time allows $\sim 80\%$ of the scintillating light to be collected within the 25 ns gap between two consecutive bunch crossings. This material is also characterised by a good radiation resistance enabling the calorimeter to operate with limited deterioration on the scale of several years. The shortcoming of this crystal is the low light production, $\sim$10 photoelectrons / MeV, which makes a light amplification system necessary. This

is achieved by using *Avalanche Photo-Diodes* (APD) in the barrel and *Vacuum Photo-Triode* (VPT) in the endcaps. Both types of photomultipliers are able to tolerate the hard radiation and strong magnetic field inside CMS. The calorimeter is subdivided into the
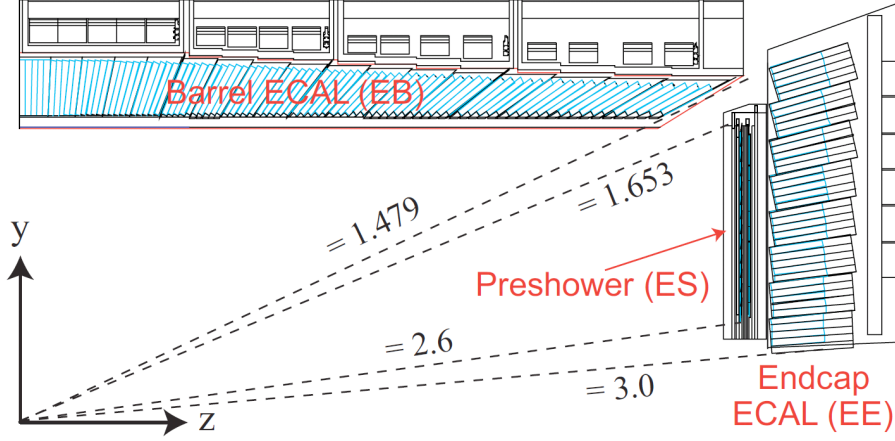

Figure 3.16: ECAL section in the $r - z$ plane [118].

following subsections (fig, 3.16):

- ECAL barrel (EB), covers the pseudorapidity region up to $|\eta| < 1.479$ and has a cylindrical symmetry;

- ECAL endcap (EE), they cover the pseudorapidity region with $1.479 < |\eta| < 2.6$;

- a preshower (ES), installed in the $|\eta| < 0.9$ and $1.65 < |\eta| < 2.61$ region.

The barrel is composed of 61200 crystals with a longitudinal dimension of 25.8 X_0, and a granularity for single crystal of $\Delta\eta \times \Delta\varphi = 0.0175 \times 0.0175$. To prevent the photons from entering the small gap between adjacent crystals and escape detection, crystals are oriented with a slightly tilted angle with respect to the nominal interaction point. Closing the barrel detectors on the sides, the EE comprise in total 7324 crystals, 24.7 X_0 long. The scintillating modules are tightly packed and the gap between them and the tracker system has been kept to a minimum. This is also done through a pre-shower detector, which further reduces the separation between detectors and helps distinguishing pions and photons. The energy resolution for a homogeneous calorimeter can be written as [119]:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} + \frac{b}{E} + c \, , \tag{3.6}$$

where: a is a stochastic term related to the statistical fluctuations of the photon collection; b is related to the electronic noise and PU effects; c dominates at high energy and takes

into account systematic effects like shower leakage and imperfections of the calorimeter. The ECAL barrel energy resolution was measured for electrons obtaining the following values: a = 2.8% GeV$^{0.5}$, b = 12% GeV and c = 0.3%.

Hadronic calorimeter

Stopping heavier, strongly interacting, particles requires materials with different characteristics. A separate calorimeter is therefore dedicated to the energy measurement for hadrons. HCAL is a sampling calorimeter that covers the region $|\eta| < 5$, and is divided into four sections (Fig. 3.17):

- Barrel Hadronic Calorimeter (HB): covers the $|\eta| < 1.4$ region and is formed by alternating layers of brass and plastic scintillator. It is formed by 2304 *towers* with a granularity of $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$. The energy reading is performed with *Hybrid Photo-Detectors* (HPDs).

- Endcap Hadronic Calorimeter (HE): covers the region with $1.3 < |\eta| < 3$, providing an overlap with HB. It comprises 2304 calorimeter towers, made of the same materials as the ones in HB. As of 2018 the HPDs used for energy reading have been replaced by *Silicon Photomultipliers* (SiPMs).

- Outer Hadronic Calorimeter (HO): placed between the solenoid magnet and the muon chambers, it provides an enhancement of the shower longitudinal containment. It is made up of plastic scintillators with a granularity of $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$. The role of absorber is instead taken by layers of the magnet return yoke. The scintillating light is collected by *wavelength shifter fibers* (WLF) and transported to SiPMs placed on the side of the return yoke.

- Forward Hadronic Calorimeter (HF): it is placed in the region beyond the endcaps and covers the region up to $|\eta| < 5$. Plastic scintillators are not suited for the high level of radiation in the forward region of the detector. The active material used are quartz fibers, where incoming particles release energy in the form of Čerenkov radiation, which is collected by *Photomultiplier Tubes* (PMT), while iron is used as absorbing material.

The energy resolution for the hadronic calorimeters in the barrel region and in the endcaps is:

$$\frac{\sigma_E}{E} = \frac{0.9}{\sqrt{E(\text{GeV})}} + 0.045 \ , \tag{3.7}$$

while for HF the energy resolution is:

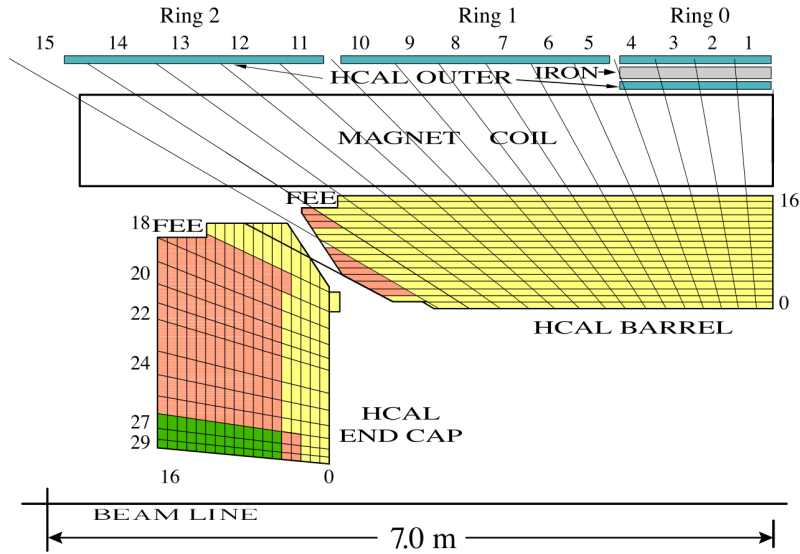$$\frac{\sigma_E}{E} = \frac{1.72}{\sqrt{E(\text{GeV})}} + 0.09 \ . \tag{3.8}$$

Figure 3.17: HCAL section in the $r - z$ plane [120].

3.2.4 Muon chambers

Muons have an average lifetime of 2.2 μs [23], their decay length is therefore much larger than the size of the CMS detector. They interact weakly with the calorimeters and can escape the detector without being stopped. Their detection, and momentum measurement are operated by the tracking system in conjunction with the muon detection system. This system, referred to as muon chambers, is placed beyond the magnet coil and is affected by the 1.8 T return magnetic field that flows through the return yoke. As shown in Fig. 3.18, muon chambers can be divided in three types:

- Drift Tubes (DT): detectors located in the $|\eta| < 1.2$ region. This subdetector is organized in four cylindrical concentric stations centered around the beam direction. The three innermost cylinders comprise 60 drift chambers, while the outermost layer is made out of 70. Drift chambers are filled with a gas mixture containing Ar (85%) and CO_2 (15%). The DTs spatial resolution is 100 μm.

- Cathode Strip Chambers (CSC): cover the region between $0.8 < |\eta| < 2.4$. In the endcaps region the magnetic field is stronger and more irregular compared to the barrel region. The radiation exposure is also higher, making DT suboptimal for the detection. CSC are used instead of the DT, they are multiwire proportional chambers in which the cathode is segmented into radial and transverse strips with respect to the wires serving as anodes. The gas used is a mixture of Ar(40%), CO_2 (50%) and CF_4 (10%).

- Resistive Plate Chambers (RPC): they are gas detector with moderate spatial resolution but higher time resolution compared to the other detectors. They are placed behind DT and CSC detectors and used in the trigger system.
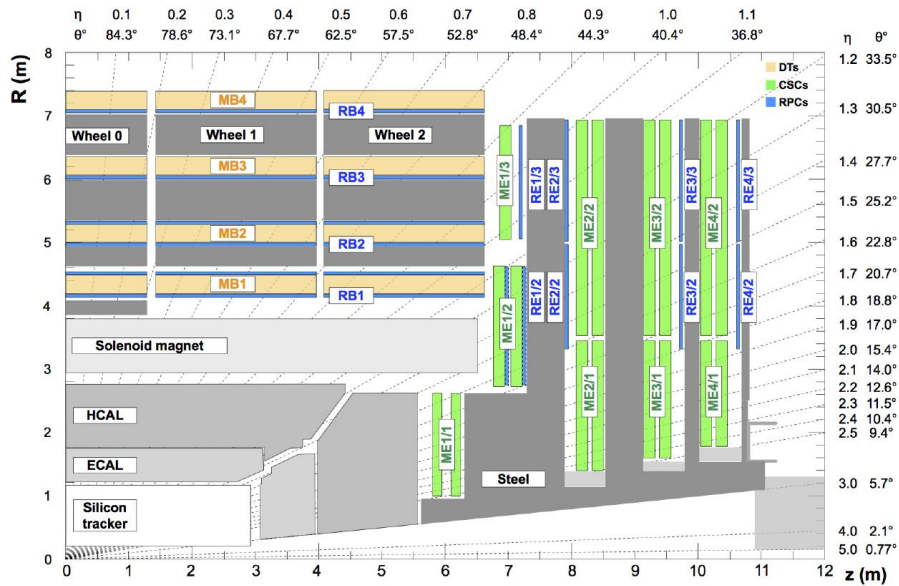

Figure 3.18: Schematic view in the r-z plane of the muon detection system [121].

3.2.5 Trigger system

At the LHC collision points proton bunches cross once every 25 ns and for each bunch crossing there can be over 50 primary interactions, with potentially hundreds of particles produced per collision. Storing all the data produced by each subdetector for each bunch crossing is nearly impossible taking into account both the amount of data and the rate at which it should be stored. This presents a challenge for the data storage at CMS. Of the proton-proton collisions most can be classified as *almost elastic* or *diffractive* events, which correspond to interactions where the colliding protons either recoil from each other without being destroyed in the process, or the proton structure is broken leading to the emission of highly forward jets. These events present low transverse transferred momentum and therefore are of minor interest when looking at processes involving exchange of on-shell gauge boson, or Higgs physics. Only a fraction of these *minimum bias* events are written on disk, and are mainly used for detector calibration or luminosity measurements.

To reconstruct a statistically significant number of events with signatures of interest a more elaborate system is required. This system is named *trigger*, as it stores data based on specific inputs received by some of the CMS detectors [122]. The trigger system is divided into two parts: *Level 1 Trigger* (L1) and *High Level Trigger* (HLT).

The L1 is a hardware based trigger which is needed to reduce the flow of information from 40 MHz to 100 kHz. Event selection is based only on inputs from calorimeters and muon chambers and has the task of identifying whether in an event a high energy electron, muon, photon or jet has been identified.

The HLT is designed to reduce the *output rate* of L1 to about 800 events/s. Events that pass the HLT are written to disk and stored in the CMS computing center at CERN, called *Tier 0*. HLT corresponds to a software level selection, which uses all the information coming from the subdetectors reducing the rate of minimum bias events while prioritising high transverse momenta objects. To process the events effectively, the HLT must have a good rejection rate of minimum bias events while keeping a good efficiency in the selection of other, rarer, phenomena. This is also required to be done in a limited amount of time, for this purpose the HLT code is organized on multiple levels:

- *Level 2 trigger* uses the full information gathered from calorimeters and muon chambers;

- *Level 2.5 trigger* adds to the algorithm the information from the pixel detector;

- *Level 3 trigger* uses the data collected by all subdetectors.

The events which pass the HLT selection are then saved on mass storage and become available for offline data analysis.

3.2.6   Data processing

The raw data from the detectors are processed and analyzed to reconstruct physical events. This requires to combine the information gathered by each subdetector in order to identify the particles produced in each pp collision (event), and reconstruct and store their properties.

In CMS, events are stored in datasets which are processed in multiple successive steps. Starting from raw data (RAW), these are combined in order to have higher level information: for example separate hits in the tracker are combined in order to reconstruct tracks and energy deposits in the calorimeters are combined in the form of clusters. These reconstructed objects form the RECO data tier, muons, electrons and jets are also reconstructed at this stage. The Analysis Object Data (AOD) is a subset of the RECO dataset, obtained by keeping only the information on higher level reconstructed objects, like track, vertices, muons, jets etc.

Most analyses use further subsets of the AOD datasets, called miniAOD and nanoAOD, which store only a fraction of the reconstructed objects. This thesis work uses miniAODs, where the data stored for each event is kept to less than 100 kB.

4

CHAPTER

Tau reconstruction in CMS

Contents

Tau leptons coming from Higgs and Z boson decays have a typical energy of the order of tens of GeV. They are unstable particles with a decay length of the order of the millimetre (see Eq. 1.30) in that energy range. The innermost layer of the CMS detector is placed at 3 cm from the beamline (see Section 3.2.2), and the probability for a 30 GeV tau lepton to travel 20 times its decay length without decaying is less than 10^{-5}. This means that a tau lepton originating from a primary interaction vertex, i.e. not produced by the decay of a long-lived particle, will produce no track hits in the pixel or strip detectors. The reconstruction of these *prompt* tau leptons must be operated via their decay products.

To describe the tau reconstruction algorithm in CMS it is important to first describe the general reconstruction algorithm for objects such as muons, electrons and jets.

4.1   Object reconstruction in CMS

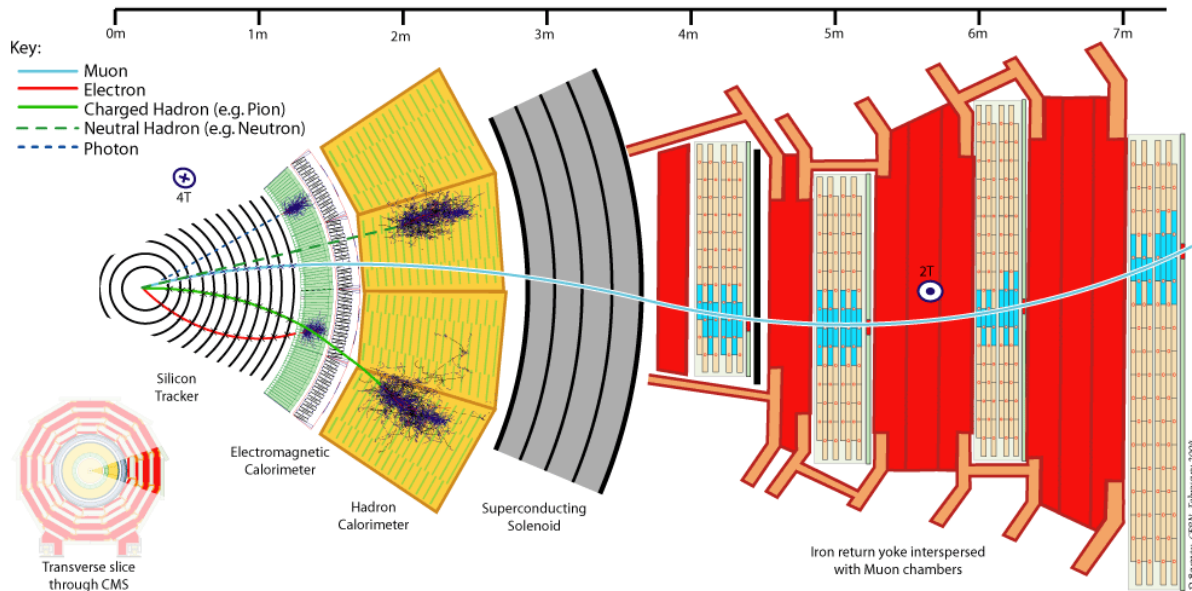## 4.1.1   The particle flow algorithm

Figure 4.1: Schematic representation of the CMS detectors with different particles traveling through the subdetectors [123].

The physics object reconstruction in CMS is accomplished using an algorithm called *Particle Flow* (PF) [124]. It uses the whole information gathered by the subdetectors to reconstruct energy, momentum and trajectory of each stable particle.

First, the PF algorithm identifies the quantities measured by each subdetector, like charged particles tracks in the silicon tracker, energy clusters in calorimeters or muon tracks in the outer section of the detector. The algorithm groups these signatures into *blocks* according to whether they could be associated to the same particle. As an example, a charged particle track pointing to an energy cluster in a section of the electromagnetic calorimeter could be associated to an electron or positron candidate. Once these initial blocks are constructed the algorithm proceeds to identify the particles in the following order:

1. **Muons:** they are identified using the hits in the muon chambers and in the silicon tracking system, and by ECAL and HCAL clusters compatible with minimum ionizing particle (MIP) signatures. The requirements for a track to be assigned to a muon is that at least a hit in a muon chamber was found. After the track hits are assigned to a muon, they are removed from other blocks.

2. **Electrons**[1]**:** the algorithm tries to pair together tracks in the silicon tracker with energy clusters in ECAL and HCAL. The matching tracks and energy deposits are removed before proceeding to the next step.

3. **Charged hadrons:** the remaining tracks in the tracker are associated with this type of particles. The tracks are matched to energy clusters in ECAL and HCAL and then removed from the list of objects.

4. **Neutral hadrons and photons:** energy deposits in HCAL, which have not been matched before, are marked as neutral hadrons, while those in ECAL are assigned to photons.

The last step of the PF algorithm involves the measurement of the total transverse energy of the event. In a pp collision the total momenta in the transverse plane must sum up to 0. When this is not verified it means that some energy in the transverse plane is missing in the event. This *missing transverse energy* (MET) could be linked to inefficiencies in the detector or to particles which travelled through the subdetectors without interacting, like the neutrinos. In this thesis work two different algorithms for the MET reconstruction have been used and are discussed in Section 4.1.6. On top of the particle identification, other algorithms are used to improve the reconstruction of certain particles or their properties. The following sections are dedicated to provide more details on the reconstruction of specific objects like muons or jets.

4.1.2   Tracks and vertices

Tracks of charged particles in CMS are reconstructed using hits in successive layers of the tracking system, with the exception of muon tracks which use hits also in the muon chambers. The reconstructed tracks are then used to identify the primary partonic interaction vertices in the various pp collisions, and the secondary vertices associated to decays of long-lived particles. The algorithms used in CMS for track reconstruction [125] is based on a Kalman Filter (KF) algorithm [126, 127]. The algorithm proceeds by approximation, assigning hits and a tentative momenta to the track after adding each layer:

1. A hit in the first layer is used as seed for the algorithm; the first track candidate is constructed as a straight line pointing outward, with a very large uncertainty.

2. On the next layer the KF hit associated to the track is taken as compromise between:

   a) the hit determined experimentally;

---

[1]In the context of object reconstruction in a pp collision the only difference between electrons and positron is their electric charge. As such it is common to use the word electron to refer both to the particle and its corresponding antiparticle.

  b) the hit predicted by extrapolating the track candidate based on the hits in the
     previous layers.

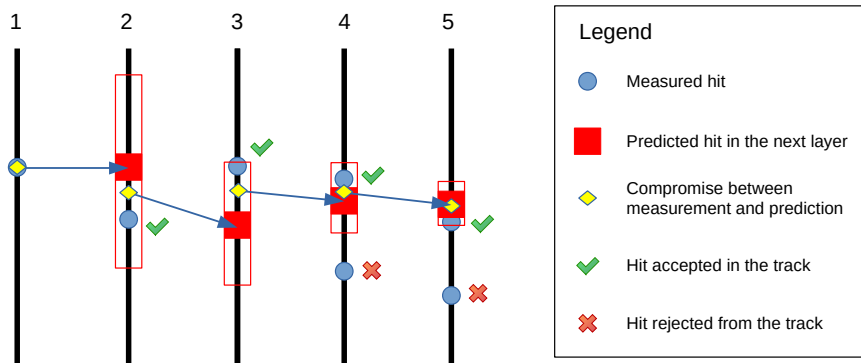3. The parameters of the track candidate are updated with the addition of every new
   hit.

Figure 4.2: Schematic depiction of a Kalman Filter algorithm: starting from the measured
point (blue circles) on layer 1 the KF algorithm predicts the position of the expected hit (red
square) onto the next layer. The compromise between the expected and measured hit (yellow
diamond) is then used to re-parametrize the track and predict the successive hit. A measured
hit is added to the track depending on whether it is found within the search window around the
predicted hit (red empty rectangle). The search window becomes smaller with each hit added
to the track.

The KF algorithm is also illustrated in Fig. 4.2. The complexity of the algorithm is
increased to account for several phenomena that can occur. These include:

- the presence of missing hits in the detector layers;

- the possibility of tracks crossing each other;

- the presence of displaced vertices, from which multiple tracks can emerge.

This is done by following an iterative approach were the hits of the best reconstructed
tracks are removed at each step, and the remaining hits are combined into new tracks [125].
   After the track candidates have been identified, they are used to reconstruct primary
and secondary interaction vertices. The beam spot (BS) position is estimated as an
average over multiple consecutive bunch crossings, as it depends primarily on the beam

focusing and is therefore approximately stable throughout an LHC fill. The BS is defined as the center of the region where the LHC beams collide at P5. It is reconstructed as a 3-D point using a combined fit [128] of over 1000 tracks in order to achieve a precision of $\mathscr{O}(\mu m)$ in the x and y coordinates. Its z coordinate is less stable across multiple bunch crossings due to the extension along the z direction of the bunches. Each track is parametrized as a helix with the following parameters:

- C : signed[2] curvature of the track;

- φ_0 : azymuthal angle of the momentum associated to the trajectory, taken at the point of closest approach (PCA) to the nominal interaction point;

- d_0 and z_p: respectively distance in the transverse plane and z direction of the PCA and origin;

- $\cot\theta$: cotangent of the polar angle of the track momentum at PCA.

The fit then is performed using the $d_0 - \varphi_0$ correlation:

$$d_0(z_p, \varphi_0) = x_0 \cdot \sin\varphi_0 + \frac{\mathrm{d}x}{\mathrm{d}z} \cdot \sin\varphi_0 \cdot z_p - y_0 \cdot \cos\varphi_0 + \frac{\mathrm{d}y}{\mathrm{d}z} \cdot \cos\varphi_0 \cdot z_p \ , \qquad (4.1)$$

where x_0 and y_0 are the x and y coordinates of the BS in the $z = 0$ plane, and $\mathrm{d}x/\mathrm{d}z$ and $\mathrm{d}y/\mathrm{d}z$ are the x and y slopes of the beam. Another measurement of the BS is operated averaging the position of the primary vertices and the final estimation of the BS is done by combining the two measurements.

In order to estimate the size of the BS reconstructed in CMS, its resolution is shown in Fig. 4.3 for the first part of the 2018 data-taking period, labeled Run2018A. The plots show the resolution as a function of the LHC fills, which are the segments of data-taking between consecutive beam injections in the LHC. During a fill the bunches become less collimated as can be observed by the increase in the uncertainty on the BS coordinates. The uncertainty over the BS coordinates can be interpreted as an estimation of the size of the region where the bunches interact. This region has an average size along the x and y coordinates of 8 μm × 7 μm, while it is larger along the z axis, with an average size between 3 and 3.5 cm. The dimension along the z axis is also the less stable during an LHC fill, increasing by over 15%.

Interaction vertices in CMS are identified with the Adaptive Vertex Reconstruction (AVR) [125]. When applying this strategy to reconstruct primary interaction vertices (PVs) the basic mechanism can be described as:

1. select tracks associated to prompt objects;

2. cluster the tracks which originate from the same region;

---

[2]The sign is taken as the charge of the corresponding charged particle.
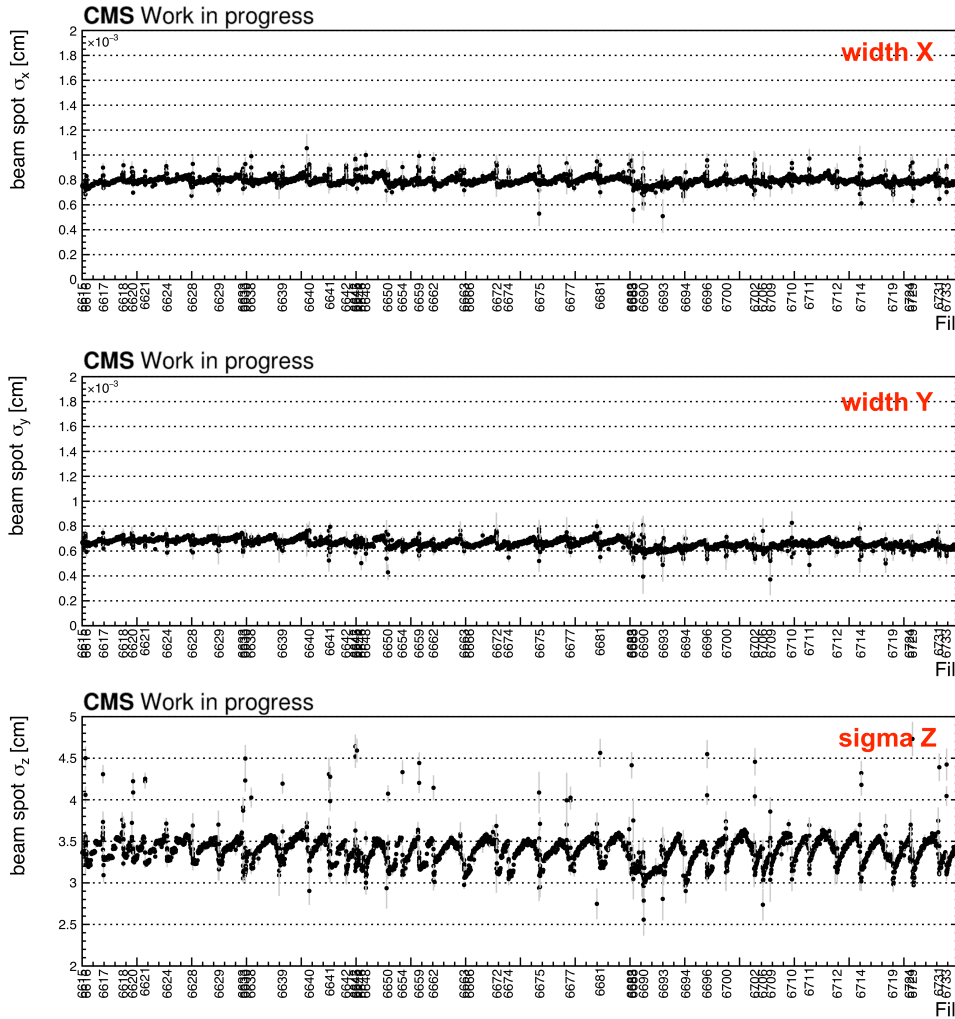
Figure 4.3: Beam spot resolution using the first part of the 2018 data-taking in CMS. The resolution on the x (**top**), y (**middle**) and z (**bottom**) coordinates of the BS position are shown with respect to the LHC fills [129].

3. fit the vertex position based on the cluster of tracks selected;

4. assign a weight to the tracks based on how they are compatible with the fitted vertex.

The selection of prompt tracks is based on track fit quality, the presence of hits in the first two PIX layers, and the impact parameter (IP) significance, defined as:

$$IP_{sig} = \frac{d_0}{\sigma_{IP}} \; , \tag{4.2}$$

with σ_{IP} uncertainty on the IP and d_0 being the aforementioned transverse component of the IP, taken with respect to the estimated BS position. The track clustering is performed with a Deterministic Annealing (DA) technique [130]. Vertex candidates selected by the DA algorithm, which are associated to at least two tracks, are then processed by an Adaptive Vertex Fitter (AVF) [131]. This robust KF-based algorithm provides the estimate for the primary vertices coordinates, their correlation matrices, and assigns a weight to each track associated to the PV. The closer the weight is to 1, the more the corresponding track is compatible with originating from the PV. Secondary vertices (SV) associated to long lived particles are generally computed with variations on the KF algorithm. In the case of tau leptons, the SV is computed in the 3 prong channel, by fitting the tracks of the three charged pions in the final states. More details on this and other techniques to measure the tau decay vertex are shown in a later section.

4.1.3 Muons

Muon tracks are reconstructed independently in the various subdetectors [132,133]. A track reconstructed in a single muon chamber is assigned to a *local muon*. The combination of multiple local muon tracks allows to construct *stand-alone muons*, and is based on a Kalman Filter algorithm [126,127], analogous to the one used for the track fitting in the inner tracker. Muon tracks in the silicon tracker are reconstructed independently of the muon chambers leading to two possible definitions for the reconstructed muons:

- *tracker muons*: they are reconstructed inside-out, with a track in the silicon tracker being extrapolated to the muon chambers and associated to a stand-alone track. This reconstruction is optimized for lower transverse momenta (p_T) muons which cannot be reliably reconstructed with only muon chambers only.

- *global muons*: they are reconstructed outside-in, a well reconstructed stand-alone track is associated to a track reconstructed in the inner tracker.

The CMS detector is optimized for muon reconstruction and achieves an efficiency of muon reconstruction as either tracker or global muons greater than 99% [133]. In several cases, a global muon is compatible with a tracker muon candidate and the two objects are merged in the reconstruction. It should be apparent that for a single recorded muon several definition of track can be put in place. This definition influences an important parameter used for the measurement of the Higgs CP properties: the impact parameter of muons. This vector is defined based on the point of closest approach (PCA) of the muon track to the primary vertex (PV). Given the importance of reconstructing the muon tracks precisely in the region closer to the beamline, the muons used for this analysis are treated as tracker muons, in the sense that their IP is reconstructed using only hits in the tracker system. More details on the IP reconstruction is postponed to Section 6.3.3.

The muon identification (ID) is applied on top of the categorization of their tracks. Its objective is to classify muon candidates based on the quality of their reconstructed trajectories and identify specific topologies, like muons originating from the decay of long lived particles. The muon ID proceeds by applying additional selections based on variables associated to the muon reconstruction, like the number of hits in the inner tracker and in the muon chambers, the degree of compatibility between hits in the inner tracker and standalone muons, and the χ^2 of the track fit quality. Other requirements used to identify muons are related to global variables, i.e. quantities that are related not only to the muon track but also to other particles in the reconstructed event. The relative isolation is one of these variables, it quantifies the amount of particles reconstructed close to the muon in the $\eta - \varphi$ plane, and can be used to identify muons originating from the decay of baryons produced within a jet. The muon relative isolation is defined as [132, 133]:

$$ I_{rel}^{\mu} = \frac{\sum E_T(\text{charged}) + \max\left(\sum E_T(\text{neutral}) - 0.5 \sum E_T(\text{charged,PU}), 0\right)}{p_T^{\mu}} \ , \qquad (4.3) $$

where p_T^{μ} is the muon transverse momentum, and E_T is the transverse energy of the particles contained in a ΔR cone of radius 0.4 around the muon direction of flight. In parentheses the type of particles is specified: *charged* for charged hadrons, *neutral* for neutral hadrons and photons, and *charged, PU* for the charged hadrons originated by pile-up (PU) interactions. The term $\sum E_T(\text{neutral}) - 0.5 \sum E_T(\text{charged,PU})$ is used to correct the energy associated to the neutral particles found around the muon based on the energy of charged particles originated by PU vertices. This correction is applied since charged particles traversing the tracker system can produce photons via bremsstrahlung radiation. This in turn allows to more accurately compute the muon isolation from objects originating from the same PV. This correction can be overestimated, resulting in the term $\sum E_T(\text{neutral}) - 0.5 \sum E_T(\text{charged,PU})$ being negative. To avoid this scenario, the corrected energy from neutral particles is set to never be lower than 0 with the use of the max() function.

Several working points (WPs) are defined in CMS:

- Loose muon ID: it labels muons selected by the PF algorithm that are also either a tracker or a global muon.

- Medium muon ID: compared to the loose WP, it additionally requires that the tracker segment of the muon track uses hits from more than 80% of the inner tracker layers it traverses.

- Tight muon ID: it is a muon which passed the medium WP and whose inner track is obtained by at least six layers of the inner tracker, one of which needs to be from the pixel detector. This WP in particular aims at reconstructing prompt muons, as it rejects muons with $d_0 > 0.2$ cm or $d_z > 0.5$ cm, with d_0 and d_z transverse and longitudinal components of the muon IP respectively.

Other WPs are defined for muons related to specific topologies, like the one targeting muons coming from long lived particles like b-hadrons which would be excluded by the Tight WP. In the selection of muons for the measurement of the CP properties of the Higgs boson in $H \to \tau\tau$ decays, the Medium WP was used. The efficiency for this WP with respect to the muon p_T and η is shown in Fig. 4.4 using 2017 data and simulation. It was calculated with respect to the reconstructed tracker muons with $p_T > 20$ GeV. The efficiency drops around $|\eta| \sim 0.2$ corresponding to the transition between the wheels 0 and 1 of the muon detectors, previously shown in Fig. 3.18.
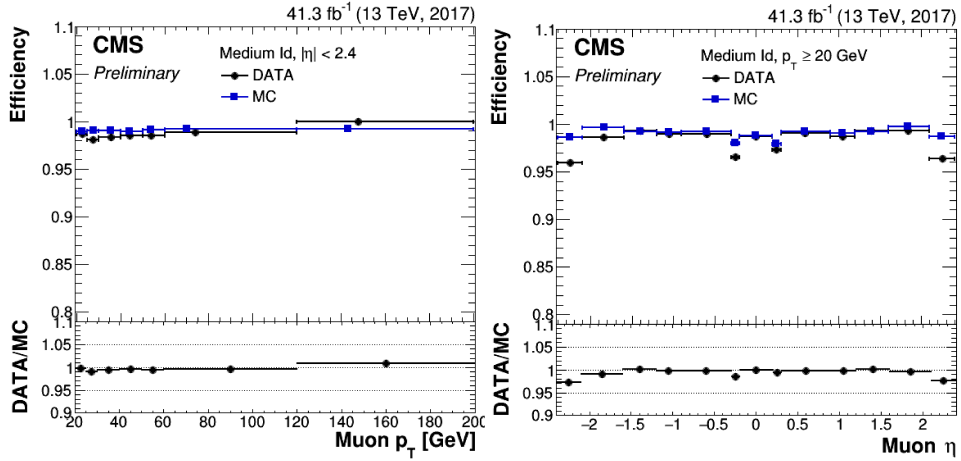

Figure 4.4: Efficiency for the muon Medium ID WP with respect to the muon transverse momentum (**left**) and pseudorapidity (**right**) using 2017 data and $Z \to \mu\mu$ events [134, 135]. The lower part of the plots shows the ratio between the Muon ID efficiency in data and the one in MC. The efficiency is calculated with respect to the tracks identified as tracker muons by the PF algorithm having $p_T > 20$ GeV.

4.1.4   Electrons and photons

Electron reconstruction is intrinsically connected to the reconstruction of photons and vice versa. Electrons lose energy while traversing the CMS tracker system due to bremsstrahlung radiation. To correctly reconstruct the initial energy of an electron it is therefore essential to collect the energy of the radiated photons. The energy spreads mainly along the φ direction because of the bending of the electron trajectory in the magnetic field. Conversely high energy photons can convert into an electron-positron pair producing two electromagnetic showers also extended along the φ direction. This means that by the time electrons and photons reach the ECAL they might be already converted into a shower of electrons and photons. To reconstruct the total energy of the primary electron or photon a clustering algorithm is used. An initial clustering is performed to find groups of crystals, where the collected energy is above a certain threshold. These

thresholds are determined via calibration and based on the electronic noise of the corresponding detectors. Typical values are ~ 80 MeV (~ 300 MeV) for EB (EE) [136]. The cluster which collected the highest energy in a certain $\eta \times \varphi$ region, is taken as a *seed* for the algorithm. Clusters in the proximity of the seed are merged together to form a *super-cluster* (SC). The super-cluster reconstruction is a combination of two algorithms:

- *mustache* algorithm: it uses information from ECAL and the preshower to group clusters spreading around the seed cluster. The name comes from the typical shape of the resulting SC (Fig. 4.5). It is optimized to find low energy clusters shifted in the φ direction with respect to the seed-cluster.

- *refined* algorithm: combines ECAL clusters with tracks reconstructed in the inner tracker. It looks for signature of early bremsstrahlung and pair production to find clusters shifted with respect to the mustache SCs.
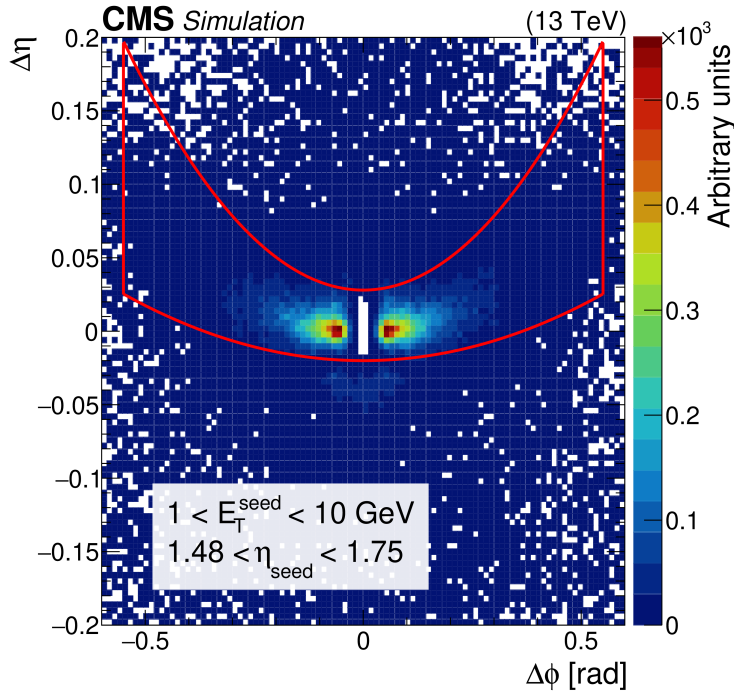
Figure 4.5: Supercluster defined by the mustache algorithm (delimited by the red line) centered around the cluster used to seed the algorithm (white area at the center of the plot) [136].

Tracks reconstructed in the inner tracker are tested for compatibility with a SC using a *Gaussian Sum Filter* (GSF) algorithm [137]. The tracks that pass the selection are named *GSFTracks* and when coupled with a SC they are used to reconstruct the properties of an electron. The association of a GSFTrack with a SC can be *tracker-driven* or *ECAL-driven*, depending on whether the seed for the algorithm is a loosely identified track or SC.

The ECAL-driven algorithm starts from the central region of a SC and extrapolates an hypothetical electron trajectory towards the PV. A track is then accepted if its first two hits in the tracker are in a certain window around the extrapolated track. The tracker-driven algorithm is instead optimized for low energy electrons[3] [136] and uses cut-based and multivariate-based requirements to check the compatibility of inner tracks with a SC. Starting from loosely selected SCs, the PF algorithm then constructs electron and photon candidates based on whether the associated GSFTrack has a hit in the first PIX layer. Fig. 4.6 shows the reconstruction efficiency for electrons in 2017 simulation and data.
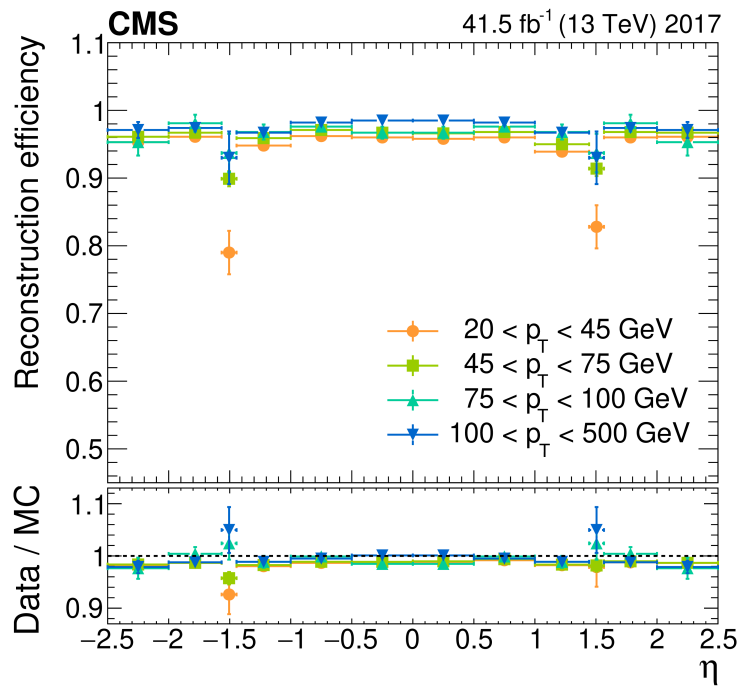

Figure 4.6: Electron reconstruction efficiency [136]. **Top:** efficiency in electron identification, measured by comparing simulated $Z/\gamma^* \to e^- e^+$ events with collected data. **Bottom:** ratio of the electron reconstruction efficiencies in simulation and real data is shown. The colors are associated to different energy range for the simulated electrons.

The final selection of electrons and photons for physics analyses is based on identification criteria, performed both with cut-based and Boosted Decision Tree (BDT) approaches. The same electron and photon ID algorithm is used for the measurements described in Chapters 5 and 6 of this thesis. It is a BDT-based algorithm which takes advantage of several properties of the reconstructed electron and photon candidates. ECAL shower parameters are shared between the photon and electron IDs:

---

[3]The tracker-driven algorithm has an efficiency of $\sim 50\%$ for electrons of $p_T \sim 3$ GeV, which decreses to $\sim 5\%$ for $p_T > 10$ GeV.

- H/E: it measures the energy deposit in an HCAL cone of radius $\Delta R < 0.15$ surrounding the SC. It is used to reject hadronic showers which started already in ECAL and estimate the leakage of electromagnetic (e.m.) showers in HCAL.

- $\sigma_{i\eta i\eta}$: the second moment of the crystal energy distribution in η, it is used to estimate the spread of the e.m. shower across the SC. It is used to reduce the effect of electronic noise in ECAL.

- R_9: the energy sum of the 3×3 crystal array centered around the most energetic crystal in the SC, divided by the total energy of the SC. This parameter is relevant to identify signs of early photo-conversion or bremsstrahlung.

On top of the aforementioned variables, the photon ID also uses the relative isolation of the photon. In Run 2 the relative isolation (ISO) for photons and electrons was defined as [136]:

$$I_{rel}^{e/\gamma} = \frac{\sum E_T(\text{charged}) + \max\left(\sum E_T(\text{neutral}) - \rho \times A_{eff}, 0\right)}{p_T^{e/\gamma}} \ , \tag{4.4}$$

where $p_T^{e/\gamma}$ is the transverse momentum of the electron or photon, and E_T is the transverse energy of the particles shown in parentheses: charged hadrons, or neutral hadrons and photons; $\rho \times A_{eff}$ is a correction based on PU: ρ is the average transverse energy density in the event and A_{eff} is the area of the isolation region corrected to account for the η-dependence of the transverse energy for PU events. The sums are performed over particles in a cone of $\Delta R = 0.3$ around the electron or photon direction of flight.

Unlike photons, reconstructed electrons are associated also to tracks via the GSF algorithm. The electron ID can therefore take advantage of the GSFTrack parameters. These include the χ^2 of the track fit, the number of valid hits, and the track compatibility with the SCs, determined using the distance in the $\eta - \varphi$ plane between the track extrapolation to ECAL and the SC center. Two versions of the electron ID are available:

- one includes the relative isolation among the input parameters;

- a separate electron ID is determined without including I_{rel}^e among its input parameters.

The analyses presented in this thesis used the latter electron ID definition. A cut on I_{rel}^e was then used at the level of event selection.

4.1.5   Jets

As mentioned in Section 1.2, quarks and gluons produced in pp collisions hadronize into cones of collimated colorless objects named jets. Before approaching their reconstruction in CMS it is important to note some features they possess. First, the hadrons

which make up a jet have energies ranging over several orders of magnitude, reaching scales where the perturbative approach for QCD calculation would fail. This would result in the divergence of the cross sections for gluon emission at very low angle (*collinear divergence*) or with very low momentum (*infrared divergence*). To account for that, in the calculations of the cross section a cut-off scale is introduced, which defines the perturbative approach used to make the calculations converge. This scale is then varied to operate calculations at different energies, without breaking the perturbative approach. Physical observables cannot depend on the scale chosen, therefore any physical jet related observable $\mathscr{O}$ dependent on the momenta P_i (i=1,...,N) of its constituents must be:

- *infrared-safe* $\longrightarrow$ $\mathscr{O}(p_1, ..., p_N) \equiv \mathscr{O}(p_1, ..., p_N, \varepsilon)$ with ε energy of another emitted particle, chosen however small;

- *collinear-safe* $\longrightarrow$ $\mathscr{O}(p_1, .., p_i, .., p_N) \equiv \mathscr{O}(p_1, .., p_{i1}, p_{i2}, .., p_N)$ with $p_i = p_{i1} + p_{i2}$ and p_{i1} and p_{i2} momenta of collinearly emitted particles.

The jet reconstruction strategy in CMS is chosen in order to satisfy both requirements and is the anti-k_t algorithm [138]. It follows an iterative procedure, which takes the various hadron candidates reconstructed by the PF algorithm and groups them into jets.

For each object i of transverse momentum $p_{T,i}$ a *distance* from the beam axis is defined as follows:

$$d_{iB} \equiv p_{T,i}^{-2} \ . \tag{4.5}$$

Another *distance* is associated to each possible pair in the following way:

$$d_{ij} \equiv \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{\mathrm{R}^2} \ , \tag{4.6}$$

where $\Delta R_{ij}^2 \equiv (\varphi_i - \varphi_j)^2 + (\eta_i - \eta_j)^2$, and R is the size of the jet cone in the $\eta - \varphi$ plane, which in CMS is chosen to be 0.4. After ordering all the *distances* d_{iB} and d_{ij} in decreasing order the algorithm works as follows:

1. the smallest *distance* is examined:

   a) if it is of type d_{iB}, then i is assigned to a jet;
   b) if it is of type d_{ij} the momenta of the particles i and j are summed.

2. if the particle has been assigned to a jet it is removed from the list, then the algorithm investigates the new smallest *distance* available.

The procedure is started anew iteratively till each particle has been assigned to a jet. The distances defined above are collinear-safe by definition, while the infrared-safety is provided by applying a cut on the jet minimum momentum.

As shown in Fig. 4.7, isolated jets reconstructed with the anti-k_t algorithm are associated a characteristic cone-like shape. In this thesis it will common therefore to refer to a "jet cone" when discussing the identification of jet sub-structures.
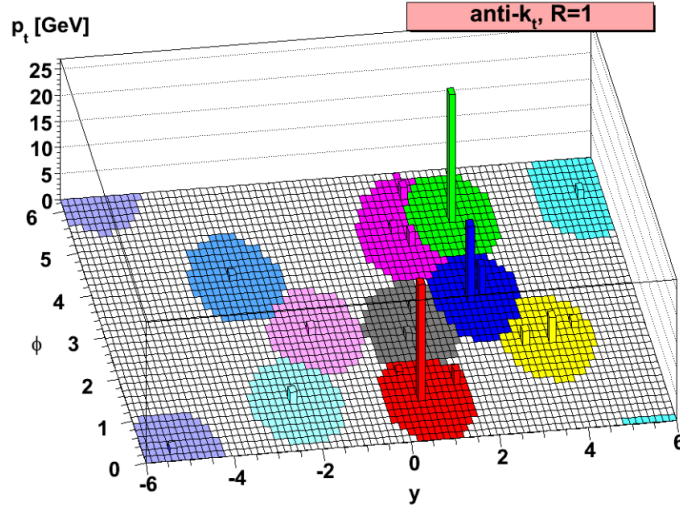
Figure 4.7: Schematic representation of jet clusters reconstructed by the anti-k_t algorithm [138]. Each jet cluster is represented by a different color. When relatively isolated, the jets reconstructed with this algorithm are characterized by a cone-like shape.

Jet energy calibration

Jet energy and transverse momenta are measured in simulation using the properties of stable particles which have been clustered into a jet by the anti-k_t algorithm. Pile-up jets, electronic noise in the calorimeters, and other factors can cause a discrepancy between the jet modeling in simulation and the reconstructed jets. A calibration of the jet energy is performed in order to achieve a more accurate description of the jet properties at reconstruction level.

The jet 4-momentum is corrected with a multiplicative factor C:

$$p_\mu^{corr} = C \cdot p_\mu^{raw} \ , \tag{4.7}$$

where $p_\mu^{corr(raw)}$ represents the μ-th component of the corrected (uncalibrated) jet 4-momentum. The correction factor C is the combination of several corrections which are applied sequentially to the uncalibrated jet 4-momentum:

$$C = \mathrm{C_{offset}}(p_T^{raw}) \cdot \mathrm{C_{MC}}(p_T', \eta) \cdot \mathrm{C_{rel}}(\eta) \cdot \mathrm{C_{abs}}(p_T'') \ , \tag{4.8}$$

where:

1. $\mathrm{C_{offset}}$: this correction aims at removing the energy in the jet which is not coming from the primary interaction vertex, e.g. pile-up jets and electronic noise;

2. C_{MC}: the transverse momentum of the jet corrected at the previous step (p'_T) is calibrated so that the average of its distribution ($\langle p_T^{reco} \rangle$) matches the average of generator level jets ($\langle p_T^{gen} \rangle$). The ratio $R = p_T^{reco}/p_T^{gen}$ is referred to as *jet response*, and the average correction, expressed as a function of $\langle p_T^{reco} \rangle$, is defined as

$$C_{MC}(p_T^{reco}) = \frac{1}{\langle R \rangle} \ , \tag{4.9}$$

and is referred to as *jet energy scale* (JES);

3. C_{rel}: the jet response is corrected based on the pseudorapidity region;

4. C_{abs}: the JES are corrected using dedicated measurements performed on $Z/\gamma^* + jets$ and multijet production events as a function of the jet pseudorapidity and the tranverse momentum corrected at the previous steps (p''_T).

The set of corrections applied to the jet energy and transverse momentum are globally referred to as jet energy corrections (JEC).

Heavy flavor tagging

Hadrons containing heavy quarks (charm and bottom) are long-lived particles, with average lifetimes of the order of $10^{-13} - 10^{-12}$ s [23]. They can travel between several millimeters (charmed hadrons) up to a few centimeters (b-hadrons) before decaying. The resulting jets are usually wider than the ones initiated by light quarks and gluons, and it is possible to reconstruct within the jet the decay vertices of heavy flavor hadrons. The process of identifying these jet topologies is usually referred to as heavy flavor tagging, and is performed by several algorithms proposed in CMS [139, 140].

In the context of analysing events where pairs of tau leptons are involved, a relevant background process is the $t\bar{t}$ production. The top quark has a mass of ~ 173 GeV allowing it to decay into a bottom quark and an on-shell W boson at tree level. This causes the top quark to have a short average lifetime ($\sim 0.5 \times 10^{-24}$ s [23]) and decay before hadronizing. This means that both di-tau and $t\bar{t}$ events involve two charged current electroweak decays, resulting in a few similarities between the processes, and making the presence of b-jets in the $t\bar{t}$ process one of its defining features. In the analysis of the CP structure of the Yukawa coupling in $H \to \tau\tau$ decays a veto against b-jets is applied. These jets are identified by a heavy flavor tagging algorithm specialized in the identification of b-jets: the b-tagging algorithm *deepCSV*, a Deep Neural Network (DNN) based identification [140, 141]. The DNN takes as input several low and high level features. The low level features are related to the tracks assigned to the jet by the anti-k_t algorithm: they include the significance for the IP components in the transverse plane and along the z axis, analogous to the quantities defined in Section 4.1.2. Higher level features include the number of SVs reconstructed by the AVR algorithm, the number of tracks associated to each SV and the fraction of

energy associated to each tracks emerging from a SV. The network has five output nodes corresponding to the following jet flavors:

- exactly one b hadron;

- two or more b hadrons;

- exactly one c hadron;

- two or more c hadrons;

- any other flavor (u, d, s or gluon jets).

A softmax activation function[4] is used for the output layer of the network, resulting in the output scores of these categories ($\mathscr{P}(f)$ with $f \in \{b, bb, c, cc, usbg\}$) being all distributed between 0 and 1, and to obey the following condition:

$$\sum_{f}^{flavors} \mathscr{P}(f) = 1 \ . \tag{4.11}$$

In the context of this thesis work the number of b hadrons is not a relevant parameter, therefore the b-tagger is defined as $B - tag = \mathscr{P}(b) + \mathscr{P}(bb)$. The b-veto is then applied by rejecting events containing jets with $B - tag$ value above a certain threshold.

4.1.6   Missing transverse energy

As previously stated in Section 1.2, partons inside a proton can be considered as freely moving, with the strong interaction becoming dominant at the Fermi scale ($\mathscr{O}(10^{-15}\mathrm{m})$). In pp collisions the hard scattering interaction can be assigned to two single partons belonging to the colliding protons, while the interaction between partons of the same proton can be neglected. Each parton has a momentum $\vec{p}_i = x_i \vec{P}$ equal to a fraction x_i of the proton momentum $\vec{P}$. The momentum fractions x_1 and x_2 carried by the colliding partons are not necessarily equal. The parton momenta $\vec{p}_1$ and $\vec{p}_2$ are therefore not required to be equal in module, but are both directed along the beam line. Therefore, the momentum associated to the partonic scattering center of mass, $\vec{p}_1 + \vec{p}_2$, has a projection in the transvere plane equal to 0. The physical variables chosen to describe a pp collision must be invariant under Lorentz boost along the z axis, e.g. the transverse momentum of

---

[4]This function can be written analytically as:

$$f(\vec{x})_i = \frac{\exp\{x_i\}}{\sum_j^5 \exp\{x_j\}} \ , \tag{4.10}$$

with $i \in \{1, 2, 3, 4, 5\}$ being the index running over the 5 output categories.

a particle. In any pp collision the transverse momenta of the particles emitted must sum up to 0:

$$\sum_i \vec{\mathrm{p}}_{T,i} = 0 \ , \tag{4.12}$$

with $\vec{\mathrm{p}}_{T,i}$ transverse momentum of the i-th particle, and the sum running on all particles produced in the hard scattering vertex. This is verified under the assumption that all particles are reconstructed and their energy is correctly measured. A more accurate depiction of what is observed experimentally is that some energy in the transverse plane is missing. This missing transverse energy (MET) is defined as:

$$\vec{p}_T^{\,mis} = - \sum_i^{reco.} \vec{p}_{T,i} \ , \tag{4.13}$$

where the sum now runs over the reconstructed particles, and $\vec{p}_{T,i}$ is the reconstructed transverse momentum of each particle. The MET itself combines detector inefficiencies effects with the presence of particles which escaped the detector without interacting, like neutrinos or neutral weakly interacting particles in BSM theories, collectively referred to as *invisible particles*. Tau leptons decay through weak charged current, and at least a neutrino is always present among the decay products. So the processes studied in this thesis work require a precise reconstruction of the MET.

Two different definitions of MET are used in this thesis:

- *particle flow MET*: PF-MET [142];

- *PileUP per Particle Identification MET*: PUPPI-MET [143].

The PF-MET is defined as in Eq. 4.13, with the sum operated over all the PF candidates. The PUPPI-MET instead attempts at reconstructing the MET component originated by the hard scattering process. This is done by scaling the 4-momenta of each particle ($\vec{p}_{T,i}$) in the event by a weight w_i, which ideally should be 1 for products of the hard scattering and 0 for particles coming from PU vertices. The MET is then calculated as the negative sum of the weighted transverse momenta of the particles:

$$\vec{p}_T^{\,mis} = - \sum_i^{reco.} w_i \vec{p}_{T,i} \ . \tag{4.14}$$

The resulting MET approximates the total transverse energy of invisible particles coming from the hard scattering process, assuming the reconstruction inefficiencies can be neglected.

The weights w_i are computed by constructing for each particle a shape α, which aims at identifying particles originated by the PV or by PU vertices. This shape has an analytical form which can vary based on which subdetectors were used to reconstruct the

particle [143]. In most cases, it is determined by the transverse momentum of a particle and the properties of the particles reconstructed in its proximity.

On an event-by-event basis the distribution of α is then constructed using the charged particles assigned to PU vertices by the algorithms described in Section 4.1.2. In the forward regions of the detector, the tracker information is not available. The α distribution for these regions is obtained by correcting the median ($\bar{\alpha}_{PU}$) and root-mean-squared (RMS_{PU}) of the corresponding distribution in the barrel region using transfer factors computed in simulation. The value α_i corresponding to the i-th particle is then compared to the α distribution for PU events, in order to calculate the likelihood of that particle being originated from PU vertices. This is done by computing a signed chi-squared:

$$\text{signed } \chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})|(\alpha_i - \bar{\alpha}_{PU})|}{(RMS_{PU})^2} \ . \tag{4.15}$$

This χ^2 term is then converted to a weight by taking the cumulative distribution of the χ^2:

$$w_i = F_{\chi^2, ndf=1}(\text{signed } \chi_i^2) \ , \tag{4.16}$$

where $F_{\chi^2, ndf=1}$ is the cumulative distribution for a χ^2 with one degree of freedom. In CMS these weights are used directly for neutral particles, while for charged particles the weights are shifted to 0 or 1 based on whether they are assigned to PU vertices or to the PV [144].

To illustrate the difference between the two definition, Fig. 4.8 shows how the transverse mass distribution for a muon and MET system changes using both definitions. The transverse mass m_T is defined as:

$$m_T = \sqrt{2 p_{T,\mu} \, p_T^{mis} \, \left(1 - \cos(\Delta\phi_{\mu,met})\right)} \ , \tag{4.17}$$

with p_T^{mis} and $p_{T,\mu}$ representing the MET, obtained with either definition and the muon momentum. The $\phi_{\mu,met}$ parameter represents the azymuthal separation between the muon and the MET:

$$\cos(\Delta\phi_{\mu,met}) = \frac{\vec{p}_{T,\mu} \cdot \vec{p}_T^{\ mis}}{p_{T,\mu} \, p_T^{mis}} \ . \tag{4.18}$$

The distributions shown in Fig. 4.8 were obtained in the context of the study of the CP properties of the Higgs Yukawa coupling to tau leptons (see Chapter 6). For the purpose of discussing the difference between the two definitions of the MET, the following differences are the most relevant:

- the system formed by muon and MET for processes involving two tau leptons, one of which decaying muonically and the other hadronically, has a transverse mass distribution peaking at 0;
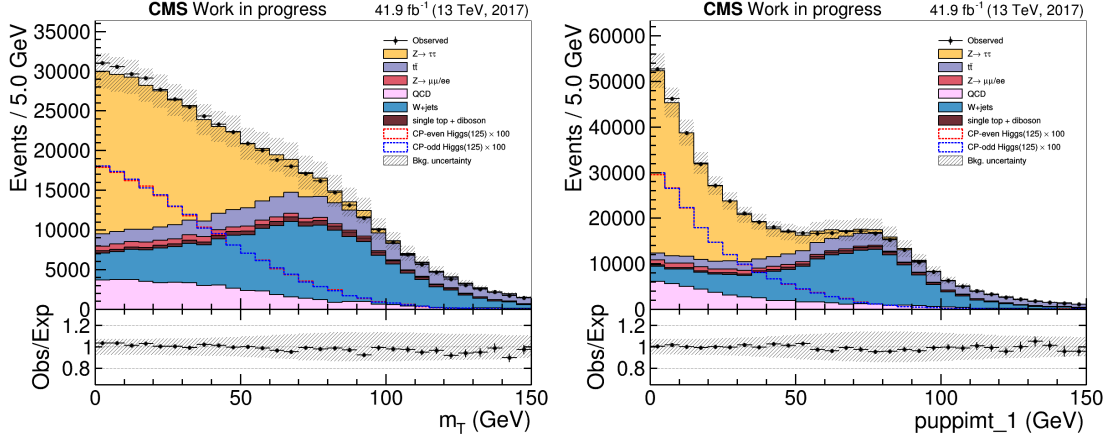
Figure 4.8: Transverse mass distribution for a muon and MET system, using events where a muon was reconstructed alongside a hadronically decaying tau lepton collected in 2017. The MET is defined on the **left** (**right**) with the PF (PUPPI) algorithm. The description of the processes represented in the plots and the event selection are presented in Section 6.2.

- for processes involving a W boson decay to muon and neutrino, produced in association with a jet, the m_T distribution peaks at ~ 80 GeV, since it represents the invariant mass of the muon and neutrino system obtained from a W boson decay.

Using the PUPPI-MET definition to reconstruct m_T (right part of Fig. 4.8), allows a higher percentage of events involving genuine tau decays (represented by the yellow histogram and by the overlaid histograms representing Higgs decays) to have a reconstructed m_T closer to 0. At the same time, events with a W boson produced in association with jets (blue histogram) present a more pronounced peak in the m_T distribution around 80 GeV. This allows to improve the separation between these different categories of events. As the Higgs boson decays belong to the former category of processes, the use of PUPPI-MET allows for a better identification of Higgs decays with respect to events involving W boson production.

Both MET definitions are affected by the inefficiencies and limited resolution of the detectors, which can be partially cured via calibration. One of the applied corrections is called *Type-1* and is performed by correcting the jet transverse momenta:

$$\vec{p}_T^{\,miss} = \vec{p}_T^{\,miss,raw} - \sum_{jets}(\vec{p}_T^{\,corr} - \vec{p}_T^{\,raw}) \ , \tag{4.19}$$

where $\vec{p}_T^{\,miss,raw}$ is the MET defined in Eq. 4.13 and $\vec{p}_T^{\,corr} - \vec{p}_T^{\,raw}$ is a correction applied to the jet transverse momenta. It includes the JEC [145] discussed in the previous section. Furthermore, muons which are aligned to a jet direction have their momenta subtracted from the jet momenta, while for the electrons a fraction of its ECAL energy deposit is subtracted.

The study of the MET resolution is usually done selecting dilepton events associated mostly to the Drell-Yan process, or photon production in association with jets. These topologies allow to construct a precise frame of reference for the decaying vector boson, and measure its recoil against the hadronic system. This allows to improve the agreement of MET resolution between simulation and data, and such corrections were applied in each measurement presented in this thesis work.

4.2 Tau identification and reconstruction

Tau leptons in CMS are reconstructed by studying their decays, previously discussed in Section 1.5.1. To facilitate the understanding of the next chapters it is useful to introduce here some notations.

Tau decays can either be fully leptonic, with the tau decaying into a muon or electron and two neutrinos, or involve hadrons, with or without intermediate mesonic resonances. Starting from this section, fully leptonic decays will be referred to simply as *leptonic*, while decays involving at least a charged hadron will be called *hadronic*. The following symbols are used:

- τ_h: hadronically decaying tau lepton;

- τ_l: leptonically decaying tau lepton.

It is common to refer to the charged decay products of tau leptons as *prongs* and divide the decay channels according to the number of prongs (Fig. 4.9). Leptonic decays are all *one prong* decays, thus they do not require other labeling aside from occasionally specifying whether the tau decayed into a muon (τ_μ) or an electron (τ_e). Hadronic decays instead are classified depending on the mesonic resonance involved, leading to the labels: τ_h with $h \in \{\pi, \rho, a_1^{1Pr}, a_1^{3Pr}\}$. The labels a_1^{1Pr} and a_1^{3Pr} are both associated to decays involving the a_1 meson as a resonance, and indicate respectively decays with one ($\tau^- \to \nu_\tau \pi^- \pi^0 \pi^0$) or three ($\tau^- \to \nu_\tau \pi^- \pi^- \pi^+$) prongs.

When talking about reconstruction of tau leptons it is usual to refer implicitly to the identification of hadronic decays. This can be understood by discussing the properties of leptonic decays. A τ_l candidate is an isolated, i.e. not part of a jet or a collimated particle shower, lepton reconstructed following the procedures discussed in sections 4.1.3 and 4.1.4. The reason for not implementing more precise reconstruction techniques stems from the difficulty of separating τ_l candidates from prompt leptons. Leptonic decays involve always two neutrinos, making them 3-body decays, in which only one decay product is actually identified. This means that the fraction of energy of the decaying tau lepton carried by the prong falls within a continuous spectrum. Energy related quantities are therefore not different between τ_l and prompt leptons. Furthermore, tau leptons have decay lengths of the order of millimetres, making the distinction between leptons coming from the PV
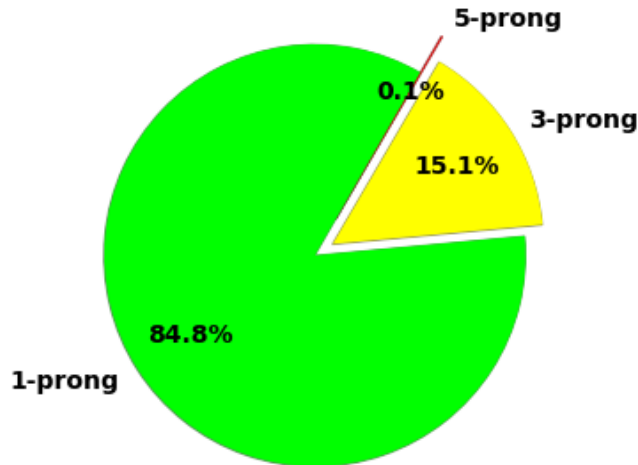
Figure 4.9: Pie chart representing the tau decay channels separated with respect to the number of charged decay products with relative branching fractions.

and the ones coming from vertices so close to the beamline extremely arduous. All these arguments apply to the reconstruction of a single τ_l candidate, however in di-tau events it is possible to access variables such as the angular separation between the prongs, the invariant mass of the system and other properties which allow to identify a τ_l candidate.

4.2.1   The hadron-plus-strip algorithm

Reconstruction of hadronically decaying tau leptons is operated in CMS with the hadron-plus-strip (HPS) algorithm [146,147]. Candidate jets, photons and electrons reconstructed by the PF algorithm are tested for compatibility with hadronic decay channels of tau leptons. A typical τ_h candidate is an isolated collimated jet with low multiplicity. The HPS algorithm aims at identifying τ_h candidates with high efficiency while rejecting the main background: quark and gluon jets coming from the QCD multijet production.

The HPS algorithm is seeded by PF jet candidates identified by the anti-k_t algorithm with a cone size of $\Delta R < 0.4$. Particles in the jets are then tested as candidates for:

- hadrons: charged particles depositing their energy in both ECAL and HCAL;

- *strips*: a cluster of electrons and photons producing in ECAL signatures compatible with a π^0 decays.

Several tau hadronic decay channels include π^0 mesons, which decay into two photons almost 100% of the times (as previously shown in Table 1.5). Photons have a chance of converting into an electron-positron pair, which then are separated along the φ direction due to the magnetic field. This results in an ECAL cluster narrow in η and extended in φ, referred to as *strip*. Photons and electrons in the jet cone are clustered to form a strip

if they are found in a certain $\Delta\eta \times \Delta\varphi$ window. In Run 1 the window size was fixed and set to $\Delta\eta \times \Delta\varphi = 0.05 \times 0.20$ [146], while in Run 2 a *dynamic* strip reconstruction was introduced [147]. The algorithm proceeds iteratively:

1. a strip is seeded by the leading photon or electron (e/γ) found in the jet not yet assigned to a strip;

2. the position and transverse momentum of the leading e/γ are assigned to the strip as the position of its center (η^{strip} and φ^{strip}) and momentum (p_T^{strip});

3. the next highest p_T e/γ candidate is then assigned to the strip if the distance between its position ($\eta^{e/\gamma}$ and $\varphi^{e/\gamma}$) and the strip center satisfies the following relations:

$$\Delta\eta(e/\gamma - strip) < max\left(f(p_T^{e/\gamma}) + f(p_T^{strip}), 0.15\right) \tag{4.20}$$

$$\Delta\varphi(e/\gamma - strip) < max\left(g(p_T^{e/\gamma}) + g(p_T^{strip}), 0.30\right) \tag{4.21}$$

with

$$f(p_T) = 0.20 \cdot p_T^{-0.66} \tag{4.22}$$

$$g(p_T) = 0.35 \cdot p_T^{-0.71} \tag{4.23}$$

4. if the e/γ candidate is included in the strip then the properties of the strip are re-evaluated based on its constituents:

$$p_T^{strip} = \sum^{e/\gamma} p_T^{e/\gamma} \tag{4.24}$$

$$\eta^{strip} = \frac{1}{p_T^{strip}} \sum^{e/\gamma} p_T^{e/\gamma} \cdot \eta^{e/\gamma} \tag{4.25}$$

$$\varphi^{strip} = \frac{1}{p_T^{strip}} \sum^{e/\gamma} p_T^{e/\gamma} \cdot \varphi^{e/\gamma} \tag{4.26}$$

5. the process continues till no other e/γ candidate is found within the strip window, and the clustering of a new strip is initiated using the unassigned e/γ candidates.

The algorithm improved strip reconstruction compared to the fixed-window one done in Run 1, as it allowed to tackle events with different topologies:

- high energy τ_h decay in highly collimated jets and the use of a smaller window size allows for better discrimination against quark and gluon jets;

- through multiple scattering in the tracker, electrons and positrons originated by the π^0 decay could fall outside the fixed window;

- charged pions could radiate low p_T particles by interacting with the tracker material reaching ECAL outside a fixed window.

The functions defined in Eq. 4.22 and 4.23 were chosen in order to include in a strip 95% of all electrons and photons coming from τ_h decay products. This was done by simulating single τ_h and reconstructing the distance in η and φ between the e/γ coming from τ_h decays and the τ_h direction of flight, used as a proxy for the strip center, and then fitting the contour representing the 95% quantile as shown in Fig. 4.10.

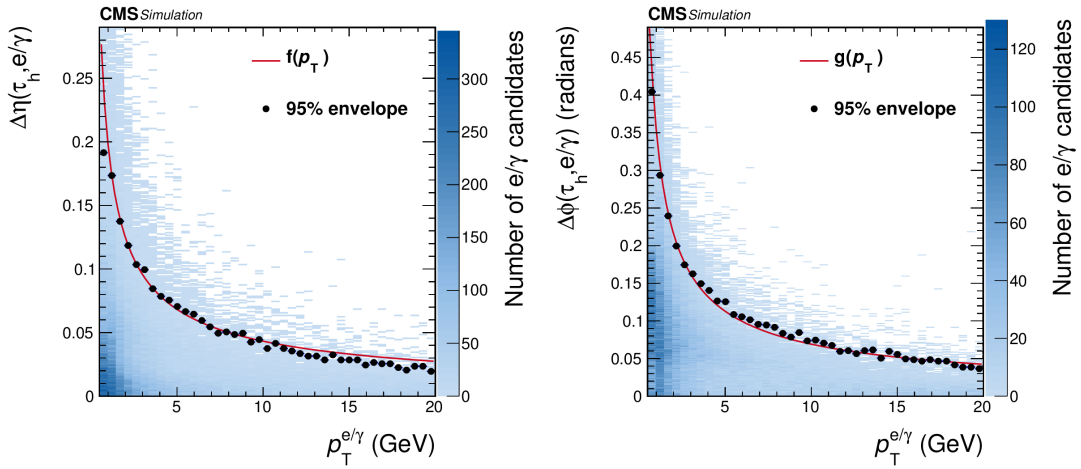**Figure 4.10:** Distance in η (**left**) and in φ (**right**) between e/γ candidates and the τ_h direction of flight, as a function of $p_T^{e/\gamma}$. The dashed lines represent the functions $f(p_T)$ and $g(p_T)$ shown in Eq. 4.22 and 4.23 which are used to fit through the dotted points, which show 95% quantile for the given p_T bin [147].

The highest energy hadron and strip candidates selected (up to six each) are then combined as potential τ_h candidates and compared to the different hadronic decay modes of tau leptons. The assignment of a decay mode is based on the number of prongs and strips available and the invariant mass of the tau decay products.

For each τ_h^i candidate, corresponding to different combinations of hadrons and strips, the invariant tau lepton mass m_{τ_h} is reconstructed. It is defined as the module of the 4-momentum $p_{\tau_h} = (E_{\tau_h}, p_x^{\tau_h}, p_y^{\tau_h}, p_z^{\tau_h})$ obtained by adding the 4-momenta of the reconstructed visible decay products. It is useful to factorize the contribution on the tau invariant mass coming from the strips when defining the decay mode. This is done since the energy associated to a strip depends on its window size, as it affects how many electrons and photons are associated to the strip. An *ad hoc* estimation of this contribution

is calculated as [147]:

$$\Delta m_{\tau_h}^{strip} = \sqrt{\left(\frac{\partial m_{\tau_h}}{\partial \eta^{strip}} f(p_T^{strip})\right)^2 + \left(\frac{\partial m_{\tau_h}}{\partial \varphi^{strip}} g(p_T^{strip})\right)^2} \ , \qquad (4.27)$$

where the two partial derivatives[5] in the equation represent the changes to the invariant mass brought by the distance of the strip from the τ_h direction of flight.

Based on the number of reconstructed hadrons and the values of m_{τ_h} [6] and $\Delta m_{\tau_h}^{strip}$ of each τ_h candidate, the HPS assigns the decay mode as follows[7]:

- **one prong** ($h^{\pm}$): This DM corresponds to cases where only one charged hadron candidate was identified, not accompanied by good strip candidates, and $0 < m_{\tau_h} < 1$. The choice of this mass window allows to target not only charged pions, but also the rarer decays to kaons. For this DM, m_{τ_h} is then assigned to be equal to the pion mass.

- **one prong plus one strip** ($h^{\pm}\pi^0$): A charged hadron is required to be identified together with a strip, and their visible invariant mass is required to be $0.3 - \Delta m_{\tau_h} < m_{\tau_h} < 1.3 \times \sqrt{p_T/100} + \Delta m_{\tau_h}$, with the upper limit constrained between 1.3 and 4.0.

- **one prong plus two strips** ($h^{\pm}\pi^0\pi^0$): This DM requires one hadron to be reconstructed alongside two strips. Their visible mass should be $0.4 - \Delta m_{\tau_h} < m_{\tau_h} < 1.2 \times \sqrt{p_T/100} + \Delta m_{\tau_h}$, with the upper limit constrained between 1.2 and 4.0.

- **two prongs** ($h^{\pm}h^{\pm}(h^{\mp})$): This DM is assigned to τ_h candidates where two charged hadrons are identified having invariant visible mass between $0 < m_{\tau_h} < 1.2$. For this decay mode and the next one, the requirement on the τ_h candidate electric charge is relaxed to allow values different from 1, the two charged hadron candidates can therefore have either equal or opposite charge.

---

[5]The two partial derivatives in Eq. 4.27 are analytically defined as:

$$\frac{\partial m_{\tau_h}}{\partial \eta^{strip}} = \frac{p_z^{strip} E_\pi - E^{strip} p_z^\pi}{m_{\tau_h}} \ , \qquad (4.28)$$

$$\frac{\partial m_{\tau_h}}{\partial \varphi^{strip}} = \frac{p_y^{strip} p_x^\pi - p_x^{strip} p_y^\pi}{m_{\tau_h}} \ , \qquad (4.29)$$

where $\mathbf{p}^\pi = (E_\pi, p_x^\pi, p_y^\pi, p_z^\pi)$ and $\mathbf{p^{strip}} = (E_{strip}, p_x^{strip}, p_y^{strip}, p_z^{strip})$ represent the 4-momenta associated respectively to the charged prong system (e.g. the pion in one prong decays) or the strip.

[6]Fig. 4.13 shows the distribution of the reconstructed τ_h mass, after the choice of which hadrons and strips are assigned to the reconstructed τ_h.

[7]The values for the tau lepton mass are all expressed in GeV.

- **two prongs plus one strip** $(h^{\pm}h^{\pm}(h^{\mp})\pi^0)$: Cases with two charged hadrons and a strip are assigned to this DM. The visible mass should fall within the range $0. < m_{\tau_h} < 1.2 \times \sqrt{p_T/100} + \Delta m_{\tau_h}$, with the upper limit constrained between 1.2 and 4.0.

- **three prongs** $(h^{\pm}h^{\pm}h^{\mp})$: This DM is associated to three charged hadrons with no additional strip, their invariant visible mass is required to be $0.8 < m_{\tau_h} < 1.4$.

- **three prongs plus one strip** $(h^{\pm}h^{\pm}h^{\mp}\pi^0)$: Three charged hadrons and a strip are required to be present, furthermore the mass of the tau candidate should be in the range $0.9 - \Delta m_{\tau_h} < m_{\tau_h} < 1.6 + \Delta m_{\tau_h}$.

At this stage the decay channels are selected with a rather loose selection, as can be noticed by the presence of the two prong channels which correspond to τ_h candidates with charge different from 1. Two prong decays can be originated by three prong decays where one of the charged hadrons was not reconstructed in the jet cone. For most analyses such decays are excluded as they are not sufficiently well reconstructed. Only one and three prong decays are considered from this point onward.

It is common to label the hadronic decays with an integer index based on the number of prongs (n_{prongs}) and strips (n_{strip}):

$$DM = 5 \times (n_{prongs} - 1) + n_{strip} . \tag{4.30}$$

With this convention one prong decays are assigned the labels 0, 1 or 2, while three prong decay correspond to the indices 10 and 11. Another thing to consider is that DMs 1 and 2 are defined with a largely overlapping mass window. Furthermore, in the decay of sufficiently boosted a_1 meson, the two π^0 can be emitted quite close to each other. This leads to a large fraction of events in which one single strip can incorporate the decay products of both π^0 mesons. To account for that, all decays with DM 2 are merged into the DM 1 ensemble after the selection operated by the HPS algorithm, forming the **one prong plus π^0s** DM.

Given that decays with one or two π^0 mesons correspond to different mesonic resonances, they have different spin-correlation between the tau leptons and their decay products. This has a large impact in terms of measuring the CP properties of the Higgs boson, therefore it made necessary the introduction of a more advanced DM reconstruction algorithm to run on top of the HPS one. This additional algorithm is described in Section 6.3.4 as it was specifically developed in the scope of that analysis. To avoid confusion between the current definition of DM and the one which is used in later chapters the four DMs which are provided by the HPS algorithm are labeled as:

HPS-DM 0 : **one prong**;

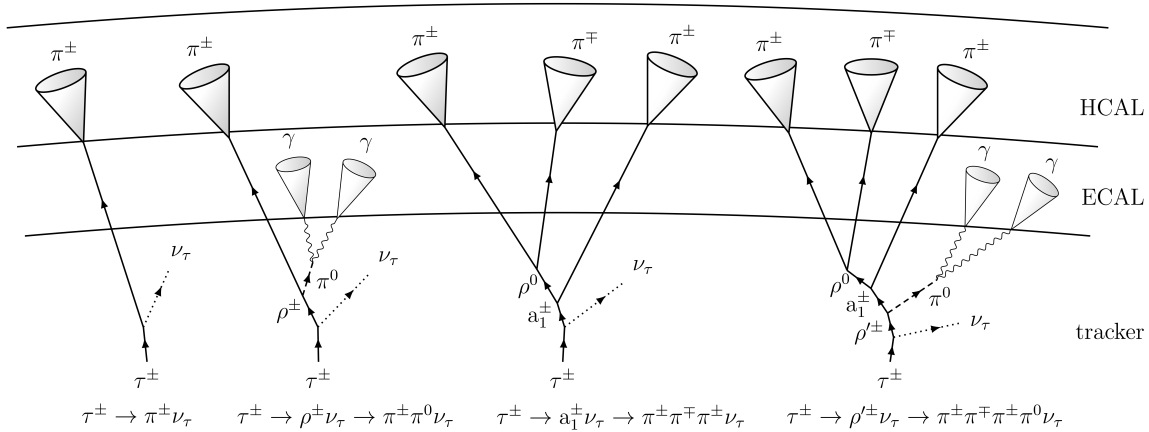HPS-DM 1 : **one prong plus π^0s**;

Figure 4.11: Schematic representation of the four HPS-DMs considered at analysis level, constructed using the latex package provided in [148]. From left to right the DMs shown are: one prong, one prong plus π^0s, three prong and three prong plus π^0.

HPS-DM 10 : **three prong**;

HPS-DM 11 : **three prong plus π^0**.

A schematic representation of these four DMs is shown in Fig. 4.11 to emphasize the expected signatures of each DM, those being the tracks and energy deposits in HCAL for the charged pions, and the electromagnetic clusters in ECAL originated by neutral pions.
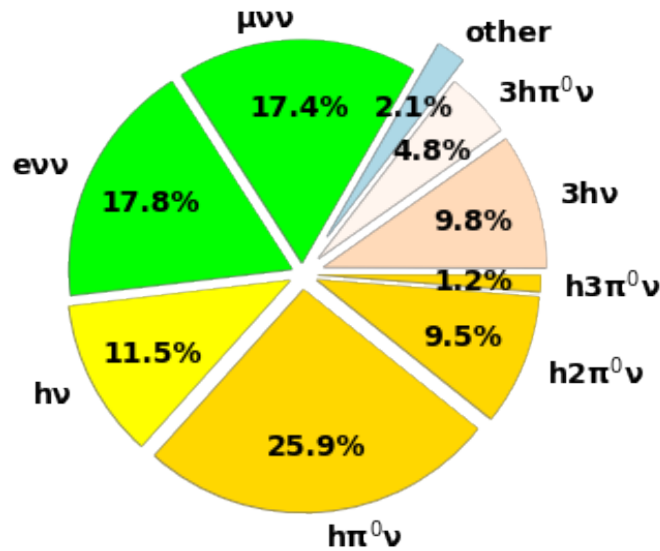

Figure 4.12: Pie chart representing the tau decay channels with relative branching fractions. The letter **h** is used to label charged hadrons.

Fig. 4.12 shows the tau decay channels with their relative branching fractions (BR). With respect to the values listed previously listed in Tab. 1.4, the BRs have been adjusted taking into account that in the CMS detector kaons appear indistinguishable from pions. The hadronic decay channels are therefore labeled based on the number of charged hadrons **h** and of π^0s. The colors assigned to the various sections of the pie chart represent the DM labels assigned to such decays after the HPS algorithm, neglecting DM misidentification. In green the leptonic decay channels are highlighted, while yellow, orange, sepia and pearl are used for the HPS-DMs 0, 1, 10 and 11 respectively.
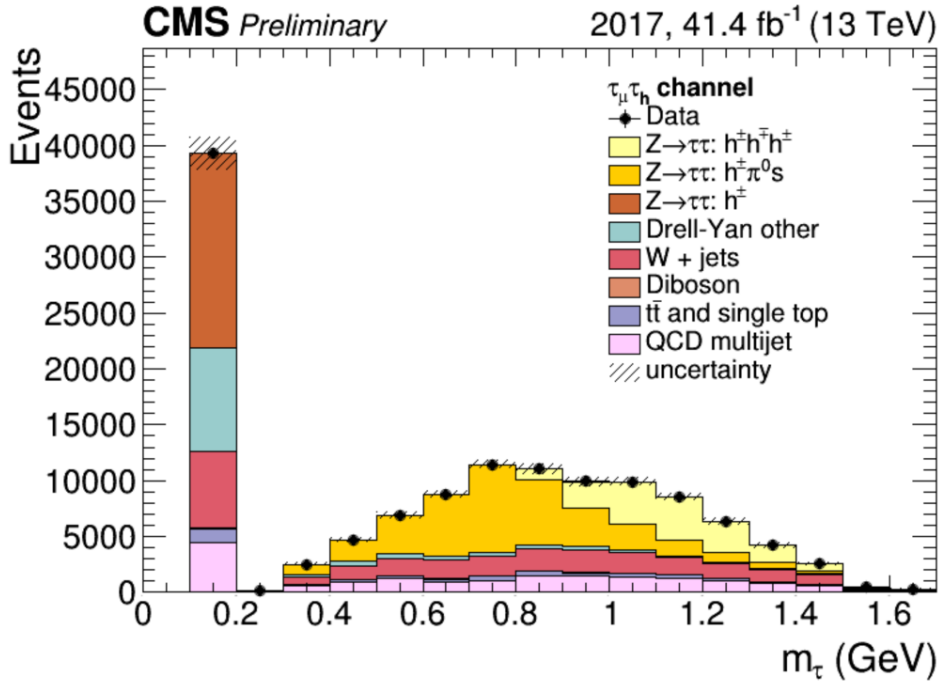
4.3 Misidentification of hadronic taus

Figure 4.13: Distribution of the invariant mass of τ_h candidates selected by the HPS algorithm [149]. These candidates are recorded in association with a muon, as part of a $\tau_\mu \tau_h$ event. For one prong decays the m_{τ_h} is set equal to the pion mass, resulting in a peak around 0.14 GeV. This plot was made before the latest updates in the HPS algorithm, as such the $h^\pm h^\pm h^\mp \pi^0$ channel is merged with the $h^\pm h^\pm h^\mp$ one.

Looking at the distribution of the reconstructed mass for the τ_h candidates identified by the HPS algorithm (Fig. 4.13) it becomes quite evident that the contributions from genuine hadronically decaying tau leptons are not sufficient to describe the recorded data.

This is because several objects can be misidentified as hadronically decaying tau leptons by the HPS algorithm:

- **jets**: a highly collimated quark or gluon jet can be mistaken for any tau decay;

- **muons**: can produce a signature similar to a one prong tau decay;

- **electrons**: can emit photons via bremsstrahlung radiation and mimic the ρ meson decay and be reconstructed as one prong plus π^0s decays.

As previously mentioned, a τ_h candidate is seeded by a PF jet in the HPS algorithm, making quark and gluon jets the highest source of contamination for the identification of τ_h. Light leptons can also be misidentified as τ_h, electrons in particular can occasionally produce electromagnetic showers that reach the HCAL, mimicking the signature of a charged pion, and emit bremsstrahlung photons which could be reconstructed as a strip by the HPS algorithm. A schematic representation of these signatures is shown in Fig. 4.14.
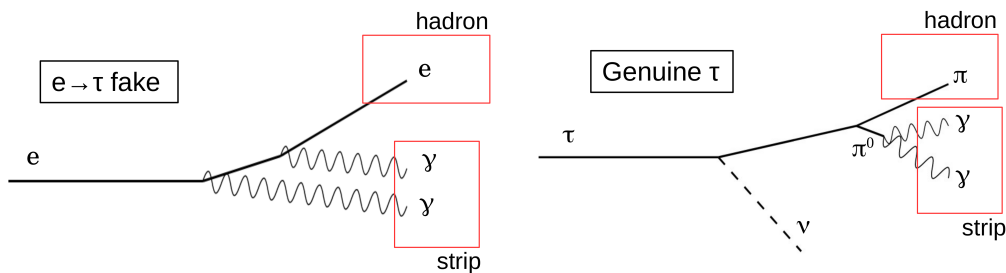

Figure 4.14: Schematic representation of the signatures identified by the HPS algorithm for an electron faking a τ_h (**left**) and a genuine τ_h (**right**).

To reduce the misidentification rate, and thus select a more pure sample of hadronically decaying tau leptons, several methods have been implemented in CMS throughout the years. During Run 1 and Run 2, the identification of genuine τ_h was generally performed separately for each misidentified object:

- jets misidentification was reduced using one of the following methods (see Appendix C.1):

  - a cut-based isolation sum discriminator,

  - a BDT-based discriminator;

- a cut-based discriminator was used to reduce the misidentification rate of muons (see Appendix C.2);

- rejection of misidentified electrons was performed via a BDT-based discriminator (see Appendix C.3).

The aforementioned methods were not used in the scope of the analyses presented in this thesis, as during Run 2 a new tau identification algorithm was deployed: the DeepTau neural network-based identification (DeepTau ID) [150, 151]. The DeepTau ID was used in both analyses presented in this thesis and is therefore described in the next section.

As it can be useful to compare the efficiency of the DeepTau ID with the discriminators previously used in CMS, the latter are briefly described in Appendix C.

4.4   DeepTau identification

The DeepTau [150, 151] identification is based on a multiclass convolutional neural network (CNN) [152] used to reduce the misidentification of jets, muons and electrons as hadronically decaying tau leptons. It takes as inputs low and high level features associated to the τ_h candidate and produces 4 output scores, one for each class:

- genuine taus (y_τ);

- jets (y_j);

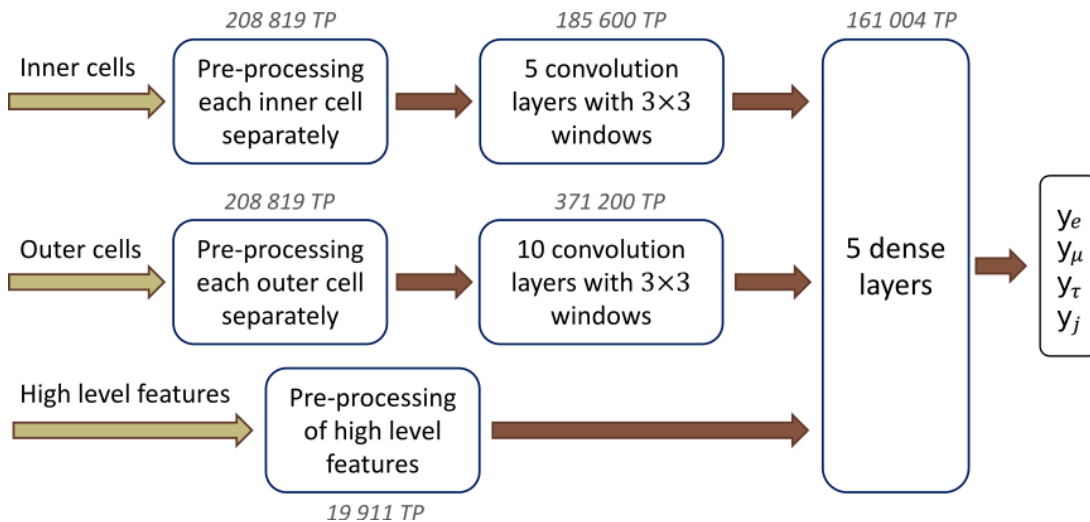- muons (y_μ);

- electrons (y_e).

Figure 4.15: Structure of the DeepTau NN [151].

The basic structure of the NN is shown in Fig. 4.15. Input features are grouped into several subsets depending on whether they are related to the signal cone, the isolation cone or they are higher level variables associated to the τ_h candidate or the event itself. Each subset of inputs are pre-processed separately and then passed through convolutional layers and finally feeded into a fully connected feed-forward structure, from which the 4 output scores are extracted. The DeepTau identification is then obtained by combining these NN scores into 3 classifiers. The following sections go into further details over each one of these parts of the NN structure.

4.4.1   Input features

The *low level* features are the tracks and energy deposits of the tau decay products and other PF candidates in the signal and isolation cones, which are defined similarly to what was done for the BDT-based tau identification. The CMS detector is divided in cells of $\eta \times \varphi$, and the properties of the leading PF candidate found in each cell are taken as inputs for the NN.

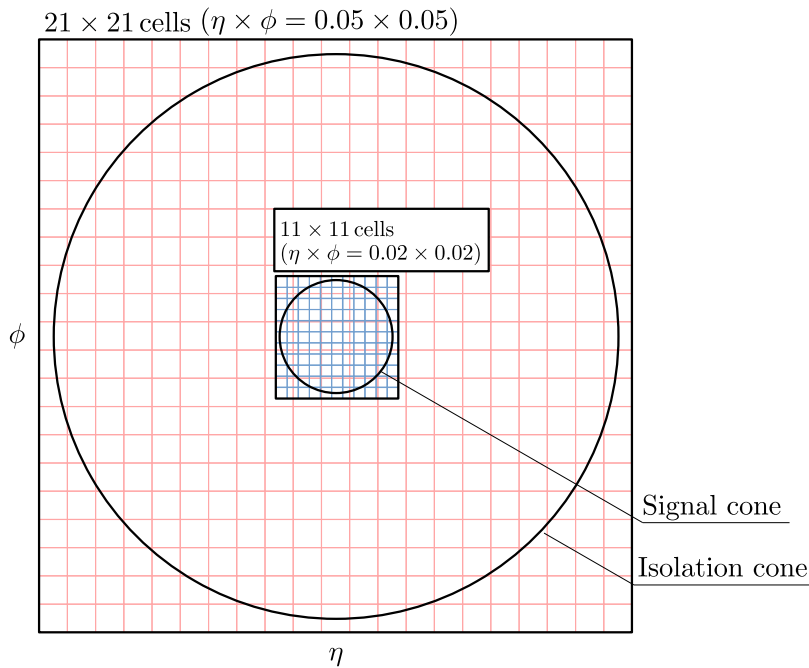A set of 21×21 cells arranged as a square around the leading tau decay product

Figure 4.16: Schematic view of the grid for input features of the DeepTau NN [151]: the cells highlighted in red are the outer cells with size $\Delta\eta \times \Delta\varphi = 0.05 \times 0.05$, while the blue ones have size $\Delta\eta \times \Delta\varphi = 0.02 \times 0.02$ and are the inner cells. The two concentric black circles represent the signal and isolation cones with radius $\Delta R = 0.1$ and $\Delta R = 0.5$ respectively.

direction of flight are defined with a granularity of $\eta \times \varphi = 0.05 \times 0.05$, and overlap with a set of 11×11 cells with a granularity of $\eta \times \varphi = 0.02 \times 0.02$ as shown in Fig. 4.16. The cells with higher granularity are named *inner cells*, and cover a signal cone of radius $\Delta R = 0.1$, while the others are named *outer cells* covering an isolation cone of radius $\Delta R = 0.5$. The two granularities allow to study more precisely the topology of the event in the region closer to the leading tau decay product, allowing for good discrimination between tau leptons and misidentified leptons and jets, while keeping the number of features manageable from the computational point of view. Having the two grids of input features overlap in the central region also helps keeping the correlation between the pre-processed features of inner and outer cells, preventing its loss before entering the fully connected layers.

Other input features are referred to as *high level* as they are related to either the tau candidate, like its transverse momenta, HPS-decay mode, η, φ and other properties, or to the recorded event in its entirety, like the average energy of the event. This results in a number of input features of the order of 100 thousands ($\mathcal{O}(100k)$).

4.4.2 Convolutional layers and training

The input features are processed via convolutional layers [152]. This technique was developed in the context of image recognition and analysis, and allows the network to look for features within a grid without accounting for where they appear.
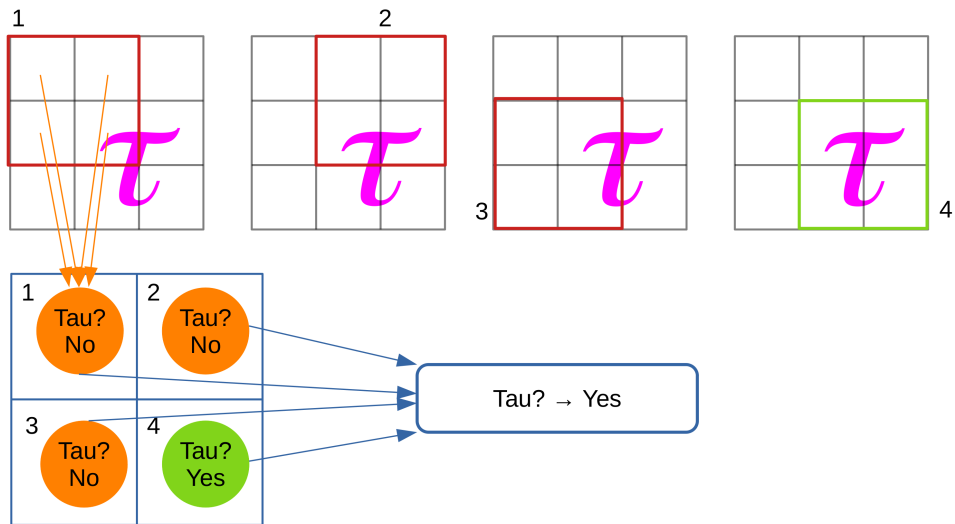

Figure 4.17: Schematic representation of a convolutional neural network: the filter of size 2×2 is moved across a 3×3 grid to determine whether the tau (in magenta) is present. The 4 ouputs form a new 2×2 grid, and a second convolutional layer is used to determine the output.

In order to explain how this layers operate, it is useful to define a simpler problem: finding a set of features within an image. For the solution of this problem, the position of the features within the image is irrelevant. In Fig. 4.17 the letter tau has been placed within a grid of size 3×3, and occupies a region of size 2×2. In order to determine whether the grid contains a tau, a *search window* is defined. The search window, called in literature *kernel* or *filter*, is then used to group a subset of the input cells. Their features are then used as input for a fully connected layer [153] which determines whether the selected window contains the tau. The search window is then moved across the grid, and the outputs of each window are combined to determine the global output of the CNN. In this example the search window has size 2×2, and the input grid has size 3×3, resulting in the convolutional layer having 4 nodes[8], which could then be used as an input for a convolutional or a fully connected layer to determine the output.

This simple example can be compared to the CNN implemented for the identification of genuine τ_h by imagining that each cell of the input grid contains a variety of features. Instead of one convolutional layer with only one filter, a sequence of convolutional layers with multiple filters is then applied. The output of these layers is then used to determine whether in the input grid electrons, photons, muons and hadrons were identified, and infer if they are originated from a genuine τ_h or a misidentified electron, muon or jet.

As shown in the left part of Fig. 4.18 low level features associated to muons, hadrons and e/γ are processed separately via 3 convolutional layers of window size 1×1, the

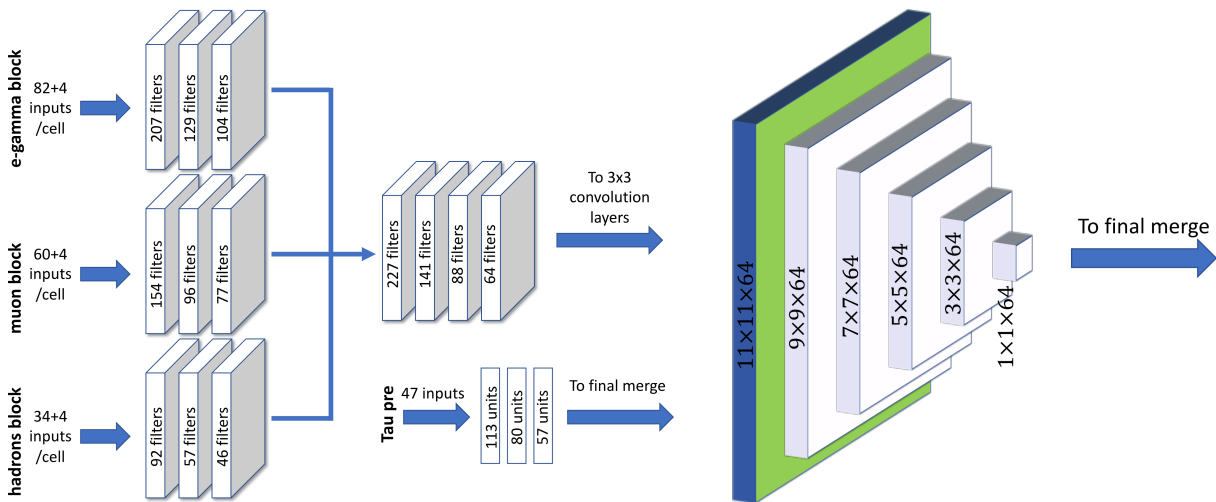**Figure 4.18: Left:** schematic of the preprocessing layers for low and high level features. **Right:** graphic representation of the convolutional layers used for inner cells: 5 convolutional layers of window size 3×3 [151].

---

[8]More generally, given a grid of size $m \times m$ and a window of size $n \times n$, the output is a grid of size $(m + 1 - n) \times (m + 1 - n)$.

outputs are then concatenated and sent to 4 convolutional layers. The low level features coming from the inner (outer) cells are then concatenated and further processed by 5 (10) convolutional layers of window size 3×3 as shown in the left part of Fig. 4.18. The pre-processing of high level features is instead carried on separately via 3 convolutional layer with window size 1×1.

All features are then concatenated and processed via 5 fully connected dense layers leading to the four output nodes. The training is performed using the NAdam algorithm [154] with $\mathcal{O}(1.5M)$ trainable parameters. It is a supervised training: the network receives, as an input, events containing either genuine τ_hs or misidentified objects, together with a label designating what each object is. The network then learns how to sort these objects into the four output classes based on their features. The training is performed using samples of simulated Drell-Yan (DY) or high mass Z (Z′) interactions for genuine τ_hs and misidentified objects, together with simulated processes like $t\bar{t}$, W+Jets, and QCD multijet production to account for the different jet topologies.

4.4.3 Classifiers

Using the four output scores, two possible strategies could in principle be established to define a classification algorithm:

a) associate any event to the class which achieved the highest output;

b) combine the scores themselves for each event to construct a classifier.

The first strategy has the benefit of providing relatively pure samples of genuine τ_h and misidentified objects. However it offers little choice in balancing the efficiency in the selection of genuine τ_h with the rejection of misidentified object. This makes strategy a) more practical for event categorization to distinguish signal and background events in measurements of cross sections or signal strengths, and is indeed chosen in the analyses presented in Chapters 6 and 7. In the context of tau identification the selection of pure samples of misidentified objects is of little interests. The 4 output scores are therefore combined to define 3 distinct classifiers:

$$P_{\tau \text{ vs } obj.} = \frac{y_\tau}{y_\tau + y_{obj.}} \ , \tag{4.31}$$

with $obj. = j, e, \mu$. Fig. 4.19 shows the misidentification rate as a function of the efficiency for the selection of genuine hadronically decaying tau leptons for the three classifiers. The DeepTau classifiers are compared to the MVA- and cut-based classifiers used in CMS before the deployment of DeepTau. For the same efficiency in the selection of genuine hadronically decaying tau leptons, the DeepTau classifiers present a noticeable reduction of the misidentification rate. For the against-electron and muon classifiers, the misidentification rate is reduced by at least a factor 3, while the against-jet classifier achieves a

reduction $\gtrsim 20\%$. As previously mentioned the DeepTau identification provides discrimination against all three contamination sources since the three classifiers are correlated with each other. In particular, events associated to a low value of y_τ will present low values also for all classifiers.

When considering the application of the DeepTau NN-based identification algorithm to a physics analysis, the efficiency in identifying genuine hadronically decaying tau leptons is increased by $\sim 20\%$, while the misidentification rate from light leptons and jets is reduced
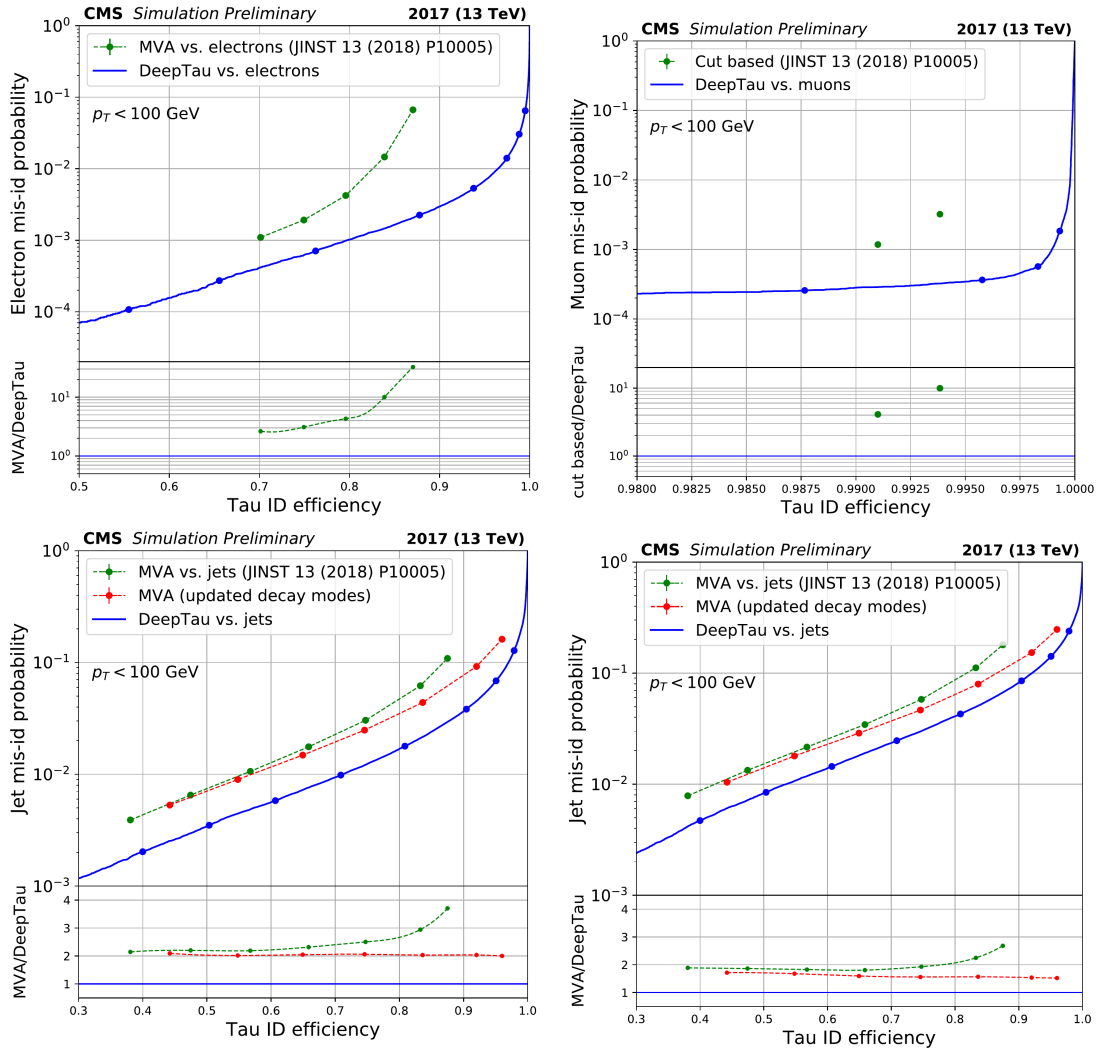

Figure 4.19: Misidentification rate as a function of the efficiency for the selection of hadronically decaying tau leptons. **Top row:** against-electron (**left**) and against-muon (**right**) classifiers validated on $Z \to ee$ and $Z \to \mu\mu$ samples. **Bottom row:** against-jet classifier validated on two different jet topologies: $t\bar{t}$ (**left**) and W+Jets (**right**) [150].

by $\sim 23\%$. As an example, Fig. 4.20 shows the improvement in the reconstruction of events where a tau lepton decays to a muon (τ_μ) and another one decays hadronically. The visible mass distribution of the $\tau_\mu\tau_h$ pair obtained with the DeepTau classifiers shows a greater contribution from the DY$\rightarrow \tau_\mu\tau_h$ process and reduced contributions from other processes, compared to the distribution obtained with the old MVA-based classifiers.
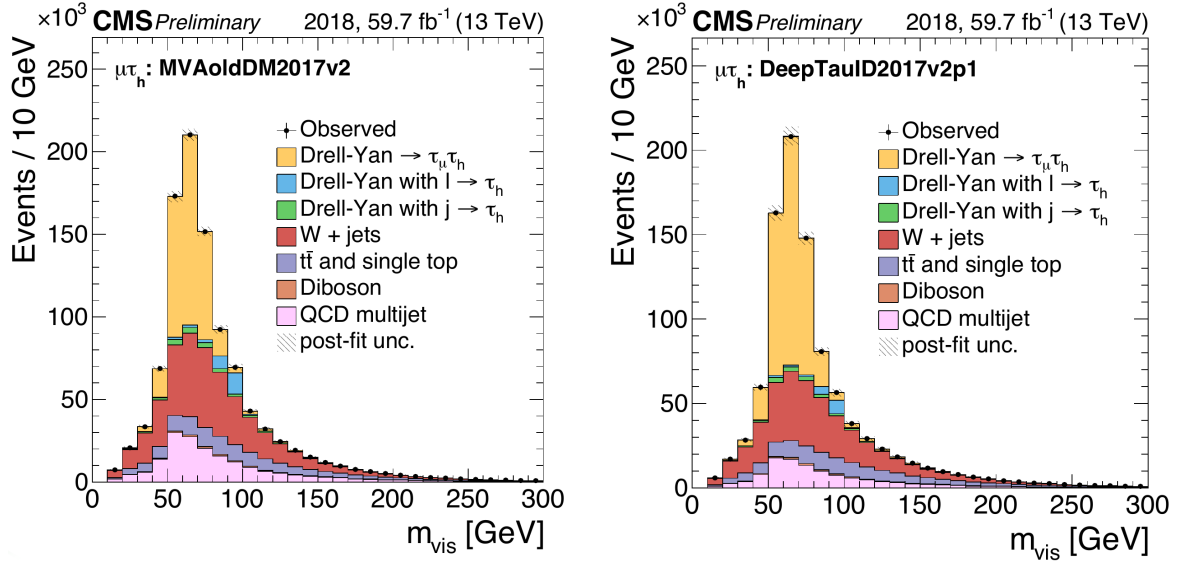

Figure 4.20: Visible mass distribution for $\tau_\mu\tau_h$ channel using the old MVA-based classifiers (**left**) and the DeepTau ones (**right**) [150].

Measurement of the $e \to \tau_h$ fake rate

Contents

A good level of agreement between simulated MC samples and real data is paramount to achieve precise and reliable measurements. In particular, the study of processes with low cross sections requires a detailed understanding of the processes which can mimic its signature. These analyses often involve the search of an excess in data compared to the total yield of known background processes, estimated from simulation. Any disagreement between data and MC samples would hinder these measurements and has to be cured with dedicated techniques. A possible source of disagreement is the mismodeling of physical object reconstruction efficiency in simulation. In the case of analyses involving tau leptons it is therefore important to correct for eventual mismodeling of the tau identification efficiency in simulation.

During Run 1 and Run 2, correction scale factors were determined [146, 147] for the MVA- and cut-based discriminators (see Appendix C) for the identification of genuine τ_h and the rejection of misidentified jets, muons and electrons.

After the deployment of the DeepTau identification (ID) algorithm, similar scale factors were determined for the DeepTau ID classifiers. As shown in Section 4.4, the DeepTau ID provides three distinct classifiers, one for each source of contamination. Correction

scale factors are required for each classifier, and are measured in regions where each contamination source becomes more relevant. These *fake enriched regions* are defined as:

- the $\tau_\mu\tau_h$ channel for the against-jet classifier, more precisely a region where the largest fraction of reconstructed events are produced by processes involving jets, like QCD multijet production, $t\bar{t}$ and W boson production in association with jets (W+Jets);

- the $\tau_\mu\tau_h$ channel for the against-muon classifier, in a region where most events come from the Drell-Yan (DY) production of a pair of muons ($Z \to \mu\mu$);

- the $\tau_e\tau_h$ channel for the against-electron classifier, using a DY enriched region similarly to what is done for the against-muon classifier, but selecting an electron pair instead of a muon one.

The studies performed to validate the DeepTau classifiers can be further divided depending on whether their efficiencies are evaluated on genuine τ_hs or on the contamination sources, i.e. their background rejection. To avoid confusion, the following notation is used in this chapter:

- *tau selection efficiency*: efficiency in the selection of genuine hadronically decaying tau leptons;

- *fake rate*: fraction of jets, electrons or muons which have been reconstructed as τ_h candidates by the HPS algorithm and pass the corresponding DeepTau classifier. They are used as an estimation for the misidentification rate of each object as τ_hs.

The measurement of the tau selection efficiency is the first one to be performed and leads to the definition of Working Points (WPs) for each classifier. These WPs are defined as the values of the classifiers such that:

$$N(P_{\tau \text{ vs } obj.} \geq \text{WP}(\varepsilon)) = \varepsilon N_{tot} \ , \tag{5.1}$$

where N_{tot} represent the number of events including genuine τ_h used to evaluate the classifier efficiency, $N(P_{\tau \text{ vs } obj.} \geq \text{WP}(\varepsilon))$ is the subset of those events for which the classifier ($P_{\tau \text{ vs } obj.}$) returned values larger than $WP(\varepsilon)$, with ε desired efficiency.

The measurements were performed using simulated $H \to \tau\tau$ events and allowed to define the WPs as shown in Table 5.1.

Using the defined WPs the misidentification rates of jets, electrons and muons as hadronically decaying tau leptons were then measured. The following sections go into further details on one of these measurements: the $e \to \tau_h$ fake rate.

WP	VVTight	VTight	Tight	Medium	Loose	VLoose	VVLoose	VVVLoose
VSjet	40%	50%	60%	70%	80%	90%	95%	98%
VSe	60%	70%	80%	90%	95%	98%	99%	99.5%
VSmu	-	-	99.5%	99.8%	99.9%	99.95%	-	-

Table 5.1: Target efficiencies for the selection of genuine τ_hs using the DeepTau classifiers (labeled VSjet, VSe and VSmu). The measurement was performed using τ_h with $p_T \in \{30, 70\}$ GeV from $H \to \tau\tau$ simulated events.

5.1 Analysis overview

As previously stated in Section 4.3, electrons can be misidentified as τ_h candidates by the HPS algorithm. The electromagnetic shower caused by the electron can extend to the hadronic calorimeter mimicking a hadron. Moreover, bremsstrahlung radiation can potentially take place due to the particle interacting with the tracker. The emitted photons can then be misidentified as the ones coming from a π^0, thus mimicking a strip. For that reason, electrons can be misidentified as hadronic tau leptons mainly for decay channels with one prong and one prong plus one strip.

The DeepTau NN is trained to discriminate genuine tau leptons from other sources of contamination, and a specific classifier is formulated to separate genuine hadronic tau leptons from electrons:

$$DeepTau2017v2p1VSe = \frac{P_\tau}{P_\tau + P_e} \, , \tag{5.2}$$

with P_τ and P_e NN scores corresponding respectively to the *genuine* τ_h and the *electrons* output nodes of the DeepTau NN. The DeepTau against-electron classifier is indicated in this chapter as $DeepTau2017v2p1VSe$, since it is the name used for storing its values inside miniAOD datasets in CMS. The name is also used to convey some additional information on the classifier: it was obtained by training the NN on 2017 data and MC simulations, and it used the version 2.1 of the training, corresponding to the NN architecture described in Section 4.4. Eight WPs are defined using this discriminator based on the rejection of electrons faking τ_h as shown in Table 5.1. In particular, a tighter WP corresponds to a higher rejection power against electrons and lower τ_h identification efficiency.

Fig. 5.1 shows the distribution of the $DeepTau2017v2p1VSe$ classifier in data and MC simulation. The black points correspond to data collected by the CMS detector during 2018 and sorted in the SingleElectron dataset, i.e. events identified by the presence of an isolated electron of sufficiently high transverse momenta. Table 5.2 shows the dataset divided into separate sets characterized by sharing the same detector conditions. The
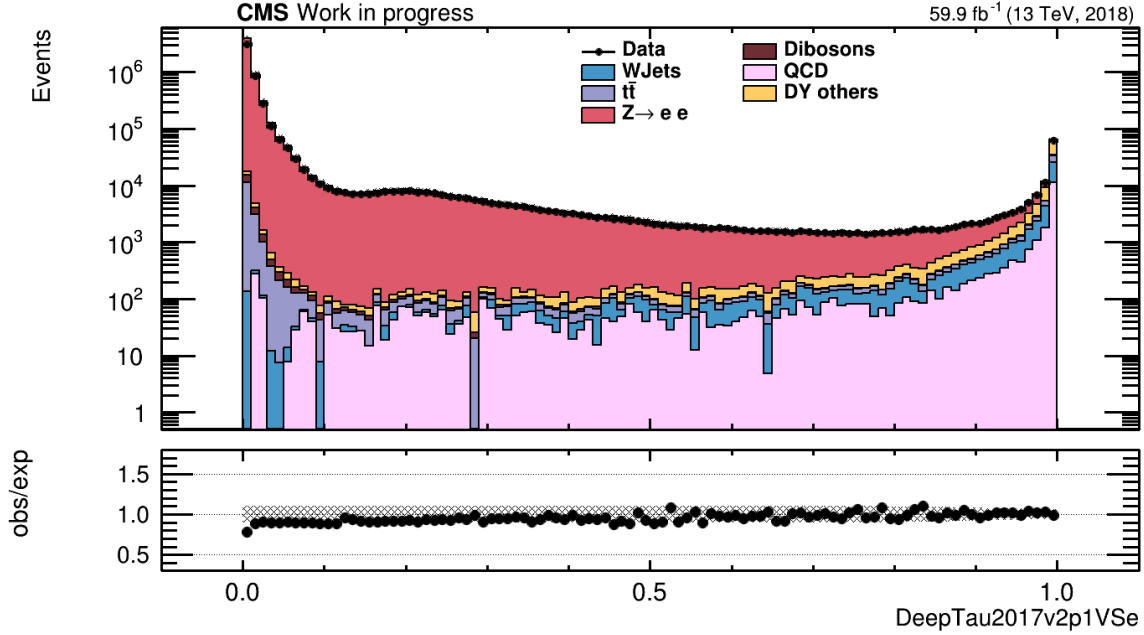
Figure 5.1: Top: *DeepTau2017v2p1VSe* distribution in data and MC for the 2018 datasets. **Bottom:** bin-by-bin ratio between the number of recorded events in data and the sum of all the simulated processes. The applied preselection is described in Section 5.2.

integrated luminosity across each dataset is also shown, together with the total integrated luminosity in the years of data-taking.

The stacked histograms correspond to the MC simulations of all relevant processes that can contribute to the final state of interest:

- $Z \to ee$ (red);

- DY others (yellow): mainly $Z \to \tau\tau$;

- WJets (blue);

- $t\bar{t}$ (purple);

- VV and single top production (brown).

All processes are estimated using MC generators, and their cross sections are listed in Table 6.4, as they are used both for the measurement presented in this chapter and for the one described in Chapter 6. The QCD multi-jet production, shown in pink, is estimated with a data-driven method, described in Section 5.2.1. The classifier shown in the figure is, as previously explained in Section 4.4, distributed between 0 and 1. A value closer to 1 corresponds to processes involving genuine τ_h, as seen by the higher relative

Table 5.2: SingleElectron datasets recorded by the CMS experiment during the Run 2 data-taking period and associated recorded luminosity. The integrated luminosity is also shown in bold for the years 2016, 2017 and 2018, and across the entire data-taking period.

Dataset	Recorded luminosity (fb^{-1})
Run2016B	5.79
Run2016C	2.57
Run2016D	4.25
Run2016E	4.01
Run2016F	3.10
Run2016G	7.54
Run2016H	8.61
Full 2016	**35.9**
Run2017B	4.82
Run2017C	9.66
Run2017D	4.25
Run2017E	9.28
Run2017F	13.54
Full 2017	**41.5**
Run2018A	14.00
Run2018B	7.10
Run2018C	6.92
Run2018D	31.92
Full 2018	**59.9**
Full Run2 data-taking	**137**

fraction of $Z \to \tau\tau$ events. Contribution from $Z \to ee$ events dominates instead the region corresponding to lower values of the classifier.

A clear trend is visible in the ratio plot shown in the lower part of Fig. 5.1: the ratio between observed events (collected data) and expected ones (obtained from simulation or data-driven methods) increases with respect to the classifier value. To correct this trend in simulation and improve its agreement with recorded data a Tag&Probe method is used to measure the misidentification rate of electrons as hadronic tau leptons in data and simulation. The ratio of the $e \to \tau_h$ fake rate (FR) in data and simulation is then applied as a scale factor (SF) to simulated events in order to correct the mismodeling observed in simulation.

The τ_h candidate found in simulated $Z \to ee$ events is an electron misidentified as an hadronically decaying tau lepton by the HPS algorithm. As such the $e \to \tau_h$ misidentification rate can be calculated as the fraction of simulated $Z \to ee$ events that pass the against-electron discriminator for a certain WP:

$$FR = \frac{N_{Z \to ee}^{pass}}{N_{Z \to ee}^{pass} + N_{Z \to ee}^{fail}} \; . \tag{5.3}$$

The labels *pass* and *fail* are in this chapter assigned to events which possess a τ_h candidate that has been respectively identified as a genuine tau or an electron by the against-electron discriminator.

Measuring the same quantity in recorded data is inherently more complicated due to not being able to assert with certainty whether a τ_h candidate was indeed a genuine τ_h or not. A Tag&Probe method provides a good strategy for such a measurement. It requires to select a reconstructed object as a *tag*, a well identified object whose nature is assumed to be known with certainty both in data and simulation. Another object, reconstructed in the same event, is then used to *probe* the efficiency of the classifier being studied. The strength of this method lies in selecting a tag and probe pair which can be easily assigned to a well identifiable physical process. The decay of a Z boson to leptons has a decay width of (83.984 ± 0.086) MeV [23] leading to a sharp resonance in the invariant mass distribution of the electron-positron pair. By measuring the $e \to \tau_h$ FR in a region enriched in $Z \to ee$ events it is possible to select a well reconstructed electron with a high degree of confidence that it was indeed an electron, and a loosely identified τ_h candidate which is in most instances a misidentified electron. These two objects serve as the tag and the probe for the measurement, and their selection is described in the following section.

5.2   Event selection

This analysis was performed by studying the $\tau_e\tau_h$ final state for data recorded during the Run 2 data-taking period. The $e \to \tau_h$ FR was measured separately for each year of data-taking in order to account for different detector conditions. The SingleElectron dataset was used for the measurement amounting to an integrated luminosity of 137 fb^{-1}. Recorded and simulated events are selected using triggers (see Section 3.2.5) which require the presence of an isolated electron of sufficiently high transverse momentum. The triggers used are the SingleElectron25 trigger in 2016 and SingleElectron35 in 2017 and 2018. The corresponding high level trigger (HLT) paths and online p_T thresholds are shown in Table 5.3.

The trigger choice is based on the need of a well reconstructed electron in the event. These triggers have high efficiency and contributions from jets misidentified as electrons are kept under control. Events are then selected by requiring a well identified electron which satisfies the following conditions:

Table 5.3: Triggers used for the measurement of the $e \to \tau_h$ misidentification rate.

Year	Name	HLT path	p_T threshold
2016	SingleElectron25	HLT_Ele25_eta2p1_WPTight_Gsf_v	$p_T > 25$ GeV
2017	SingleElectron35	HLT_Ele35_WPTight_Gsf_v	$p_T > 35$ GeV
2018	SingleElectron35	HLT_Ele35_WPTight_Gsf_v	$p_T > 35$ GeV

- $p_{T,e} > 36$ GeV, $|\eta_e| < 2.1$;

- it should pass the Tight WP of the BDT-based electron identification (described in Section 4.1.4) and relative isolation $I_{rel}^e < 0.1$ (defined in Eq. 4.4);

- the tag should match the electron which fired the trigger within $\Delta R < 0.5$;

- $d_z < 0.2$ cm, $d_0 < 0.045$ cm, with d_z and d_0 longitudinal and transverse components of the impact parameter;

- $m_T < 30$ GeV, with $m_T = \sqrt{2p_{T,e}E_T^{mis}(1 - \cos \Delta\phi_{e,met})}$, with E_T^{mis} missing transverse energy (MET) computed with the PF algorithm and $\phi_{e,met}$ azymuthal separation between electron and MET, this is done to reject electrons produced in W+Jets events.

The probe is then selected as a loosely identified τ_h candidate identified by the HPS algorithm. It should also pass the following selection:

- have opposite charge with respect to the electron;

- $p_{T,\tau} > 20$ GeV, $|\eta_\tau| < 2.3$;

- pass the following WPs for the DeepTau classifiers: Medium, VLoose and VVVLoose for the against- jet, muon and electron respectively;

- have angular distance $\Delta R > 0.5$ from the electron direction of flight.

For events which have not passed the VVVLoose WP of the against-electron classifier a strong discrepancy was observed between data and simulation in 2016 and 2017. These events are characterized by a low NN score for the τ_h class in the DeepTau NN and are therefore sensitive to the WPs used for the against-jet and against-muon classifiers and the corresponding SFs applied. The discrepancy observed hinted at a mismodeling which could not be solved by the application of a SF and could worsen the precision of the SF calculation for the tighter WPs.

The VVVLoose against-electron was therefore applied as a preselection for the measurement. This was done also taking into account that the measured SFs are meant to be applied in physics analysis targeting a final state with an electron and a τ_h. The VVLoose WP has similar efficiency for the selection of genuine τ_h and can reduce the misidentification rate by a factor 2 compared to the loosest WP. The decision was therefore taken to recommend the use of the VVLoose WP instead of the VVVLoose one, and the SFs computed applying the loosest WP as a preselection.

As previously mentioned, the strength of a Tag&Probe method comes from targeting well identifiable physical processes. The measurement is therefore carried out in the region where the highest contribution from the $Z \to ee$ process is present. The selected region is defined based on the invariant mass of the tag and the visible decay products of the probe, i.e. the visible mass (m_{vis}). A region is selected, centered around the Z boson mass, between 60 and 120 GeV, since for $Z \to ee$ events the invariant mass of the two electrons peaks at 91 GeV, while for $Z \to \tau\tau \to \tau_e\tau_h$ events the invariant mass of the visible decay products peaks at lower values due to the presence of neutrinos in the decay.

The measurement has the goal of providing correction scale factors to be applied on MC simulation to correct the rate of electron misidentification with respect to data. Other discrepancies between data and MC need to be taken into consideration before performing the measurement. This is done by applying the following scale factors on simulated events:

- pile-up correction, which allows to correct the pile-up distribution in simulated samples in order to better describe the number of pile-up interactions in data;

- electron identification and isolation, obtained by measuring the efficiency of electron identification and isolation in di-electron events selected in simulation and data;

- trigger efficiency, measured in the same final state as the previous correction;

- scale factor to correct for the DeepTau vs jet discriminator efficiency, this is measured in the $\mu\tau_h$ final state as a function of $p_{T,\tau}$;

- Z p_T reweighting, measured using the $Z \to \mu\mu$ process and used to correct the $Z \to \tau\tau$ simulation around the Z peak;

- recoil correction for missing transverse energy (MET), used to correct mainly the jet contribution to the MET.

5.2.1   Data-driven estimation of QCD background

Before describing the measurement of the misidentification rate of electrons as hadronically decaying tau leptons, it is useful to mention how the QCD multijet production process is treated. Unlike other processes which are taken directly from simulation, the QCD multijet production is estimated with a data-driven method.
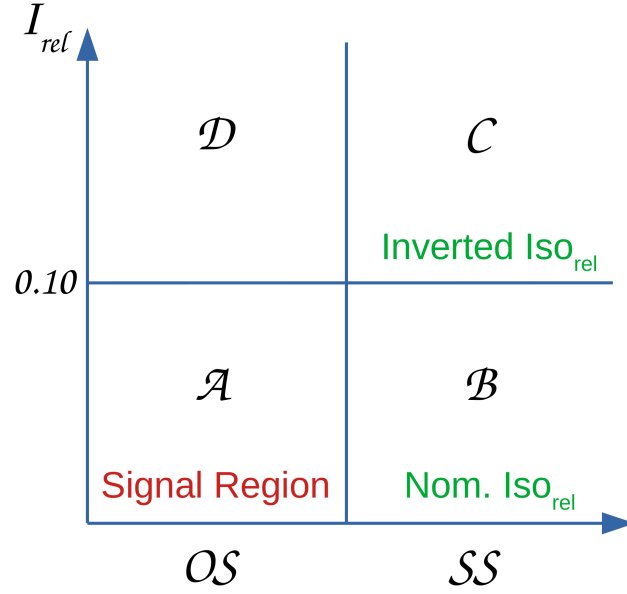
Figure 5.2: Schematic view of the ABCD method used for data-driven estimation of the QCD multijet production process.

The method used is commonly referred to as ABCD method: based on two uncorrelated variables the phase space is divided in four distinct region as shown in Fig. 5.2. For illustration purposes let us define the regions in the following way:

A) Signal Region (SR): which requires an isolated electron and a τ_h candidate having opposite electric charge;

B) Same Sign (SS) isolated region: obtained by selecting an isolated electron and a τ_h candidate having the same electric charge;

C) SS anti-isolated region: similar selection compared to region B, but with inverted cut on electron isolation;

D) Opposite Sign (OS) anti-isolated region: same selection as for region A inverting the cut on electron isolation.

Since the variables are uncorrelated with each other, any selection performed on recorded or simulated events using one variable will select the same fraction of events regardless of what selection is applied on the other variable. By indicating with A, B, C, D the number

of events in each region for a certain physical process, the following relation is derived:

$$\frac{A}{B} = \frac{D}{C} = SF_{SSOS} \ .$$

(5.4)

By measuring the value of $SF_{SSOS} = D/C$ in data, the QCD multijet production contribution in the SR can be estimated as:

$$QCD_{SR} = SF_{SSOS} \left(Data_B - \sum^{bkg.} MC_B \right) \ .$$

(5.5)

In practice, contributions from simulated MC samples (MC_B) are subtracted from the recorded data ($Data_B$) in region B. The obtained distribution is then scaled to estimate the QCD multijet production in the SR by multiplying it by the scale factor SF_{SSOS} .

5.3   Fit of the $e \rightarrow \tau_h$ fake rate in data

Eq. 5.3 provides an estimation of the $e \rightarrow \tau_h$ misidentification rate when performed on simulated $Z \rightarrow ee$ events. In order to measure the misidentification rate in real pp collision it is necessary to estimate the number of $Z \rightarrow ee$ events in data. This is done as a Maximum Likelihood (ML) fit of the modeled processes of interest to recorded data, performed using the Combine statistical toolkit [155]. The analytic expression of the likelihood function is provided in the next Section while here the categories and distributions used in the fit are described.

As previously mentioned, the processes having an isolated electron and an τ_h candidate in the final state are:

- $Z \rightarrow ee$ (ZEE);

- $Z \rightarrow \tau\tau$ (ZTT);

- Drell-Yan (DY) production of two leptons in association with a jet misidentified as a hadronically decaying tau lepton (ZJ);

- W boson production in association with jets (W);

- $t\bar{t}$ production;

- di-boson and single top production, grouped together in the VV template;

- QCD multijet production (QCD).

The ZEE process is treated as a signal in the fit, as it is the one used to estimate the $e \to \tau_h$ FR in data. The other processes are instead treated as a background for the measurement. The fit is performed using the distributions of these processes with respect to selected variables, which will be later specified, referred to as *templates* in this chapter.

Signal and background templates are then fit to the recorded data, separately for:

- each year of data-taking;

- each WP of the against-electron classifier;

- two regions in pseudorapidity defined by the τ_h direction of flight:

    - *Barrel*: $|\eta_\tau| < 1.460$;
    - *Endcap*: $1.558 < |\eta_\tau| < 2.3$;

The division between years is done based on the upgrades and maintenance performed on the CMS detectors between different years, previously described in Section 3.1.1.

The separation between WPs is done for two main reasons: one mainly concerning the physics measurement, and the other based on how these scale factors are used in physics analyses. As previously shown in Fig. 5.1, the agreement between simulation and recorded data can vary according to well defined trends. Measuring the scale factor to correct the FR for different WPs allows to properly correct for such trends in simulation. From the practical point of view, the scale factors provided by this measurement are used by physics analyses which require the identification of hadronic tau leptons. In each of these physics analyses, a certain WP of the against-electron discriminator is applied at the level of event selection. To ensure the best possible agreement between data and simulation the measured scale factors are therefore provided separately for each WP.

Lastly, the fit is performed separately for τ_h candidates reconstructed in the Barrel and in the Endcap regions to account for the different ECAL geometry. The pseudorapidity region $1.460 < |\eta| < 1.558$ is excluded from the measurement, as the tau identification efficiency becomes noticeably lower, leading to a decrease in the overall precision of the measurement.

The main parameter of interest (POI) for the fit is the scale factor needed to correct the mismodeling of the misidentification rate in simulation: ε_{FR}. It is defined as the ratio between the FR estimated in data via the ML fit (FR'), and the one estimated in $Z \to ee$ simulation (FR):

$$\varepsilon_{FR} = \frac{FR'}{FR} \ . \tag{5.6}$$

Both FR and FR' are evaluated based on the number of events that pass and fail the selection operated by a certain WP of the against-electron classifier. Therefore the ML fit is performed simultaneously in two regions:

- PASS: the event has passed the against-electron discriminator, the $Z \to ee$ events in this region correspond to instances when an electron is misidentified as a τ_h;

- FAIL: the event was rejected by the against-electron discriminator.

The FR is estimated as the fraction of ZEE events which pass a certain WP of the against-electron classifier. A correction applied to the misidentification rate does not affect the number of preselected events, i.e. the denominator of Eq. 5.3,

$$N^{tot} = N^{pass}_{Z \to ee} + N^{fail}_{Z \to ee} \ . \tag{5.7}$$

The total number of reconstructed events is kept constant by varying ε_{FR}. A correction applied to the misidentification rate is therefore interpreted as a migration of events between the two regions: an increase to the FR leads to a lower number of events being assigned to the FAIL region and more events be assigned to the PASS one. This means that the ε_{FR} parameter is anti-correlated between the two regions.

Taking into account that ε_{FR} does not affect the number of preselected events, Eq. 5.6 can be rewritten as

$$\varepsilon_{FR} = \frac{FR'}{FR} = \frac{N'^{\,pass}_{Z \to ee}}{N^{pass}_{Z \to ee}} \ , \tag{5.8}$$

with $N'^{\,pass}_{Z \to ee}$ representing the number of ZEE events estimated in data in the PASS region. In the fit the ZEE normalization in the PASS region is therefore scaled by ε_{FR} in order to measure the misidentification rate.

In simulation, the number of ZEE events in the FAIL region can be calculated by inverting Eq. 5.3, and is equal to

$$N^{fail}_{Z \to ee} = (1 - FR)N^{tot}_{Z \to ee} \ . \tag{5.9}$$

Varying ε_{FR} alters the ZEE normalization in the FAIL region as

$$N'^{\,fail}_{Z \to ee} = \frac{1 - FR'}{1 - FR} N^{fail}_{Z \to ee} \ , \tag{5.10}$$

with $N'^{\,fail}_{Z \to ee}$ representing the number of ZEE events estimated in data in the FAIL region.

The total number of ZEE events across the two regions is then assigned an uncertainty, estimated with a parameter which scales coherently the ZEE normalization in the PASS and FAIL regions. This and other systematic uncertainties for the measurement are detailed in Section 5.3.2.

To accurately measure the correction to the misidentification rate, the uncertainty on the ZEE normalization has to be constrained. This is done in the fit by using the number of events[1] in the FAIL region, which is dominated by ZEE events as shown by the plot on the right in Fig. 5.3.

---

[1]For practical reasons, the number of events in the failed region was obtained by filling a histogram of the m_{vis} distribution using only 1 bin. The label used on the x axis of the related plots is a consequence of this fact.

The m_{vis} distribution was instead chosen for the fit in the PASS region. The ZEE template peaks around 90 GeV in this distribution, allowing higher sensitivity to the $Z \to ee$ process compared to the various backgrounds. This distribution also allows to constrain, especially for the tighter WPs, the normalization and systematic uncertainties of background templates. This is due to the regions at low or high m_{vis}, in the considered range, having higher contributions from background processes, ZTT in particular. This can be observed in the left plot in Fig. 5.3, which shows the m_{vis} distribution in the range between 60 and 120 GeV for the Tight WP of the against-electron classifier.
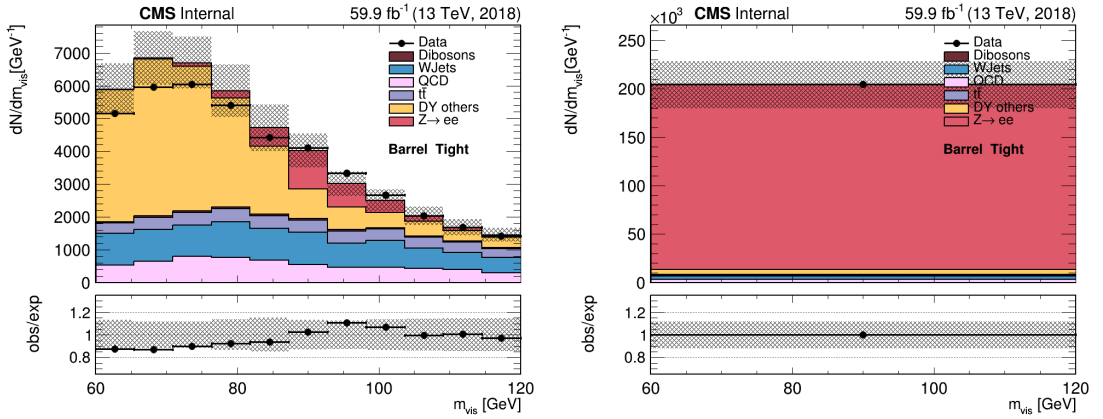
Figure 5.3: Distributions for the Tight against-electron WP, Barrel region in 2018. **Left:** m_{vis} distribution in the PASS region. **Right:** Number of events in the FAIL region.

5.3.1 Fit model

The main parameter of interest in the fit is ε_{FR}, which is anti-correlated between the PASS and FAIL regions. It affects the normalization of the ZEE template in the two regions and is treated as a *rate parameter*: it is allowed to vary freely within a defined range. This range was initially set to $[0,2]$ and occasionally adapted for specific WPs.

As previously mentioned, the normalization of the ZEE template is also altered by a parameter which is correlated between the two regions. It is a rate parameter allowed to vary freely within the range $[0,10]$, and labeled ε_{vsJet}. The shape of the template is also altered by two parameters representing the uncertainty on the energy scale of isolated electrons (the tag), and electrons misidentified as τ_h (the probe). The latter parameter affects only the ZEE template and is referred to as Fake Energy Scale (FES). Other parameters in the fit are also introduced to alter the shape or normalization of background templates.

These parameters are usually referred to as *nuisance* parameters ($\vec{\theta}$) and they represent the systematic uncertainties of the measurement. The likelihood can be written as a

function of the aforementioned parameters as in [156]:

$$L(\varepsilon_{FR}, \varepsilon_{vsJet}, FES, \vec{\theta}) =$$

$$\prod_{j=\{pass,fail\}} \prod_{i}^{N_{bin}^{j}} P\left(n_{i,j} | S_{i,j}(\varepsilon_{FR}, \varepsilon_{vsJet}, FES, \vec{\theta}) + B_{i,j}(\vec{\theta})\right) \times \prod_{m}^{N_{nuis}} C_m(\theta_m | \tilde{\theta}_m) , \qquad (5.11)$$

where P is the Poisson distribution, representing the probability of measuring a number of events ($n_{i,j}$) in the bin i of the category j based on the expected number of events for the signal ($S_{i,j}$) and the total background ($B_{i,j}$). The value of $B_{i,j}$ depends only on the nuisance parameters, while $S_{i,j}$ depends also on the POI and the ε_{vsJet} and FES parameters. The factors $C_m(\theta_m | \tilde{\theta}_m)$ represent the *a priori* distributions for the N_{nuis} nuisance parameters, with $\tilde{\theta}_m$ representing the nominal value of the θ_m parameter. The first product acts on the PASS and FAIL categories, while the second affects the bins used in each category. The FAIL category represents the number of events, and therefore contains only 1 bin ($N_{bin}^{fail} = 1$), while $N_{bin}^{pass} = 11$.

Instead of maximizing the likelihood, the Combine statistical toolkit minimizes the negative log-likelihood

$$NLL \equiv -\log\left(L(\varepsilon_{FR}, \varepsilon_{vsJet}, FES, \vec{\theta})\right) , \qquad (5.12)$$

using Minuit [157], a minimization routine present in the toolkit.

5.3.2   Systematics

The following notation is used in this section in order to discuss the systematic uncertainties:

- **rate parameter**: it alters the normalization of a process. The corresponding nuisance parameter is assigned a probability density function (pdf) uniform within a defined range.

- **lnN**: log-normal uncertainty, it alters the normalization of a process and the corresponding nuisance parameter is assigned a log-normal pdf.

- **shape**: dedicated templates are created representing the systematic variation with respect to the nominal distribution, it is used to treat uncertainties which do not simply alter the normalization of a template. The corresponding nuisance parameter is assigned a Gaussian pdf.

- **shapeU**: uniform shape uncertainty, it alters the shape of the distribution like the shape uncertainties. The associated nuisance parameter is instead assigned a uniform pdf within a given range.

There are two uncertainties which only affect the ZEE template:

- ε_{vsJet}: the correction to the against-jet classifier efficiency on ZEE events;

- FES: the energy scale of electrons misidentified as hadronically decaying tau leptons.

The first parameter was not used in the determination of the correction SF using the previously used MVA- and cut-based Tau ID classifiers (see Appendix C for a description of the classifiers and [146, 147] for the associated SF measurements). It was introduced specifically for this measurement, and is therefore explained in greater detail compared to the other uncertainties.

The three DeepTau ID classifiers have been defined in Eq. 4.31, reported here for convenience of the reader:

$$P_{\tau \text{ vs } obj.} = \frac{y_\tau}{y_\tau + y_{obj.}} \ , \tag{5.13}$$

with y_τ representing the output score of the genuine tau class of the NN, and $y_{obj.}$, with $obj. = j, e, \mu$, the three distinct output scores for each misidentified object. The four output scores of the NN sum to 1, therefore events having $y_\tau \sim 1$ are associated to $y_{obj.} \sim 0$ for all sources of contamination. In these cases $P_{\tau \text{ vs } obj.} \sim 1$ for all classifiers. A similar consideration applies to events associated to $y_\tau \sim 0$, for which one can expect all three classifiers present low values. When a cut is applied on a certain WP of a classifier, it can have different efficiency in data and MC not only on the genuine τ_h but also on misidentified objects. The ε_{vsJet} was therefore introduced to account for a possible mismodeling of the against-jet classifier efficiency on ZEE events.

The other parameter which alters exclusively the ZEE template is a shapeU uncertainty on the energy scale (ES) of the electrons misidentified as τ_h. The dedicated templates representing the FES are defined as follows:

- the upward fluctuation of the FES is represented by the m_{vis} distribution obtained when the probe 4-momentum is scaled by 1.25;

- the downward fluctuation is obtained similarly, with the 4-momentum being multiplied by 0.75.

The ε_{vsJet} and FES parameters are both associated to uniform distributions. In the fit, they float freely within a defined range: $[0, 10]$ for ε_{vsJet} and $[0.75, 1.25]$ for the FES.

The normalization of the MC simulated processes have theoretical and experimental uncertainties, they are listed in Tab. 5.4 and are labeled "X normalization" with X being the process to which the uncertainty is assigned. The QCD template is estimated with the data-driven method described in Section 5.2, and is also assigned a lnN uncertainty which accounts for the uncertainty on the scale factor between same sign and opposite sign regions (see Section 5.4). The W template normalization is instead corrected using

Table 5.4: Summary table of the systematic uncertainties for the $e \to \tau_h$ FR measurement.

Description	Value	Templates affected	Type
τ identification efficiency	10%	ZTT, $t\bar{t}$, VV	lnN
Against-jet discriminator efficiency on ZEE	freely-floating	ZEE	rate parameter
Electron isolation identification and trigger	5%	All except QCD and W	lnN
Jet$\to \tau_h$ fake rate	30%	ZJ	lnN
Normalization for data-driven estimated/corrected processes	20%	QCD and W	lnN
DY normalization	5%	ZJ, ZEE, ZTT	lnN
$t\bar{t}$ normalization	10%	$t\bar{t}$	lnN
VV normalization	10%	VV	lnN
Luminosity uncertainty	2.6%	All except W and QCD	lnN
Electrons faking tau leptons energy scale	25%	ZEE	shapeU
τ energy scale	5%	ZTT	shape
Electron energy scale	1%(B) 2.5%(E)	All except QCD	shape
Visible mass resolution	10%	ZEE	shape

a SF determined in the region $m_T > 80$ GeV, and assigned an uncertainty similar to the QCD process.

Other normalization uncertainties include:

- The electron identification, isolation and trigger: an uncertainty of 5% is applied to all processes except QCD and W. It represents the uncertainty on the electron reconstruction, identification and on the trigger efficiency.

- Luminosity uncertainty: approximated to 2.6% for all years, it is applied on all templates except QCD and W.

- $j \to \tau_h$ FR uncertainty: 30% uncertainty for the jets misidentified as τ_h for DY production in association with jets (ZJ).

Shape systematics are computed by altering the nominal template as a function of the associated nuisance parameter. As such, the numerical value for these uncertainties

varies from bin to bin and it is best to provide just a general description of how the altered templates are determined for the m_{vis} distribution:

- τ energy scale (ES): the corresponding templates are obtained similarly to the FES. The probe 4-momenta for processes involving genuine τ_h is scaled by 1.05 or 0.95 to determine the upward and downward fluctuations of the m_{vis} distribution. The 5% uncertainty is interpreted as the standard deviation of the Gaussian pdf associated to the τ ES.

- Electron ES: this parameter varies the 4-momentum of the tag momenta, scaling it by 1% in the Barrel region (labled B in Table 5.4) and by 2.5% in the Endcap (corresponding to the label E), and is also constrained to follow a Gaussian distribution.

- Visible mass resolution: $m'_{vis} = \alpha_{res}(m_{vis} - 91) + 91$, the distribution is widened or tightened by the parameter α_{res} with respect to the position of the Z peak.

5.3.3 Results

As previously stated, the aim of the measurement is to calculate corrections for the ZEE MC simulations to better describe the misidentification of electrons as hadronic tau leptons observed in data. This section thus reports the fit results for the three parameters which exclusively affect the ZEE template:

- ε_{FR};

- against-jet discriminator efficiency for ZEE (ε_{vsJet});

- $e \to \tau_h$ energy scale, or Fake Energy Scale (FES).

Of these parameters, only the ε_{FR} corresponds to the parameter of interest (POI) of the fit and is measured with the highest precision as a result. The other two are treated as nuisance parameters, and thus are measured with lower precision. They are however freely floating within their corresponding ranges. Their post-fit values can therefore still be used to calculate more precise corrections for the MC simulation.

As previously mentioned, results are presented separately for:

- each year of data-taking: 2016, 2017 and 2018;

- each WP of the against-electron classifier[2]: VVLoose, VLoose, Loose, Medium, Tight, VTight and VVTight;

- two regions in pseudorapidity defined by the τ_h direction of flight:

---

[2]Excluding the VVVLoose one which is applied as a preselection.

    − *barrel*: $|\eta_\tau| < 1.460$;

    − *endcap*: $1.558 < |\eta_\tau| < 2.3$;

To describe the results of the ML fit, an example of pre-fit and post-fit distributions in the PASS and FAIL regions is presented in Fig. 5.4. In these distributions it is possible to observe the events for 2018 data and MC in the Barrel region, after a cut on the Medium WP of the against-electron discriminator. The left side of the figure shows the pre-fit distributions for the visible mass distribution in the PASS region (in the top row) and the number of events in the FAIL region (bottom row), the corresponding distributions post-fit are shown on the right. The post-fit values of the three parameters affecting the ZEE template corresponding to these distributions can be found in Table 5.7 and are also reported here: $\varepsilon_{FR} = 1.16 \pm 0.05$, $\varepsilon_{vsJet} = 1.08 \pm 0.07$ and FES$=1.035 \pm 0.008$.

The positive value of the $e \rightarrow \tau_h$ ES corresponds to an increase to the probe ES for the ZEE template. This causes a reduction of the number of events at lower m_{vis} and an
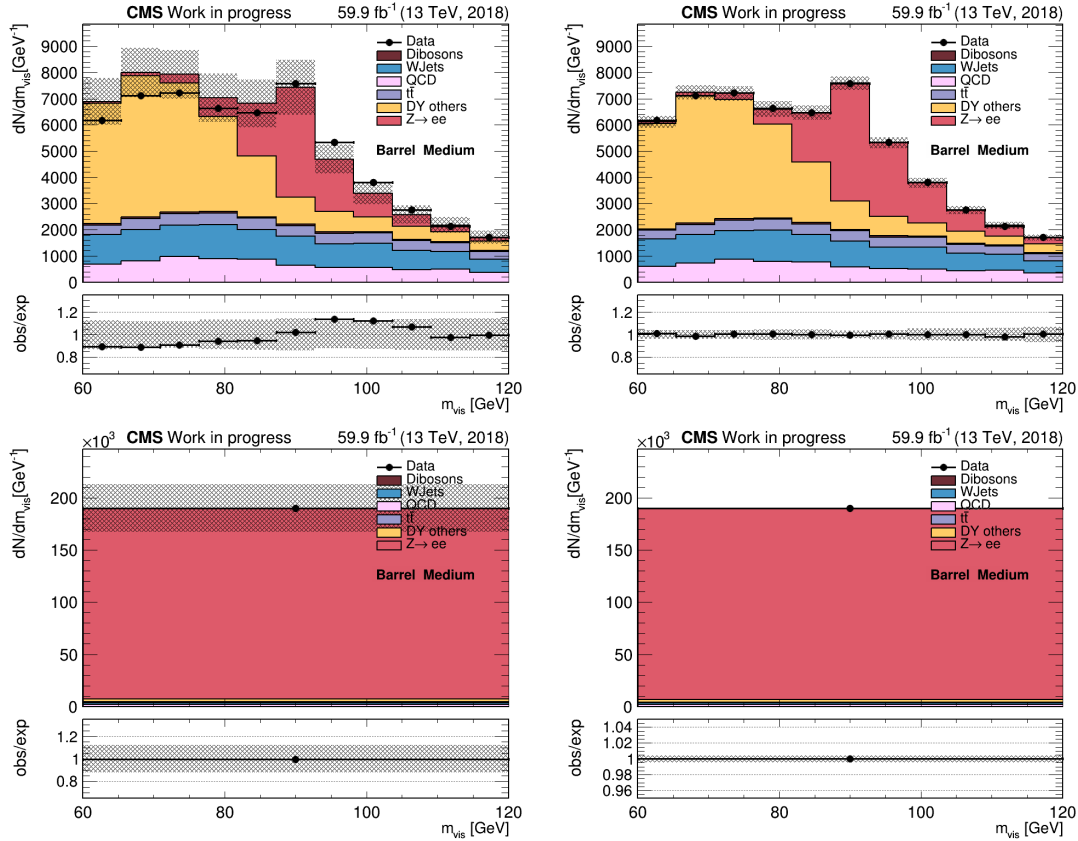
Figure 5.4: Distributions for the Medium WP, Barrel region in 2018: m_{vis} distribution in the PASS region pre-fit (**top left**) and post-fit (**top right**). Number of events in the FAIL region pre-fit (**bottom left**) and post-fit (**bottom right**).

increase at higher m_{vis}. In this case it compensates the deficit of MC compared to data which is found around 100 GeV.

The other two parameters both affect the normalization of the ZEE template. The ε_{FR} parameter is anti-correlated between the two regions, and does not alter the overall number of ZEE events. The ε_{vsJet} scales instead the ZEE normalization coherently in the two regions. This results in the ZEE normalization in the PASS region being scaled by the product of the two factors,

$$\varepsilon_{PASS} = \varepsilon_{FR} \times \varepsilon_{vsJet} = 1.25 \pm 0.14 \ , \tag{5.14}$$

while the in the FAIL region the two parameters interfere with each other and the normalization scales by

$$\varepsilon_{FAIL} = \frac{1 - \varepsilon_{FR} \times FR}{1 - FR} \times \varepsilon_{vsJet} = 1.07 \pm 0.07 \ , \tag{5.15}$$

with $FR = 0.06$ being the pre-fit FR measured in MC, and the uncertainty being computed assuming the two parameters as uncorrelated.

Tables 5.5, 5.6 and 5.7 show the results obtained for 2016, 2017 and 2018 respectively. The values shown in the three tables are the pre-fit FR and the post-fit values of the fit parameters affecting the ZEE template, with an uncertainty expressed as the symmetrized 68% CL interval. For some of the tightest WPs the number of ZEE events which pass the selection is lower than the bin-by-bin statistical fluctuation of the background processes. This means that the fit can converge even assuming that there are no ZEE events in the PASS region, as any discrepancy between the background models and the recorded data can be replicated by the systematic uncertainties of the backgrounds. This can be observed in Fig. 5.5 and is also shown by the post-fit value of the the FR scale factor ($\varepsilon_{FR} = 0.5 \pm 1.0$) being compatible with 0 within 1 standard deviation. Given that ε_{FR} is anti-correlated between the PASS and the FAIL regions, it is a good tool to assess how relevant the ZEE process is for making the fit converge in the PASS region. For contrast, ε_{vsJet} is correlated between the PASS and FAIL regions and its value (0.92 ± 0.07 in this example) is mostly constrained from the FAIL region.

As previously stated, the goal of this measurement is to determine scale factors to correct the modeling of electron misidentification as τ_h in simulation. These scale factors are meant to be applied to all analyses in CMS which study final states involving hadronically decaying tau leptons. A certain WP of the DeepTau against-electron classifier is applied in these analyses at the level of event selection. The WP is chosen in each of these analyses depending on the specific final state of interest, with a tighter WP being usually applied to final states with an electron identified alongside a τ_h. The SFs are applied to events in simulation which passed the selection operated with the against-electron classifier, and where the τ_h is a misidentified electron.

The SFs aim at correcting the normalization of processes involving misidentified electrons which passed the corresponding WP of the against-electron classifier. These SFs
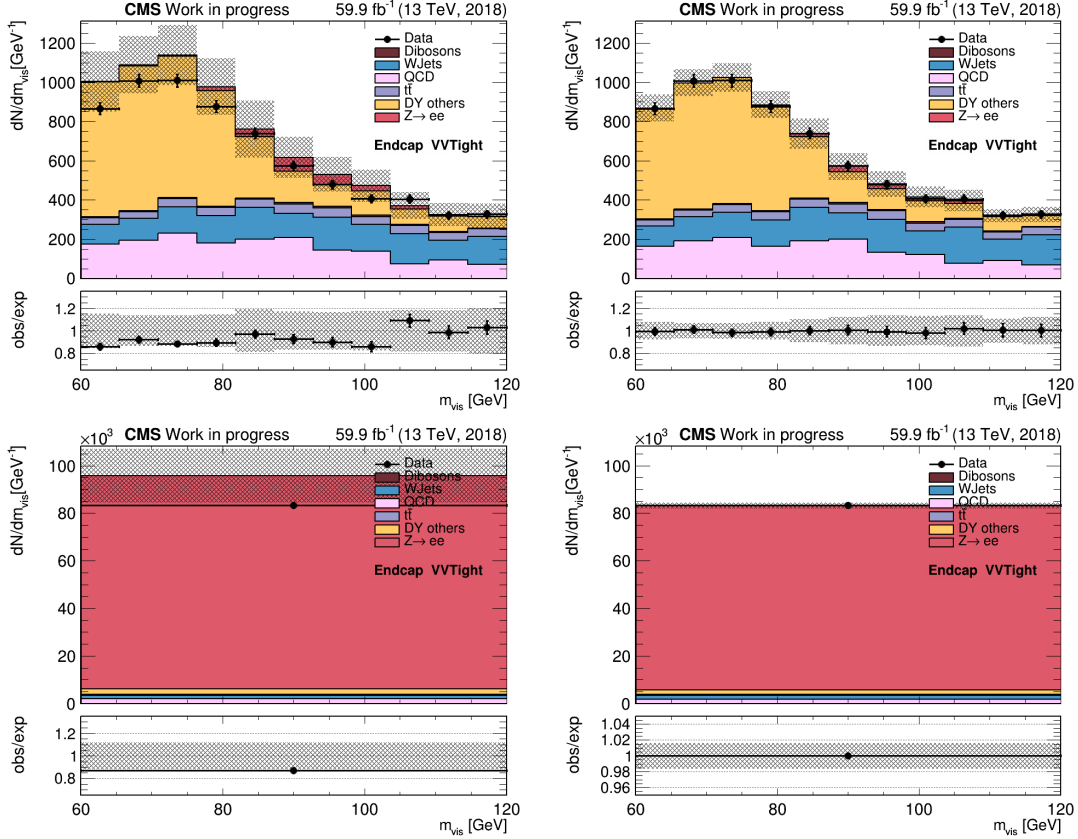
Figure 5.5: Distributions for the VVTight WP, Endcap region in 2018: m_{vis} distribution in the PASS region pre-fit (**top left**) and post-fit (**top right**). Number of events in the FAIL region pre-fit (**bottom left**) and post-fit (**bottom right**).

correspond to the correction applied to the ZEE normalization in the PASS region. As a function of the fit parameters, this is determined as:

$$SF = \varepsilon_{FR} \times \varepsilon_{vsJet} \; , \tag{5.16}$$

while the uncertainty on the SF is determined by treating the two fit parameters are uncorrelated:

$$\frac{\Delta SF}{SF} = \frac{\Delta \varepsilon_{FR}}{\varepsilon_{FR}} + \frac{\Delta \varepsilon_{vsJet}}{\varepsilon_{vsJet}} \; . \tag{5.17}$$

The values of these SFs are presented in the next section, where a comparison between this measurement and the ones performed by changing the against-jet WP is shown. The SFs which have been recommended for the use in analyses preformed by the CMS experiment can be found in the column labeled *Medium* of Tables 5.8, 5.9 and 5.10 for 2016, 2017 and 2018 respectively.

The measured SFs have been used in all analyses in CMS which study the Run 2 data-taking period and use the DeepTau ID for the identification of hadronically decaying tau leptons [65, 158].

Table 5.5: $e \to \tau_h$ FR estimated in simulation, presented together with the scale factors obtained to correct the $e \to \tau_h$ FR, ES and against-jet efficiency for 2016. The uncertainties represent the 68% confidence level on the measured fit parameter.

	Barrel			
WP	pre-fit FR	ε_{FR}	ε_{vsJet}	ES
VVLoose	0.67 ± 0.03	0.999 ± 0.003	1.38 ± 0.11	1.020 ± 0.003
VLoose	0.33 ± 0.03	0.976 ± 0.010	1.25 ± 0.09	1.030 ± 0.003
Loose	0.138 ± 0.015	0.97 ± 0.03	1.32 ± 0.09	1.033 ± 0.005
Medium	0.054 ± 0.007	1.05 ± 0.06	1.37 ± 0.10	1.020 ± 0.015
Tight	0.016 ± 0.002	0.93 ± 0.29	1.31 ± 0.09	1.06 ± 0.04
VTight	0.0066 ± 0.0008	1.14 ± 0.33	1.33 ± 0.09	1.10 ± 0.06
VVTight	0.0026 ± 0.0003	1.82 ± 0.66	1.33 ± 0.09	1.03 ± 0.10
	Endcap			
WP	pre-fit FR	ε_{FR}	ε_{vsJet}	ES
VVLoose	0.66 ± 0.03	0.990 ± 0.006	1.30 ± 0.1	1.010 ± 0.003
VLoose	0.30 ± 0.03	0.90 ± 0.02	1.25 ± 0.09	1.023 ± 0.008
Loose	0.119 ± 0.013	0.80 ± 0.07	1.24 ± 0.09	1.043 ± 0.018
Medium	0.051 ± 0.006	0.84 ± 0.11	1.29 ± 0.10	1.07 ± 0.04
Tight	0.016 ± 0.002	1.12 ± 0.27	1.31 ± 0.10	1.100 ± 0.018
VTight	0.0055 ± 0.0007	1.25 ± 0.66	1.27 ± 0.09	1.12 ± 0.07
VVTight	0.0021 ± 0.0003	1.83 ± 1.77	1.31 ± 0.10	1.11 ± 0.10

Table 5.6: $e \to \tau_h$ FR estimated in simulation, presented together with the scale factors obtained to correct the $e \to \tau_h$ FR, ES and against-jet efficiency for 2017. The uncertainties represent the 68% confidence level on the measured fit parameter.

WP	pre-fit FR	ε_{FR}	ε_{vsJet}	ES
	Barrel			
VVLoose	0.68 ± 0.03	0.973 ± 0.006	1.14 ± 0.09	1.025 ± 0.003
VLoose	0.35 ± 0.03	0.960 ± 0.016	0.97 ± 0.06	1.035 ± 0.003
Loose	0.136 ± 0.015	0.94 ± 0.04	1.02 ± 0.07	1.038 ± 0.008
Medium	0.055 ± 0.007	1.09 ± 0.14	1.08 ± 0.08	1.015 ± 0.018
Tight	0.016 ± 0.002	1.11 ± 0.29	1.10 ± 0.07	1.04 ± 0.05
VTight	0.0063 ± 0.0008	1.07 ± 0.43	1.10 ± 0.07	1.04 ± 0.09
VVTight	0.0024 ± 0.0003	0.76 ± 1.12	1.12 ± 0.08	1.07 ± 0.19
	Endcap			
WP	pre-fit FR	ε_{FR}	ε_{vsJet}	ES
VVLoose	0.64 ± 0.03	0.990 ± 0.006	0.99 ± 0.08	1.008 ± 0.003
VLoose	0.30 ± 0.03	0.94 ± 0.03	1.06 ± 0.09	1.018 ± 0.008
Loose	0.126 ± 0.014	0.84 ± 0.09	1.08 ± 0.10	1.03 ± 0.02
Medium	0.057 ± 0.007	0.80 ± 0.11	1.08 ± 0.08	1.04 ± 0.04
Tight	0.019 ± 0.002	0.84 ± 0.26	1.11 ± 0.08	1.11 ± 0.06
VTight	0.0077 ± 0.0010	0.83 ± 0.59	1.14 ± 0.08	1.12 ± 0.14
VVTight	0.0029 ± 0.0004	0.92 ± 1.23	1.16 ± 0.09	1.08 ± 0.18

Table 5.7: $e \to \tau_h$ FR estimated in simulation, presented together with the scale factors obtained to correct the $e \to \tau_h$ FR, ES and against-jet efficiency for 2018. The uncertainties represent the 68% confidence level on the measured fit parameter.

WP	pre-fit FR	ε_{FR}	ε_{vsJet}	ES
		Barrel		
VVLoose	0.70 ± 0.03	0.976 ± 0.003	0.93 ± 0.06	1.028 ± 0.003
VLoose	0.36 ± 0.03	0.971 ± 0.011	0.98 ± 0.06	1.040 ± 0.003
Loose	0.143 ± 0.016	1.03 ± 0.02	1.03 ± 0.07	1.048 ± 0.005
Medium	0.058 ± 0.007	1.16 ± 0.05	1.08 ± 0.07	1.035 ± 0.008
Tight	0.018 ± 0.002	1.34 ± 0.16	1.10 ± 0.07	1.04 ± 0.03
VTight	0.0075 ± 0.0009	1.61 ± 0.28	1.11 ± 0.07	1.05 ± 0.03
VVTight	0.0033 ± 0.0004	2.18 ± 0.66	1.13 ± 0.07	1.06 ± 0.06
		Endcap		
WP	pre-fit FR	ε_{FR}	ε_{vsJet}	ES
VVLoose	0.64 ± 0.03	1.011 ± 0.004	0.90 ± 0.07	0.989 ± 0.003
VLoose	0.30 ± 0.03	0.96 ± 0.04	0.90 ± 0.07	0.995 ± 0.005
Loose	0.128 ± 0.014	0.89 ± 0.07	0.88 ± 0.07	1.000 ± 0.015
Medium	0.062 ± 0.007	0.76 ± 0.11	0.86 ± 0.07	1.02 ± 0.03
Tight	0.022 ± 0.003	0.74 ± 0.17	0.89 ± 0.07	1.03 ± 0.06
VTight	0.0073 ± 0.0009	0.98 ± 0.46	0.93 ± 0.07	1.15 ± 0.08
VVTight	0.0028 ± 0.0003	0.5 ± 1.0	0.92 ± 0.07	1.05 ± 0.21

5.4   Correlation with the against-jet classifier

As previously mentioned, the ε_{vsJet} nuisance parameter was introduced in the fit model described in Section 5.3, to account for a possible mismodeling of the against-jet classifier in the selection of electrons misidentified as hadronically decaying tau leptons. This mismodeling could emerge as a result of the correlation between the DeepTau classifiers, and introduce a dependency in the measured $e \to \tau_h$ SFs on the against-jet WP used to measure them. The ε_{vsJet} parameter corrects the normalization of the ZEE templates coherently in the PASS and FAIL regions, and is intended to compensate for the dependence on the against-jet WP used.

The measured SFs for the $e \to \tau_h$ misidentification are supposed to be applied on all analyses targeting a final state with a hadronically decaying tau lepton. Therefore, a check was required to determine whether a residual dependence of the measured SFs on the against-jet WP was present. If a sizable dependence was indeed present, the measured SF would not be efficient in correcting the simulation agreement with data in analyses which use an against-jet WP different from Medium[3].

The measurement described in the previous section was therefore repeated for different WPs of the against-jet classifier, in order to determine whether the resulting SFs are consistent within uncertainties with the ones determined for the Medium against-jet WP. These measurement were performed using the same recorded data, MC simulations, fit and uncertainty models, and definitions for the distributions and regions where the fit is performed. The only change applied in these measurement was in the event selection of the probe. Candidates of hadronically decaying tau leptons were selected using the following against-jet classifier WPs: Loose, Medium and Tight.

This study was performed as a preliminary investigation of the correlation between the correction SFs for the against-electron classifier efficiency and the against-jet WP. The against-jet WPs were instead chosen based on the ongoing analyses performed in CMS at the time. The comparison was performed directly on the normalization SFs applied to the PASS regions, which as a function of the fit parameters is

$$SF = \varepsilon_{FR} \times \varepsilon_{vsJet} \ . \tag{5.18}$$

As previously discussed, the measurement was performed separately for two regions of pseudorapidity, each of the against-electron classifier WP, and each year of data-taking during Run 2.

Tables 5.8, 5.9 and 5.10 report the values for corrections SFs measured for 2016, 2017, and 2018 respectively. Each table is divided in two sections, the one on top is associated to the measurement performed in the Barrel region ($|\eta_\tau| < 1.460$) while on the bottom the SFs for the Endcap region ($1.558 < |\eta_\tau| < 2.3$) are listed. Each column is associated

---

[3]For example, one of the final states studied in the analysis of the $H \to \tau\tau$ process in the SM [159] uses the Tight WP of the against-jet classifier.

to a different against-jet WP (labeled *VsJet WP*), while the rows correspond to different against-electron WPs.

The uncertainties for the SFs are computed, assuming the two fit parameters as uncorrelated, in the following way:

$$\frac{\Delta SF}{SF} = \frac{\Delta\varepsilon_{FR}}{\varepsilon_{FR}} + \frac{\Delta\varepsilon_{vsJet}}{\varepsilon_{vsJet}} \ .\tag{5.19}$$

The choice was made based on the following observations:

- ε_{vsJet} fixes the overall normalization of the ZEE template in the fit and is mostly uncorrelated with the against-electron WP;

- the value of ε_{vsJet} is mostly constrained by the FAIL region of the fit, since it is the one dominated by $Z \to ee$ events.

The SFs measured using different WPs of the against-jet classifier appear to be mostly compatible with each other within uncertainties. A hint to a possible trend in the SFs dependency on the against-jet classifier WP can be observed in the measurements done with 2016 and 2017 data and simulated samples. In particular it appears that for a large fraction of measurements the central value of the measured SF decreases by choosing a tighter against-jet WP. This trend is however not found in the measurements performed with Run 2018. Since overall the measurements are in good agreement with each other, further investigation of the classifiers correlation was deemed unnecessary. Analyses using different WPs of the against-jet classifier were recommended to use the SFs measured using the Medium against-jet WP.

Table 5.8: Scale factors calculated for the $e \to \tau_h$ misidentification rate using different against-jet WP in 2016.

Barrel			
		VsJet WP	
WP	Loose	Medium	Tight
VVLoose	1.26 ± 0.09	1.38 ± 0.08	1.23 ± 0.10
VLoose	1.29 ± 0.12	1.22 ± 0.08	1.13 ± 0.09
Loose	1.29 ± 0.12	1.28 ± 0.10	1.20 ± 0.13
Medium	1.39 ± 0.19	1.44 ± 0.13	1.34 ± 0.20
Tight	1.28 ± 0.27	1.22 ± 0.38	1.25 ± 0.29
VTight	1.78 ± 0.60	1.52 ± 0.36	1.37 ± 0.46
VVTight	2.43 ± 1.32	2.42 ± 0.43	2.12 ± 0.91
Endcap			
		VsJet WP	
WP	Loose	Medium	Tight
VVLoose	1.31 ± 0.11	1.29 ± 0.08	1.24 ± 0.11
VLoose	1.17 ± 0.13	1.13 ± 0.09	1.16 ± 0.13
Loose	1.13 ± 0.19	0.99 ± 0.16	1.03 ± 0.15
Medium	1.18 ± 0.26	1.08 ± 0.21	1.04 ± 0.19
Tight	1.76 ± 0.53	1.47 ± 0.32	1.25 ± 0.37
VTight	2.13 ± 1.14	1.59 ± 0.60	1.12 ± 0.71
VVTight	4.99 ± 4.00	2.40 ± 1.04	0.94 ± 1.36

Table 5.9: Scale factors calculated for the $e \to \tau_h$ misidentification rate using different against-jet WP in 2017.

Barrel			
	VsJet WP		
WP	Loose	Medium	Tight
VVLoose	0.97 ± 0.08	0.97 ± 0.08	0.97 ± 0.08
VLoose	0.95 ± 0.08	0.95 ± 0.08	0.95 ± 0.08
Loose	0.96 ± 0.10	0.96 ± 0.11	0.97 ± 0.10
Medium	0.99 ± 0.15	1.18 ± 0.20	1.18 ± 0.20
Tight	1.53 ± 0.50	1.22 ± 0.32	1.60 ± 0.64
VTight	1.43 ± 1.00	1.18 ± 0.47	1.34 ± 0.50
VVTight	0.00 ± 2.23	0.85 ± 1.54	1.46 ± 0.96
Endcap			
	VsJet WP		
WP	Loose	Medium	Tight
VVLoose	1.20 ± 0.11	1.03 ± 0.09	1.12 ± 0.10
VLoose	1.09 ± 0.13	1.00 ± 0.12	0.96 ± 0.11
Loose	1.00 ± 0.17	0.91 ± 0.20	0.86 ± 0.14
Medium	0.91 ± 0.19	0.86 ± 0.21	0.89 ± 0.17
Tight	1.44 ± 0.42	0.93 ± 0.38	1.02 ± 0.29
VTight	2.01 ± 0.78	0.95 ± 0.78	1.05 ± 0.53
VVTight	3.43 ± 1.66	1.07 ± 1.41	1.07 ± 1.41

Table 5.10: Scale factors calculated for the $e \to \tau_h$ misidentification rate using different against-jet WP in 2018.

Barrel		
	VsJet WP		
WP	Loose	Medium	Tight
VVLoose	1.00 ± 0.07	0.91 ± 0.06	0.98 ± 0.07
VLoose	1.04 ± 0.08	0.95 ± 0.07	0.99 ± 0.08
Loose	1.10 ± 0.10	1.06 ± 0.09	1.09 ± 0.09
Medium	1.14 ± 0.13	1.25 ± 0.14	1.27 ± 0.14
Tight	1.61 ± 0.30	1.47 ± 0.27	1.42 ± 0.34
VTight	1.69 ± 0.36	1.79 ± 0.42	1.79 ± 0.42
VVTight	3.34 ± 0.98	2.46 ± 0.90	2.61 ± 0.78

Endcap		
	VsJet WP		
WP	Loose	Medium	Tight
VVLoose	0.93 ± 0.07	0.91 ± 0.07	0.90 ± 0.07
VLoose	0.87 ± 0.09	0.86 ± 0.10	0.89 ± 0.09
Loose	0.78 ± 0.12	0.78 ± 0.12	0.78 ± 0.12
Medium	0.68 ± 0.16	0.65 ± 0.15	0.62 ± 0.13
Tight	0.58 ± 0.21	0.66 ± 0.20	0.84 ± 0.27
VTight	0.70 ± 0.48	0.91 ± 0.50	0.76 ± 0.41
VVTight	1.02 ± 1.06	0.46 ± 1.00	1.44 ± 0.90

6

CP measurement in $H \to \tau_\mu \tau_h$ decays

Contents

Since the discovery of the Higgs boson in 2012 [1, 2] the measurement of its properties has become one of the main objectives for the CMS and ATLAS collaborations. Several analyses have been performed throughout the years to measure its mass [40], spin [60], production cross-section and decay width [3–5, 38]. Among these properties is also the behavior under charge conjugation (C) and parity (P), referred together as CP-symmetry. As discussed in Chapter 2, several strategies have been deployed to tackle the CP properties of the Higgs boson and its couplings to SM particles. The direct couplings to vector bosons were studied by looking at the Higgs decays to four leptons [54, 60] and its production via vector boson fusion [63, 64]. The direct coupling to top quarks and effective coupling to gluons were instead investigated through the topology of Higgs production mechanisms [38, 65].

This chapter is dedicated to the study of the CP properties of the Yukawa coupling of the Higgs boson to tau leptons. This coupling is unique compared to the previous ones as it can only be accessed by looking at the decays of the Higgs boson, as it does not contribute to the production mechanisms. To study the CP properties of this coupling its structure is altered to account for both a CP-even and a CP-odd component. Assuming that the Yukawa coupling to tau leptons is the only altered Higgs coupling with respect to the SM, the Higgs production mechanism can be treated according to the SM prediction. The topology of the initial state, including its production cross-sections via ggH, VBF, and VH, and the topology of jets coming from the initial state radiation, is treated as associated to a SM-like Higgs boson. The CP structure of the Yukawa coupling is investigated using exclusively the Higgs decay products topology.

The analysis was performed by studying three different final states for the $H \to \tau\tau$ decays:

- $\tau_\mu\tau_h$: the semileptonic final state where a tau lepton decays into a muon and the other one decays hadronically;

- $\tau_e\tau_h$: a final state similar to the precedent but with a tau lepton decaying to an electron instead of a muon;

- $\tau_h\tau_h$: the final state where both tau leptons decay hadronically.

This chapter focuses on the characteristics which are common among the three channels, while providing more details on the channel which was the main focus of this thesis project:

the $\tau_\mu\tau_h$ channel. The scope of this thesis was to study the CP properties of the Higgs Yukawa coupling to tau leptons in the $\tau_\mu\tau_h$ channel, and contribute to the combination of the measurement with the analogous one in the $\tau_h\tau_h$ channel first [158], and later with the $\tau_e\tau_h$ one. This chapter provides the results obtained with the $\tau_\mu\tau_h$ channel exclusively, while the combination with the other channels is presented in Chapter 7.

6.1 Analysis overview

The aim of the measurement is to estimate the CP properties of the Higgs Yukawa coupling to tau leptons. As previously shown in Section 2.2, the Yukawa Lagrangian can be rewritten to account for a CP-mixing in the following way:

$$\mathscr{L}_{Y,\tau} = -\frac{m_\tau}{v}(\kappa_\tau\bar{\tau}\tau + \tilde{\kappa}_\tau\bar{\tau}i\gamma_5\tau)h \; . \tag{6.1}$$

Since the $H \to \tau\tau$ process has been measured, the decay width of the process is an established physical observable which constrains the admitted values for the reduced couplings κ_τ and $\tilde{\kappa}_\tau$. Having already written the Lagrangian by factorizing the value of the SM Yukawa coupling (m_τ/v), this is equivalent to the condition:

$$\kappa_\tau^2 + \tilde{\kappa}_\tau^2 = 1 \; . \tag{6.2}$$

This was already discussed in Section 2.2.2, however when performing the experiment this condition must be altered. The $H \to \tau\tau$ decay width has not been measured with infinite precision, and constraining the squared sum of the two couplings to be exactly 1 carries over the assumption that the measured decay width is exactly equal to the SM one. Eq. 6.2 can be rewritten to include another parameter, μ, representing the signal strength for the $H \to \tau\tau$ process, i.e. the ratio between its measured cross-section (σ_{meas}) and SM prediction (σ_{SM}):

$$\mu = \frac{\sigma_{meas}}{\sigma_{SM}} = \kappa_\tau^2 + \tilde{\kappa}_\tau^2 \; . \tag{6.3}$$

The signal strength can be measured alongside the Yukawa couplings, since, to estimate the latter, it is necessary to estimate the number of $H \to \tau\tau$ events reconstructed among the recorded data.

Instead of measuring the individual couplings, it is possible to define an angle representing their admixture. This was previously shown in Eq. 2.17, which can be rewritten to include the $H \to \tau\tau$ signal strength:

$$\begin{aligned}
\kappa_\tau &= \sqrt{\mu}\cos\varphi_{\tau\tau} \; , \\
\tilde{\kappa}_\tau &= \sqrt{\mu}\sin\varphi_{\tau\tau} \; .
\end{aligned} \tag{6.4}$$

The angle $\varphi_{\tau\tau}$ is referred to as "CP mixing angle" in this text and is the main target of the measurement together with the $H \to \tau\tau$ signal strength, μ. The SM prediction of a

pure CP-even coupling corresponds to $\kappa_\tau = \sqrt{\mu}$ and $\tilde{\kappa}_\tau = 0$, or equivalently $\varphi_{\tau\tau} = 0°$. A pure CP-odd coupling is instead represented by $\varphi_{\tau\tau} = 90°$ ($\kappa_\tau = 0$ and $\tilde{\kappa}_\tau = \sqrt{\mu}$), while $\varphi_{\tau\tau} = 45°$ corresponds to the maximum mixing between CP-even and CP-odd couplings ($\kappa_\tau = \tilde{\kappa}_\tau = \sqrt{\mu/2}$). The measurement of the CP mixing angle is performed by studying the spin correlation between the two tau leptons, which is reconstructed based on their decay products. The $H \to \tau\tau$ process cross-section can be written as in Eq. 2.23, here reported for convenience of the reader:

$$
\mathrm{d}\sigma_{H\to\tau\tau} / \mathrm{d}\cos\left(\theta^+\right) \mathrm{d}\cos\left(\theta^-\right) \mathrm{d}\cos\left(\phi^+\right) \mathrm{d}\cos\left(\phi^-\right) \propto
$$
$$
\left(1 + \cos\left(\theta^+\right)\cos\left(\theta^-\right) - \sin\left(\theta^+\right)\sin\left(\theta^-\right)\cos\left(\varphi_{CP} - 2\varphi_{\tau\tau}\right)\right) ,
$$

(6.5)

where $\theta^\pm$ and $\phi^\pm$ represent the polar and azymuthal angles defining the direction of each polarimetric vector with respect to the corresponding tau lepton direction of flight, $\varphi_{CP} = \phi^+ - \phi^-$ and $\varphi_{\tau\tau}$ is the CP mixing angle. The cross-section depends on the angle between the decay planes of the two tau leptons (φ_{CP}) via a sinusoidal function, as can be observed in Fig. 6.1. In the presence of CP mixing, a phase shift can be observed in the *acoplanarity angle* distribution, equal to two times the CP mixing angle.

With simulated samples of signal and background processes, it is possible to define all the observables in the event which are relevant for the measurement of μ and $\varphi_{\tau\tau}$.

As the measurement targets the differential cross-section of the Higgs decays to tau lepton with respect to φ_{CP}, it is necessary to reconstruct this angle on an event-by-event basis. The methods used in this analysis to measure φ_{CP} in each event based on the tau decay channel have been described in Section 2.2.3. A description of how the CP mixing is modelled in simulated $H \to \tau\tau$ events is instead presented in Section 6.4.2, together with some details on the other simulated processes used in this analysis. In order to improve the reconstruction of φ_{CP} several methods have been introduced and are presented in Section 6.3.

As mentioned in Section 2.2.3, in tau decays involving only one visible decay product (one prong) the acoplanarity angle is reconstructed using the impact parameter (IP) of the lepton or charged pion. As the resolution on the IP has to be extremely high to allow for a precise measurement of the acoplanarity angle, it was improved with several techniques, described in Sections 6.3.2 and 6.3.3. Furthermore, the spin-correlation between the tau lepton and its decay products varies depending on the decay channel, which must be accurately identified in order to properly reconstruct the acoplanarity angle. This was done by implementing an optimized tau decay mode identification, described in Section 6.3.4.

The $H \to \tau\tau$ process has a relatively low cross-section compared to other processes which can produce similar signatures in the detector. Since the aim of the project is to measure the CP mixing angle in reconstructed Higgs decays, such processes constitute a background for the measurement. In order to correctly identify $H \to \tau\tau$ events in
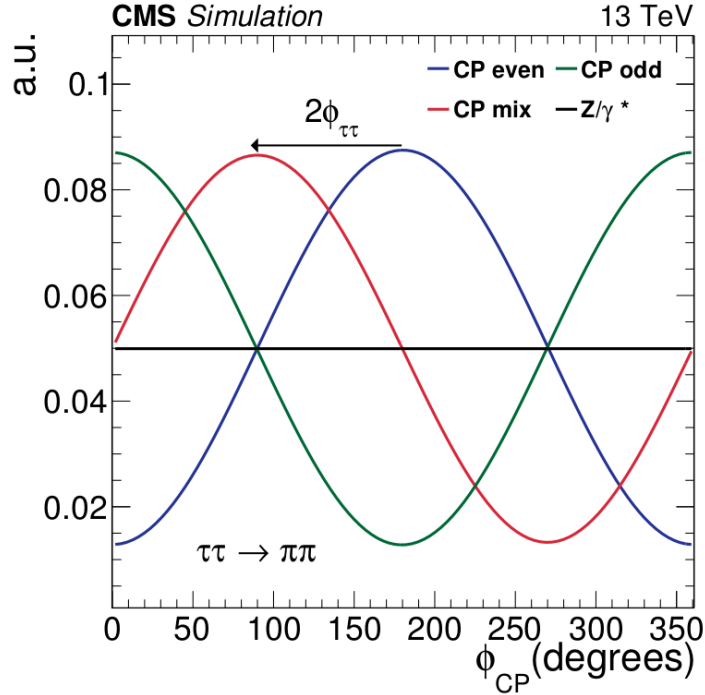
Figure 6.1: Acoplanarity angle distribution in the $\tau_\pi\tau_\pi$ channel [158]. The distributions are shown for simulated $H \to \tau\tau$ processes, according to three different CP hypotheses, and faithfully reproduce the analytical distributions expected from theory. The acoplanarity distribution is shown for the DY process in gray and appears as a flat line.

recorded data two conditions must be fulfilled:

- achieve a good modelling of data using Monte Carlo simulation or data-driven methods;

- define a phase-space region where the number of events for the targeted process ($H \to \tau\tau$ in this case) is sufficiently high to not be mimicked by statistical fluctuations of the background processes.

The starting point for satisfying both conditions is to select events having an experimental signature similar to the one expected for the $H \to \tau_\mu\tau_h$ process: an isolated muon and a τ_h candidate. Sections 6.2 and 6.6 describes how this selection is done and how the agreement between the background modelling and recorded data is ensured. This includes the use of data-driven methods to estimate the most relevant contributions to the background, discussed in Section 6.5. This allowed to accurately model the recorded data, while the ratio of signal over background events (S/B) was increased by categorizing events with a Neural Network, as described in Section 6.7. Three categories were defined: one for the signal, and two for the dominant background processes.

The measurement of the parameters of interest is then performed using a negative log-likelihood minimization, as a fit of the signal and background models to the recorded data. The fit is performed simultaneously on multiple signal and background categories to obtain a better estimation of the signal properties and constrain uncertainties on the background and signal models. Section 6.8 describes the fit model and the treatment of the uncertainties. All these procedures were needed to achieve the results shown in Section 6.9.

6.2   Trigger and preselection
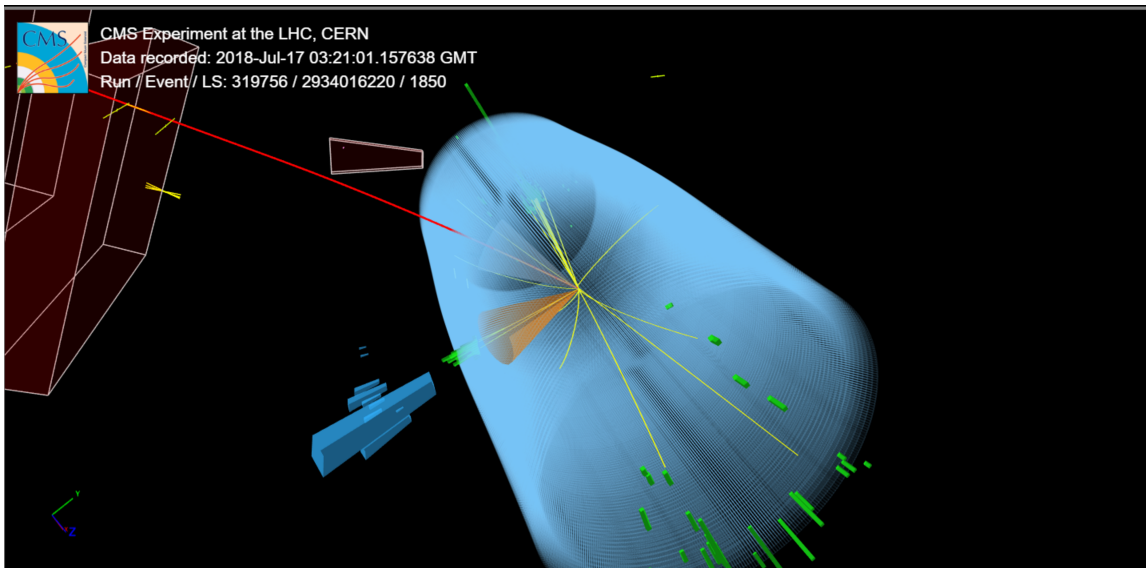
Figure 6.2: Event display for an event identified as a $H \to \tau_\mu\tau_h$ decay [158].

At the LHC, bunches of protons are circulated in the accelerator and collide with a frequency of ~ 40 MHz. This rate is extremely high, making storing all information recorded by the detectors for each pp collision unfeasible. As discussed in Section 3.2.5, in order to record events which are of particular physical interest a trigger system selects pp collisions which produced specific signatures in the detector. For the process of interest, $H \to \tau\tau \to \tau_\mu\tau_h$, the identifying signature is an isolated muon and a hadronically decaying tau lepton, as shown in Fig. 6.2. The CMS detector is optimized for the reconstruction of muons, which are the main objects used to identify the events of interest.

The triggers used in this analysis are listed in Table 6.1, they are divided based on the year of data-taking, and on whether they require a single muon of sufficiently high transverse momentum, or a muon identified alongside a hadronically decaying tau. The former triggers are labeled *SingleMuonX* while the latter are labeled *MuonXTauY*, with

X and Y being the p_T threshold required respectively for the identified muons and tau candidates. The third column of the table shows the path associated to the High Level Trigger (HLT), which lists more details related to the trigger, like the range in pseudorapidity where the object is reconstructed, or the type of particle reconstructed, e.g. tracker or global muon. For example HLT_IsoMu20_eta2p1_LooseChargedIsoPFTauHPS27_-eta2p1_CrossL1_v is one of the HLT paths for the Muon20Tau27 trigger used for 2018. It requires an isolated global muon reconstructed with $p_T > 20$ GeV, together with a loosely isolated charged τ_h candidate identified by the HPS algorithm with $p_T > 27$ GeV, both reconstructed in the pseudorapidity region $|\eta| < 2.1$. The fourth column lists instead the offline requirements for the muon and τ_h depending on the trigger that got fired.

Events recorded by triggers are then selected by requiring a muon which satisfies the following conditions:

- the muon transverse momentum must satisfy the requirements listed in the fourth column of Table 6.1 depending on which trigger got fired;

- $|\eta| < 2.1$, as this is the range used to reconstruct muons at the trigger level;

- the muon must match the one that fired the trigger, either in single muon or muon+tau triggers, within a cone of $\Delta R < 0.5$;

- $d_0 < 0.045$ cm and $d_z < 0.2$ cm, with d_0 and d_z transverse and longitudinal IP of the muon, as defined in Section 4.1.2;

- pass the Medium WP muon identification algorithm, discussed in Section 4.1.3;

- $I_{rel} < 0.15$, with the relative isolation defined in Eq. 4.3.

The events should also include a τ_h candidate identified by the HPS algorithm and passing the following selection criteria:

- $p_T > 20$ and $|\eta| < 2.3$;

- if the event was selected by a muon+tau trigger, the aforementioned requirements are tightened to the selection shown in the fourth column of Table 6.1 and $|\eta| < 2.1$;

- $d_z < 0.2$ cm,

- pass the loosest WP for the DeepTau classifiers defined in Section 4.4.

The two objects are further required to have opposite electric charge, and be separated by $\Delta R > 0.5$. The latter condition is implemented to exclude cases where the muon originates from the leptonic decay of a baryon produced inside a jet, misidentified as a τ_h candidate. If an event includes multiple muons or τ_h candidates which satisfy these conditions, the pair selected is the one having the most isolated muon and tau lepton.

Year	Trigger	High Level Trigger path	Offline selection (GeV)
2016	SingleMuon22	HLT_IsoMu22_v, HLT_IsoMu22_eta2p1_v HLT_IsoTkMu22_v, HLT_IsoTkMu22_eta2p1_v	$p_{T,\mu} > 23$
	Muon19Tau20	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_v HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1_v	$p_{T,\mu} > 20, p_{T,\tau} > 25$
2017	SingleMuon24	HLT_IsoMu24_v	$p_{T,\mu} > 25$
	SingleMuon27	HLT_IsoMu27_v	$p_{T,\mu} > 28$
	Muon20Tau27	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1_v	$p_{T,\mu} > 21, p_{T,\tau} > 32$
2018	SingleMuon24	HLT_IsoMu24_v	$p_{T,\mu} > 25$
	SingleMuon27	HLT_IsoMu27_v	$p_{T,\mu} > 28$
	Muon20Tau27	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1_v HLT_IsoMu20_eta2p1_LooseChargedIsoPFTauHPS27_eta2p1_CrossL1_v	$p_{T,\mu} > 21, p_{T,\tau} > 32$

Table 6.1: Triggers used in the $\tau_\mu \tau_h$ channel.

The most isolated muon is the one having the lowest relative isolation, as defined by Eq. 4.3, while the τ_h candidates isolation is equal to their score for the against-jet DeepTau classifier, meaning that a higher score corresponds to a more isolated tau lepton.

These requirements are applied as a preselection for the analysis, and are used to select events with an isolated muon and a hadronically decaying tau lepton.

6.3 Event reconstruction

Observables used for the measurement of the CP mixing angle are defined by the properties of objects reconstructed by the Particle Flow (PF) algorithm, previously described in Section 4.1.1. In particular, properties like the transverse momenta and energy of particles, and their direction of flight in the $\eta - \varphi$ plane, have been already used throughout the years in several CMS analyses. Several high level features are instead specific for this work and as such require a more detailed description.

6.3.1 Estimation of the tau lepton pair invariant mass

One of the most relevant variables used to identify $H \to \tau\tau$ decays with respect to similar processes is the invariant mass of the tau lepton pair. Due to the presence of neutrinos in tau decays, part of their energy is not reconstructed, causing a reduced resolution on their invariant mass. The SVFIT algorithm [160] was introduced in order to estimate the invariant mass of a tau lepton pair based on the reconstructed visible tau decay products and the missing transverse energy (MET) in the event. The algorithm
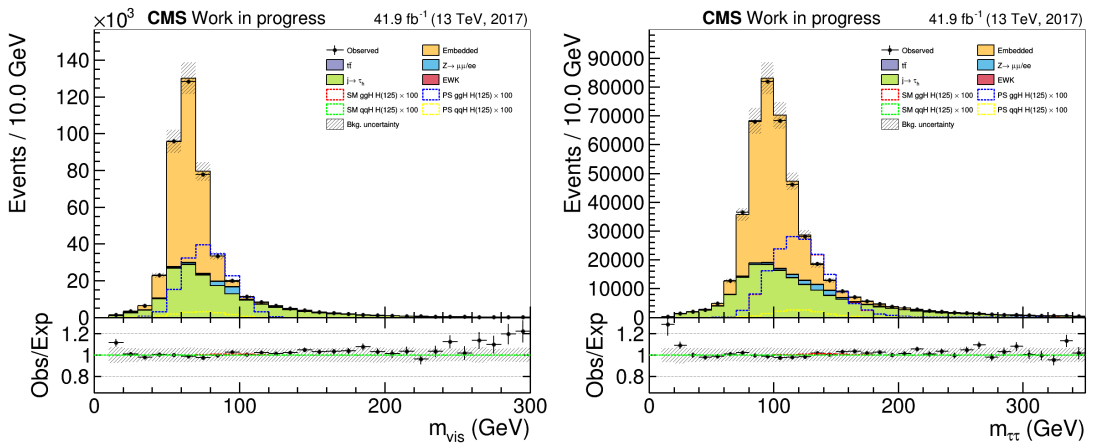
Figure 6.3: Distribution of the invariant tau lepton pair mass reconstructed using only the visible decay products (**left**) and with the SVFit algorithm (**right**). Event selection and background composition is detailed in Section 6.6.

takes as inputs the MET and its covariance matrix, together with the kinematic properties of the tau candidates or their visible decay products:

- for τ_μ or τ_e decays p_T, η, φ and mass of the reconstructed muon or electron are used;

- for τ_h decays, p_T, η, φ, m_τ and HPS-DM are used for the tau candidate, obtained by combining its hadronic decay products.

The algorithm uses a *Dynamical Likelihood Method* (DLM) [161, 162] in order to estimate the invariant mass of the tau lepton pair ($m_{\tau\tau}$). It estimates the probability that the reconstructed values for the MET and visible decay product kinematics are observed based on a certain value of $m_{\tau\tau}$. The probability is estimated for multiple values of $m_{\tau\tau}$ in a range between 5 GeV and 2 TeV, with higher granularity compared to the expected resolution of the invariant mass. The best estimate for $m_{\tau\tau}$ is the one which maximized the probability to observe the reconstructed quantities.

This method allows to estimate the mass of the boson which decays into the tau lepton pair. This is illustrated in Fig. 6.3 where the invariant mass of the $\tau_\mu \tau_h$ system is shown on the left using the visible decay products (m_{vis}), and on the right using the SVFIT algorithm ($m_{\tau\tau}$). The plots include backgrounds modelled using the data-driven techniques illustrated in Section 6.5 and the signals described in Section 6.4.2. For the $Z \to \tau\tau$ process, labeled *Embedded* in the plot, the peak shifts from 70 to around 90 GeV between m_{vis} and $m_{\tau\tau}$ showing that the invariant mass of the tau lepton pair is now closer to the Z mass. The same is observed for the $H \to \tau\tau$ process, divided in its different production mechanisms and CP hypotheses in the figure, as its peak shifts from 80 to around 125 GeV. This allows an improvement in the separation between reconstructed Higgs and Z boson decays.

6.3.2   Primary vertex reconstruction

In the one prong channel the acoplanarity angle is measured using the IP of the charged particle trajectory. The resolution in the measurement of this IP is crucial to reconstruct the acoplanarity angle. The IP resolution depends on two main factors:

- PV resolution;

- resolution of the point of closest approach (PCA) of the track to the PV.

The resolution on the PCA depends on the algorithm used to reconstruct it, as discussed in Section 6.3.3. The resolution on the PV has a large impact on the IP reconstruction and is instead detailed in this section.

Fig. 6.4 shows a schematic view of the collision point in CMS keeping the elements relevant for the IP reconstruction:
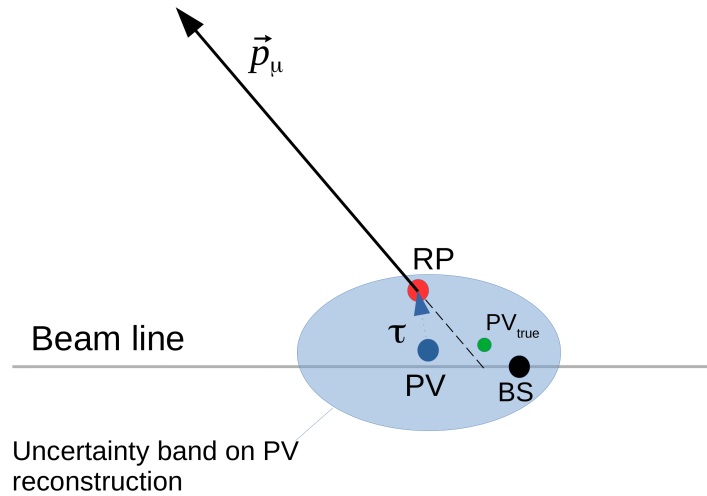
Figure 6.4: Schematic representation of the effect of the PV uncertainty on the IP reconstruction.

- the momentum of a charged particle (a muon in the example);

- the PV and its uncertainty;

- the innermost hit in the tracker, which is used as a reference point (RP) for the track extrapolation to the PV;

- the beam line and beamspot (BS).

As shown in Fig. 2.7, the IP is defined as the component of the vector connecting PV and RP which is orthogonal to the charged particle momentum. The reconstructed PV (blue circle) is measured with a certain precision, enlarged in Fig. 6.4 for illustrative purposes. Depending on the direction of flight of the charged particle, this uncertainty can be comparable with the size of the IP. If the true position of the PV (green circle) is on the opposite side of the line which extends the particle trajectory towards the beamline, the reconstructed IP has opposite sign with respect to the real one. Changing the sign of the IP results in a phase shift of π in the acoplanarity angle distribution, which would deteriorate the precision of the measurement. This effect can be reduced by improving the PV resolution, which was optimized in the scope of this analysis.

The PV reconstruction algorithm described in Section 4.1.2 uses tracks from the various recorded particles to measure the position of the PV. In this analysis the fit was altered in two ways:

1. the position and uncertainty of the beam spot was included in the fit;

2. the tau decay products tracks are excluded from the fit.

As previously discussed in Section 4.1.2, the BS is defined as the center of the region where the proton bunches collide. Its position is determined by studying tracks collected over multiple bunch crossings and is reconstructed with higher precision compared to the PV of a hard scattering process [128]. Its inclusion in the fit helps constraining the x and y coordinates of the PV, improving their resolution.

The exclusion of the tau decay product tracks from the vertex fit can instead lead either to an improvement or a deterioration of the PV resolution depending on the specific case studied. When a highly boosted tau lepton decays either to a lepton or hadronically, the tracks of its charged decay products appear displaced with respect to the PV. These tracks are associated generally to high momentum particles, and are produced further from the beam line compared to the decay products of lower energy tau leptons. High momentum tracks are weighted more in the PV fit, and can therefore bias the PV reconstruction in the direction of the tau decay vertex. This would result in a deterioration of the PV resolution in the transverse plane. The exclusion of tau decay product tracks can therefore improve the resolution of the x and y coordinates of the PV when tau leptons are highly boosted.

Low energy tau leptons decay closer to the PV and have less collimated decay products. The inclusion of their tracks in the fit can help constraining the PV position. This effect can be quite limited when the tau lepton decays to an electron or charged pion, as their tracks are reconstructed using only hits from the CMS tracking system (see Section 4.1.2 and 4.1.4). Muon tracks are instead reconstructed with greater precision, using both hits in the tracker system and in the muon chambers (see Section 4.1.3). Their inclusion in the fit can therefore improve the PV resolution for events involving low energy tau leptons.

Fig. 6.5 shows the improvement in the PV resolution coming from the exclusion of the tau decay product tracks and the inclusion of the BS constraint in the fit. The resolution noticeably improves for the x and y coordinates, while along the z axis it is mildly reduced. As shown in Section 4.1.2, the x and y coordinates of the BS are measured with a precision of $\sim 6 - 8$ μm, allowing for a more precise estimation of the PV coordinates in the transverse plane. The z coordinate resolution of the BS is less stable during data-taking, resulting in a higher uncertainty (~ 3.5 cm). Its inclusion in the PV fit results in a slight reduction of the resolution in the z direction.

This vertex fitting algorithm improves the PV resolution, and is chosen to define the PV in this analysis. The impact of the vertex fitting algorithm choice in the measurement of $\varphi_{\tau\tau}$ is shown in Section 6.9.2.

6.3.3 Impact parameter reconstruction

As previously mentioned in Section 4.1.2, the track of a charged particle can be parametrized as a helix, and characterized by five parameters:
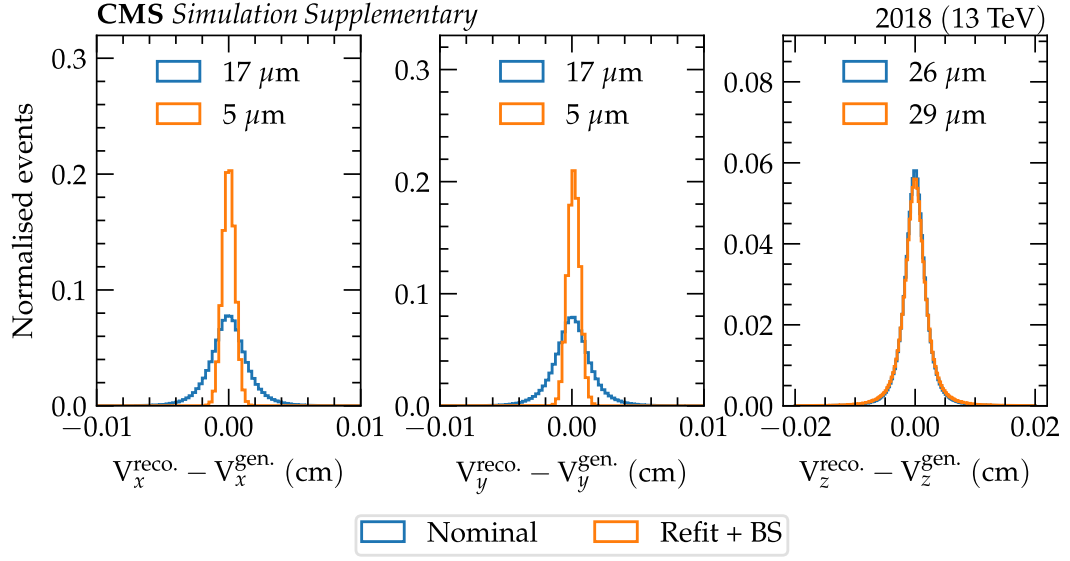
Figure 6.5: Resolution of different PV definitions, measured using simulated $H \to \tau\tau \to \tau_\mu\tau_h$ as the difference between the generator-level coordinates and the reconstruction level ones [158].

- $C = q/p$: charge divided by momentum, representing the curvature of the track;

- d_0 and d_z: respectively distance in the transverse plane and z direction between the PCA and the nominal interaction point;

- λ and ϕ: polar and azymuthal angles of the track at the PCA.

The d_0 and d_z parameters represent the projection of the PCA in the transverse plane and towards the beam axis, but computing the IP as a three-dimensional vector requires knowing all three coordinates of the PCA. Using the aforementioned parameters it is possible to reconstruct the helix associated to the particle trajectory:

$$\vec{x}(t) = \overrightarrow{BS} + \begin{pmatrix} -d_0 \sin\phi + \dfrac{\cos\lambda}{BC} \cdot \operatorname{sign}(C) \cdot [\sin\phi(1 - \cos t) + \cos\phi \sin t] \\[2mm] d_0 \cos\phi + \dfrac{\cos\lambda}{BC} \cdot \operatorname{sign}(C) \cdot [\sin\phi \sin t + \cos\phi(1 - \cos t)] \\[2mm] \dfrac{d_z}{\cos\lambda} + \dfrac{\sin\lambda}{BC} \cdot \operatorname{sign}(C)\, t \end{pmatrix}, \qquad (6.6)$$

where t is the curve parameter which allows to describe all helix points, B is the magnetic field and ϕ and λ are now defined with respect to the BS position ($\overrightarrow{BS}$). This curve parametrization can then be used to minimize the distance

$$\delta = \left| \vec{x}(t) - \overrightarrow{PV} \right| \qquad (6.7)$$

between the curve points and the PV. The IP is then defined as the vector connecting PV and the point which minimized the distance δ:

$$\overrightarrow{IP} = \vec{x}(t_{min}) - \vec{PV} \ . \tag{6.8}$$

Having the IP measured in this way also allows to compute its resolution (σ_{IP}) analytically as a combination of the covariance matrices depending on the helix parameters and PV coordinates. This in turn allows to compute the IP significance as the ratio of the IP magnitude and its resolution:

$$IP_{sig} = \frac{\left|\overrightarrow{IP}\right|}{\sigma_{IP}} \ . \tag{6.9}$$

A cut on the IP significance can help rejecting events where the IP was reconstructed with less precision, and define a purer region in which to measure the acoplanarity angle. To stay consistent between different decay channels, the same cut was applied on each object whose IP parameter was used, namely π in one prong decays, muons and electrons. The chosen cut was $IP_{sig} > 1.5$, and was determined as described in Section 6.9.2.

This variable is generally not properly modelled in simulation. It was therefore calibrated in simulation in order to maintain a good agreement with recorded data. The calibration was performed in two regions defined to be orthogonal to the ones on which the measurement of the CP mixing angle is performed. The calibration of pion and muon IP significance is performed using the $Z \to \mu\mu$ sideband region, defined by requiring two oppositely charged muons satisfying the following conditions:

- the highest momentum muon should be the one that fired the trigger (HLT_-IsoMu22_v for 2016, HLT_IsoMu27_v in 2017 and 2018);

- $p_{T,\mu}^{lead} > 28$ GeV and $p_{T,\mu}^{trail} > 25$ GeV, with $p_{T,\mu}^{lead(trail)}$ being the transverse momentum of the leading (trailing) muon;

- both muons should be reconstructed within a pseudorapidity range $|\eta| < 2.4$ in 2017 and 2018, and $|\eta| < 2.1$ in 2016;

- both muons should pass the Medium WP of the muon ID algorithm, as defined in Section 4.1.3;

- their invariant mass ($m_{\mu\mu}$) should be within the range $70 < m_{\mu\mu} < 110$ GeV, in order to select events where the muons are originated from a Z boson decay.

The reason behind using the same region for calibrating the IP significance of pions and muons is that their IP resolution is similar. One of the main contributors to the resolution of the PCA, and therefore of the IP, is the multiple scattering process [163].

If the nuclear interactions are neglected, the angular spread of particles scattered by the interactions with the tracker material depends on the particle mass and charge. Muons and pions have both unitary charge, and their masses are close in value [23], resulting in them being similarly affected by the multiple scattering effect. As shown in Fig. 6.6, these approximations work and the resolution of the IP is similar between muons and pions. Electrons require a different treatment due to their lower mass and higher chance of emitting bremsstrahlung radiation while traversing the tracker.
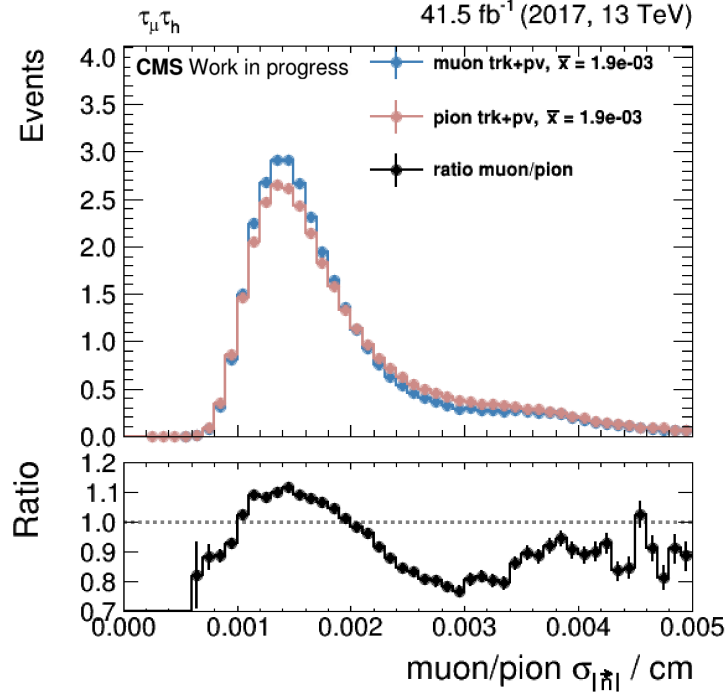
Figure 6.6: Resolution of the muon and pion IP.

The calibration is initially performed on simulated events presenting a pair of prompt leptons ($Z \rightarrow \mu\mu$), the corrected IP ($\overrightarrow{IP}_{corr}$) is determined through a quantile mapping method:

$$\overrightarrow{IP}_{corr} = F_{data}^{-1}(F_{sim}(\overrightarrow{IP})) \ , \tag{6.10}$$

where $\overrightarrow{IP}$ is the reconstructed IP, F_{sim} the cumulative distribution of the IP in simulation, and F_{data}^{-1} the inverse of the cumulative distribution, i.e. the quantile function, in data. The corresponding corrections for events involving non-prompt leptons or hadrons, like decays of a tau lepton pair system, are applied by subtracting from the reconstructed IP its generator-level value ($\overrightarrow{IP}_{gen}$), and then using the cumulative and quantile functions determined for prompt leptons:

$$\overrightarrow{IP}_{corr} = \overrightarrow{IP}_{gen} + F_{data}^{-1}(F_{sim}(\overrightarrow{IP} - \overrightarrow{IP}_{gen})) \ . \tag{6.11}$$

Fig. 6.7 shows the effect of the IP calibration: on the left the IP significance of muons is shown uncalibrated, while on the right the effect of the calibration is shown with the improvement of the simulation agreement with data.
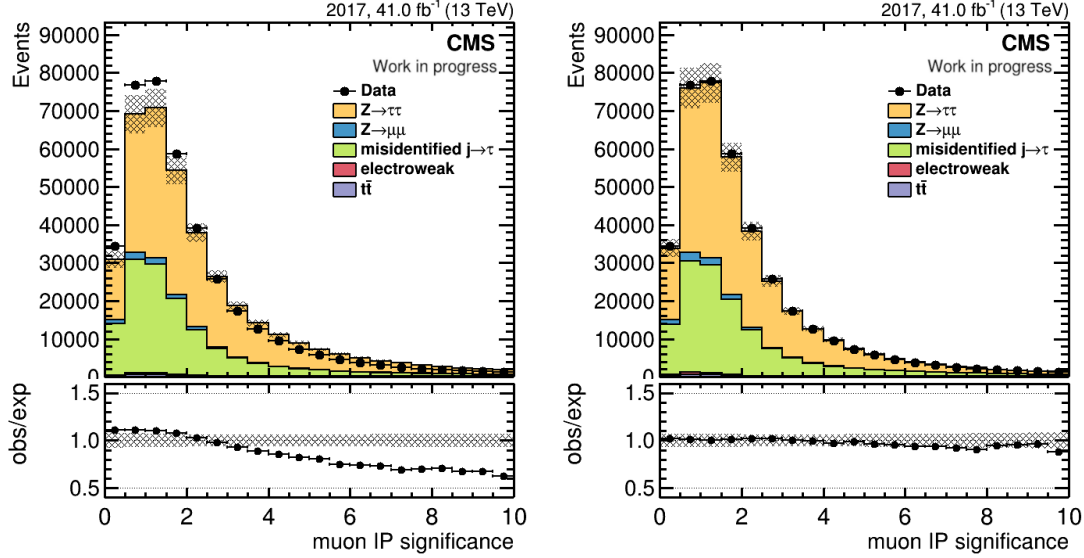
Figure 6.7: Calibration of the IP significance using PV fitted excluding the tau decay product tracks and including the BS constraint. On the left the uncalibrated IP_{sig} is shown to be mismodelled in simulation, the calibration is able to fix the mismodelling as shown by the calibrated distribution on the right.

6.3.4   MVA-based τ decay mode

As shown in Section 2.2.3, the methods used to reconstruct the acoplanarity angle vary depending on the tau decay channels. No reconstruction method was described for the 3 prong + π^0s channel, as the corresponding tau decay chain lacks an intermediate mesonic resonance. This results in a loss of spin correlation between the tau lepton and its decay products, making the $\tau \to \pi^\pm \pi^\pm \pi^\mp \pi^0 \nu_\tau$ decay unusable for the purpose of measuring the acoplanarity angle. Decays belonging to this channel need to be discarded, while events associated to the reconstructed 3 prong channel should be kept.

Furthermore even in cases where the same method is used, like in the one prong + π^0s channels, the spin correlation between the tau lepton and its decay products changes depending on the intermediate mesonic resonance. The ρ and a_1 mesons are both involved into these decays, but have different properties (see Tab. 1.5) and thus their decay products present different angular distributions. The $\tau \to \pi^\pm \pi^0 \nu_\tau$ can be split into two consecutive two-body decays: $\tau \to \rho^\pm \nu_\tau \to \pi^\pm \pi^0 \nu_\tau$. The ρ meson has a spin-parity $J^P = 1^-$ and decays into two different bosons of spin-parity $J^P = 0^-$. The a_1

meson, appearing as an intermediate state in $\tau \to \pi^{\pm}\pi^0\pi^0\nu_\tau$ decays, has instead a spin-parity $J^P = 1^+$ and decays into a $\rho\pi$ pair resulting in a final state with two identical pions and one with different charge. It was shown [50, 164] that the correlation between the tau polarization and the a_1 direction of flight changes depending on whether the a_1 meson is longitudinally or transversely polarized. Mixing together decays involving a ρ resonance with the ones involving the a_1 meson would therefore dilute the resolution of the acoplanarity angle.

The HPS algorithm (Section 4.2.1) is in principle able to distinguish events where one or two strips are reconstructed. However, the mass windows for the DM 1 and 2 largely overlap, and the strip size is determined dynamically resulting in a huge fraction of $\tau \to \pi^{\pm}\pi^0\pi^0\nu_\tau$ decays to be reconstructed as DM 1. This in turn justified the choice of merging these two decay modes identified by the HPS algorithm.

A multivariate analysis (MVA) based τ decay mode identification algorithm [165] was used to properly separate these two decay modes, and in general improve the decay mode identification. Two algorithms have been developed in order to target one prong and three prong channels separately, and are multi-class Boosted Decision Tree (BDT) which use the XGBoost library [166]. The BDT targeting the one prong channel has four classes ($\pi^{\pm}$, $\pi^{\pm}\pi^0$, $\pi^{\pm}\pi^0\pi^0$ and *other*) while the three prong BDT has three ($\pi^{\pm}\pi^{\pm}\pi^{\mp}$, $\pi^{\pm}\pi^{\pm}\pi^{\mp}\pi^0$ and *other*). The category named *other* is used to drop events which are not well reconstructed and is not used in the practical application of this MVA algorithm. The two algorithms then take a variety of input features coming from the tau decays, such as the angular variables of the two leading e/γ candidates in the reconstructed strip, the HPS-decay mode and the energy of each charged pion and strip. Fig. 6.8 shows the improvements in purity and efficiency brought by the MVA algorithm compared to the HPS one. Each decay mode is identified with increased purity, resulting in a higher percentage of events which correspond to their assigned decay mode. As for the efficiency this appears overall similar to the HPS one with the exception of DM 2.

The algorithm performs an identification of the tau decay modes based on the information provided by the HPS algorithm. A τ_h identified as having DM 0 by the HPS algorithm is associated to a single hadron in the event. Among its constituents no electrons or photons are present which could be used to reconstruct a strip. If the τ_h is then identified as having MVA-DM 1 or 2, it means that a cluster of electrons and photons was found in proximity to the charged pion, but not assigned to a strip. In order to not dilute the precision in the reconstruction of the acoplanarity angle[1], events having MVA-DM 1

---

[1]As described in Section 2.2.3, the decay plane for $\tau \to \pi^{\pm}\pi^0$ decays (i.e. corresponding to DM 1) is reconstructed using the momenta of the charged and neutral pions. The latter are reconstructed as shown in Eq. 2.47 reported here for convenience of the reader,

$$|\vec{p}_{\pi^0}| = \left| \sum^{e/\gamma} \vec{p}^{\,e/\gamma} \right| .$$
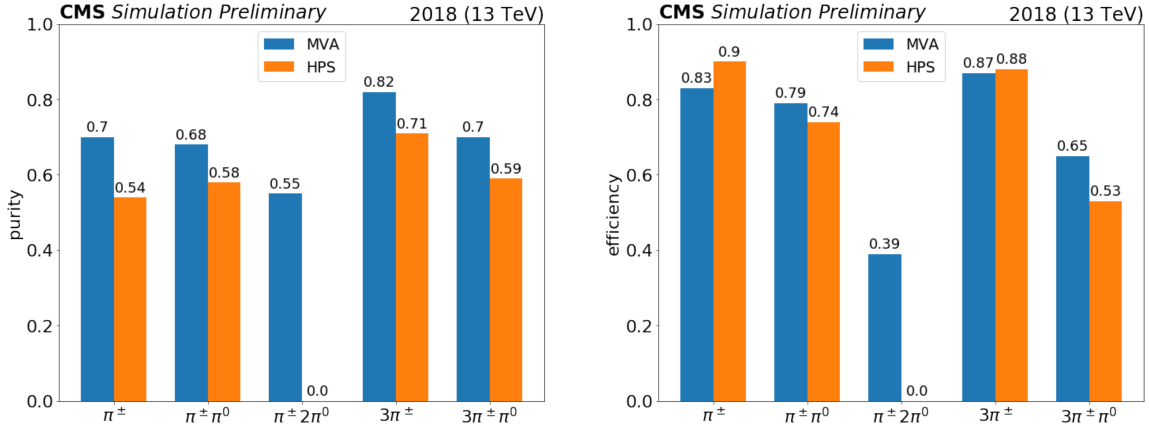
Figure 6.8: Left (Right): purity (efficiency) in the selection of simulated tau decays in the HPS (orange) and MVA (blue) tau identification algorithms [165].

or 2 and HPS-DM 0 are dropped

6.4 Simulation of signal and background processes

In order to measure the CP mixing angle of the Higgs Yukawa coupling to tau leptons, Monte Carlo (MC) simulation of $H \to \tau\tau$ decays and processes with similar signatures are necessary. The $H \to \tau\tau$ signal is simulated for the three dominant production mechanisms: ggH, VBF and VH. In order to properly simulate the CP-admixture of the Yukawa couplings, a dedicated method is required, and is detailed in Section 6.4.2. For a $\tau_\mu\tau_h$ final state the signature of interest in the detector is an isolated muon reconstructed alongside a hadronically decaying tau lepton. This signature can occur for other physical processes, which are considered as a background for the measurement:

- Drell-Yan (DY) production of tau leptons: $Z \to \tau\tau$;

- DY production of light leptons: $Z \to \mu\mu$;

- $t\bar{t}$ events with either a tau lepton pair final state, or jets misidentified as τ_h;

- single top (or anti-top) production;

- di-boson production (VV=WZ,WW,ZZ);

It is the sum of the momenta of electrons and photons associated to the strip by the HPS algorithm. A τ_h with HPS-DM 0 is not associated to a strip, and therefore cannot be used to reconstruct a decay plane with the same method applied to other events with MVA-DM 1 or 2 and HPS-DM 1.

- W boson production in association with jets (W+Jets);

- QCD multijet production.

These background processes can be grouped together in four categories:

- genuine taus;

- $\mu\tau_h$ events;

- $l \to \tau_h$ fakes;

- $j \to \tau_h$ fakes.

The first one represents processes having in the final state a tau lepton pair in which one tau has decayed hadronically and the other one to a muon. It comprises both *irreducible backgrounds*, i.e. processes having exactly the same final state as the signal such as $Z \to \tau\tau \to \tau_\mu\tau_h$, and processes having a final state of the form $\tau_\mu\tau_h + X$, with X representing additional objects reconstructed in the final state. The $t\bar{t} \to \tau_\mu\tau_h b\bar{b}$ process is a good example of the latter, presenting two b-jets in the final state in addition to the $\tau_\mu\tau_h$ pair. Due to the importance of correctly modelling these background processes, their estimation and agreement with recorded data was improved with a data-driven method, as shown in Section 6.5.2.

Some processes can also produce a muon in addition to a hadronically decaying tau lepton, without including an intermediate tau lepton pair state. For example in processes involving two W bosons, it is possible for a W boson to decay to a muon, and another one into a tau lepton which then decays hadronically. These processes are estimated directly from MC simulation.

The other two background categories include processes where the reconstructed τ_h candidate identified by the HPS algorithm, is a misidentified lepton or jet. Processes like W+Jets, $Z \to \mu\mu$, and di-boson, can produce an isolated muon, coming from the Z or W boson decays, while in the case of QCD multijet production the muon can be produced via the leptonic decays of a hadron within a jet. Only a minor fraction of these events include muons which are isolated with respect to the jets, but due to the large cross-section of this process it still provides a significant contribution to the $\tau_\mu\tau_h$ final state. The τ_h candidate found in QCD multijet production and W+Jets events is usually a misidentified et. Section 6.5.1 describes a data-driven method used to improve the estimation of the background, where a jet is misidentified as a τ_h. Processes with a lepton misidentified as a τ_h are instead estimated with MC.

6.4.1 Datasets used

This analysis uses data recorded during the Run 2 data-taking period. The integrated luminosity of the analyzed data is 137 fb^{-1}, using the SingleMuon dataset. This dataset

is divided in different subsets depending on the detector conditions, and the integrated luminosity of each dataset is listed in Table 6.2.

Table 6.2: SingleMuon datasets recorded by the CMS experiment during the Run 2 data-taking period and associated recorded luminosity. The integrated luminosity is also shown in bold for the years 2016, 2017 and 2018, and across the entire data-taking period.

Dataset	Recorded luminosity (fb^{-1})
Run2016B	5.79
Run2016C	2.57
Run2016D	4.25
Run2016E	4.01
Run2016F	3.10
Run2016G	7.54
Run2016H	8.61
Full 2016	**35.9**
Run2017B	4.79
Run2017C	9.63
Run2017D	4.25
Run2017E	9.31
Run2017F	13.54
Full 2017	**41.5**
Run2018A	14.03
Run2018B	7.06
Run2018C	6.90
Run2018D	31.74
Full 2018	**59.7**
Full Run2 data-taking	**137**

Processes involving Higgs decays to tau leptons were simulated for three production mechanisms: ggH, VBF and VH. They were simulated using the Powheg MC generator [167], used together with the Pythia8 [168] toolkit in order to model the CP-admixture of the Yukawa coupling as detailed in the next section. Their cross-sections are listed in Table 6.3, where the VH process has been split into ZH, W$^+$H and W$^-$H.

Simulated background processes are instead listed in Table 6.4 together with their cross-sections calculated either at leading order (LO) or at the level of calculation specified in the table: (N)NLO for (next-to-)next to leading order. For each process, part of

Table 6.3: List of Higgs production mechanisms considered in this analysis and corresponding cross-sections.

Higgs production mechanism	cross-section (pb)
ggH	3.046
VBF	0.237
ZH	0.0594
W$^+$H	0.0527
W$^-$H	0.0358

the simulation is performed using the Pythia8 [168] toolkit, which includes modules dedicated to the simulation of tau lepton decays. The main hard scattering process is instead simulated by another MC generator, chosen depending on which physical process is being simulated. The MC generators used are Powheg [167], MadGraph [169] or MadGraph5_aMC@NLO [170], an upgraded version of MadGraph5 able to calculate corrections at NLO and are listed in the middle column of Table 6.4. In the dataset column the following notation is used: l, ν, q represent respectively a charged lepton (including τ leptons), a neutrino and a light quark (u,d,s,c), while jet is used to refer to both quark and gluon jets.

6.4.2 Simulation of CP-mixing

This analysis required the use of dedicated MC sample to account for the CP-admixture in the Higgs Yukawa coupling to tau leptons. As this is not the main focus of this thesis work here only a general description of a simulated event is provided, in order to explain how it is altered to account for CP-mixing in Higgs decays to tau leptons.

In general a simulated proton-proton collision event recorded in a particle detector can be represented as in Fig. 6.9. It is useful to separate the event simulation in three separate steps:

1. simulation of the hard scattering vertex and decay of outgoing objects;

2. parton shower and hadronization of outgoing partons;

3. simulation of the detector response.

The first two steps are generally grouped together to describe a *generator-level* event. At this stage, all particles are simulated regardless of whether they are reconstructed by the

---

[2]This process is simulated inclusively for any number of jets.

[3]Events are simulated with an intermediate virtual photon/Z boson of mass between 10 and 50 GeV.

Table 6.4: List of Monte Carlo simulated used in this analysis and corresponding cross-sections. For cross-sections calculated above the LO, the labels NLO and NNLO are used.

Dataset	MC generator	Cross-Section (pb)
$t\bar{t} \to 2l2\nu$	Powheg	88.29
$t\bar{t} \to 2q$	Powheg	377.96
$t\bar{t} \to l\nu + q$	Powheg	365.35
$t+W$ production	Powheg	35.85
$\bar{t}+W$ production	Powheg	34.97
single top production	Powheg	136.02
single anti-top production	Powheg	80.95
DY inclusive[2]	MadGraph	6077.22 (NNLO)
DY + 1 jet	MadGraph	877.8
DY + 2 jet	MadGraph	304.4
DY + 3 jet	MadGraph	111.5
DY + 4 jet	MadGraph	44.03
Low mass[3] DY	MadGraph	21,658.0 (NLO)
W+Jets inclusive[2]	MadGraph	61,526.7 (NNLO)
W + 1 jet	MadGraph	8104.0
W + 2 jet	MadGraph	2793.0
W + 3 jet	MadGraph	992.5
W + 4 jet	MadGraph	544.3
EWK W^+ + 2jets	MadGraph	25.62
EWK W^- + 2jets	MadGraph	20.25
EWK Z + 2jets	MadGraph	3.987
$ZZ \to 2l2q$	MadGraph5_aMC@NLO	3.688 (NLO)
$ZZ \to 2l2\nu$	MadGraph5_aMC@NLO	0.6008 (NLO)
$ZZ \to 4l$	MadGraph5_aMC@NLO	1.325 (NLO)
$WW \to 2l2\nu$	Powheg	11.08
$WW \to l\nu 2q$	Powheg	45.99
$WZ \to 2l2q$	MadGraph5_aMC@NLO	6.331 (NLO)
$WZ \to l3\nu$	MadGraph5_aMC@NLO	3.293 (NLO)
$WZ \to l\nu 2q$	MadGraph5_aMC@NLO	11.66 (NLO)
$WZ \to 3l\nu$	MadGraph5_aMC@NLO	5.052 (NLO)
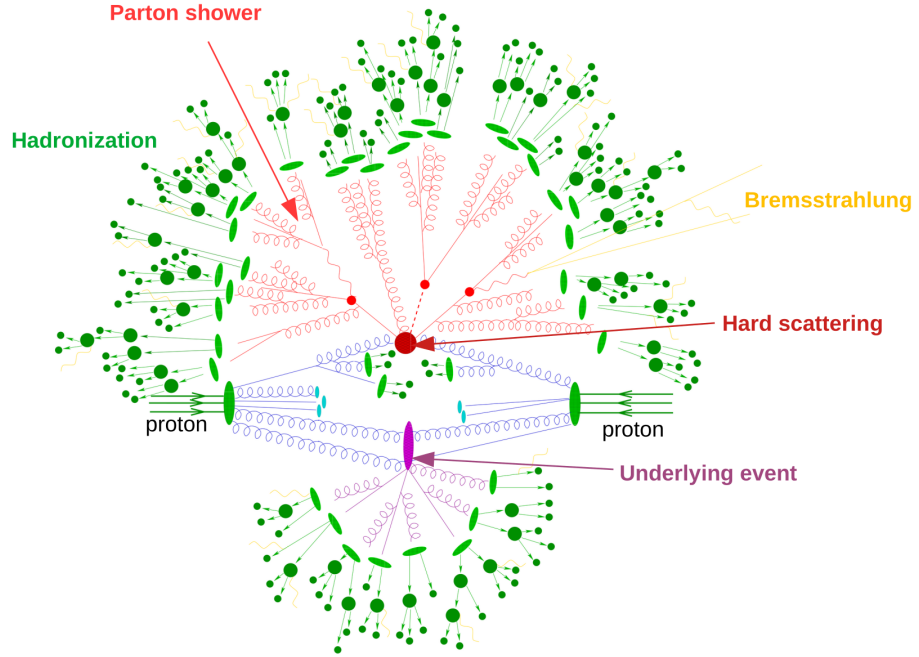
Figure 6.9: Schematic view of a proton-proton collision from a MC generator point of view (original image taken from [171]): starting from the incoming protons, two partons are selected to simulate the hard scattering process (dark red circle), other partons can either remain close to the beamline and fall out of the detector acceptance, or produce soft scattering processes forming the underlying event (purple). The hard scattering process is then simulated, generating the outgoing particles. Depending on the type of simulated process, these outgoing particles can then produce parton shower (light red), decay (light red circles) or reach the detector and be reconstructed. Partons showers are then converted via the hadronization process into streams of color neutral baryons which are reconstructed (green circles). Eventual initial and final state radiation, including emission of bremsstrahlung photons (yellow), are then simulated.

particle detector. For example, at this stage neutrinos are present and their momentum and direction of flight is accessible. This is relevant for simulating the CP-mixing in Higgs to tau decays.

Section 2.2.2 already introduced the concept of polarimetric vectors to describe how the differential cross-section of a tau decay depends on the angular distribution of its decay products:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega_{decay}} = \frac{\mathrm{d}\sigma^{unpol}}{\mathrm{d}\Omega_{decay}} \left(1 + \vec{s} \cdot \vec{h}\right) \; , \tag{6.12}$$

with σ^{unpol} representing the cross-section for an unpolarized tau decay, $\vec{s}$ the tau polarization and $\vec{h}$ its polarimetric vector. Both the tau polarization and its polarimetric vector are well defined at generator level. This means that the spin-correlation between the tau

and its decay products can be accessed by means of the weight $w_T = \vec{s} \cdot \vec{h}$. Based on this, the TAUOLA [73, 83] library for MC generation has routines capable of computing event weights, based on calculations provided by the CLEO collaboration [51], to properly model the decay of the polarized tau lepton.

The CP mixing angle $\varphi_{\tau\tau}$ affects the spin-correlation between the tau leptons coming from the Higgs decay. To properly model this effect, such spin-correlation needs to be propagated from the outgoing particles produced in the hard scattering process to the tau decay products. This is done by the TAUSPINNER tool [77, 172] which specializes in the calculation of weights to properly encode the longitudinal and transversal polarization correlation of tau leptons in simulated events. The TAUSPINNER tool is interfaced with the PYTHIA8 [168] MC simulation tool, acting at the stage of parton shower.

The particle interaction with the detectors is then simulated. At this stage, each particle is assigned its reconstruction level properties, like energy and momentum, and the MET is calculated.

In principle, this allows to construct MC samples of $H \to \tau\tau$ decays for any CP mixing angle. However $\varphi_{\tau\tau}$ can vary continuously between -π and π meaning that covering the whole range would in principle require a large number of simulated samples. This would be computationally expensive and comes with the complication of having to interpolate between the simulated mixing angles. In this particular analysis a different route is available: as shown in Fig. 6.1, the differential cross-section of the $H \to \tau\tau$ with respect to the acoplanarity angle has a sinusoidal shape of the form:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\varphi_{CP}} \propto const - \cos\left(\varphi_{CP} - 2\varphi_{\tau\tau}\right) . \tag{6.13}$$

Ignoring the normalization, Eq. 6.13 can be rewritten, using the properties of trigonometric functions, as:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\varphi_{CP}} \sim -\cos\left(2\varphi_{\tau\tau}\right)\cos\left(\varphi_{CP}\right) - \sin\left(2\varphi_{\tau\tau}\right)\sin\left(\varphi_{CP}\right)$$
$$= -\left(\cos^2(\varphi_{\tau\tau}) - \sin^2(\varphi_{\tau\tau})\right)\cos\left(\varphi_{CP}\right) - 2\sin\left(\varphi_{\tau\tau}\right)\cos\left(\varphi_{\tau\tau}\right)\sin\left(\varphi_{CP}\right) . \tag{6.14}$$

This new form is the weighted sum of sine and cosine functions of φ_{CP}. It becomes therefore apparent that a combination of signal samples simulated for specific values of $\varphi_{\tau\tau}$ could be used to recreate the signal template for any mixing angle. The choice taken, was to use the three cases shown in Fig. 6.1, that is: the CP-even ($\varphi_{\tau\tau} = 0$), CP-odd ($\varphi_{\tau\tau} = \pi/2$) and the maximum mixing ($\varphi_{\tau\tau} = \pi/4$, labeled CP-mix) cases. For the $\pi\pi$ channel, and ignoring their normalization, they take the form:

$$\frac{\mathrm{d}\sigma^{CP-even}}{\mathrm{d}\varphi_{CP}} \sim -\cos(\varphi_{CP}) , \tag{6.15}$$

$$\frac{\mathrm{d}\sigma^{CP-odd}}{\mathrm{d}\varphi_{CP}} \sim \cos(\varphi_{CP}) , \tag{6.16}$$

$$\frac{\mathrm{d}\sigma^{CP-mix}}{\mathrm{d}\varphi_{CP}} \sim \sin(\varphi_{CP}) \ . \tag{6.17}$$

It is easy to demonstrate that for these particular distributions, and the ones obtained in all other channels (which might be phase-shifted by π due to the different spin-correlation), the following combination allows to describe the signal for a generic mixing angle:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\varphi_{CP}} = const + (\cos^2(\varphi_{\tau\tau}) - \sin(\varphi_{\tau\tau})\cos(\varphi_{\tau\tau}))\frac{\mathrm{d}\sigma^{CP-even}}{\mathrm{d}\varphi_{CP}}$$

$$+ (\sin^2(\varphi_{\tau\tau}) - \sin(\varphi_{\tau\tau})\cos(\varphi_{\tau\tau}))\frac{\mathrm{d}\sigma^{CP-odd}}{\mathrm{d}\varphi_{CP}} \tag{6.18}$$

$$+ 2\sin(\varphi_{\tau\tau})\cos(\varphi_{\tau\tau})\frac{\mathrm{d}\sigma^{CP-mix}}{\mathrm{d}\varphi_{CP}} \ .$$

This allowed to create a parametric expression for a general signal template based on one unpolarized signal template and three weights computed by the TAUSPINNER module.

6.5   Background estimation

This analysis aims at determining the differential cross-section of a signal process which has low inclusive cross-section with respect to the background processes. The yields of each process must be therefore measured differentially in different regions of the phase-space, here referred as *bins* for simplicity. In order to achieve a good identification of the signal events, the signal yield in each bin must be higher than the statistical fluctuations of the background processes. Constraining the statistical fluctuations of simulated processes in each bin is important to achieve a precise measurement of the signal process properties. For a process simulated just with a MC generator, the statistical fluctuations in a bin depend on the number of simulated events reconstructed in the phase-space region of interest. The plots shown at the end of Section 6.3.3 (Fig. 6.7) can be used for a practical example of the limitation of simulated MC events. The uncertainties shown in the plot for simulation (gray area) include just a few percent uncertainty on the various processes cross-sections and the statistical uncertainties of MC simulations. The uncertainty for recorded data is instead purely statistical and the vertical bars placed on the data points are basically non visible and covered by the size of the marker. It is quite clear that the uncertainty for simulation is much larger than what the bin content would suggest. This is because simulated events are weighted to the theoretical prediction of the corresponding process cross-section, while the number of simulated events in each bin is much smaller. This leads to a non negligible MC statistical uncertainty. To reduce

these bin-by-bin statistical fluctuations, and the systematic uncertainties related to the simulation of detector effects, some data-driven methods can be used. An example of a data-driven method was given in Section 5.2.1, and is based on the idea of measuring the data yields in a side-band region and use them to infer the yield of a specific process in the phase-space region of interest for the analysis. This idea is also at the core of the data-driven method used to predict the yield of processes with jets misidentified as hadronically decaying tau leptons: the Fake Factor (FF) method [173, 174]. A different data-driven method was instead adopted to describe processes with two genuine tau leptons decaying into the final states of interest. This technique is based on the idea of embedding simulated tau decays into events selected in recorded data [175]. An overview of both methods is presented in the following sections.

6.5.1   Fake Factor method

As previously discussed in Section 4.3, it is not uncommon for a quark and gluon jet to be misidentified as a τ_h candidate by the HPS algorithm. Even after applying a selection using the DeepTau against-jet discriminator (see Section 4.4) a non negligible contribution of $j \rightarrow \tau_h$ fakes is still present when selecting a pair of tau leptons. Together with the $Z \rightarrow \tau\tau$ process, the $j \rightarrow \tau_h$ fakes make one of the major contributions to the phase-space of interest, and in the case of the $\tau_h \tau_h$ channel they are actually the dominant background. An accurate estimation of this process is crucial for a successful analysis of the Higgs production differential cross-section with tau decays, and is therefore done with a data-driven method to improve its agreement with recorded data and reduce bin-by-bin simulation statistical fluctuations.

The main idea of the method is to measure the yields of processes with $j \rightarrow \tau_h$ in a region enriched in such processes and orthogonal to the phase-space region of interest, i.e. the signal region (SR). This region is selected in the $\tau_\mu \tau_h$ channel using the event preselection discussed in Section 6.2 and applying the cuts described in Section 6.6. The rejection of $j \rightarrow \tau_h$ fakes is instead performed applying a cut on the against-jet discriminator provided by the DeepTau ID. Inverting this cut defines a phase-space with more jet fakes and less genuine τ_hs. This region is named application region (AR) and is used to measure the yields that are then scaled to the SR.

The events reconstructed in the AR are weighted by *fake factors* (F_F), which are calculated using dedicated regions orthogonal to both SR and AR, named *determination regions* (DR). For the $\tau_h \tau_h$ channel most jets faking hadronic tau leptons come from a QCD multijet production process, and the fraction of τ_h that pass the against-jet classifier is computed in a region where both τ_h candidates have the same sign. The F_F is then computed similarly to what described in Section 5.2.1:

$$F_\mathrm{F} = F_\mathrm{F}^{\mathrm{QCD}} = \frac{N_{pass}}{N_{fail}} \ , \tag{6.19}$$

where $N_{pass/fail}$ corresponds to the number of events that pass/fail the selection of the against-jet classifier in the DR. This factor is then corrected to account for the presence of other processes with misidentified jets.

For $\tau_l \tau_h$ the $j \to \tau_h$ fakes contributions have several competitive sources:

- QCD multijet production;

- $t\bar{t}$;

- W+Jets.

$$F_F = \sum_i w_i \, F_F^i$$

$$w_i = \frac{N_{AR}^i}{\sum_j N_{AR}^j}$$

$$i, j \in \{QCD, W+jets, t\bar{t}\}$$

$$^\dagger \text{Taken from simulation}$$
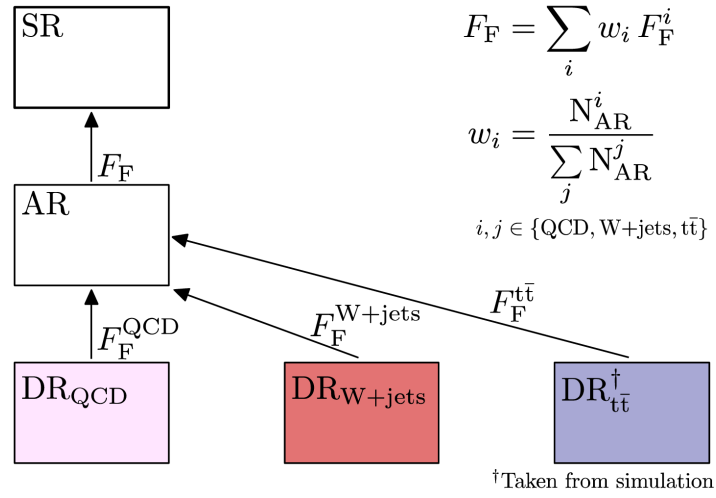
Figure 6.10: Schematic depiction of the FF method for semi-hadronic channels ($\tau_l \tau_h$) [176].

As shown in Fig. 6.10 the fake factor applied in the AR is computed as a combination of process specific fake factors (F_F^i with $i \in$ QCD,W+Jets,$t\bar{t}$). These factors are determined as the ratio of events that pass or fail the against-jet discrimination and are estimated in a dedicated DRs:

- DR_{QCD}: the lepton and hadronic tau candidate are selected to have the same sign, and the relative isolation of the light-lepton (I_{rel}^l, defined in Eq. 4.3 and 4.4) is set to be larger than 0.05;

- DR_{W+Jets}: $m_T > 70$ GeV, where $m_T = \sqrt{2 p_{T,l} E_T^{mis} (1 - \cos \Delta \phi_{l,met})}$ with E_T^{mis} missing transverse energy (MET) computed with the PUPPI-MET algorithm (see Section 4.1.6) and $\phi_{l,met}$ azymuthal separation between lepton and MET.

As shown in Section 6.6, a b-jet veto is applied in order to reduce the number the $t\bar{t}$ events in the SR. The events that pass the selection are those where the b-jets produced from the top quark decays are not correctly identified as originating from a b-quark. In order to calculate the FF for this process a dedicated DR should be defined where the event topology is sufficiently similar to the one of SR and AR, and with high enough statistics. A DR satisfying these conditions was not identified, as such $F_{\mathrm{F}}^{t\bar{t}}$ are measured directly in simulation using the events that pass the preselection shown in Section 6.2.

The fake factor applied on events in the AR is then the weighted sum of the ones for the separate processes:

$$F_{\mathrm{F}} = \sum_i \left(\frac{N_{AR}^i}{\sum_j N_{AR}^j} \right) , F_{\mathrm{F}}^i , \tag{6.20}$$

with $i, j \in$ QCD,W+Jets,$t\bar{t}$, and N_{AR}^j representing the number of events of the process j selected in the AR. $F_{\mathrm{F}}^{\mathrm{QCD}}$ and $F_{\mathrm{F}}^{\mathrm{W+Jets}}$ are determined in bins of:

- 3 bins based on the number of jets: they are defined by a number of jet equal to 0, 1 or ≥ 2;

- 5 bins of MVA-DM (see Section 6.3.4);

- 2 bins of IP_{sig}: one for $IP_{sig} \geq 1.5$ and one for $IP_{sig} < 1.5$.

The $F_{\mathrm{F}}^{t\bar{t}}$ factor is instead measured in bins of MVA-DM and IP significance, with the same bin definition as the other fake factors. Corrections to account for residual disagreement between simulation and data are then applied as function of the τ_h p_T, PUPPI-MET, m_T and I_{rel}^l.
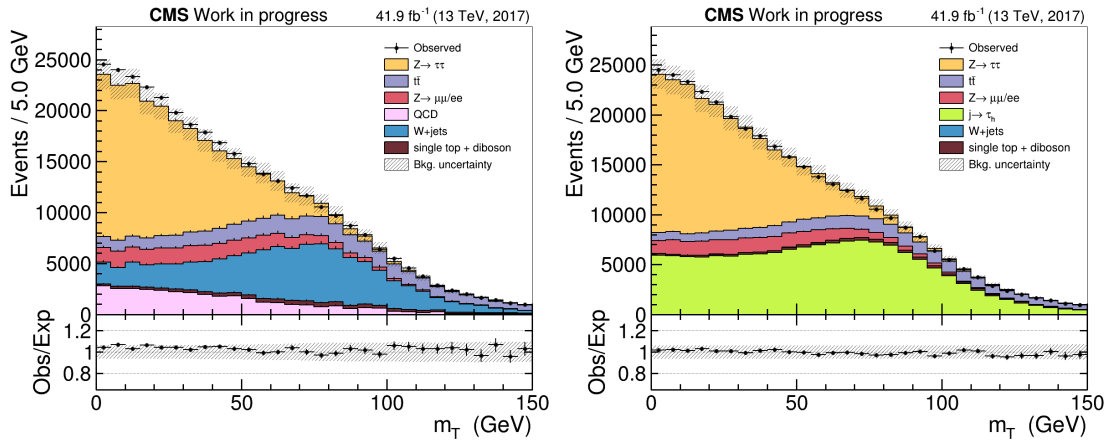
Figure 6.11: Transverse mass distribution obtained by using MC estimation for the $j \to \tau_h$ fakes (**left**) and by using the FF method (**right**). Events are selected according to the preselection in Section 6.2.

Fig. 6.11 shows the transverse mass distribution for data and MC simulation in 2017. On the right the QCD multijet production process (pink histogram) is estimated using the data-driven method described in Section 5.2.1, while on the left the FF method is applied to estimate the contribution coming from $j \to \tau_h$ fakes (green histogram). As can be observed, especially in the region $m_T < 40$ GeV, the statistical fluctuations of the background processes have been noticeably reduced improving the simulation agreement with recorded data.

6.5.2 τ embedding technique

The $Z \to \tau\tau$ process is an irreducible background for the study of Higgs decays to tau leptons. The two processes share the same final state and their distinction is one of the main obstacles in performing the analysis. The cross-section for the production of a Z boson is roughly 3 orders of magnitude larger than the gluon-gluon fusion production of the Higgs boson [3, 177, 178]. Having a background process with such a large cross-section, and very similar signature to the signal is a serious obstacle. The uncertainties on this process yield should be kept as low as possible in order to have the signal yields larger than the statistical uncertainties of the background prediction. This is done using a data-driven method: the embedding of tau simulation in recorded data [175].
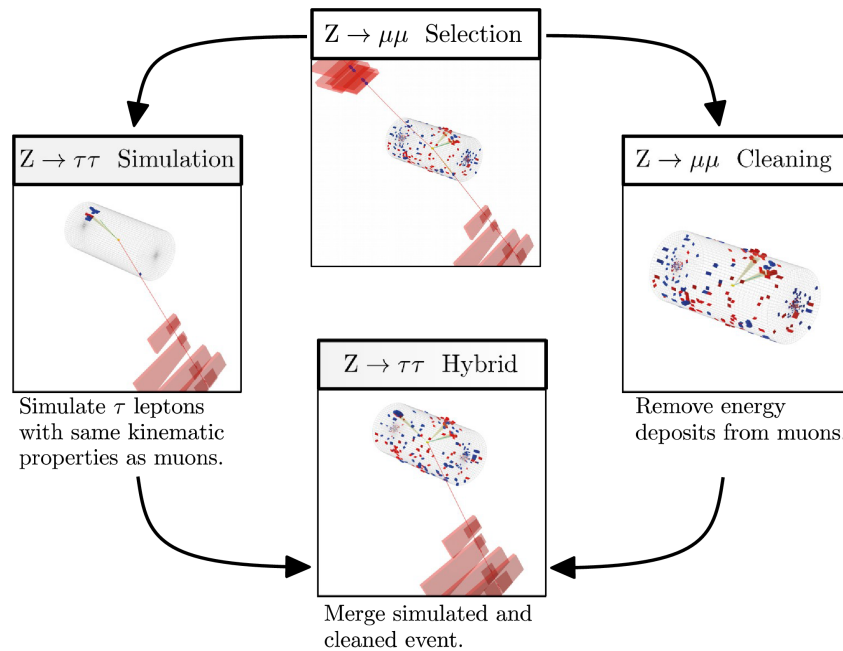
Figure 6.12: Schematic representation of the tau embedding technique [175].

The basic principle of this method is illustrated in Fig. 6.12:

1. events where two muons have been identified are selected in recorded data;

2. the event is cleaned from the signatures associated to the muon: hits in the tracker and muon chambers, and energy deposits in the calorimeters;

3. tau leptons are simulated in an empty detector and assigned the same kinematic properties of the removed muons;

4. the simulated tau leptons are embedded into the recorded events, and the signatures of their decays are encoded into the event.

The method is then validated by comparing the resulting kinematic distributions with the ones obtained by simulation. The main source of smearing of the mass resolution comes from the missing transverse energy caused by the neutrinos in the decay, while effects such as final state radiation (FSR) are negligible [175].

 The method is applied to all processes reconstructed in data with two muons in the final state, allowing to estimate the corresponding processes involving two tau leptons. This includes, aside from DY, also processes such as di-boson production (VV) and $t\bar{t}$. The possibility of estimating $t\bar{t}$ events with two genuine tau leptons is particularly notable, as the b-jet which accompany the $t\bar{t}$ production are in these events not simulated, but reconstructed. This has the benefit of a more accurate description of this jet topology.

 The data-driven estimation of the aforementioned processes benefit from the large amount of di-muon events recorded in CMS, allowing to estimate these processes with
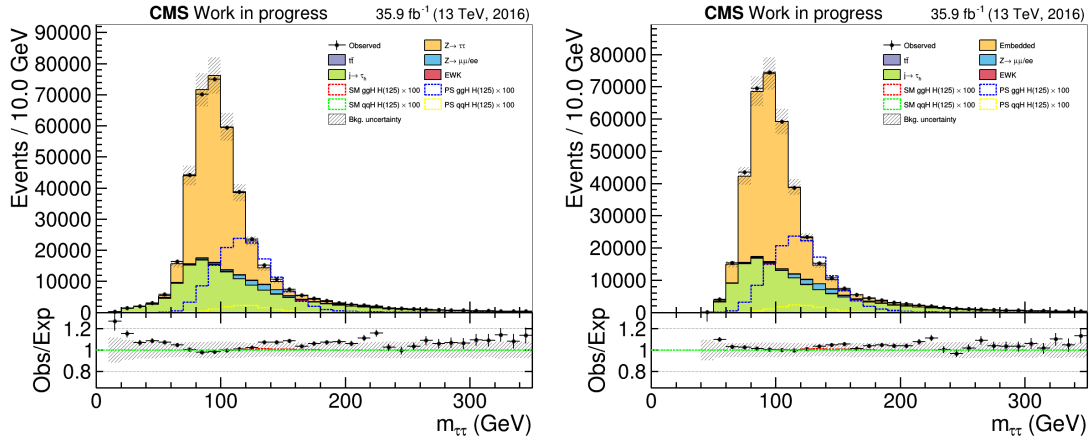
Figure 6.13: Invariant tau lepton pair mass distribution for 2016 data and MC. The $j \to \tau_h$ fakes have been estimated with the FF method, while the genuine taus contributions are estimated using MC (**left**), and tau embedding (**right**). Events selection is shown in Section 6.6. The plot on the right is obtained using the additional requirement $m_{vis} > 40$ GeV.

lower statistical fluctuations compared to the one associated to MC simulation. This data-driven estimation also allows for a more accurate description of the detector response, and more accurate modelling of the PU.

Fig. 6.13 shows the invariant mass of the tau lepton pair where the contributions coming from genuine taus are estimated with MC on the left, and tau embedding on the right. The use of tau embedding allows to reduce the statistical uncertainties in modeling of genuine tau lepton pairs and improve the model agreement with the recorded data.

Proper modeling of τ polarization in the embedded sample is a prerequisite for this analysis. Muon polarization cannot be reconstructed in CMS and consequently cannot be assigned to the simulated tau leptons. As a result, polarization effects are not inherently present in the simulation process and have to be injected in the embedded events by applying appropriate weights to the event.

As shown in Section 2.2.3, when integrating over all possible direction of flight for the two tau leptons the acoplanarity angle distribution is flat for DY events. The event selection described in Section 6.6 does not manifestly lead to acceptance effects which would cause a modulation in the acoplanarity angle distribution to become discernible for DY. This was investigated by studying the acoplanarity angle distribution with DY MC (Appendix B), leading to the conclusion that MC simulation offers an accurate description of recorded data and that acceptance effects are below statistical fluctuations.

Artificially adding a transverse polarization correlation in embedded events was therefore deemed unnecessary for this analysis. The possibility of an accidental polarization correlation being introduced in simulation was however investigated. This was done by
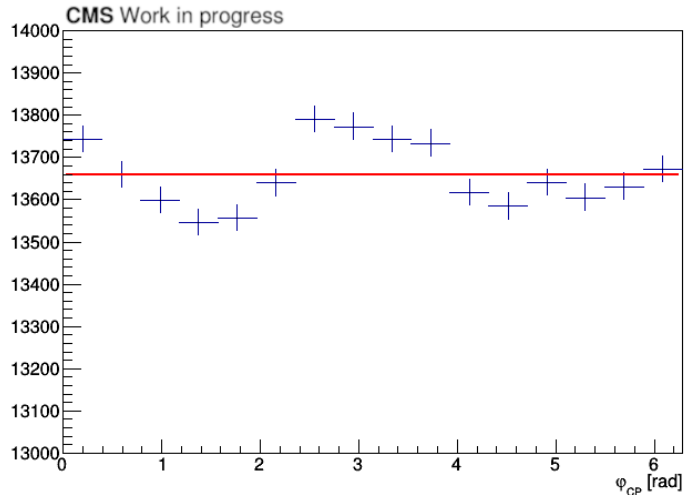
Figure 6.14: Distribution of the acoplanarity angle for tau embedded events. The 2018 dataset is used for this particular plot, events have been selected in a $Z \to \tau\tau$ enriched region. The red horizonal line represents the average of the bin contents.

looking at the distribution of the acoplanarity angle for embedded events. As shown in Fig. 6.14, the distribution is not flat but presents a modulation which is not explained by statistical fluctuations.

The acoplanarity angle distribution both for generator-level embedded events and reconstructed MC events was found to be flat, hinting at this modulation being an artificial effect related to the detector simulation. In this data-driven method the tau leptons are simulated in an empty detector and then embedded into recorded events. This can potentially introduce a misalignment between the simulated detector and the real one, and be the likely cause of the modulation found in the acoplanarity angle distribution.

This modulation was therefore interpreted as a non-physical effect, and fixed at the stage of measuring the CP mixing angle as discussed in Section 6.8.

6.6   Event selection

As previously discussed in Section 6.2, several processes contribute to the final state of interest for this analysis. With respect to the preselection previously defined, several additional requirements are used to select the events on which the measurement is performed:

- τ_h candidates are required to have MVA-DM $\in \{0, 1, 2, 10, 11\}$: this allows to select only the DM which have been identified by the MVA-DM algorithm.

- τ_h candidates with MVA-DM $\in \{1, 2\}$ are required to have HPS-DM different from 0: events with HPS-DM 0 have no reconstructed strip and therefore cannot be used to reconstruct the acoplanarity angle with the technique for defined one prong + π^0s decays.

- τ_h candidates are further required to pass the following WP for the DeepTau classifiers:

    - Medium against-jet WP: used to reduce the contamination from $j \to \tau_h$ fakes;

    - Tight against-muon WP: used reject events where a muon is misidentified as τ_h, mainly affecting the $Z \to \mu\mu$ process.

- $m_T < 50$ GeV: this cut is placed to reduce the contribution from W+Jets process, for which the m_T distribution peaks around 80 GeV.

- $m_{vis} > 40$ GeV: this is done to slightly reduce the contribution from non-resonant DY production of genuine taus, as very few events involving Higgs decays have visible mass lower than 40 GeV.

- In order to reduce the contribution from the $Z \to \mu\mu$ process a veto is applied to reject events where a second muon is found having opposite electric charge to the selected one and:

  - $p_T > 10$ GeV and $|\eta| < 2.4$;

  - pass the Medium muon identification WP;

  - $d_0 < 0.045$ cm and $d_z < 0.2$ cm;

  - $I_{rel} < 0.3$.

- A veto is applied on events where a pair of loosely reconstructed electrons or muons separated by $\Delta R > 0.15$ are reconstructed.

- b-jet veto: if at least a jet of $p_T > 20$ GeV passes the Medium WP of the deepCSV b-jet tagging algorithm the event is discarded. This cut allows to drastically reduce the contribution from $t\bar{t}$ events.

Some additional selections are applied only when performing the statistical inference for the CP mixing angle:

- Exclusion of MVA-DM 11: this DM corresponds to decays into 3 charged pions and a π^0, which do not posses an intermediate mesonic resonance, as such it is excluded in later stages of the analysis.

- $IP_{sig}^{\mu} > 1.5$: this allows to remove events with poorly reconstructed muon impact parameter, as previously mentioned in Section 6.3.3.

- $IP_{sig}^{\pi} > 1.5$ for MVA-DM 0: similar to the previous selection, events with poorly reconstructed charged pion IP are removed when the IP of pions is used for the acoplanarity angle reconstruction, i.e. in the $\tau_{\mu}\tau_{\pi}$ channel.

All the aforementioned selections are summarized in Table 6.5, including the event preselection described in Section 6.2. The fourth column of the table is used to describe which background processes are mainly affected by the performed selection. For the event preselection and the requirements based on the quality of the event reconstruction, the label *All* is used.

6.6.1 Corrections

Several corrections are then applied to simulated samples in order to obtain a good agreement between simulation and recorded data. More specifically the following corrections are applied to simulated samples:

Table 6.5: Summary table of the event selection.

Name	Year	Selection	Target process
Trigger	2016	$p_{T,\mu} > 23$ GeV $p_{T,\mu} > 20$ GeV, $p_{T,\tau_h} > 25$ GeV, $|\eta_{\tau_h}| < 2.1$	All
	2017, 2018	$p_{T,\mu} > 25$ GeV $p_{T,\mu} > 21$ GeV, $p_{T,\tau_h} > 32$ GeV, $|\eta_{\tau_h}| < 2.1$	
η range	All	$|\eta_\mu| < 2.1, |\eta_{\tau_h}| < 2.3$	All
Opposite sign	All	$q_\mu \cdot q_{\tau_h} < 0$	QCD multijet
$\eta - \varphi$ separation	All	$\Delta R_{\mu,\tau_h} > 0.5$	$j \to \tau_h$ fakes
Prompt muon	All	$d_0 < 0.045$ cm and $d_z < 0.2$ cm	$j \to \mu$ fakes
Muon ID and Iso	all	Medium ID and $I_{rel}^\mu < 0.15$	$j \to \mu$ fakes
DeepTau ID	All	Medium against-jet	$j \to \tau_h$ fakes
		VVLoose against-electron	$e \to \tau_h$ fakes
		Tight against-muon	$\mu \to \tau_h$ fakes
MVA-DM selection	All	MVA-DM $\in \{0, 1, 2, 10\}$	All
Transverse mass	All	$m_{T,\mu} < 50$ GeV	W+Jets
Visible mass	All	$m_{vis} > 40$ GeV	Low mass DY
b-veto	All	Medium WP of deepCSV tagging	$t\bar{t}$
Cut on the IP	All	$IP_{sig}^\mu > 1.5$ $IP_{sig}^\pi > 1.5$ (only if MVA-DM=0)	All

- Pile-up correction: the pile-up distribution is taken from MC and reweighted in order to match the one in recorded data, this allows for a more accurate description of quantities that strongly depend on the PU, e.g. the quality of track reconstruction and the uncertainty on the energy reconstruction in calorimeters.

- Trigger efficiency: this correction targets the efficiency of the trigger in simulation to match the one in recorded data, and is measured in di-muon events for triggers targeting an isolated muon, and in $\tau_\mu \tau_h$ events for triggers requiring both a muon and a tau lepton. For triggers requiring a muon and hadronic tau pair, the applied

corrections depend on both $p_{T,\mu}$ and p_{T,τ_h}.

- Muon isolation and identification: the efficiency is corrected to account for the mismodeling of muon reconstruction and identification in simulation. The corrections are measured in di-muon events like the previous one.

- Corrections related to the DeepTau classifiers and energy scales of genuine τ_hs and objects misidentified as τ_hs:

  - against-jet classifier: determined in $\tau_\mu\tau_h$ events, they correct the tau selection efficiency, and their energy scale (ES);

  - against-lepton classifiers: the scale factors (SFs) to correct the $e \to \tau_h$ fake rate (FR) are described in Chapter 5, similar methods are used to correct the $e \to \tau_h$ ES and $\mu \to \tau_h$ FR.

- b-jet identification: a mismodeling of the b-tagging efficiency would result in additional b-jets being identified or in the rejection of some of the ones that passed the Medium WP for the deepCSV algorithm. This alters the amount of $t\bar{t}$ events selected in the analysis and is corrected using SF calculated in a $t\bar{t}$ enriched region.

- MET recoil corrections: the MET distribution is reconstructed in events where neutrinos are not expected, like DY production of two light leptons. The distribution in simulation is corrected to match the one in data to improve the MET modeling in simulations for events involving W, Z and Higgs bosons.

- Corrections for the Z and top p_T distributions: both distributions are reconstructed in simulations and recorded data by selecting phase-space regions dominated by $Z \to \mu\mu$ and $t\bar{t}$ events respectively. SFs are then calculated to correct the simulated distribution so as to match the one in recorded data, improving the modeling of DY and $t\bar{t}$ simulations.

- Pre-firing weights: during Run 2 an increasing in the offset of the ECAL timing pulse was observed, which led occasionally to the Level-1 trigger relying on ECAL to fire on the previous bunch crossing. In 2016 and 2017, weights to correct for this effect were calculated and are applied on MC simulation [122]. This was solved in 2018.

Of the aforementioned corrections only the ones affecting simulated muons and hadronic tau leptons are applied to the tau embedded samples.

A good agreement between simulation and data was achieved after applying the aforementioned corrections. This can be seen in Fig. 6.15, 6.16 and 6.17 where the MVA-DM and the lepton pseudorapidity distributions are shown respectively for 2016, 2017 and

2018. Fig. 6.19 and 6.20, shown in the Section 6.7, and Fig. D.1 to D.10 in Appendix D also show good agreement between the background modeling and recorded data.

The CP-even and CP-odd hypotheses (red and blue histograms for gluon-gluon fusion, and green and yellow for vector boson fusion) appear indistinguishable in the plots, hinting at the absence of biases towards a particular CP hypothesis due to the selection. The number of simulated events selected after all selection cuts are applied was verified to be independent of the CP hypothesis. The event yields for signal processes are multiplied by a factor 100 in the plots, as they are negligible with respect to the backgrounds in several phase-space regions. It is necessary to implement a method to improve the event classification and construct a signal enriched region. As described in the next section, this is done with a multi-class neural network.
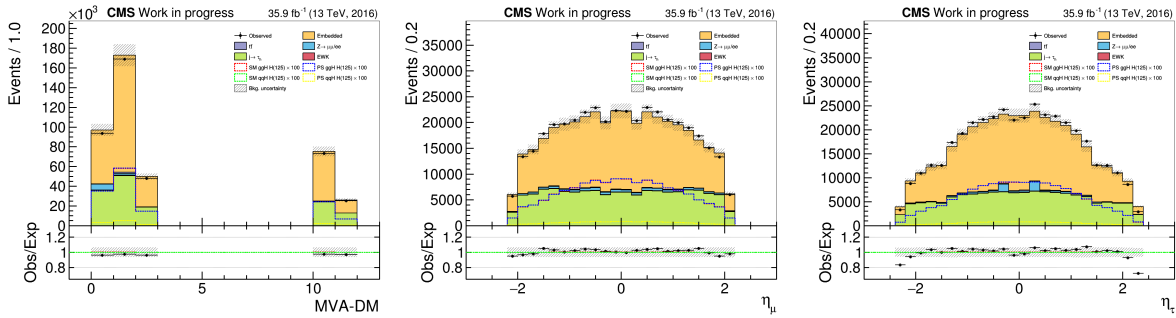
Figure 6.15: Distribution of tau MVA-DM (**left**) and pseudorapidity of muon (**middle**) and tau (**right**), using data and MC for 2016.
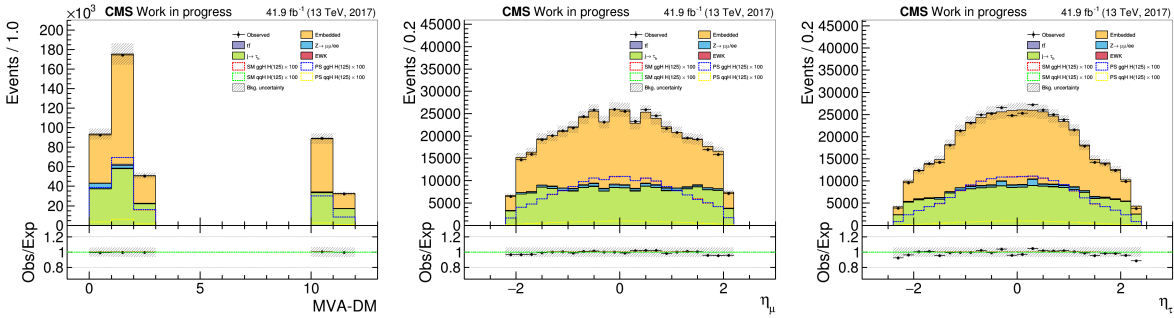
Figure 6.16: Distribution of tau MVA-DM (**left**) and pseudorapidity of muon (**middle**) and tau (**right**), using data and MC for 2017.
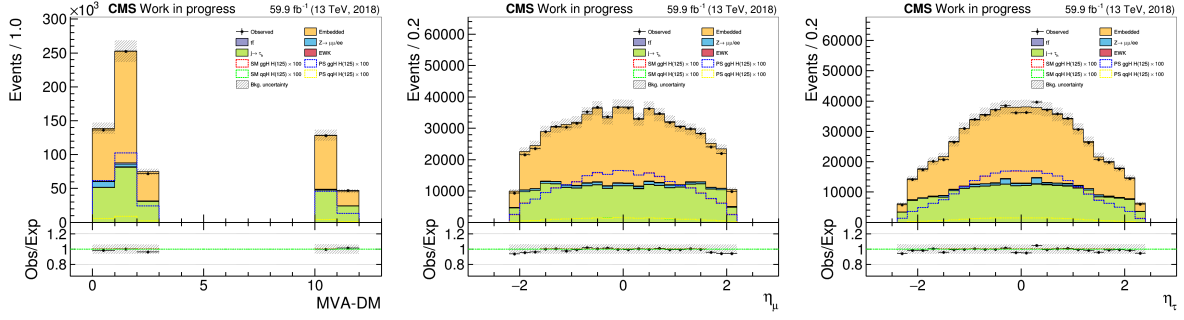
Figure 6.17: Distribution of tau MVA-DM (**left**) and pseudorapidity of muon (**middle**) and tau (**right**), using data and MC for 2018.

6.7 Event classification with neural network

Improving the signal separation from background processes can be treated as a classification problem, i.e. a procedure which aims at classifying events into different categories. This method can be tackled by machine learning, more specifically by supervised learning where a machine learning algorithm is trained using events of known classes in order to make predictions on a separate dataset. For the purpose of this analysis, a multi-class neural network (NN) was developed, based on the one previously used for the measurement of the $H \to \tau\tau$ cross-section [174]. The structure of the NN is shown in Fig. 6.18,
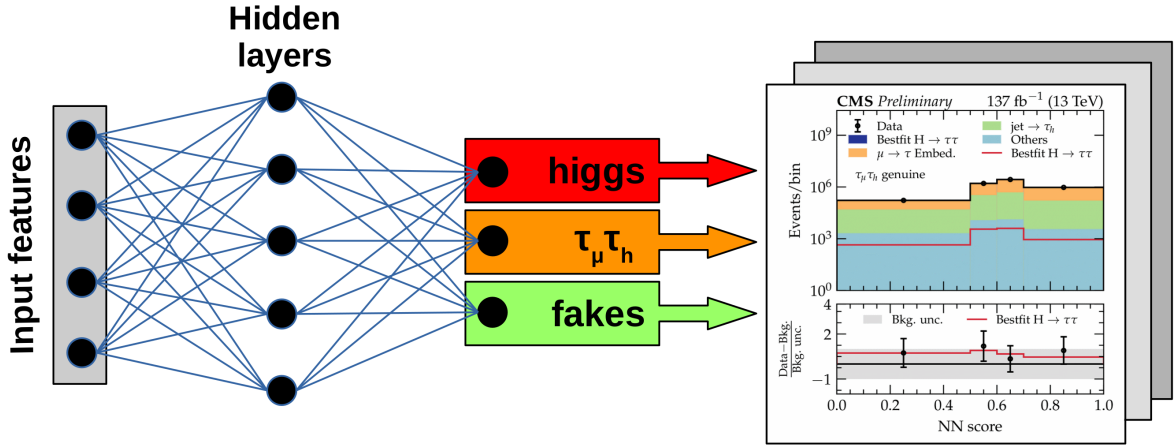
Figure 6.18: Schematic representation of the NN structure (image done by the thesis author taking inspiration from [179]): input features are taken from simulated or recorded events and their values are used to estimate the probability for each event to be assigned to the signal category "higgs" or the background classes "$\tau_\mu \tau_h$" or "fakes".

it is a fully connected feed-forward NN [153] with three output classes:

- **higgs**: the category targeting the signal processes, namely the Higgs boson production via gluon-gluon fusion (ggH), vector boson fusion (VBF) and Higgs-strahlung from Z or W bosons (ZH and WH, or more generally VH);

- **taus** or $\tau_\mu\tau_h$: targeting processes involving genuine τ leptons not originated by Higgs decays, e.g. $Z \to \tau\tau$, $t\bar{t} \to \tau\tau + 2$b-jet and $VV \to \tau\tau$;

- **fakes**: all processes involving leptons (mostly muons) or jets misidentified as hadronically decaying tau leptons ($l/j \to \tau_h$) are assigned to this category.

6.7.1   Input features

The input layer comprises 13 nodes, corresponding to the 13 input variables chosen for the NN classification:

- transverse momentum of muon and tau: $p_{T,\mu}$ and $p_{T,\tau}$;

- visible mass and transverse momentum of the $\tau_\mu\tau_h$ system: m_{vis} and $p_{T,\tau\tau}$;

- invariant mass (see Section 6.3.1) for the $\tau_\mu\tau_h$ system: $m_{\tau\tau}$;

- jet multiplicity: n_{jets};

- leading and subleading jet p_T[4]: $p_{T,lead.j}$ and $p_{T,trail.j}$;

- transverse momentum and invariant mass of the jet system[4]: m_{jj} and $p_{T,jj}$;

- separation in pseudorapidity between the two jets[4]: $\Delta\eta_{jj}$;

- MET and m_T, computed with the PUPPI-MET algorithm (see Section 4.1.6).

The variables have been chosen based on their potential in separating events between the three output classes. The invariant mass of the two tau leptons, shown in Fig. 6.19, provides a good discrimination power between $H \to \tau\tau$ and $Z \to \tau\tau$ events, which differ mainly due to the invariant mass of the intermediate bosonic resonance. The signal templates are shown overlaid onto the background ones and peak around 125 GeV, compared to the genuine tau pair background (yellow histogram) which peaks around 90 GeV. Other differences between these decays stem from the spin nature of the Higgs and Z bosons, and have not been included as NN inputs in order to avoid introducing biases towards

---

[4]If available: if one or no jets are found in the event ($n_{jets} \leq 1$) quantities related to the di-jet system or the trailing jet are set to -10 ($m_{jj} = p_{T,jj} = \Delta\eta_{jj} = p_{T,trail.j} = -10$), if $n_{jets} = 0$ the momentum of the leading jet is also set to -10.

a particular CP mixing angle for the Higgs Yukawa coupling to tau leptons. The $m_{\tau\tau}$ variable is evaluated using the SVFit algorithm as discussed in Section 6.3.1.

Generally, a NN can take advantage of the correlation between input variables to improve the efficiency of the classification. Indeed it was observed that the inclusion of both $m_{\tau\tau}$ and some of the quantities used to derive it allow to improve the NN performance. As shown in Fig. 6.20, the distributions for the transverse momentum of the two leptons, are different for Higgs decays and for the dominant background processes. This is particularly evident in the $p_{T,\tau}$ distribution. The m_{vis}, $p_{T,\tau\tau}$ and MET variables were introduced following the same logic, in order to take advantage of their correlation with $m_{\tau\tau}$.
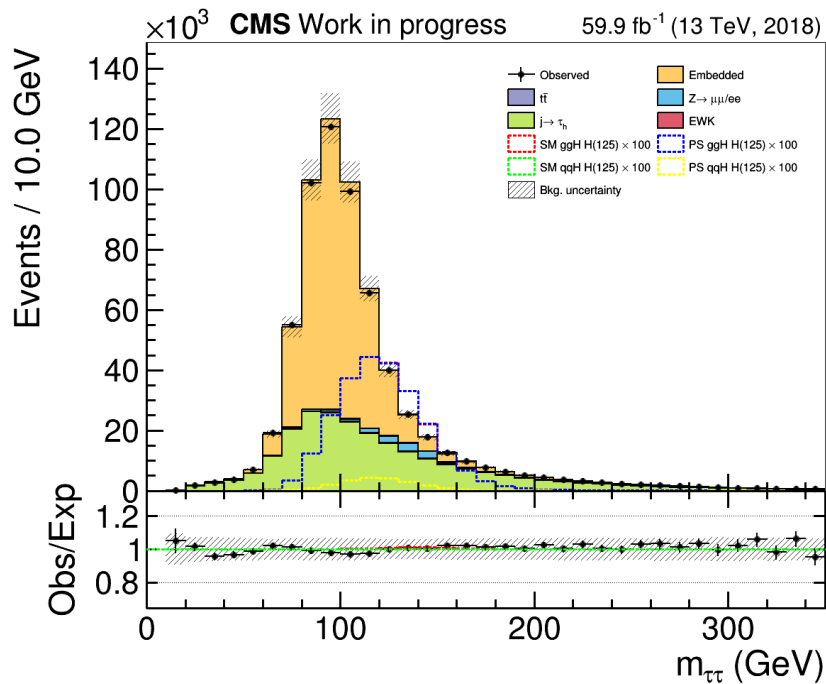
Figure 6.19: Most relevant input variable for the NN training: invariant tau lepton pair mass in 2018.

The variables targeting the jet system were instead introduced to target specific phase-space regions:

- events with one jet can be associated to Higgs bosons recoiling against a jet radiated from the initial state (ISR): in these cases the leading jet p_T (labeled $p_{T,lead.j}$ in Fig. 6.20) is correlated to the Higgs transverse momentum;

- events with two or more jets can be used to target the VBF production mechanism: observables like the invariant mass, transverse momentum, and pseudorapidity separation of/between the two leading jets can be used to study this event topology and identify Higgs decays from the relevant backgrounds.
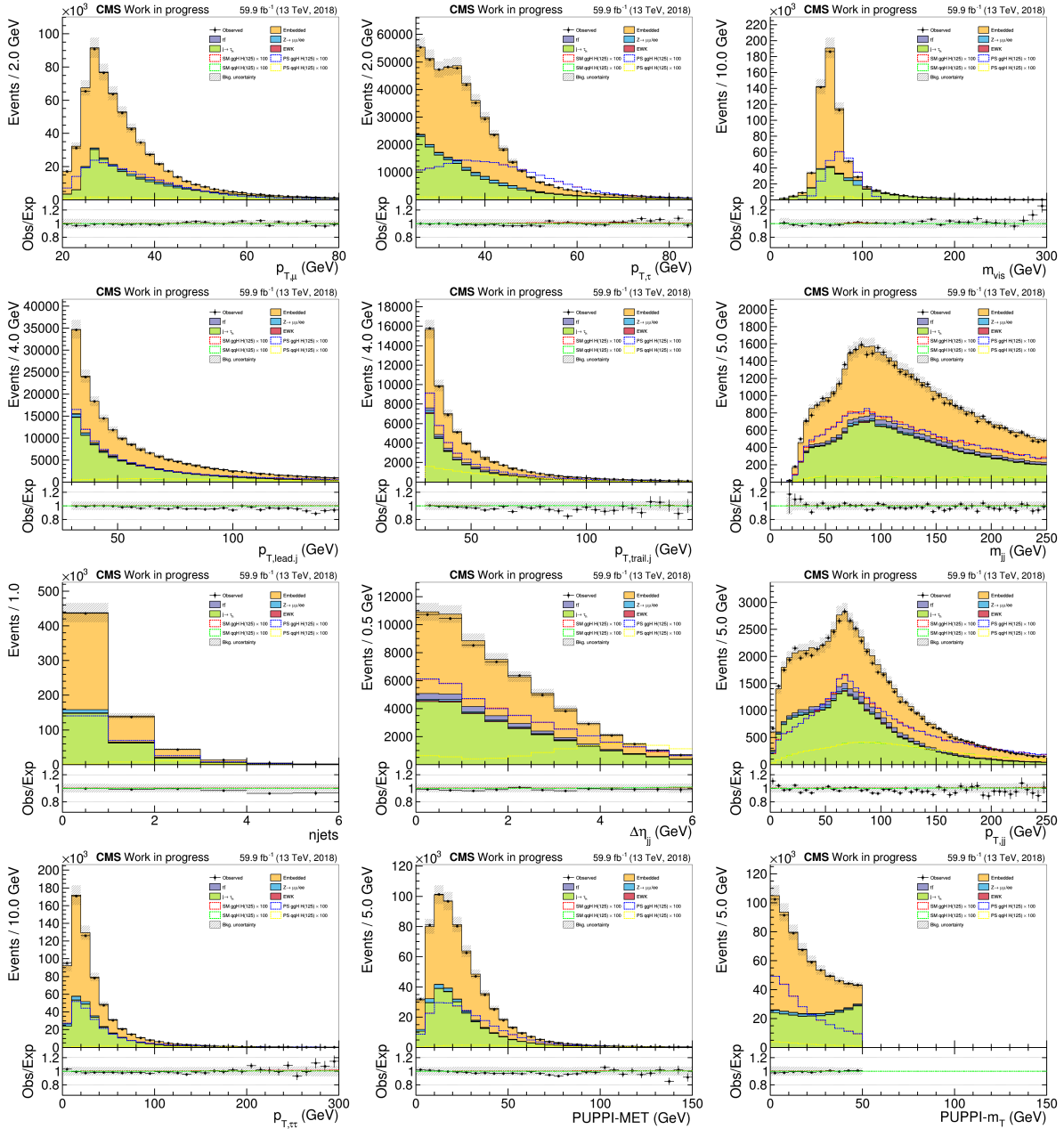
Figure 6.20: Input variables for the NN classification using 2018 signal and background models, and data. **First row:** transverse momentum of muon (left) and τ_h (middle), and their visible mass (right). **Second row:** transverse momentum of leading (left) and subleading (middle) jets, and their invariant mass (right). **Third row:** number of (left), eta separation between (middle) and transverse momentum of (right) the jets. **Bottom row:** transverse momentum of the lepton pair (left), missing transverse energy (middle), and transverse mass distribution (right).

The number of jets (n_{jets}) is used in order to define which jet related variables are defined in an event. The leading jet p_T is included in the training only when $n_{jets} \geq 1$, while m_{jj}, $p_{T,jj}$, $\Delta\eta_{jj}$ and $p_{T,trail.j}$ are considered when $n_{jets} \geq 2$.

The inclusion of m_T was motivated by the possibility of improving the separation between W+Jets events and other processes.

Twelve input variables are shown in Fig. 6.20 while $m_{\tau\tau}$ is shown separately in Fig. 6.19 to improve the readability of the plot. All input variables are shown for 2018 data, and signal and background models. The ones obtained for the other years are presented in Appendix D.

6.7.2   NN architecture and training

The NN approaches the classification problem by minimizing the *loss function* which is defined as the cross entropy

$$L(\{y_{ik}\}, \{y'_{jk}\}) = \sum_{k=1}^{N} c_{kj} y'_{jk} \log(y_{ik}(\{h_e\}, \{\vec{x}_k\})) \ , \tag{6.21}$$

where

$$\{y'_{jk}\} = \begin{cases} 1, & \text{if category } i \text{ for event } k \text{ coincides with event class } j \\ 0, & \text{otherwise} \end{cases} \ , \tag{6.22}$$

the index k runs over the events used for the training and $\{\vec{x}_k\}$ are their corresponding input features. The indices i and j represent respectively the predicted class for the k-th event and the true class it belongs to. The term y'_{jk} is 1 when the event is correctly assigned to the class it should belong to, and 0 otherwise. The y_{ik} represent the output score for the event k in the output class i and depend on the input features, and the trainable parameters $\{h_e\}$ of the NN, which are varied during the training. The parameters c_{kj} are used instead to weight each class so that they are equally valued by the NN. These weights are applied since the training datasets are noticeably unbalanced in size, with the signals, in particular, having proportionally many more simulated events with respect to their cross-sections. To account for this, each event k is weighted in the training according to the target class j. The weights are computed based on the total amount of input data used and the cross-section of the associated physical processes. This allows the network to value all classes equally.

The input datasets provided to the NN are all taken from MC simulation, aside from the QCD multijet process, which is estimated with the data-driven method detailed in Section 5.2.1. At each training cycle (or *epoch*) the loss is evaluated on two separate subsets of the input data:

- training dataset: it is used to train the trainable parameters of the network;

- validation dataset: predictions for this dataset are performed using the trained NN parameters in order to evaluate the NN performance.

Of the input events, 90% were randomly selected to be used for training and 10% for validation. Training is performed separately for different subsets of the input data, referred to as batches, which contain 1000 randomly selected events among the input datasets.

The network has 3 hidden layers with 100 nodes each and uses the Rectified Linear Unit (ReLU) as an activation function. The analytic expression for this function is

$$f(x) = \max(0, x) \ , \tag{6.23}$$

and has the advantage of a constant gradient which makes training faster [153]. In the minimization of the loss function, it also reduces the likelihood of a vanishing gradient, i.e. of progressing towards the loss minimization asymptotically, slowing down the learning rate. In order to interpret the three output scores as probabilities, a softmax activation function is used for the output layer. This function takes the form:

$$f(\vec{x})_i = \frac{\exp\{x_i\}}{\sum_j^3 \exp\{x_j\}} \ , \tag{6.24}$$

with $i \in \{1, 2, 3\}$ being the index running over the three output categories. This function returns values between 0 and 1, and can be interpreted as the probability for an event to be assigned to the i-th class. The three output scores also sum up to 1, allowing for a more robust probabilistic interpretation of the outputs.

At each training cycle, 50 randomly chosen node outputs per layer are dropped. This method [180] prevents nodes from becoming either unnecessary or fundamental for the classification, as the network is forced to train and use all nodes of each layer in order to solve the classification problem. This is commonly referred to as using a *drop-out rate* of 0.5. The loss function minimization is performed by the Adam optimizer [181] which follows a stochastic gradient descent approach with a learning rate of 10^{-4}.

The training is stopped if no improvement is found in the span of 20 epochs or if no convergence is found in 300 epochs. In practice, the second condition was put in place to keep the training-time limited, but the training reached convergence earlier in most cases.

The training is performed in two folds, with input data being divided in two equally large datasets. For clarity sake, let us label the datasets as 1 and 2: dataset 1 is used to train a network, whose predictions are evaluated on dataset 2. A second network is then trained on dataset 2 and used to make a prediction for dataset 1. This allows the entire statistics to be used in the analysis and avoids making predictions on the same data used for training.

All hyper-parameters and characteristics of the NN architecture are summarized in Table 6.6 for convenience of the reader.

Table 6.6: Set of optimized parameters and architecture choices for the neural network used in this analysis.

Parameter	Value
Number of layers	3
Nodes per layer	100
Drop-out rate	0.5
Activation function	ReLU
Output function	Softmax
Optimizer	Adam
Learning rate	0.0001
Batch size	1000
Early stop	20 epochs
Max epochs	300 epochs

6.7.3 Output of the NN

Events are assigned to the class which has the highest output score of the three. Since a softmax function is used for the last layer of the network, the output scores:

- are distributed between 0 and 1;

- sum up to 1.

The highest of the three scores must therefore always be larger than $1/3$.

In Fig. 6.21 the output score of each class is shown for the events which have been assigned to it and is referred to as *NN Score*. Its binning has been chosen in order to avoid empty bins in the plot: the distributions start from 0.3, as no values lower than $1/3$ are possible, and the bins are merged in the region *NN Score* > 0.7 for the $\tau_\mu\tau_h$ category (bottom left plot of Fig. 6.21), as the number of events with *NN Score* > 0.85 was found to be quite low. As can be observed in the *NN Score* distribution for the higgs category (top plot of Fig. 6.21), the high *NN Score* region is relatively pure in $H \to \tau\tau$ events, as the signal over background ratio reaches $\sim 15\%$. The background categories presented in the bottom row show that the NN can correctly classify a large fraction of the background events. Each category is dominated by the associated process, i.e. the genuine tau background for the $\tau_\mu\tau_h$ category and processes involving $j \to \tau_h$ misidentification for the fakes class. The *NN Score* distributions shown in Fig. 6.21 have
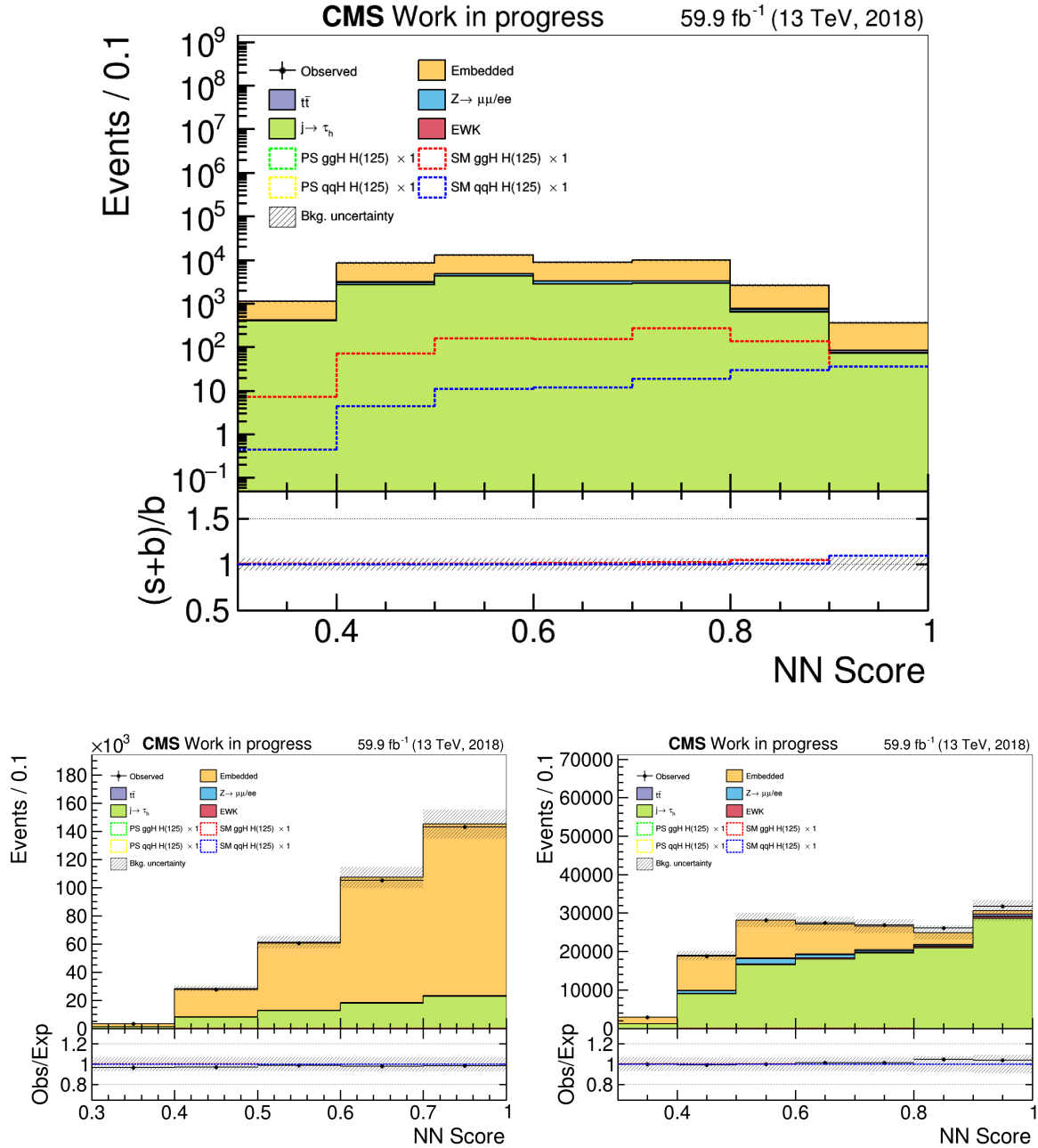
Figure 6.21: Output scores for the three NN classes: higgs class (**top**), $\tau_\mu \tau_h$ (**bottom left**) and fakes (**bottom right**). As explained in the text, the *NN Score* of an event cannot be lower than 1/3, therefore all plots start from 0.3 instead of 0. For the *NN Score* distribution in the $\tau_\mu \tau_h$ category the number of events with *NN Score* > 0.85 was quite low, all events with *NN Score* $\in \{0.7, 1.0\}$ have been merged in the same bin for illustrative purposes.

been determined using simulation and recorded data for 2018, while the corresponding scores for the other years of data-taking are shown in Fig. 6.22.

A test was also performed in order to verify that the NN did not introduce a bias towards a particular CP hypothesis. The NN was trained separately using as signal simulations:

1. unpolarized tau decays;

2. tau decays under a CP-even Higgs hypothesis;

3. tau decays under a CP-odd Higgs hypothesis.

The NN outputs of the different trained networks were then compared and no noticeable difference was observed between them. As shown in the *NN Score* distributions in Fig. 6.21 and 6.22, the histograms representing the different CP hypotheses for the same Higgs production mechanisms are indistinguishable, further hinting at the absence of biases towards particular CP hypotheses.

6.8 $\varphi_{\tau\tau}$ extraction from data

After the event categorization in signal and background categories it is possible to measure the CP mixing angle in recorded data. This is done with a Maximum Likelihood (ML) fit of the signal and background models to recorded data. The fit is performed simultaneously in the signal and background categories defined by the NN classification, using the Combine statistical toolkit [155]. The exact expression of the likelihood function to be maximized is postponed to the next section, as its understanding can benefit from discussing first which categories and distributions are used in the fit. The two parameters of interest targeted by the measurement are:

- the signal strength of the $H \to \tau\tau$ process, which in terms of the reduced Yukawa coupling takes the form $\mu = \kappa_\tau^2 + \tilde{\kappa}_\tau^2$;

- the CP mixing angle: $\varphi_{\tau\tau} = \arccos\left(\tilde{\kappa}_\tau / \kappa_\tau\right)$.

The regions and variables used to perform the ML fit are therefore chosen in order to gain access to these parameters.

The higgs category is characterized by the highest signal over background ratio, and is therefore more sensitive to the signal and its properties. The ML fit in this category is performed using a two-dimensional distribution: *NN Score vs φ_{CP}*. In order to visualize the distribution of this variable both for recorded data and signal or background models it is illustrated in Fig. 6.23 unfolded into one dimension. Each bin in this variable distribution is defined by having both the *NN Score* and the acoplanarity angle in a determined range. Fig. 6.23, for example, is divided in six bins of non uniform size in
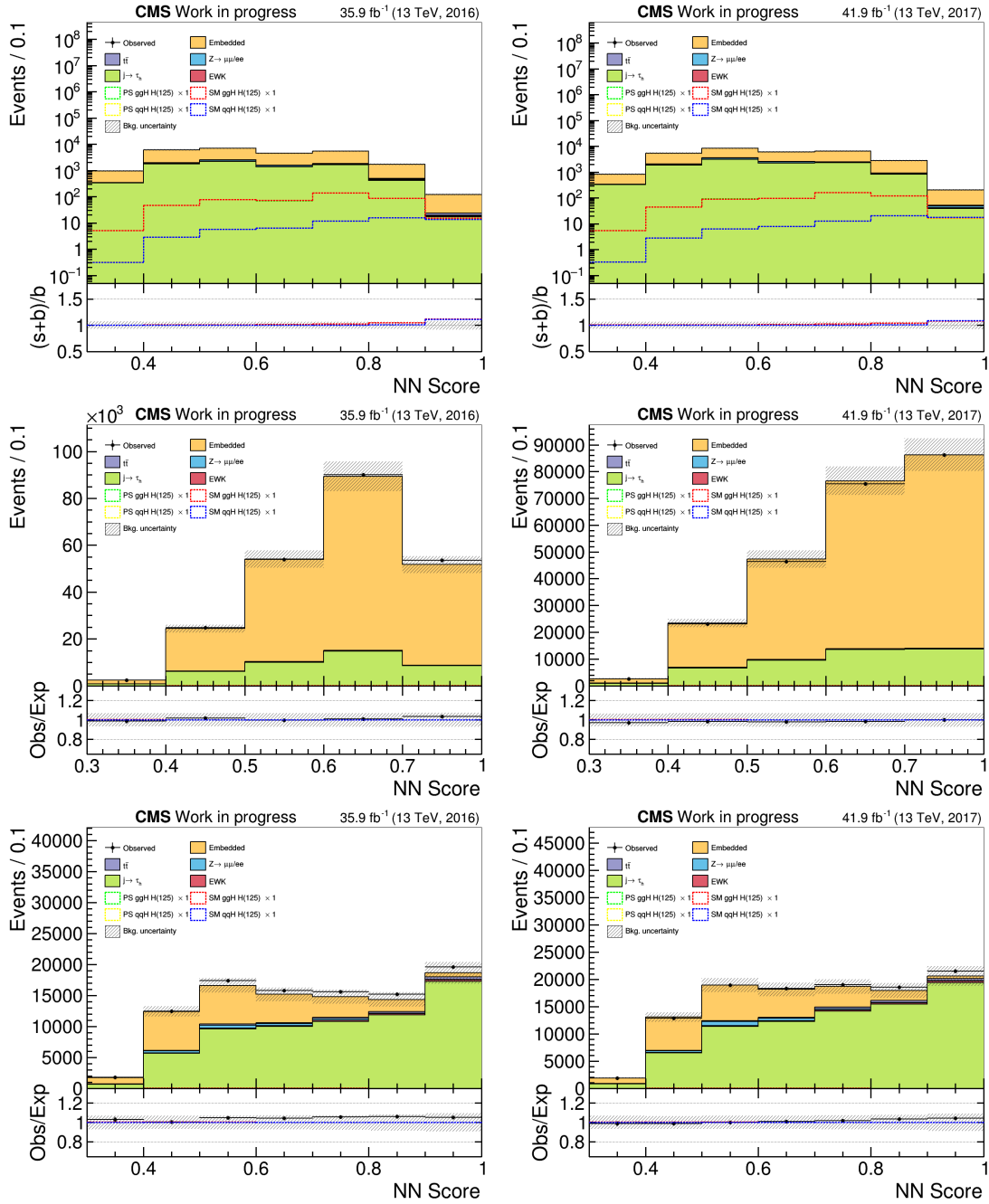
Figure 6.22: Output scores for the three NN classes in 2016 (**left**) and 2017 (**right**): higgs class (**top row**), $\tau_\mu \tau_h$ (**middle row**) and fakes (**bottom row**). The same considerations made in Fig. 6.21 regarding the binning of *NN Score* apply also to the presented distributions.

NN Score ([0,0.45,0.6,0.7,0.8,0.9,1.0]) and eight bins of size $\Delta\varphi_{CP} = \pi/4$ [rad]= 45° for the acoplanarity angle. In order to visualize the bins in the two variables dashed vertical lines are placed on top of the distribution. The region between two orange lines corresponds to a *window* in *NN Score*, containing events for which the *NN Score* of the higgs category is in the range written in orange. For example, the first window from the left includes events with *NN Score* < 0.45, while events with 0.45 < *NN Score* < 0.6 are assigned to the second window. Each window is then divided in bins of acoplanarity angle. This type of distribution is referred to as *unrolled* in the following sections.
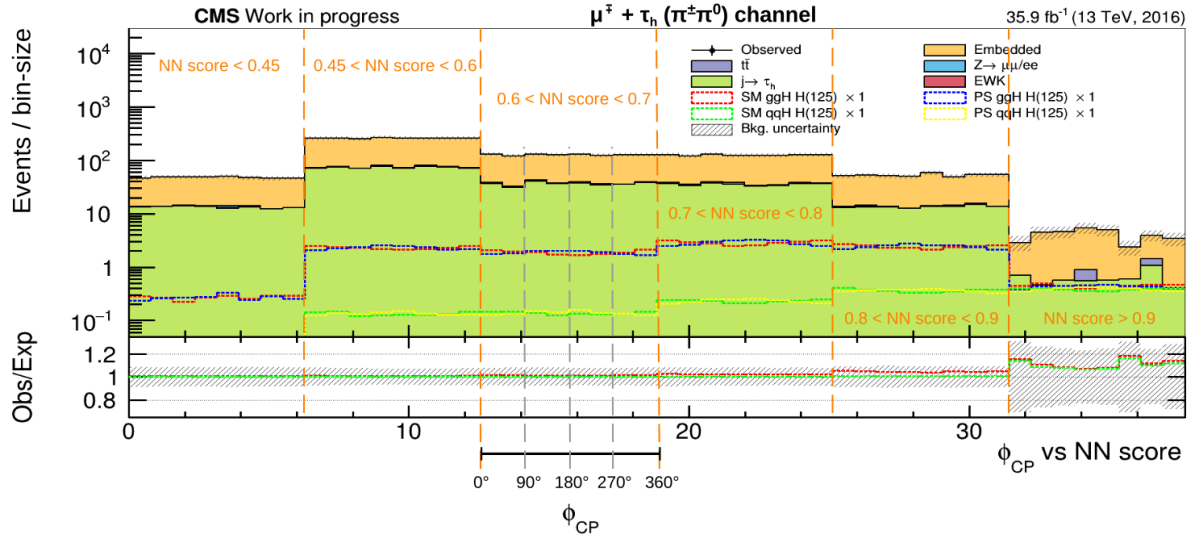
Figure 6.23: Signal category for the $\tau_\mu\tau_\rho$ channel (labelled "$\mu^\mp + \tau_h(\pi^\pm\pi^0)$ channel" on top of the figure). As explained in the text it is the unfolding of the two dimensional *NN Score vs* φ_{CP} distribution. The bins in *NN Score* are represented by the orange vertical dashed lines, while the binning in acoplanarity angle in emphasized by the vertical gray dashed lines in order to show the variation of the acplanarity angle between 0° and 360° using the black line in the lower part of the plot.

In the background categories, which are less sensitive to the signal, the ML fit is performed with the one-dimensional distribution of *NN Score*. The purpose of including the background categories in the fit is to constrain nuisance parameters affecting predictions of both background and signal and transfer these constraints to the signal class.

Due to the different modulations in acoplanarity angle between different tau decay channels, the signal category must be further divided. Fig. 6.23 shows the unrolled distribution for the $\tau_\mu\tau_\rho$ channels, and similar distributions are created for each final state. Several different binning choices were attempted, leading to fixing the binning for the *NN Score* in each category as follows:

- 6 bins for the *NN Score* of the higgs category: [0,0.45,0.6,0.7,0.8,0.9,1], they are

the same as the ones used in Fig. 6.23;

- 4 bins for the *NN Score* of the $\tau_\mu\tau_h$ category: [0,0.5,0.6,0.7,1], similar to the one used for the bottom left plot of Fig. 6.21 but with the first two bins merged;

- 5 bins for the *NN Score* of the fakes category: [0,0.6,0.7,0.8,0.9,1], corresponding to what shown in the bottom right plot of Fig. 6.21 merging the first 3 bins.

For the signal categories, the range between 0° and 360° in acoplanarity angle is divided in bins of equal size. The number of bins was adapted to account for the different number of events in each decay channel, in particular 8, 10, 4 and 4 bins are used for the $\tau_\mu\tau_\pi$, $\tau_\mu\tau_\rho$, $\tau_\mu\tau_{a_1^{1Pr}}$ and $\tau_\mu\tau_{a_1^{3Pr}}$ channels respectively. Due to the comparatively lower statistic, in the $\tau_\mu\tau_{a_1^{1Pr}}$ channel the *NN Score* bins were reduced to 4 merging the third and fourth bin and the last two.

Table 6.7: Distribution and number of bins used for each category in the ML fit for the $\tau_\mu\tau_h$ channel. The same distributions and binning have been used for all years of data-taking.

Category	Distribution	*NN Score* bins	*NN Score* binning	φ_{CP} bins
$\tau_\mu\tau_h$	*NN Score*	4	[0,0.5,0.6,0.7,1]	-
fakes	*NN Score*	5	[0,0.6,0.7,0.8,0.9,1]	-
$\tau_\mu\tau_\rho$	*NN Score vs* φ_{CP}	6	[0,0.45,0.6,0.7,0.8,0.9,1]	10
$\tau_\mu\tau_\pi$	*NN Score vs* φ_{CP}	6	[0,0.45,0.6,0.7,0.8,0.9,1]	8
$\tau_\mu\tau_{a_1^{3Pr}}$	*NN Score vs* φ_{CP}	6	[0,0.45,0.6,0.7,0.8,0.9,1]	4
$\tau_\mu\tau_{a_1^{1Pr}}$	*NN Score vs* φ_{CP}	4	[0,0.45,0.6,0.8,1]	4

In summary, for the $\tau_\mu\tau_h$ channels the following distributions are used in the fit for the extraction of the CP mixing angle for each year:

- *NN Score* for $\tau_\mu\tau_h$ and fakes categories;

- unrolled *NN Score vs* φ_{CP} for the signal category, separated by tau decay channels: $\tau_\mu\tau_h$ with $h \in \{\pi, \rho, a_1^{1Pr}, a_1^{3Pr}\}$.

These 18 distinct distributions (corresponding to 2 background and 4 signal categories for each of the 3 years of data-taking) are referred to as *templates* in the following section, where the construction of the fit model and the treatment of the uncertainties are detailed. The templates used for the ML fit in the $\tau_\mu\tau_h$ channel are also listed in Table 6.7 together with the binning used for the *NN Score* and, if used, the number of bins for the acoplanarity angle.

6.8.1 Fit model

As detailed in Section 6.4.2, the distribution of the acoplanarity angle for $H \to \tau\tau$ for a generic mixing angle $\varphi_{\tau\tau}$ can be obtained as the combination of the acoplanarity angle distributions for three specific mixing angles. Using the same notation as in Eq. 6.18, for each Higgs production mechanism the following signal templates are defined:

- **SM**: $\mathrm{d}\sigma^{CP-even}/\mathrm{d}\varphi_{CP}$;

- **PS**: $\mathrm{d}\sigma^{CP-odd}/\mathrm{d}\varphi_{CP}$;

- **MM**: $\mathrm{d}\sigma^{CP-mix}/\mathrm{d}\varphi_{CP}$.

The signal template for the extraction of the CP mixing angle is then defined as:

$$S(\vec{\mu}, \varphi_{\tau\tau}) = \mathcal{L} \cdot \sum_i \mu_i \Bigg[\Big(\cos^2(\varphi_{\tau\tau}) - \sin(\varphi_{\tau\tau}) \cos(\varphi_{\tau\tau}) \Big) \mathbf{SM}_i$$

$$+ \Big(\sin^2(\varphi_{\tau\tau}) - \sin(\varphi_{\tau\tau}) \cos(\varphi_{\tau\tau}) \Big) \mathbf{PS}_i \qquad (6.25)$$

$$+ 2 \sin(\varphi_{\tau\tau}) \cos(\varphi_{\tau\tau}) \mathbf{MM}_i \Bigg] ,$$

with i running over the different Higgs production mechanisms and $\mathcal{L}$ integrated recorded luminosity. The μ_i parameters represent the signal strengths for each production mechanism, they can be factorized in two terms

$$\mu_i = \mu_{ggH/V} \times \mu^{\tau\tau} , \qquad (6.26)$$

where:

- $\mu^{\tau\tau}$ represents the branching fraction of $H \to \tau\tau$ decays, it scales the coupling strengths of the Higgs boson to tau leptons, and is shared between different production mechanisms;

- μ_{ggH} and μ_V are the signal strengths for the specific production mechanism, with μ_{ggH} affecting the normalization of the ggH production mechanism and μ_V being used for the VBF and VH ones.

The main parameter of interest (POI) in the fit is $\varphi_{\tau\tau}$, which affects the overall shape of the signal templates, it is a freely floating parameter within the range $[-90°, 90°]$. The parameters μ_{ggH} and μ_V are instead treated as *rate parameters* in the fit, i.e. they are free to vary within a certain range, chosen for the analysis to be $[0, 10]$. The $\mu^{\tau\tau}$ parameter affects the normalization of all signal templates and is highly correlated to the two Higgs production signal strengths. It is therefore set to 1 for the measurement of

the CP mixing angle. The aforementioned parameters act only on the signals, there are however several other parameters which alter the shape or normalization of signal and background templates.

These parameters are usually referred to as *nuisance* parameters ($\vec{\theta}$) and they represent the systematic uncertainties of the measurement. The likelihood function can be written as a function of the aforementioned parameters as in [156]:

$$L(\varphi_{\tau\tau}, \mu^{\tau\tau} = 1, \mu_{ggH}, \mu_V, \vec{\theta}) =$$

$$\prod_{j}^{N_{cat}} \prod_{i}^{N_{bin}^j} P\left(n_{i,j} | S_{i,j}(\varphi_{\tau\tau}, \mu^{\tau\tau} = 1, \mu_{ggH}, \mu_V, \vec{\theta}) + B_{i,j}(\vec{\theta})\right) \times \prod_{m}^{N_{nuis}} C_m(\theta_m | \tilde{\theta}_m) , \qquad (6.27)$$

where P is the Poisson distribution, representing the probability of measuring a number of events ($n_{i,j}$) in the bin i of the category j based on the expected number of events for the signal ($S_{i,j}$) and the total background ($B_{i,j}$). The factors $C_m(\theta_m | \tilde{\theta}_m)$ represent the *a priori* distributions for the N_{nuis} nuisance parameters, with $\tilde{\theta}_m$ representing the nominal value of the θ_m parameter. N_{cat} represents the number of categories of the fit which, as shown in Table 6.7, is 18 in the $\tau_\mu \tau_h$ channel. N_{bin}^j represents instead the number of bins in the j-th category.

While $B_{i,j}$ depends only on the nuisance parameters, $S_{i,j}$ depends on the signal strengths and CP mixing angle as in Eq. 6.27 and on the nuisance parameters via the simulated templates SM, PS and MM. Instead of maximizing the likelihood, the Combine statistical toolkit minimizes the negative log-likelihood

$$NLL \equiv -\log\left(L(\varphi_{\tau\tau}, \mu^{\tau\tau} = 1, \mu_{ggH}, \mu_V, \vec{\theta})\right) , \qquad (6.28)$$

using Minuit [157], a minimization routine present in the toolkit.

6.8.2   Systematics

This section describes the systematic uncertainties and how they affect the various templates. The following notation is used in this section in order to discuss the nuisance parameters:

- **lnN**: log-normal uncertainty, it alters the normalization of a process and the corresponding nuisance parameter is assigned a log-normal probability density function (pdf).

- **shape**: dedicated templates are created representing the systematic variation with respect to the nominal shape, it is used to treat uncertainties which do not simply alter the normalization of a template. The corresponding nuisance is assigned as a Gaussian pdf.

- **MC**: label for processes taken directly from simulation, without the use of data-driven method, i.e. DY (ZLL), di-boson (VV), single top (ST), Z production by vector boson fusion (EWKZ) and t$\bar{\text{t}}$ events which include neither a $j \to \tau_h$ nor a genuine $\tau_\mu\tau_h$ pair.

- **fakes**: label used for the $j \to \tau_h$ misidentfied processes estimated with the FF method.

- **embedded**: label used for the processes involving genuine tau leptons estimated with the tau embedding data-driven method.

Starting from the theory uncertainties for the signal samples, the ones affecting only their normalization are divided in two groups: the uncertainties on the parameters used to compute the production cross-section, like α_S, and the theory uncertainties related to the missing higher order QCD and EW corrections, loop interference between heavy flavor quarks and other factors. Table 6.8 shows the values used for these uncertainties, which are all applied as lnN parameters. The branching fraction of Higgs to tau leptons also has its own theoretical uncertainties: 1.7% coming from higher order corrections, 0.99% and 0.62% as parametric uncertainties on quark masses and α_S respectively.

Table 6.8: Theoretical and parametric uncertainties for the Higgs production mechanisms [39].

Template	Parametric uncertainty	Theory uncertainty
ggH	3.2%	3,9%
VBF	2.1%	0.4%
WH	1.9%	1.9%
ZH	1.6%	1.6%

Some theoretical uncertainties propagate as shape uncertainties for the signal templates. The Higgs nominal production cross-sections are computed by placing renormalization (μ_R) and factorization scales (μ_F) to the value $\mu_R = \mu_F = m_H/2$ [39], with m_H being the Higgs boson mass. Varying those scales not only affects the overall normalization of the signal templates, but also the kinematic distributions of the Higgs boson and its accompanying jets, leading to the shapes of the signal templates being altered. Such effects are observed also at the stage of parton shower simulation. As such, all these uncertainties are treated by taking dedicated weights provided by the MC generators corresponding to multiplying or dividing μ_R and μ_F by 2.

The normalization of the MC simulated processes are also assigned both theoretical and experimental uncertainties. For tau embedding a normalization uncertainty is also

assigned, related to the trigger and muon identification uncertainties in the selection of di-muon events in recorded data. Table 6.9 provides the value for these uncertainties.

Table 6.9: Normalization uncertainties for the background models.

Template	Normalization uncertainty
ZLL	2%
$t\bar{t}$	4.2%
VV, ST	5%
EWKZ	4%
embedded	4%

Other lnN nuisance parameters are:

- Luminosity uncertainty: 2.5% for 2016, 2.3% for 2017 and 2.5% for 2018 applied on all MC templates;

- muon identification and trigger efficiencies: respectively 1% and 2%, applied on all MC;

- b-tag efficiency: applied on $t\bar{t}$ and ST, its values range between 1-9% based on the number of jets;

- IP_{sig} calibration uncertainty: between 1-5% depending on the MVA-DM, calculated by varying the IP calibration correction (see Section 6.3.3) by 25%

Shape systematics are computed by producing dedicated variation with respect to the nominal template. As such, the numerical value for these uncertainties varies from bin to bin and it is best to provide just a general description of the uncertainties and on which variables they depend:

- τ_h reconstruction and identification: these uncertainties have been calculated in bins of $p_{T,\tau}$ and MVA-DM and applied on MC and embedded templates, their value varies between 2-3%;

- τ_h trigger efficiency: this uncertainty is applied exclusively on events selected by the muon+tau trigger, it depends on $p_{T,\tau}$ and MVA-DM;

- τ_h energy scale (ES): it affects templates with simulated τ_h and depends on the tau HPS-DM, for MC (embedded) it varies between 0.8%-1.1% (0.2%-0.5%);

- μ ES: 0.4-2.7% depending on η_μ and applied to MC and embedded;

- $\mu \to \tau_h$ fake rate (FR) uncertainties: between 10 and 40% depending on η_{τ_h} (calculated with a method similar to the one described in Chapter 5);

- $\mu \to \tau_h$ ES: 1% variation on the τ_h candidate energy in the ZLL template;

- jet energy scale (JES): they are a total of 27 uncertainties, which alter the ES of jets based on their kinematic reconstruction, PU conditions, nature of the parton which initiated the jet and other properties;

- MET uncertainties for samples to which recoil corrections are not applied: they are based on the corrections applied to the JES and the unclustered ES which comes from the PF candidates that are not clustered within jets [142];

- MET recoil uncertainties: this systematic variation is obtained by varying within the measured uncertainties the recoil corrections, in samples which involve a Higgs, W or Z boson;

- non DY contributions to embedded: they are estimated by adding and subtracting 10% of the $t\bar{t}$ and VV events with genuine $\tau_\mu\tau_h$ final states from the embedded template;

- Z and top p_T reweighting uncertainty: shifted templates for both uncertainties are obtained by changing the correction SFs applied to correct the Z and top transverse momentum distributions, previously mentioned in Section 6.6.1. The SFs are altered as follows:

  - the top p_T correction is removed or applied two times to obtain the shifted templates;

  - the Z p_T correction is varied by $\pm 10\%$;

- fake factor uncertainties: it is a group of 80 uncertainties, which include the statistical fluctuations of QCD, $t\bar{t}$ and W+Jets templates in the determination regions, non-closure corrections for p_T and MET spectra and other uncertainties related to the FF calculations;

- pre-firing weights uncertainties: applied on all MC templates and extracted from the MC generators [122];

- bin-by-bin statistical fluctuation of background processes.

Table 6.10: Summary table of the systematic uncertainties affecting the background processes.

Description	Value	Templates affected	Type
Luminosity uncertainty	2016: 2.5% 2017: 2.3% 2018: 2.5%	MC	lnN
μ identification	1%.	MC	lnN
μ trigger	2%	MC	lnN
τ_h trigger	p_T dep.	MC	shape
b-jet veto	1-9%	$t\bar{t}$, ST	lnN
$\mu \to \tau_h$ FR	η_{τ_h} dep. (20-40%)	MC with $l \to \tau_h$	shape
Muon and pion IP_{sig} calib.	25%	MC	lnN
τ_h identification	p_T/DM dep. (2-3%)	MC, embedded	shape
τ_h ES	1% 1.5%	MC embedded	shape
μ ES	0.4–2.7%	MC, embedded	shape
$\mu \to \tau_h$ ES	1%	MC	shape
Jet ES	event-dep.	MC	shape
MET recoil corr.	event-dep.	MC	shape
MET unclustered ES	event-dep.	MC	shape
non-DY in embedded	10%	embedded	shape
top p_T reweighing	10%	$t\bar{t}$, ST	shape
Z p_T reweighing	10%	DY	shape
FF uncertainties	event-dep.	fakes	shape
Pre-firing	event-dep.	MC	lnN
Bin-by-bin stat. uncertainty	event-dep.	All	shape

The use of data-driven methods to model the dominant backgrounds allows to strongly constrain this last uncertainty. Additionally, most processes involved in this analysis have well defined modulations with respect to the acoplanarity angle. This allows to further reduce this uncertainty as shown in the next section.

Table 6.10 summarizes the systematic uncertainties affecting the backgrounds, complementing the ones listed in Table. 6.9.

6.8.3 Template smoothing

As previously discussed in Section 2.2.2, the acoplanarity distribution is expected to be flat[5] for the DY process and have a sinusoidal shape for the $H \rightarrow \tau\tau$. This is true for all decay channels besides $\tau_\mu\tau_\pi$, where the IP is used to define both decay planes, as shown in Fig. 2.8. In this channel, however, the background templates are supposed to be symmetric with respect to $\varphi_{CP} = \pi$ [rad], or, equivalently, invariant under a transformation $\varphi_{CP} \rightarrow 2\pi - \varphi_{CP}$. Non flat or asymmetric behaviors in backgrounds and signals are therefore investigated assuming they are originated by:

- bin-by-bin statistical fluctuations;

- mismodelling in simulation.

For signal templates, all behaviors which deviate from the sinusoidal distribution can be interpreted as bin-by-bin statistical fluctuations. For the tau embedding, as shown in Fig. 6.14, a modulation was found in the acoplanarity angle distribution. Such modulation is not present in DY MC simulation, nor in generator-level distributions for embedded tau leptons. Such a modulation was interpreted as a mismodelling originated from the tau leptons being simulated into an empty detector, not properly aligned with the real one.
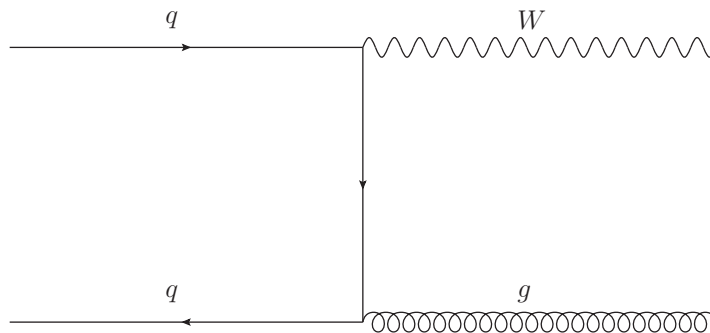
Figure 6.24: Feyman diagram for W boson production in association with a jet.

---

[5]Ignoring acceptance effects in the event selection which could lead to a modulation to become discernible, as shown in Appendix B.

For processes with $j \to \tau_h$ fakes the acoplanarity angle distribution shows a modulation in φ_{CP} in all channels. This modulation is observed both using the FF method and MC simulation of W+Jets. For the latter, the modulation is present also at generator level, hinting to a possible physical effect. For W+Jets events (Fig. 6.24), the muon and the jet are often emitted back-to-back. The IPs are orthogonal to the charged particle momenta by definition, resulting in the acoplanarity angle peaking at $\pi/2$ and $3/2\,\pi$. However, the symmetry of the distribution with respect to π is still maintained.

Taking advantage of these established modulations, all background and signal templates have been smoothed to reduce the effect of bin-by-bin statistical fluctuation. In the case of embedded samples, this also allowed to correct for a non physical effect. To explain how the smoothing is performed, let $n(\phi)$ $(n'(\phi))$ be the number of events in the bin corresponding to the acoplanarity angle $\varphi_{CP} = \phi$ before (after) the smoothing. The templates associated to $j \to \tau_h$ fakes have been *symmetrized*: the bin contents corresponding to the angles symmetric with respect to π are averaged together (top plot in Fig. 6.25):

$$n'(\phi) = n'(2\pi - \phi) = \frac{n(2\pi - \phi) + n(\phi)}{2} \ . \tag{6.29}$$

The same operation is performed on the CP-even and CP-odd signal templates in all channels, and, additionally, on all background processes in the $\tau_\mu\tau_\pi$ channel.

In all other categories, the processes involving genuine $\tau_\mu\tau_h$ and $l \to \tau_h$ are *flattened*, meaning that the template is averaged over the whole range in acoplanarity angle (middle plot in Fig. 6.25).

The CP-mix signal template is *anti-symmetrized* (bottom plot in Fig. 6.25): by defining $\hat{n}$ as the average of the CP-mix template over the range $[0, 2\pi]$, the difference between the bin content and $\hat{n}$ can be averaged between bins placed symmetrically with respect to π:

$$\Delta n_\phi = \frac{|n(\phi) - \hat{n}| + |n(2\pi - \phi) - \hat{n}|}{2} \ . \tag{6.30}$$

The bin content is then set to:

$$n'(\phi) = \hat{n} + \text{sign}(n(\phi) - \hat{n}) \times \Delta n_\phi \ , \tag{6.31}$$

$$n'(2\pi - \phi) = \hat{n} + \text{sign}(n(2\pi - \phi) - \hat{n}) \times \Delta n_\phi \ . \tag{6.32}$$

The smoothing of both signal and background distributions allowed to noticeably reduce the bin-by-bin statistical uncertainty, and led to an improvement of 6% in the results of the analysis.
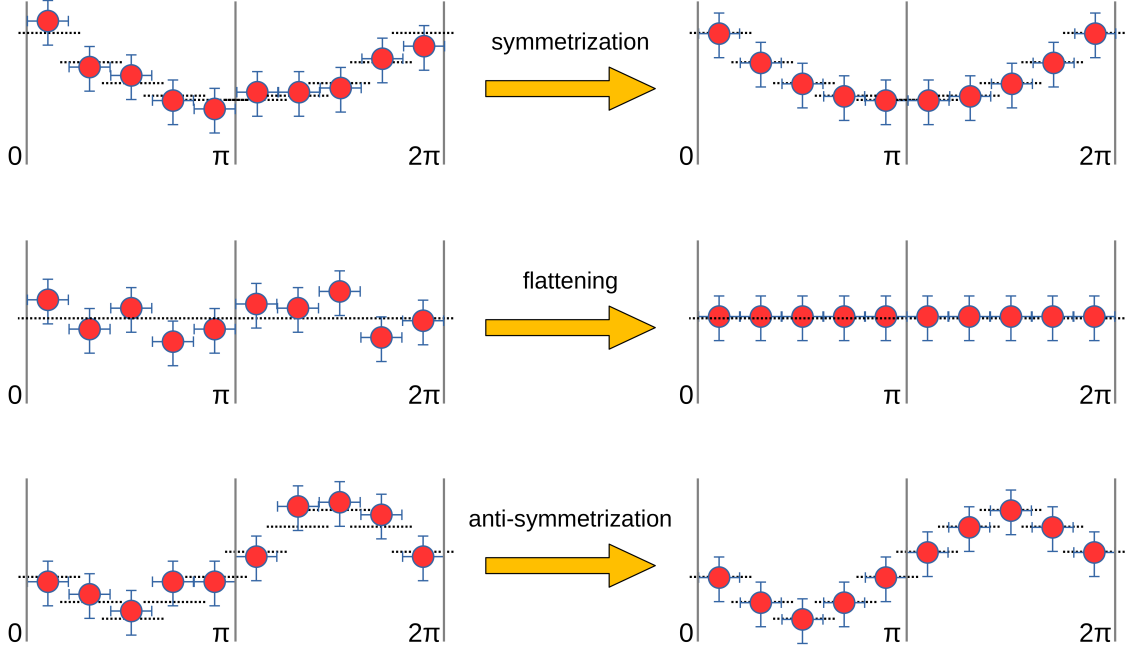
Figure 6.25: Smoothing for background and signal templates: symmetrization (top plot), flattening (middle plot) and anti-symmetrization (bottom plot) of the acoplanarity angle distribution. The red circles with error bars represent the contents of the various bins in φ_{CP} and the black horizontal dashed lines represent the averages computed as described in the text. The effect of the smoothing is shown by the red circles, which have random fluctuation on the left, assuming their expected position on the right.

6.9 Results

It is common for analyses in experimental high energy physics to follow a *blinding policy*: the measurement should be first performed using only the background and signal models without looking at data. This analysis strategy started spreading in the last decades as a way to ensure the reproducibility of an analysis and make the results more solid against scrutiny. This helps making the results unbiased with respect to the statistical fluctuations of data, by preventing the optimization of an analysis for the specific dataset being used. The measurement of the CP mixing angle targets the properties of a relatively small signal with respect to a large background, and is also the first measurement performed on this physical observable in $H \rightarrow \tau\tau$ decays. The analysis was therefore optimized in order to maximize the distinction between the extreme cases, represented by the Yukawa couplings being either purely CP-even or purely CP-odd. The negative log-likelihood minimization is performed on an Asimov dataset [182], defined as

the one that, when used to evaluate the best estimates for the $\vec{\mu}$ and $\vec{\theta}$ parameters of the likelihood function, returns their nominal values.

6.9.1   Expected results

This section summarizes the results obtained for the extraction of the CP mixing angle using an Asimov dataset. Since the likelihood function depends on a large number of parameters it is difficult to visualize its maximum. The results are therefore presented as a likelihood *profile*: the minimization of the negative log-likelihood is performed on all but one parameter, which is assigned a fixed value. This results in a series of estimated values for the negative log-likelihood which can be plotted as a function of the fixed parameter values. Since the aim of the measurement is to find the CP mixing angle that best agrees with the recorded data, a test statistic can be used to distinguish between
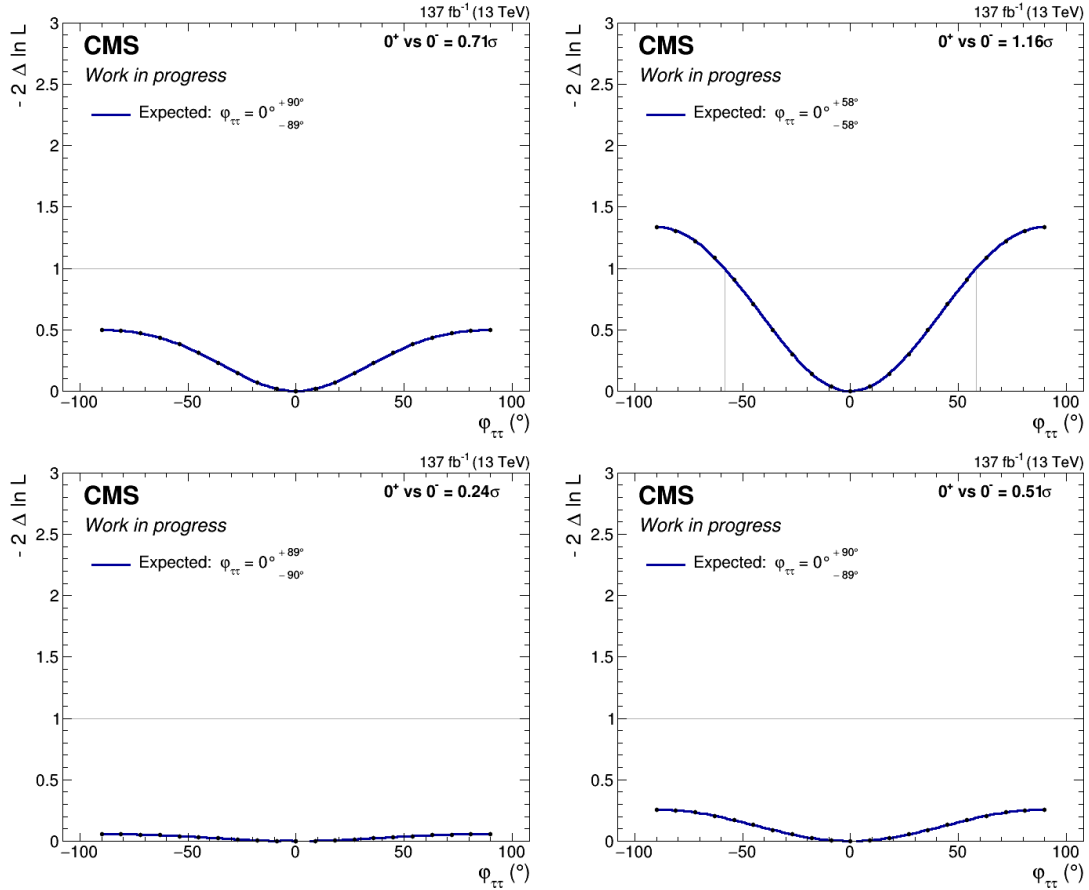
Figure 6.26: Profile likelihood scan with respect to the $\varphi_{\tau\tau}$ angle using Asimov dataset for $\tau_\mu\tau_\pi$ (**top left**), $\tau_\mu\tau_\rho$ (**top right**), $\tau_\mu\tau_{a_1^{1Pr}}$ (**bottom left**) and $\tau_\mu\tau_{a_1^{3Pr}}$ (**bottom right**).

different CP hypothesis. Following the Neyman-Pearson lemma [62], such test statistic can be defined as:

$$\lambda(\varphi_{\tau\tau}) = -2\ln\frac{L(\varphi_{\tau\tau})}{L(\widehat{\varphi_{\tau\tau}})} \ , \tag{6.33}$$

where the likelihood function has been calculated for a generic mixing angle $\varphi_{\tau\tau}$, while $\widehat{\varphi_{\tau\tau}}$ represents the mixing angle which allowed to obtain a global minimum for the likelihood. For the calculation of the expected result, the Asimov dataset is constructed in order to have $\widehat{\varphi_{\tau\tau}} = 0°$.

Fig. 6.26 shows the negative log-likelihood profile with respect to the CP mixing angle for each studied tau decay channel. The separation between CP-even and CP-odd hypotheses is represented in the figure as the difference in negative log-likelihood for $\varphi_{\tau\tau} = 0°$ and $\varphi_{\tau\tau} = 90°$. In the plots this is indicated by the value in the top right corner, namely the significance of the separation between CP hypotheses [62]. With respect to the test statistic $\lambda(\varphi_{\tau\tau})$ a significance of 1, 2 and 3 standard deviations (σ) is taken to equate the confidence levels (CLs) for the rejection of the CP-odd hypothesis with 68, 95 and 99.7% respectively. The most sensitive channel is the $\tau_\mu\tau_\rho$, as it has the
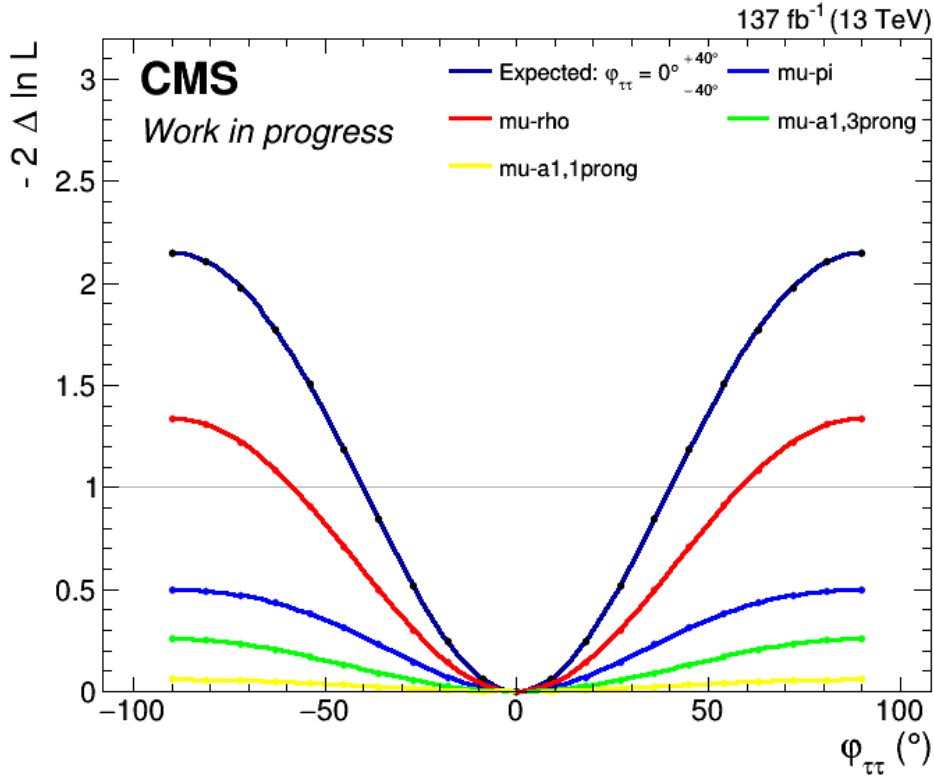
Figure 6.27: Profile likelihood scan with respect to the $\varphi_{\tau\tau}$ angle using Asimov dataset for the combined $\tau_\mu\tau_h$ channel and the ones separated by tau decay channels.

largest branching fraction, and therefore number of reconstructed events. The separation between CP-hypothesis in this channel is good, and allows to exclude a pure CP-odd hypothesis at 68% CL. The $\tau_\mu \tau_\pi$ channel also shows a high discrimination power between the two CP hypotheses, despite the difficulties of precisely reconstructing the muons and pions impact parameters. The channels corresponding to the a_1 meson resonance have lower sensitivity due to the lower branching fractions and the fact that the acoplanarity angle is reconstructed using only the intermediate ρ resonance. The combination of all channels leads to an expected limit at 68% CL on the CP mixing angle of

$$\varphi_{\tau\tau} = (\mathbf{0 \pm 40})° \,, \tag{6.34}$$

corresponding to a $\mathbf{1.47\sigma}$ significance for the exclusion of the CP-odd hypothesis.

Fig. 6.27 shows the combined fit of all different decay channels in dark blue, together with the likelihood profiles associated to the individual channels: $\tau_\mu \tau_\pi$ in light blue, $\tau_\mu \tau_\rho$ in red, $\tau_\mu \tau_{a_1^{3Pr}}$ in green and $\tau_\mu \tau_{a_1^{1Pr}}$ in yellow.

6.9.2 Vertex choice study

The aforementioned results have been obtained by measuring the IP of muons and pions using a helical extrapolation towards the PV. The PV was fitted by excluding the tau decay product tracks and including the BS coordinates into the fit algorithm. Furthermore a cut $IP_{sig} > 1.5$ was used to reject poorly reconstructed IPs. Both choices were made based on the study to which this section is dedicated.

The measurement of the exclusion significance for a CP-odd hypothesis was performed for the following PV definitions:

- $PV_{BS+!\tau \ trk}$: the PV used to obtain the results shown in the previous section and defined in Section 6.3.2;

- PV_{BS}: PV obtained by including the tau decay product tracks and the BS constraint in the fit;

- $PV_{!\tau \ trk}$: PV with tau decay product tracks excluded and no BS constraint;

- PV_{nom}: nominal PV, with tau decay product tracks and no BS.

A cut on the corresponding IP significance to be larger than 1 was used in all cases. A proper study would have required a calibration of the IP significance for each PV, which would be relatively time consuming. Therefore the measurement was performed first using the uncalibrated IPs, with the idea of performing the IP calibration only for the most promising PV definitions.

The PV_{nom} and $PV_{!\tau \ trk}$ vertex definitions led to an estimated significance of 1.12 and 1.18σ respectively, while $PV_{BS+!\tau \ trk}$ and PV_{BS} led to 1.40 and 1.41σ respectively. The

inclusion of the BS in the PV definition was therefore proved to be the main contributor to the increase of the expected significance. The IP significance was therefore calibrated for the vertex definitions including the BS. The significance was then estimated by varying the cut on IP_{sig}. The tested values for the IP significance cut were 0.5, 1, 1.5, 2 and 2.5, and the same cut was applied consistently on both muon and pion IP_{sig}. The highest measured values for the expected significance in the extraction of $\varphi_{\tau\tau}$ were: 1.47σ for $PV_{BS+!\tau\ trk}$ and $IP_{sig} > 1.5$ and 1.49σ for PV_{BS} and $IP_{sig} > 2$.

The two results are quite similar to each other, but a similar study performed in the $\tau_h\tau_h$ channel showed a noticeably higher sensitivity by using $PV_{BS+!\tau\ trk}$ as the PV vertex definition. The most likely explanation of this effect lays in the precision of the muon trajectory reconstruction in the CMS detector. Muon tracks allow to accurately constrain the PV coordinates, especially along the z axis, and improve the PV resolution.

A common vertex definition was adopted between the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels, i.e. $PV_{BS+!\tau\ trk}$. This prevented the introduction of additional uncertainties coming from the different vertex definitions, and led to the results presented in this thesis.

6.9.3 Unblinded results

The unblinding of the recorded data was performed in parallel with the $\tau_h\tau_h$ channel. Most of the figures shown in this section were made in the context of the combined fit [158]. The numerical values provided for the results refer instead only to the $\tau_\mu\tau_h$ channel. The unblinding was performed in subsequent stages:

1. unblinding of the background categories and the less sensitive bins of the signal categories;

2. unblinding of the less significant signal categories (the ones involving the a_1 meson resonance);

3. unblinding of all categories.

Each step was performed in order to check that the systematic uncertainties were neither overconstrained or pulled away from their nominal values, possibly indicating mistakes in the uncertainty model. As an example, Fig. 6.28 shows the pulls for the 30 most relevant nuisance parameters in the combination of the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels.

Goodness-of-fit (GOF) tests[6] were also performed to estimate the compatibility of the recorded data with the signal and background templates. The tests performed on the systematics and the GOF tests results are shown in Appendix E for the combination of the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels.

Fig. 6.29-6.33 show the unblinded distributions for all signal and background categories in the $\tau_\mu\tau_h$ channels. In all figures the red overlaid histogram and the dark blue

---

[6]The test performed were a saturated model [183] and Kolmogorov-Smirnov tests [184].
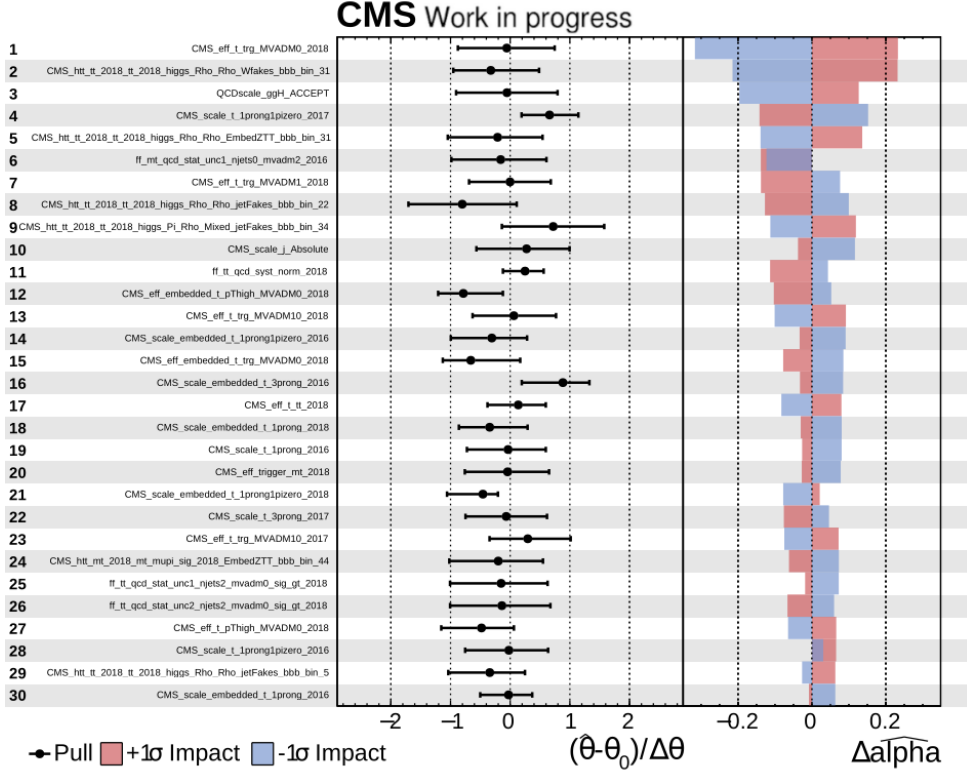
Figure 6.28: Most relevant 30 nuisance parameters and associated pulls for the combination of the $\tau_\mu \tau_h$ and $\tau_h \tau_h$ channels.

contribution to the stacked one represent the best estimate of the $H \to \tau\tau$ process determined by the combined fit. The histogram is made taking already into account the combination with the $\tau_h \tau_h$ channel, which is presented in Chapter 7. The observed CP mixing angle obtained by combining all categories in the $\tau_\mu \tau_h$ channel is

$$\varphi_{\tau\tau} = (-8 \pm 37)^\circ \, , \tag{6.35}$$

corresponding to an observed significance for the separation between CP-even and CP-odd hypotheses of

$$0^+ \text{ vs } 0^- = 1.56\sigma \, . \tag{6.36}$$

Table 6.11 summarizes the expected and observed sensitivities for the exclusion of a pure CP-odd hypothesis. The first 4 rows show the significance for each individual decay channel, while the combination is shown at the bottom row.

The $\tau_\mu \tau_\rho$ and $\tau_\mu \tau_\pi$ channels exhibit little difference between the observed significance and the predicted one. The channels involving the a_1 meson resonance exhibit instead large differences, with the 3 prong channel having more than 3 times its expected significance, and the 1 prong one having none. These discrepancies can be investigated by

Table 6.11: Expected and observed significance for the exclusion of the pure CP-odd hypothesis for the combined $\tau_\mu\tau_h$ channel and the ones separated by tau decay channels.

Channel	Exp. signif. (σ)	Obs. signif. (σ)
$\tau_\mu\tau_\pi$	0.71	0.66
$\tau_\mu\tau_\rho$	1.16	1.10
$\tau_\mu\tau_{a_1^{3Pr}}$	0.51	1.72
$\tau_\mu\tau_{a_1^{1Pr}}$	0.24	0
All channels	**1.47**	**1.56**

looking at the corresponding unblinded signal categories. In the $\tau_\mu\tau_{a_1^{1Pr}}$ channel (Fig. 6.30) it appears that the data points are very close to the background-only hypothesis, allowing no separation between CP-hypotheses. As for the $\tau_\mu\tau_{a_1^{3Pr}}$ channel (Fig. 6.31), the statistical fluctuation of data points in the second to last *NN Score* window ($0.8 < NN\ Score < 0.9$) are noticeably further from the CP-odd template (labeled PS $H \to \tau\tau$) than the expected CP-even signal, leading to a much higher observed significance.

When combining these channels with the $\tau_\mu\tau_\rho$ and $\tau_\mu\tau_\pi$ ones, the statistical fluctuations in data can be constrained better, resulting in an observed significance relatively close to the predicted one. This is shown in Table 6.11 and Fig. 6.34, where the combined observed significance is noticeably lower than the one obtained for the $\tau_\mu\tau_{a_1^{3Pr}}$ channel. The statistical fluctuation observed in the 3 prong channel has also the effect of pulling the measured CP mixing angle towards negative values.

It is possible to profile the likelihood also with respect to the $H \to \tau\tau$ signal strengths. If done using the μ_{ggH} and μ_V parameters this would allow to measure the signal strengths for the various production mechanisms. By fixing instead $\mu_{ggH} = \mu_V = 1$ and making $\mu^{\tau\tau}$ a freely floating parameter, the latter parameter takes the role of inclusive signal strength, as it scales coherently all signal templates. The expected and observed values for the $H \to \tau\tau$ inclusive signal strength is shown in Table 6.12. The observed signal strength in the $\tau_\mu\tau_{a_1^{1Pr}}$ is indeed compatible with 0 within 68% CL, as previously deduced when discussing Fig. 6.30. All other observed signal strength are instead compatible with 1 within 95% CL.

The likelihood function profile with respect to $\mu^{\tau\tau}$ is shown instead in Fig. 6.35 for the combination of all final states. The value of the test statistic for $\mu^{\tau\tau} = 0$ is ~ 18, which could be interpreted as a more than 4σ evidence for the $H \to \tau\tau$ process in the $\tau_\mu\tau_h$ final state.

Table 6.12: Expected and observed value of the inclusive $H \to \tau\tau$ signal strength for each separate decay channel and their combination.

Channel	Exp. signal strength	Obs. signal strength
$\tau_\mu \tau_\pi$	$1.0^{+0.4}_{-0.4}$	$0.4^{+0.4}_{-0.4}$
$\tau_\mu \tau_\rho$	$1.0^{+0.3}_{-0.3}$	$0.8^{+0.3}_{-0.3}$
$\tau_\mu \tau_{a_1^{3Pr}}$	$1.0^{+0.4}_{-0.4}$	$1.6^{+0.4}_{-0.4}$
$\tau_\mu \tau_{a_1^{1Pr}}$	$1.0^{+0.8}_{-0.8}$	$0.0^{+0.2}_{-0.0}$
All channels	$\mathbf{1.00 \pm 0.21}$	$\mathbf{0.74 \pm 0.19}$
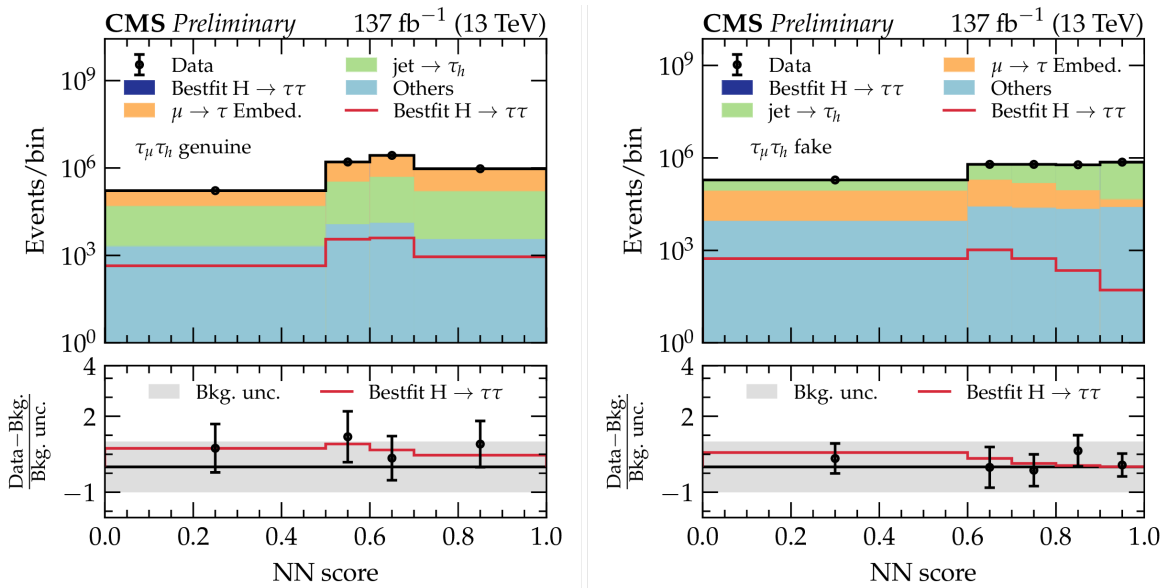
Figure 6.29: Unblinded *NN Score* distribution for the background categories [158].
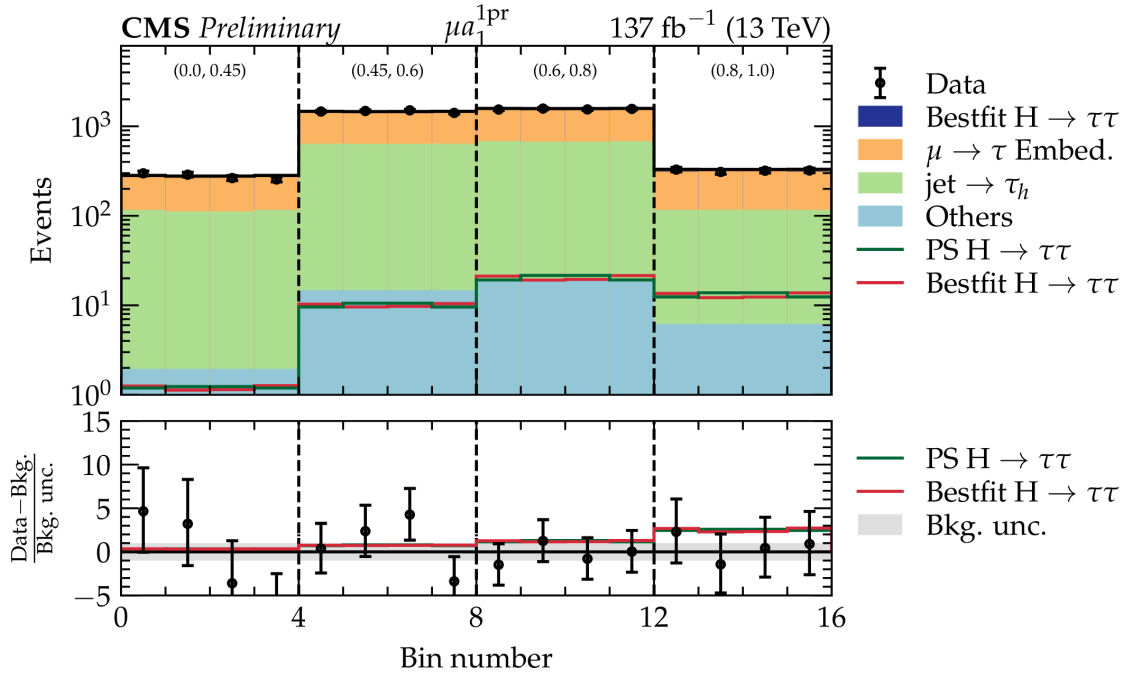
Figure 6.30: Unblinded *NN Score vs* φ_{CP} distribution for the $\tau_\mu\tau_{a_1^{1Pr}}$ signal category [158].
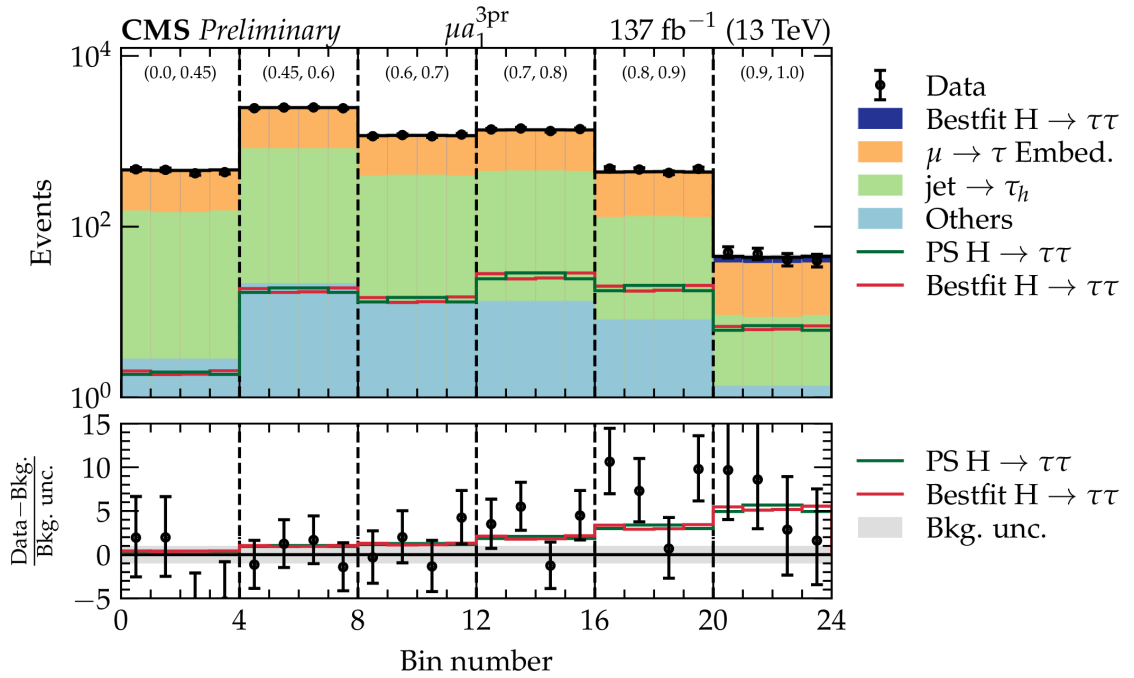
Figure 6.31: Unblinded *NN Score vs* φ_{CP} distribution for the $\tau_\mu\tau_{a_1^{1Pr}}$ signal category [158].
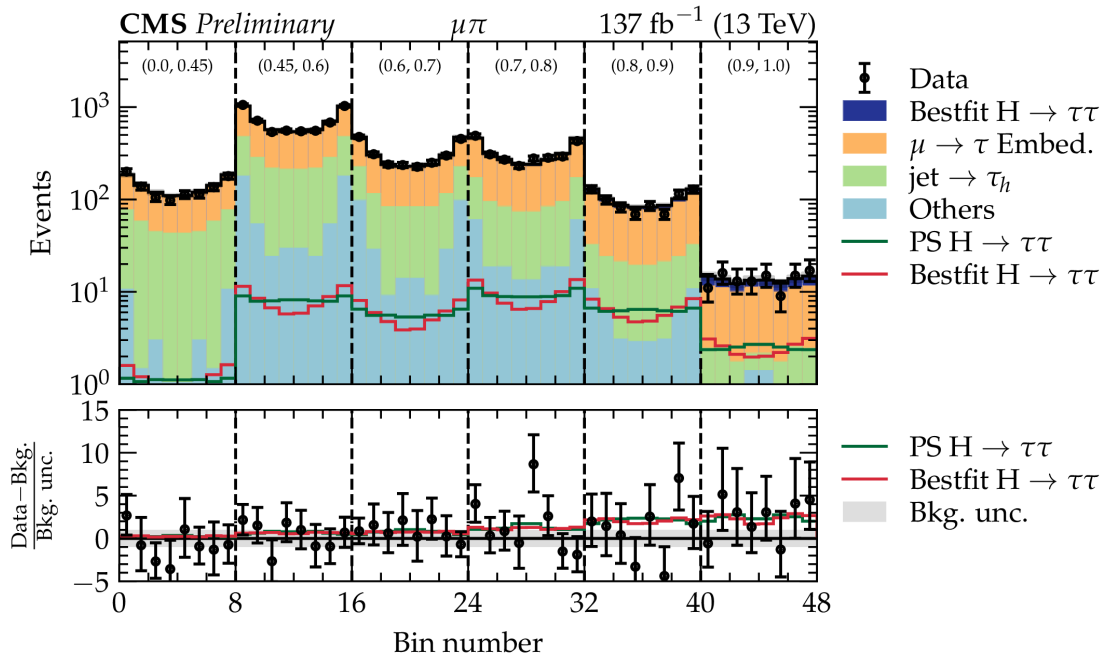
Figure 6.32: Unblinded *NN Score vs* φ_{CP} distribution for the $\tau_\mu \tau_{a_1^{1Pr}}$ signal category [158].
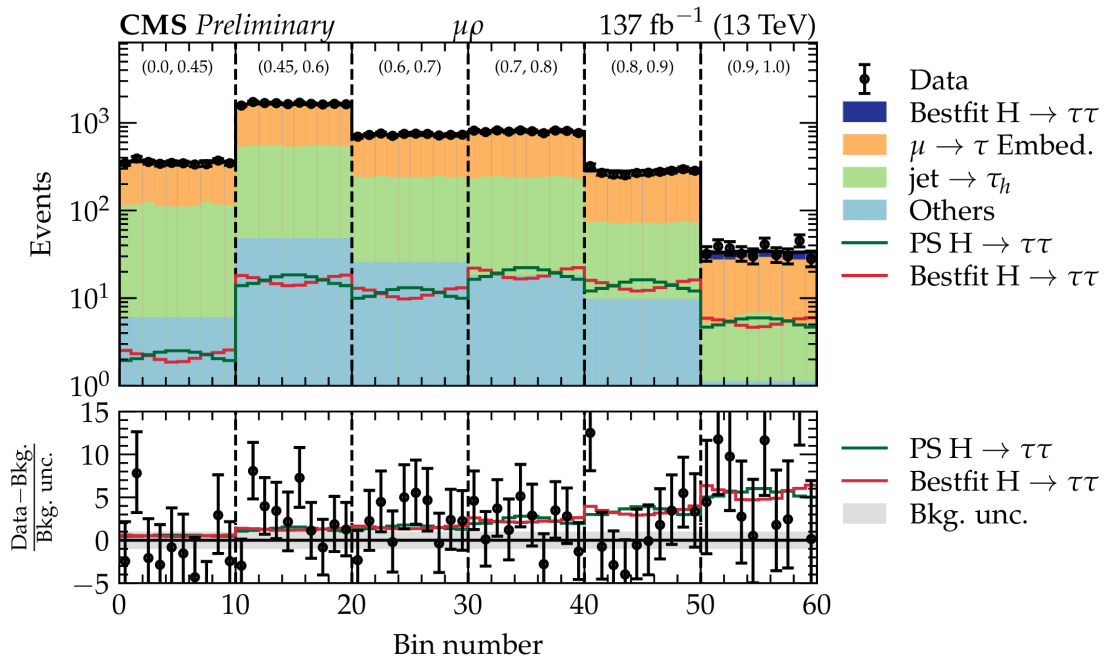
Figure 6.33: Unblinded *NN Score vs* φ_{CP} distribution for the $\tau_\mu \tau_{a_1^{1Pr}}$ signal category [158].
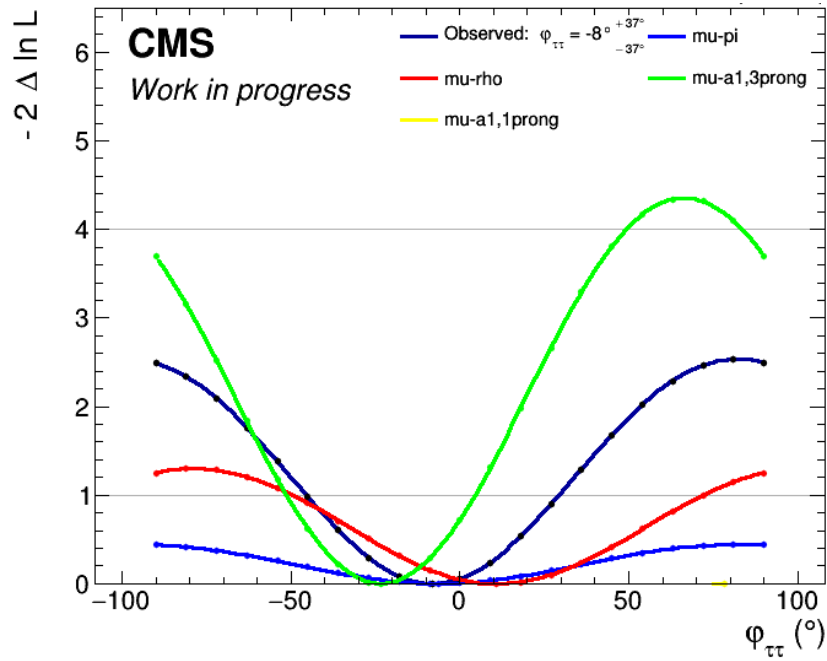
Figure 6.34: Profile likelihood scan with respect to the $\varphi_{\tau\tau}$ angle using recorded data for the combined $\tau_\mu\tau_h$ channel and the ones separated by tau decay channels.
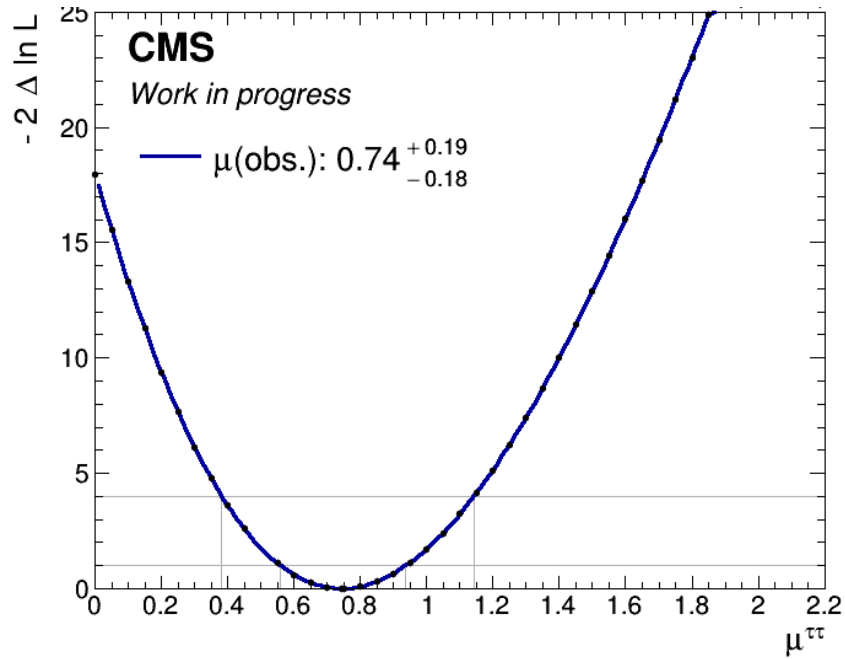
Figure 6.35: Profile likelihood scan with respect to the $\mu^{\tau\tau}$ inclusive signal strength using recorded data for the combined $\tau_\mu\tau_h$ channel.

Combination of $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels

Contents

The previous chapter showed the measurement of the CP properties of Higgs Yukawa coupling to tau leptons in the $\tau_\mu\tau_h$ channel. The same measurement was performed also in the $\tau_h\tau_h$ and $\tau_e\tau_h$ channels. The measurements performed in these other channels followed the same analysis strategy used for the $\tau_\mu\tau_h$ channel (see Section 6.1). Only minor changes were made to the background modeling and the categories of fit, and are presented in the next section.

The first measurements to be combined were the ones in the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels. A simultaneous fit was performed using the signal and background categories of both final states. The resulting measurement of the CP mixing angle is shown in Section 7.2 and were presented in the CMS Public Analysis Summary (PAS) "Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV" [158]. The results were also shown for the first time at the "International Conference on High Energy Physics" in July 2020 [185].

In a second stage the $\tau_e\tau_h$ channel was also added. At the time of this thesis writing, the results of this combination have yet to be made public and are therefore not reported in this document.

7.1   CP properties in the $\tau_h\tau_h$ channel

Compared to the semileptonic channels, which investigated four distinct signal categories, the fully hadronic one has to contend with a larger number of final states. Nine

combinations of tau decay modes were considered: $\rho \times \rho$, $\rho \times \pi$, $\rho \times a_1^{3Pr}$, $\rho \times a_1^{1Pr}$, $\pi \times \pi$, $\pi \times a_1^{3Pr}$, $\pi \times a_1^{1Pr}$, $a_1^{3Pr} \times a_1^{3Pr}$ and $a_1^{3Pr} \times a_1^{1Pr}$, corresponding to nine distinct signal categories. Events belonging to the $a_1^{1Pr} \times a_1^{1Pr}$ final state were grouped together with the $\rho \times a_1^{1Pr}$ one due to the low statistics of both channels.

As a fully hadronic final state, events are selected using a di-tau trigger, which requires the presence of two τ_h candidates with $p_T > 35$ GeV being reconstructed in the detector. For the offline selection the τ_h candidates are required to have $p_T > 40$ GeV and be reconstructed in a pseudorapidity range $|\eta| < 2.1$. The background composition is also noticeably different, with the QCD multijet production becoming the dominant background. This reflects in a different choice of triggers and event selection. As mentioned in Section 6.5.1, the data-driven estimation of the $j \to \tau_h$ background is altered in order to use only one determination region, obtained by requiring two τ_h candidates with the same sign.

Similarly to what is done for the $\tau_l\tau_h$ channels, a multivariate analysis technique is used to sort events into one signal class and two background categories. The machine learning architecture chosen for this channel was a multi-class boosted decision tree (BDT) built using the XGBoost library [166]. The three output classes are defined similarly to the ones for the semileptonic channel: one targeting the Higgs production, one for processes involving genuine $\tau_h\tau_h$ and one for $j \to \tau_h$ fakes. Nine variables are used as inputs, most of which in common with the $\tau_l\tau_h$ channel, as listed in Table 7.1, together with the ones used in the neural network for the $\tau_\mu\tau_h$ channel.

The BDT score distribution for the background categories is shown in Fig. 7.1 after executing the profile likelihood fit to measure the CP mixing angle. The $H \to \tau\tau$ process is included in the plots scaled by the best-fit value. Each category is dominated by the targeted process, with genuine $\tau_h\tau_h$ events dominating the region close to 1 in the BDT score distribution on the left of Fig. 7.1. The same can be observed in the BDT score for $j \to \tau_h$ fakes in the plot on the right. The excellent agreement between data and estimated background contribution shows that a good modelling of data was achieved.

For the signal categories, an unrolled distribution of the acoplanarity angle in regions of BDT score was used. It is defined similarly to the one presented in Section 6.9 for the $\tau_\mu\tau_h$ channel. Fig. 7.2 and 7.3 show the most sensitive signal categories for the $\tau_h\tau_h$ channel: the $\rho \times \rho$ (Fig. 7.2) and $\pi \times \rho$ (Fig. 7.3) channels. Both distributions are defined with 4 bins in BDT score ([0.0,0.7,0.8,0.9,1.0]) and 10 in acoplanarity angle. The highest bins in BDT score provide a relatively high ratio between the number of signal and background events. This in turn enhances the separation power between CP-even and CP-odd hypothesis and the sensitivity to the CP mixing angle.

---

[1]Defined using the SVFit algorithm [160] (Section 6.3.1).
[2]Defined using the PUPPI-MET algorithm [143] (Section 4.1.6).

Table 7.1: Input variables used for the event categorization via NN in the $\tau_\mu\tau_h$ channel and via BDT in the $\tau_h\tau_h$ one [158].

Observable	$\tau_\mu\tau_h$	$\tau_h\tau_h$
p_T of leading τ_h	✓	✓
p_T of trailing τ_h	✗	✓
p_T of τ_μ	✓	✗
p_T of visible di-τ	✓	✓
p_T of di-$\tau_h + p_T^{mis}$	✗	✓
Visible di-τ mass	✓	✓
di-τ mass[1]	✓	✓
Leading jet p_T	✓	✓
Trailing jet p_T	✓	✗
Jet multiplicity	✓	✓
Dijet invariant mass	✓	✓
Dijet p_T	✓	✗
Dijet $|\Delta\eta|$	✓	✗
p_T^{mis} [2]	✓	✓
$m_T(\tau_\mu, p_T^{mis})$ [2]	✓	✗
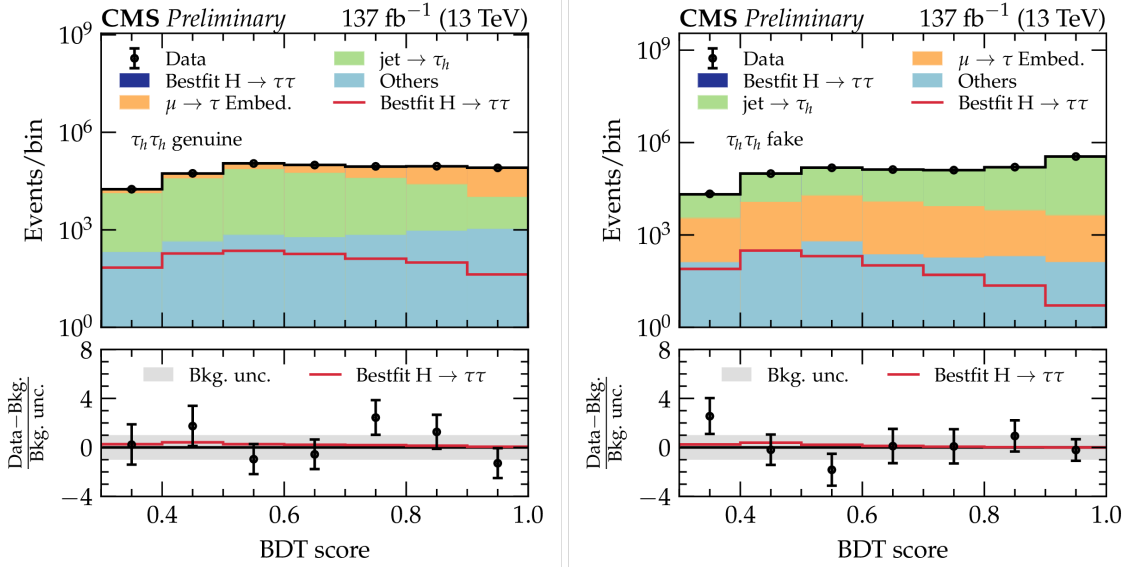
Figure 7.1: BDT score distribution for the background categories: genuine $\tau_h\tau_h$ (**left**) and misidentified τ_h (**right**) [158].
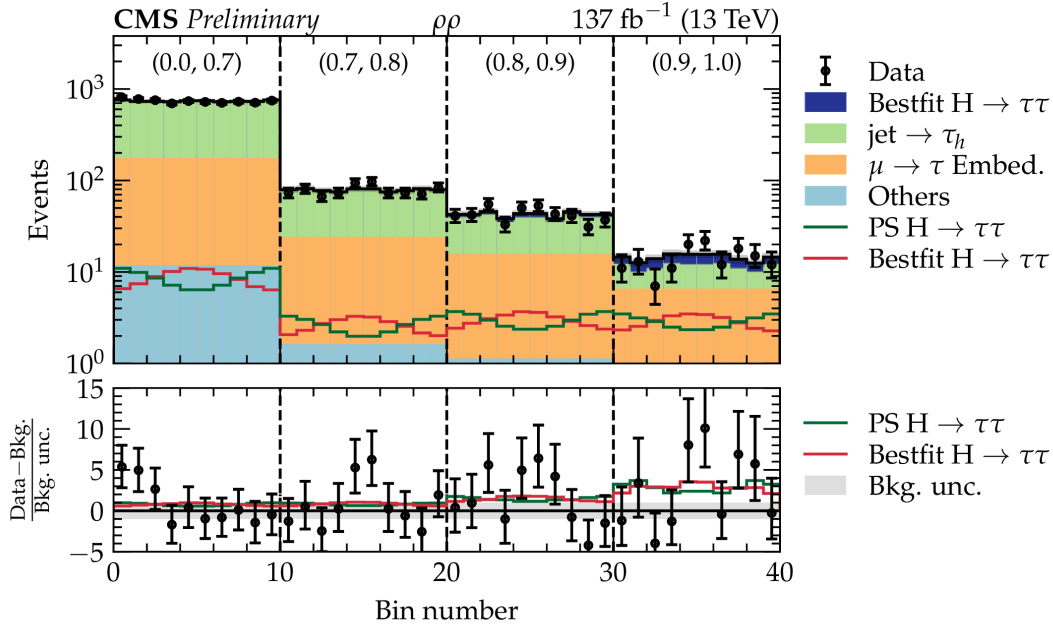
Figure 7.2: Unblinded unrolled distribution of the acoplanarity angle in windows of BDT score for the $\tau_\rho\tau_\rho$ channel [158].
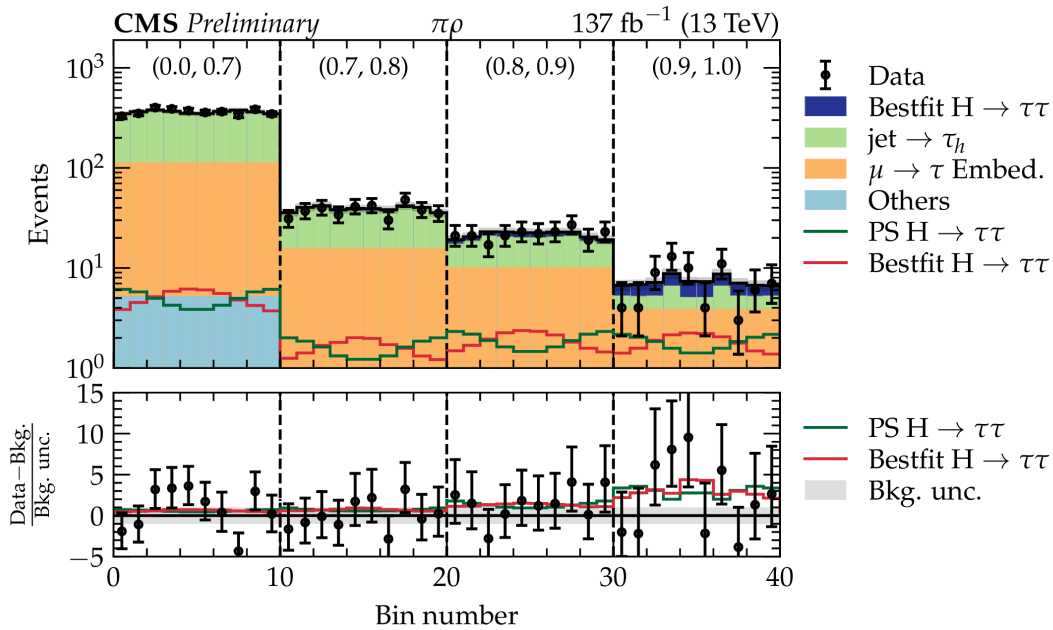
Figure 7.3: Unblinded unrolled distribution of the acoplanarity angle in windows of BDT score for the $\tau_\pi\tau_\rho$ channel [158].

7.2 $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels combination

This section describes the combination of the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels, which was the first one to be made public [158].

The fit model is the same as the one described in Section 6.8. The likelihood function shown in Eq. 6.27, reported here for convenience of the reader,

$$L(\varphi_{\tau\tau}, \mu^{\tau\tau} = 1, \mu_{ggH}, \mu_V, \vec{\theta}) =$$

$$\prod_j^{N_{cat}} \prod_i^{N_{bin}^j} P\left(n_{i,j} | S_{i,j}(\varphi_{\tau\tau}, \mu^{\tau\tau} = 1, \mu_{ggH}, \mu_V, \vec{\theta}) + B_{i,j}(\vec{\theta})\right) \times \prod_m^{N_{nuis}} C_m(\theta_m | \tilde{\theta}_m) \, , \tag{7.1}$$

depends on several parameters of interest (POIs), which are the ones associated to the $H \to \tau\tau$ process:

- $\varphi_{\tau\tau}$: the CP mixing angle;

- μ_{ggH}: the signal strength for the ggH production mechanism;

- μ_V: the signal strength for the Higgs production mechanism involving vector bosons, i.e. the VBF and VH ones;

- $\mu^{\tau\tau}$: the $H \to \tau\tau$ branching ratio (BR) relative to the SM prediction.

The technique used to measure the POIs in recorded data consists in minimizing the negative log-likelihood. The combination between channels is performed by including the signal and background categories of both channels in the fit. The signal template for the $H \to \tau\tau$ process is a weighted sum of the SM, PS and MM templates defined in Eq. 6.25, separately for each decay channel. This allows the signal templates to be scaled coherently between different decay channels when measuring the CP mixing angle, and to combine different final states in the fit. The nuisance parameters in the combined fit are then treated as correlated or uncorrelated between channels depending on whether the physical objects affected by the uncertainty is present or not in both channels.

The combined fit is performed using 17 categories for each year of data-taking (2016, 2017 and 2018):

- 6 categories for the $\tau_\mu\tau_h$ channel:

    - 2 background categories: one for processes involving genuine $\tau_\mu\tau_h$ and one for $l/j \to \tau_h$ fakes;

    - 4 signal categories[3]: $\mu \times \pi, \mu \times \rho, \mu \times a_1^{1Pr}, \mu \times a_1^{3Pr}$;

---

[3]Here the label τ_X has been removed when listing the final states in order to make them more readable.

- 11 categories for the $\tau_h\tau_h$ channel:

  - 2 background categories: one for processes involving genuine $\tau_h\tau_h$ and one for $j \to \tau_h$ fakes;

  - 9 signal categories: $\rho \times \rho$, $\rho \times \pi$, $\rho \times a_1^{3Pr}$, $\rho \times a_1^{1Pr}$, $\pi \times \pi$, $\pi \times a_1^{3Pr}$, $\pi \times a_1^{1Pr}$, $a_1^{3Pr} \times a_1^{3Pr}$ and $a_1^{3Pr} \times a_1^{1Pr}$.

Most systematic uncertainties are common between the two channels and have been introduced in Section 6.8.2. The full list of systematics considered in the combined measurement is provided in Table 7.2 to account for the ones specific to the fully hadronic channel.

The main POI of this measurement is the CP mixing angle: $\varphi_{\tau\tau}$. The profile of the likelihood with respect to this variable is shown in Fig. 7.4. The minimum of the profile likelihood observed for recorded data is found at a CP mixing angle of

$$\varphi_{\tau\tau}^{\mathbf{obs}} = (4 \pm 17)^\circ \ , \tag{7.2}$$

compared to its expected value of

$$\varphi_{\tau\tau}^{\mathbf{exp}} = (0 \pm 23)^\circ \ , \tag{7.3}$$

obtained with an Asimov dataset where the signal is produced with a pure CP-even coupling. This corresponds to an observed (expected) significance for the exclusion of a pure CP-odd hypothesis of

$$\mathbf{0^+ \ vs \ 0^-} = \mathbf{3.2}\sigma \ (\mathbf{2.3}\sigma) \ . \tag{7.4}$$

The uncertainties expressed in the aforementioned results for $\varphi_{\tau\tau}$ represent the 68% confidence level (CL) interval around the value which maximises the likelihood. This is represented in Fig. 7.4 as the values of $\varphi_{\tau\tau}$ where the likelihood function crosses the gray horizontal dashed line labeled *68%*. In order to find which are the major contributions to the uncertainty of the measurement, the likelihood can be maximized by fixing some of the nuisance parameters to their *a posteriori* values after the combined fit. This can be done to split the total uncertainty into several contributions. In this analysis the fit was performed by fixing:

- all nuisance parameters, in order to measure the pure statistical contribution to the uncertainty;

- all nuisance parameters besides the bin-by-bin statistical uncertainties of the signal and background processes;

- all nuisance parameters besides the theoretical uncertainties.

Table 7.2: Systematic uncertainties used for the combination of $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels [158]. The third column indicates if the source of uncertainty was treated as correlated between the years, while the fourth column shows the distribution used to vary the nuisance parameter.

Uncertainty	Magnitude	Correlation	Parameter distrib.
τ_h ID	p_T/DM-dep. (2–3%)	no	Gaussian
Muon reconstruction	1%.	yes	Log-normal
$e \to \tau_h$ FR	10%	no	Gaussian
$\mu \to \tau_h$ FR	20–40%	no	Gaussian
μ ID	1%	yes	Gaussian
b-jet veto	1–9%	no	Log-normal
Luminosity	2.5%	partial	Log-normal
Trigger	2% for μ, p_T-dep. for τ_h	no	Gaussian
Embedded yield	4%	no	Log-normal
$t\bar{t}$ cross section	4.2%	yes	Log-normal
Diboson cross section	5%	yes	Log-normal
Single top cross section	5%	yes	Log-normal
W+Jets cross section	4%	yes	Log-normal
Drell-Yan cross section	2%	yes	Log-normal
Signal cross sections	[39]	yes	Log-normal
top p_T reweighing	10%	yes	Gaussian
Z p_T reweighing	10%	partial	Gaussian
Prefiring (2016, 2017)	Event-dep. (0–4%)	yes	Log-normal
τ_h energy scale	1% (sim), 1.5% (emb.)	no	Gaussian
$e \to \tau_h$ energy scale	0.5–6.5%	no	Log-normal
$\mu \to \tau_h$ energy scale	1%	no	Log-normal
Muon energy scale	0.4–2.7%	yes	Gaussian
Jet energy scale	Event-dep.	partial	Gaussian
Jet energy resolution	Event-dep.	no	Gaussian
p_T^{miss} unclustered scale	Event-dep.	no	Gaussian
p_T^{miss} recoil corrections	Event-dep.	no	Gaussian
Jet$\to \tau_h$ mis-ID	Event-dep.	partial	Gaussian
$t\bar{t}$/diboson in embedded	10%	yes	Gaussian
S_{IP} in μ and π decays	25%	no	Gaussian

The systematic uncertainty is then retrieved as the missing contribution needed to obtain the total uncertainty[4]. This allowed to decompose the uncertainty as:

$$\varphi_{\tau\tau}^{obs} = (4 \pm 17(\text{stat.}) \pm 2(\text{bin-by-bin}) \pm 1(\text{syst.}) \pm 1(\text{theory}))° \,, \tag{7.5}$$

showing that the dominating uncertainty is the statistical one, meaning that the main limitation to the precision of this measurement comes from the available statistics.
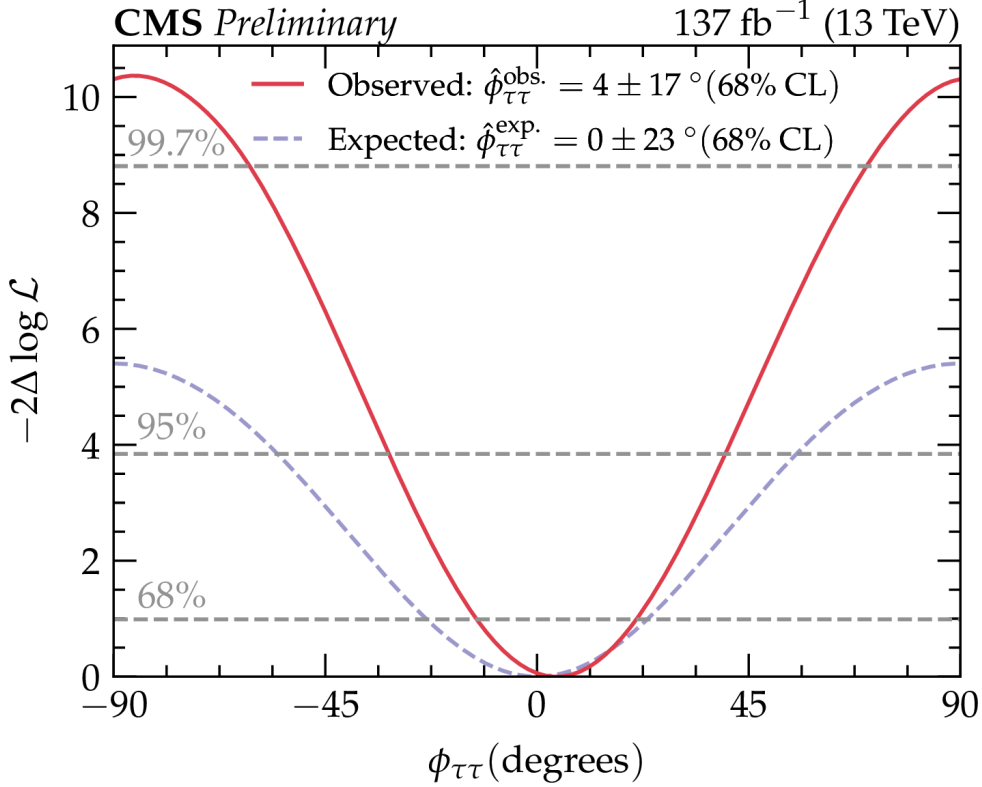
Figure 7.4: Values of the likelihood measured in data (full red line) or on Asimov dataset (dashed gray line) for different values of the $\varphi_{\tau\tau}$ angle [158].

It is evident that the observed significance is noticeably larger than the expected one. This is due to a statistical fluctuation in data which made the observed modulation in the acoplanarity angle less compatible with the CP-odd hypothesis than what could be expected from a pure CP-even hypothesis. This can be observed in Fig. 7.5, where events have been weighted in order to better visualize the compatibility with the CP-even and CP-odd hypotheses. The weight applied to each event is

$$weight = A \times \frac{S}{S + B} \,, \tag{7.6}$$

---

[4]The combination of different uncertainties is performed by taking their squared sum: $\delta_{tot} = \sqrt{\delta_{stat}^2 + \delta_{bbb}^2 + \delta_{syst}^2 + \delta_{theory}^2}$.

which depends on:

- the number of signal (S) and background (B) events in a determined bins of NN/BDT score (*NN score* for the $\tau_\mu\tau_h$ channel and *BDT score* for the $\tau_h\tau_h$ one);

- the average asymmetry (A) between CP-even and CP-odd hypothesis in that bin of NN/BDT score:

$$A_j = \frac{1}{N_{bins}^j} \sum_i^{N_{bins}^j} \frac{\left|CP_i^{even} - CP_i^{odd}\right|}{CP_i^{even} + CP_i^{odd}} \,, \tag{7.7}$$

where the sum runs on the bins in acoplanarity angle defined for a given NN/BDT bin (e.g. 10 bins for the $\tau_\mu\tau_\rho$ signal category), and $CP_i^{even(odd)}$ is the number of expected CP-even(odd) events in the i-th φ_{CP} bin of the j-th NN/BDT bin.

As can be observed in the Fig. 7.5, the data points are mostly compatible within uncertainties (which are mostly of statistical nature) with the pure scalar coupling predicted by the SM (blue histogram). They are however placed further apart from the pseudoscalar expectation (green histogram) than the scalar prediction, and are therefore less compatible with a pure CP-odd hypothesis than what was expected.

Together with the CP mixing angle, several other parameters are estimated in the negative log-likelihood minimization. The signal template normalizations are scaled as $\mu_{ggH} \times \mu^{\tau\tau}$ for the ggH production and $\mu_V \times \mu^{\tau\tau}$ for VBF and VH. Since varying the $\mu^{\tau\tau}$ parameter scales the normalization of all signal templates coherently, its effect on the templates can be replicated by altering μ_{ggH} and μ_V. In the fit to measure $\varphi_{\tau\tau}$ the parameter $\mu^{\tau\tau}$ is fixed to 1, while μ_{ggH} and μ_V are allowed to vary freely withing a fixed range.

The likelihood function can be profiled with respect to the μ_{ggH} and μ_V parameters in order to estimate the agreement of the Higgs production signal strengths with their SM expectations, i.e. $\mu_{ggH} = \mu_V = 1$. When profiling the likelihood with respect to the Higgs production signal strength $\varphi_{\tau\tau}$ is left as a freely floating parameter, while $\mu^{\tau\tau}$ is kept fixed to 1. Fig. 7.6 shows this likelihood profile, the best fit values for the two signal strengths are shown in the two-dimensional plot as a black cross of coordinates

$$\mu_{ggH} = 0.72^{+0.33}_{-0.33}$$
$$\mu_V = 1.03^{+0.55}_{-0.56} \,, \tag{7.8}$$

with the uncertainties representing the 68% CL interval. The observed values are in good agreement both with the SM expectation and with previously published measurements of the Higgs production signal strength obtained with a subset of the data collected during Run 2 [3, 174].
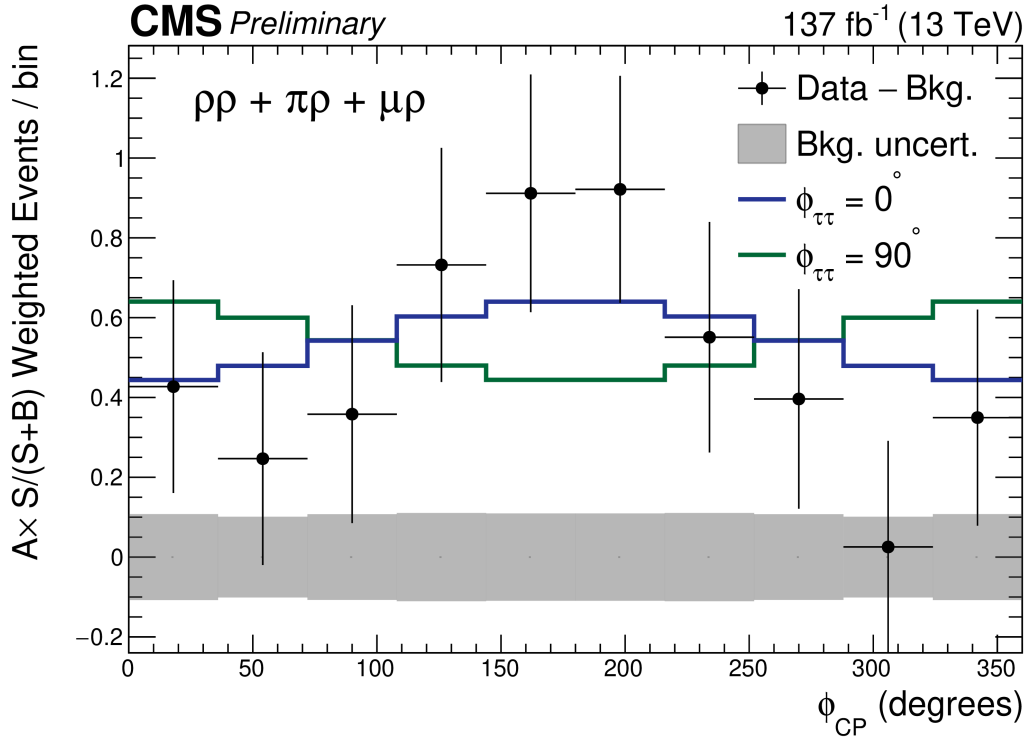
Figure 7.5: Comparison of the weighted distributions of the acoplanarity angle for observed data (black points) and the pure CP-even and CP-odd hypotheses [158]. The events are weighted after subtracting the expected background yields to enhance the separation between CP hypotheses, the weight definition is described in the text.

The signal template used in the profile likelihood fit is obtained as the sum of the different Higgs production mechanism. A decrease in the signal strength of one production mechanism can be partially compensated by increasing the signal strength of other processes in order to achieve the same total event yield. This results in the two variables being anti-correlated, as shown in Fig. 7.6.

Instead of fixing $\mu^{\tau\tau}$ to 1 it is possible to perform a fit keeping this parameter as freely floating, while constraining the μ_{ggH} and μ_V signal strengths. This was done to construct the two-dimensional likelihood profile with respect to $\varphi_{\tau\tau}$ and the $H \to \tau\tau$ branching fraction $\mu^{\tau\tau}$. Having fixed μ_{ggH} and μ_V to their SM expectations results in the $\mu^{\tau\tau}$ parameter taking the place of the inclusive Higgs production signal strength times the decay BR to tau leptons. This is shown in Fig. 7.7, where the best fit (black cross) for the CP-mixing angle,

$$\varphi_{\tau\tau} = 4 \pm 17° \, , \tag{7.9}$$

and for the inclusive Higgs production signal strength,

$$\mu^{\tau\tau} = 0.82 \pm 0.15 \, , \tag{7.10}$$
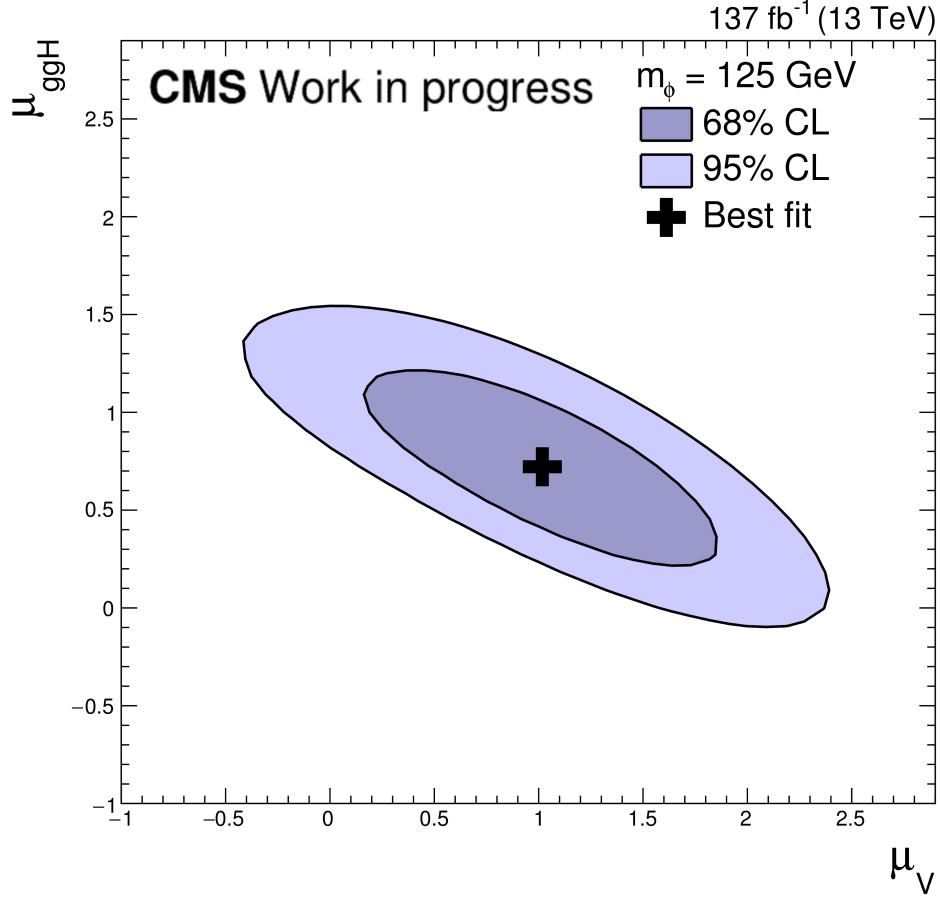
Figure 7.6: Likelihood distribution profile with respect to two variables: the production signal strength for the ggH mechanism (x axis) and the vector boson associated mechanism (y axis).

are compared to the SM expectation (red star). The best fit agrees with the SM expectation within 68% CL, as already observed for the fit done by keeping $\mu^{\tau\tau}$ fixed to 1 (Fig. 7.4). The effect of looking at the two-dimensional likelihood profile, instead of a one-dimensional one, can be noted by looking at the contour line representing the 99.7% CL (line alternating dots and dashed segments). This contour reaches the region $\varphi^{\tau\tau} = \pm 90°$, meaning that a pure CP-odd hypothesis is not excluded at 99.7% CL. The $\mu^{\tau\tau}$ parameter affects the normalization of all different signal templates, and is correlated to the amplitude of the cross section modulation with respect to φ_{CP}. This results in a relaxed limit on the compatibility with the CP-odd hypothesis when $\mu^{\tau\tau}$ is treated as a POI and allowed to freely float in the fit.

A similar procedure can be done for the two separate Yukawa couplings κ_τ and $\tilde{\kappa}_\tau$ defined in Eq. 2.14, as they can be substituted in the likelihood function via the inclusive
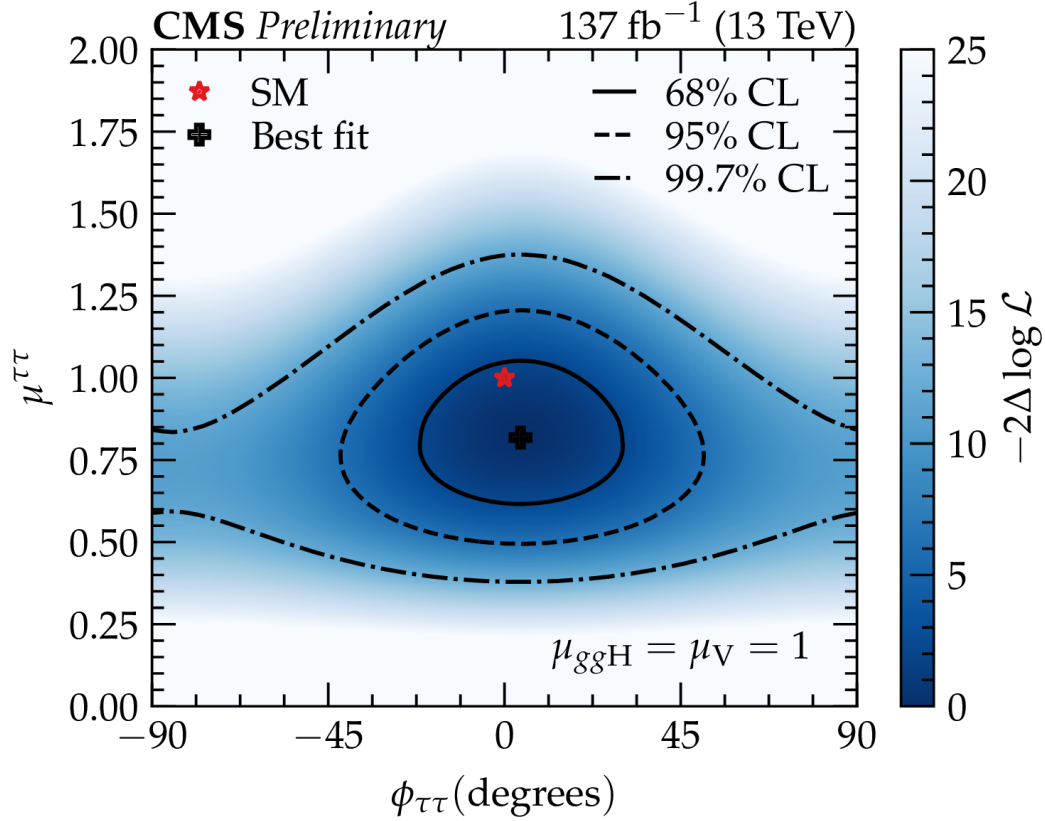
Figure 7.7: Likelihood distribution profile with respect to two variables: the CP mixing angle (x axis) and the $H \to \tau\tau$ branching fraction ($\mu^{\tau\tau}$, y axis) [158].

signal strength $\mu = \mu_{ggH/V} \times \mu^{\tau\tau}$ and $\varphi_{\tau\tau}$ parameters:

- the CP mixing angle is expressed as a function of the two couplings as

$$\varphi_{\tau\tau} = \arctan(\tilde{\kappa}_\tau/\kappa_\tau) \ ; \tag{7.11}$$

- the inclusive Higgs production signal strength is instead related to the CP-even and CP-odd couplings as

$$\mu_{ggH/V} \times \mu^{\tau\tau} = \kappa_\tau^2 + \tilde{\kappa}_\tau^2 \ . \tag{7.12}$$

The resulting two-dimensional likelihood distribution is shown in Fig. 7.8. As shown in Eq. 7.11, this analysis is sensitive only to the relative sign of the two couplings, resulting in the distribution in Fig. 7.8 being invariant under symmetry with respect to the origin of the axes.

The best fit value with respect to the couplings can be calculated by inverting Eq.
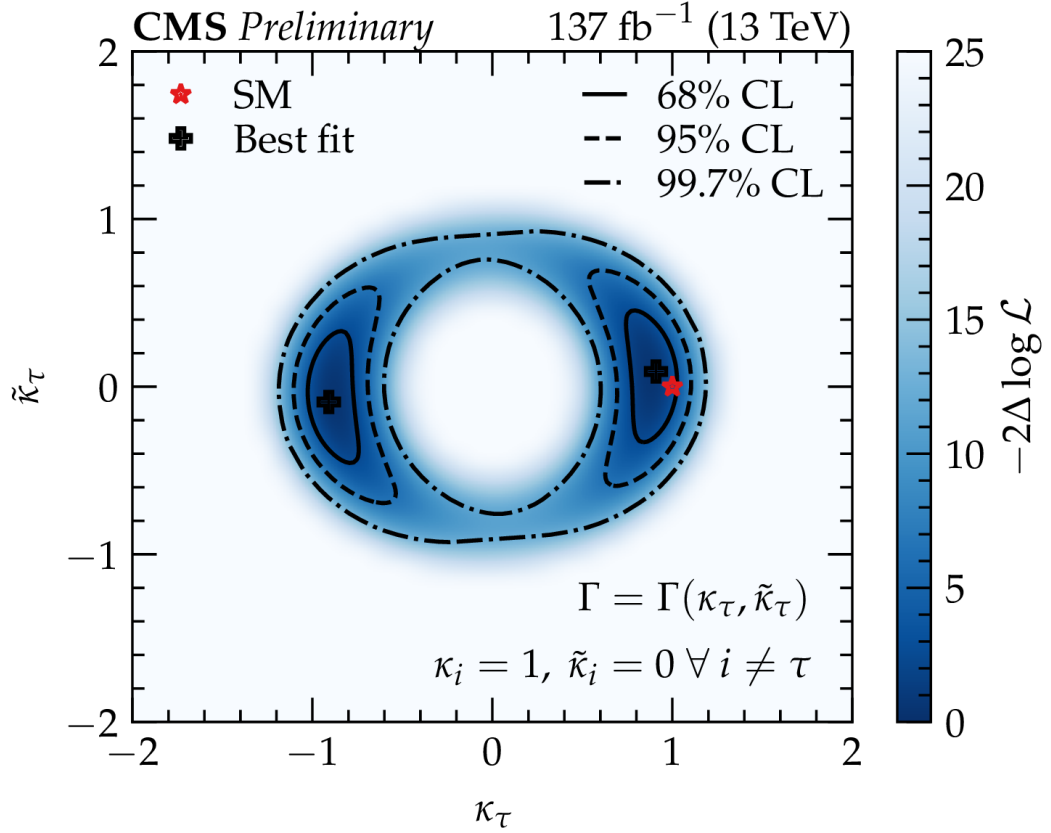
Figure 7.8: Likelihood distribution profile with respect to the separate Yukawa couplings: the CP-even κ_τ (x axis) and the CP-odd $\tilde{\kappa}_\tau$ (y axis) [158].

7.11 and 7.12:

$$\kappa_\tau = \sqrt{\mu^{\tau\tau}} \cos(\varphi_{\tau\tau}) \simeq 0.90 \ ,$$

$$\tilde{\kappa}_\tau = \sqrt{\mu^{\tau\tau}} \sin(\varphi_{\tau\tau}) \simeq 0.06 \ . \tag{7.13}$$

As shown in the figure, the 95% CL contour line extends up to the main diagonal of the plot, corresponding to the region $\tilde{\kappa}_\tau/\kappa_\tau \sim 1$. This can be interpreted as a 95% CL exclusion for models which predict CP-even and CP-odd couplings of similar magnitude.

7.3 Interpretation of the results

The study presented in Chapters 6 and 7 allows to set limits on the amount of CP mixing in the Yukawa couplings between the Higgs boson and tau leptons.

The possibility of a non pure CP-even coupling can be interpreted as a violation of the CP symmetry in the Higgs couplings. Several beyond standard model (BSM) theories [186,

187] consider how a CP-violation occurring as part of the electroweak symmetry breaking (EWSB) could be used to justify the baryon asymmetry in the Universe (BAU). As shown in [188], several experiments have presented solid evidence that the known Universe is composed mainly of matter particles. The fraction of antimatter in the Universe is found to be of the order of 10^{-9} [188], and is currently not explained within the SM. In 1967, Sakharov proposed [189] three conditions which should be met in order to justify the BAU:

1. baryon number violation,

2. C and CP violation,

3. a deviation from thermal equilibrium.

Additional sources of CP violation, like in the Yukawa interaction, can be investigated as possible contributors to the observed BAU. The possible interpretation of a CP-odd component in the Yukawa coupling between the Higgs boson and tau leptons as a source for BAU is presented in [186, 187]. In the scope of the (next to-) minimal super-symmetric extension to the standard model, (N)MSSM, additional Higgs fields are introduced. One NMSSM theory predicts a Higgs field originated by 2 doublets and one singlet. After EWSB a total of 7 physical Higgs bosons are introduced, 5 of which of neutral charge and able to couple to the tau leptons via Yukawa interaction. These interactions can be written in the form

$$\mathscr{L}_{Y,\tau}^i = -K_i \tilde{\tau}(\cos(\phi_i) + i\gamma^5 \sin(\phi_i))H_i\tau , \qquad (7.14)$$

with $i \in \{1, ..., 5\}$ representing the index associated to the 5 neutral Higgs bosons, and K_i and ϕ_i the coupling constant and the CP mixing angle for these Yukawa interactions.

Fig. 7.9 shows the available phase space of this model assuming that the Higgs boson labeled H_2 corresponds to the 125 GeV resonance identified at the LHC by the CMS and ATLAS experiments [1, 2]. The figure shows a scatter plot representing the theory predictions in this model of the CP mixing angle ϕ_i expressed in [rad] units, with respect to the mass m_{H_i} of the associated Higgs boson. As the measurement presented in this thesis uses simulated signal samples for a Higgs boson of mass 125 GeV, its results can be interpreted as limits on the available phase space for the H_i bosons having mass close to 125 GeV. The limit placed by this analysis is $\phi_i < 40°$ at 95% CL, i.e. $\phi_i \lesssim 0.22\pi$ [rad], and can be used to exclude models above this value. At 68% CL the analysis can already be used to probe the phase space $\phi_i \lesssim 0.12\pi$ [rad] instead.

In the MSSM theory, a model with 2 Higgs doublets is introduced, originating 5 physical Higgs bosons of which 3 are of neutral electric charge. Under the assumption that one of these bosons is the one identified at 125 GeV, it is possible to interpret its CP mixing angle as a source of BAU as shown in [187]. On the left in Fig. 7.10 the expected limits at 95% CL on the CP-even and CP-odd couplings are shown for several
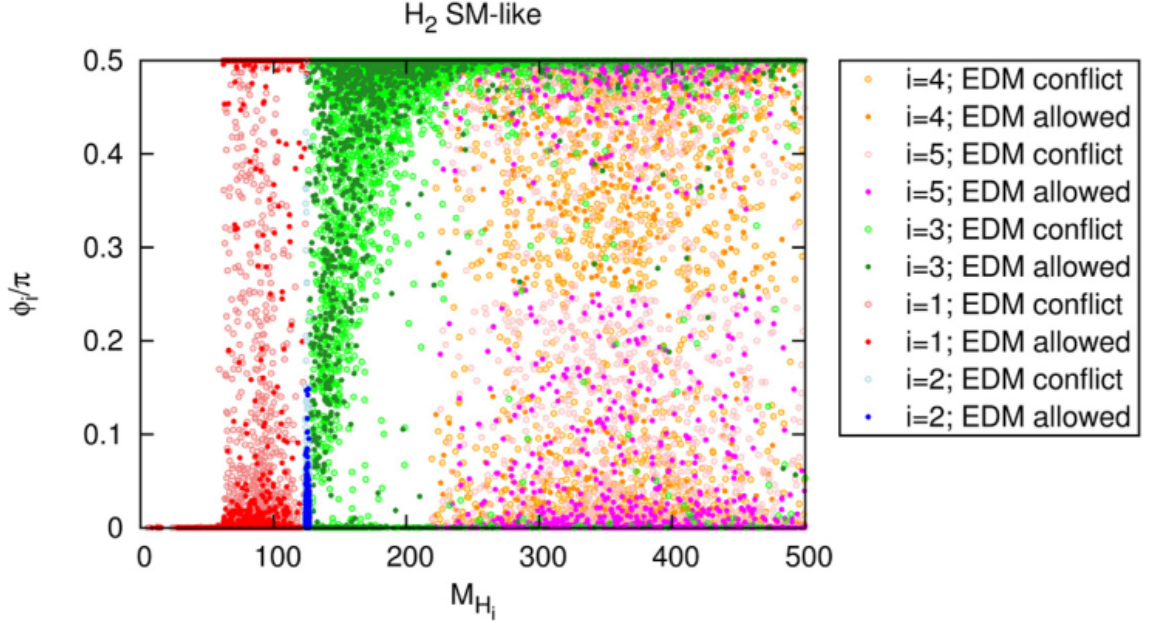
Figure 7.9: CP mixing angles ϕ_i estimating the CP violation in $H_i \rightarrow \tau\tau$ couplings as a function of the Higgs boson H_i mass [186]. Each empty or full circle corresponds to a theory prediction, and their density represents the available phase space for the theory models. The empty colored circles are in conflict with observed electro-dipole moments (EDM), while the full colored circles are allowed based on EDM bounds.

experiments. The notation used in [187] for the Yukawa couplings can be compared to the one used in this thesis with the following transformation:

$$\kappa_\tau = k_\tau \cos(\Delta) \ ,$$
$$\tilde{\kappa}_\tau = k_\tau \sin(\Delta) \ ,$$

(7.15)

with $\Delta = \varphi_{\tau\tau}$ representing the CP mixing angle. The lines labeled $\mu_{\tau\tau}$ show the limits on the coupling strength (defined in this chapter as $\mu^{\tau\tau}$ by fixing $\mu_{ggH} = \mu_V = 1$) while the other ones are associated to two-dimensional limits calculated using k_τ and Δ as parameters. Limits on the coupling strength are shown for the measurements performed by the ATLAS [190] and CMS [159] experiments, together with their expected limits estimated on a collected luminosity of 3000 fb^{-1} at the HL-LHC [191, 192]. Limits on the separate CP-even and CP-odd couplings are instead estimated in [193] for HL-LHC, and in [187] for the following proposed electron-positron colliders experiments: the Circular Electron-Positron Collider (CEPC) [194], the Future Circular Collider (FCC)-ee [195], and the International Linear Collider (ILC) [196].

These limits can be compared to the ones provided for the Yukawa CP-even and CP-odd couplings κ_τ and $\tilde{\kappa}_\tau$ calculated in [158] and previously presented in Fig. 7.8. They

are also shown on the right side of Fig. 7.10, for convenience of the reader. The couplings associated to the best-fit of this study would appear in the plot on the left as a point of coordinates (0.90, 0.06). The 68% CL contour would instead cover the region with $0.7 \lesssim k_\tau \cos(\Delta) \lesssim 1$ and $-0.3 \lesssim k_\tau \sin(\Delta) \lesssim 0.5$, already probing part of the phase space available to justify the BAU. The 95% CL contour corresponds to a region with $0.5 \lesssim k_\tau \cos(\Delta) \lesssim 1.1$ and $-0.6 \lesssim k_\tau \sin(\Delta) \lesssim 0.7$, covering a good fraction of the phase space available based on the inclusive coupling analyses, and the region at lower values of $k_\tau \cos(\Delta)$. Repeating this analysis with higher recorded statistics would allow to further constrain the phase space which allow the BAU generation.
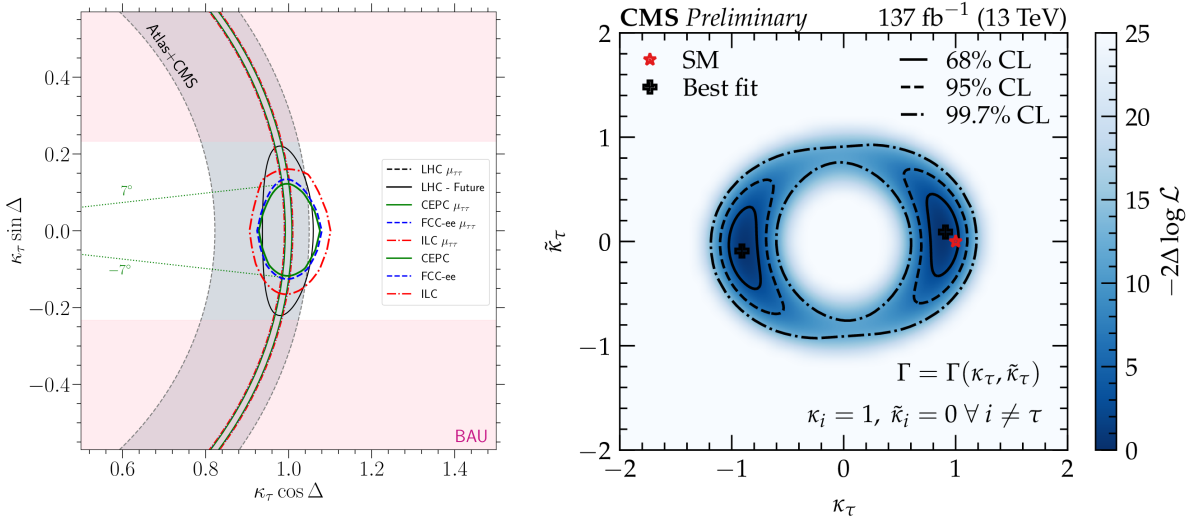
Figure 7.10: Limits on the CP-even and CP-odd couplings Yukawa couplings between the Higgs boson and tau leptons. **Left:** 95% CL limits are shown on the right [187] for inclusive coupling analyses using CMS and ATLAS data [159, 190], and projections at the HL-LHC [191, 192] and future electron-positron colliders [187]. **Right:** limits at 68, 95 and 99.7% CL obtained from the analysis [158] described in Chapters 6 and 7.

7.4   Summary and prospects

The analysis presented in Chapters 6 and 7 shows that the Yukawa coupling of the Higgs boson to tau leptons is consistent with the SM prediction within 68% CL, and is therefore mostly CP-even. It is also possible to constrain at 95% CL the contribution of a CP-odd coupling to be:

$$\left| \frac{\tilde{\kappa}_\tau}{\kappa_\tau} \right| \lesssim 1 \ . \tag{7.16}$$

Finally, the Higgs production signal strengths also appear to be consistent with the SM prediction within 68% CL and with the results obtained by dedicated analyses targeting

the Higgs production mechanisms [3, 174].

In summary, an analysis was performed to investigate the properties of the Yukawa coupling between the Higgs boson and tau leptons under CP symmetry. The analysis targeted the final states with one tau decaying hadronically and the other one either to a muon ($\tau_\mu\tau_h$ channel) or hadrons ($\tau_h\tau_h$ channel) using data recorded by the CMS experiment during the Run 2 data-taking period at the LHC. The observed Higgs production signal strengths are found to be in good agreement with their SM expectation within 68% CL. As previously stated, the observed CP mixing angle is:

$$\varphi_{\tau\tau}^{obs} = (4 \pm 17(\text{stat.}) \pm 2(\text{bin-by-bin}) \pm 1(\text{syst.}) \pm 1(\text{theory}))^\circ \ , \qquad (7.17)$$

with the uncertainties split in various contributions and all expressed as a 68% CL interval. The precision of the measurement is currently limited mainly by the statistical uncertainty, which is expected to scale as the square root of the collected luminosity. As the expected significance of the measurement is 2.3σ, with an integrated luminosity of 137 fb^{-1}, a $\sim 3.5\sigma$ expected significance for the rejection of the CP-odd hypothesis could realistically be achieved with an integrated luminosity of ~ 300 fb^{-1}, assuming the same analysis techniques are used. Looking at the other contributions to the uncertainty, the second largest one is the bin-by-bin uncertainty related to the background and signal models. The estimation of the dominant background processes via data-driven methods and the smoothing of background and signal templates is proven to be efficient in constraining this contribution to the uncertainty. Finally, the systematic and theoretical uncertainties provide only a small contribution to the overall uncertainty.

The analysis has found all POIs to be in agreement with the SM expectations within 68% CL, and, consequently, no noticeable deviation from the SM predictions. The Yukawa coupling of the Higgs boson to tau lepton appears to be mostly of CP-even nature, with a pure CP-odd coupling being excluded at 99.7% CL. Models beyond the SM that predict a mixing of CP-even and CP-odd couplings of similar amplitude can also be excluded at 95% CL.

Conclusions

The discovery of the Higgs boson, announced on July the 4th 2012 by the CMS and ATLAS collaborations [1, 2], opened a new path for research at particle colliders: the investigation of the Higgs sector. The discovery was made in channels that depend on the Higgs couplings to vector bosons, while later studies investigated the couplings to other particles in the Standard Model (SM). Direct couplings to third generation fermions have been observed with the discovery of the top associated production [38] and the decays to tau leptons [3] and b-quarks [4]. More recently, evidence for Higgs decays to a pair of muons has also been observed by the CMS experiment [5]. All studies performed so far found a good agreement between the measured couplings and the ones predicted in the SM. The coupling strengths were found to be proportional to the mass of the fermion or to the square of the vector boson mass.

The spin and CP nature of the Higgs boson was also investigated. All studies performed on the Higgs couplings to vector bosons showed no significant deviations from the SM expectation of a pure scalar Higgs boson. The properties of the Higgs couplings to top quarks under CP symmetry have also been investigated, by studying the Higgs production mechanism via top associated production.

This thesis presents the first study performed on the structure under CP symmetry of the Higgs Yukawa coupling to τ leptons. The Higgs boson coupling to τ leptons was assumed to be an admixture of CP-even and CP-odd couplings, with the ratio between the two couplings been parametrized as the CP mixing angle $\varphi_{\tau\tau}$. Higgs decays to τ leptons were studied in order to measure this angle. My contribution consisted in measuring $\varphi_{\tau\tau}$ in the $\tau_\mu\tau_h$ final state, where a Higgs boson decays into two τ leptons, one of which then decays hadronically and the other one muonically. The angle between the τ decay planes was reconstructed using the momenta and impact parameters of the τ decay products, and its distribution was used to measure $\varphi_{\tau\tau}$.

The measured [expected] CP mixing angle is:

$$\varphi_{\tau\tau} = (-8 \pm 37)^\circ \ \ [(0 \pm 40)^\circ] \, ,$$

corresponding to an observed (expected) significance for the separation between CP-even and CP-odd hypotheses of 1.56 (1.47)σ.

This study was combined with one performed in the fully hadronic final state ($\tau_h\tau_h$), achieving the following estimation of the CP admixture:

$$\varphi_{\tau\tau} = (4 \pm 17)^\circ \ \ [(0 \pm 23)^\circ] \ ,$$

corresponding to an observed (expected) significance for the exclusion of a pure CP-odd hypothesis of

$$0^+ \ vs \ 0^- = 3.2\sigma \ \ (2.3\sigma) \ .$$

The results are in good agreement with the SM prediction of a pure CP-even couplings, and allow to exclude at 95% confidence level models which predict CP-even and CP-odd Yukawa couplings of comparable magnitude.

This result was released to the public as a physics analysis summary of the CMS collaboration in July 2020 [158] and presented at the "International Conference on High Energy Physics" soon after [185]. It required the use of several experimental techniques in order to estimate background processes, improve the reconstruction of various properties of tau leptons and of its decay products.

It is also among of the first studies performed in CMS using the DeepTau neural network-based identification algorithm [150, 151] for hadronically decaying τ leptons (τ_h). This algorithm was introduced during the Run 2 data-taking period in order to identify genuine τ_hs with respect to objects mimicking their signatures, like quark and gluon jets, muons and electrons. I contributed by measuring the misidentification rate of electrons as hadronically decaying tau leptons using this identification algorithm. The study will be part of a publication from the CMS collaboration, which describes the DeepTau identification algorithm, its efficiency for the selection of genuine τ_h and its rejection of misidentified objects.

Part of the thesis project also involved working on the alignment of the CMS tracker, on tools to study the time dependence of the alignment. This allowed to improve the estimation of detector aging effects. The work done was not included in this document but is part of a paper to be published soon.

The Run 2 data-taking period for the experiments at the LHC provided a large statistics of proton-proton collisions. The main study presented in this thesis showed that the Higgs coupling to tau leptons is dominated by a CP-even component, and that CP-odd components of magnitude similar to the CP-even are excluded at 95% confidence level. The measurement precision is currently limited by the amount of recorded data, and will benefit from the expected increase in recorded luminosity from the Run 3 data-taking period. The total recorded luminosity at the highest center-of-mass energy should, by the end of Run 3, reach 300 fb^{-1}, bringing an increase in sensitivity of $\sim \sqrt{2}$.

Furthermore, under the assumption that the structure under CP-symmetry of the Higgs coupling to fermions is the same for all fermions, this study could be combined

with the one performed in the top-associated production [38, 65], further increasing the understanding of the Higgs boson properties under CP-symmetry.

In 2027 the LHC will enter the High Luminosity-LHC [103] phase and start recording data at a much higher instantaneous luminosity. In the subsequent years it is expected to collect a total integrated luminosity of up to 4000 fb^{-1} [103], increasing the available statistics by more than a factor 10. This will allow measuring Higgs couplings with a much higher precision, potentially reaching an uncertainty of 2-5% for couplings to third generation fermions [198]. Processes for which evidence was already found by the end of Run 2, like the $H \rightarrow \mu\mu$ and $H \rightarrow Z\gamma$ decays [5, 199] will reach a 5σ significance in this phase, while evidence for di-Higgs production could potentially be found [198]. The challenge of reconstructing events in a high pile-up environment will be contrasted by the planned detector upgrades [107]. The introduction of dedicated timing detectors will improve the event reconstruction and improve the pile-up mitigation. This, coupled with improvements in object reconstruction and identification, will allow more precise studies of already observed processes and further the search for hints of new physics.

The Standard Model remains to date a theory with a wide range of correctly verified predictions. While no signs of new physics have been found, the advancements in methods and the increase in recorded data will allow for more precise measurements of Higgs couplings. More precise measurements of CP properties would be particularly interesting, as any deviation found from theory would open the doors to new physics.

Anomalous couplings in HVV interactions

Section 2.1 provided a general description of how the couplings between the Higgs boson and the vector bosons (HVV couplings) are studied. The focus was mostly kept on the study of anomalous CP-odd couplings in order to establish a comparison with the CP-mixing studied in the Yukawa interactions between the Higgs boson and fermions. This Appendix is dedicated to provide further a more general description of the study of the anomalous HVV couplings, and show the most recent limits obtained by the CMS experiment.

As previously shown in Section 2.1, the scattering amplitude describing the interaction between a spin-zero H boson and two spin-one gauge bosons VV can be written as [60]:

$$
\begin{aligned}
\mathscr{A}(HVV) \simeq & \left[a_1^{VV} + \frac{k_1^{VV} q_1^2 + k_2^{VV} q_2^2}{(\Lambda_1^{VV})^2} + \frac{k_3^{VV}(q_1 + q_2)^2}{(\Lambda_Q^{VV})^2} \right] m_V^2 \varepsilon_{V1}^* \varepsilon_{V2}^* \\
& + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \bar{f}^{*(2)\mu\nu} \,,
\end{aligned}
\tag{A.1}
$$

where $f^{(i)\mu\nu} = \varepsilon_{Vi}^\mu q_{Vi}^\nu - \varepsilon_{Vi}^\nu q_{Vi}^\mu$ is the field strength tensor of a gauge boson with momentum q_{Vi} and polarization ε_{Vi}, $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma} f^{(i)\rho\sigma}$ is the dual field strength tensor with m_{Vi} pole mass of a gauge boson, Λ_1^{VV} and Λ_Q^{VV} are the scales of BSM physics.

This general expression can be used to describe both the $H \to WW, ZZ$ couplings previously discussed in section 1.3 and the effective couplings $H \to gg, Z\gamma, \gamma\gamma$. When considering couplings only to electro-weak bosons, this approach is equivalent to writing

the following effective Lagrangian:

$$\mathcal{L}_{HVV}^{eff} \simeq a_1^{ZZ} \frac{m_Z^2}{2} H Z_\mu Z^\mu - \frac{k_1^{ZZ}}{(\Lambda_1^{ZZ})^2} m_z^2 H Z_\mu \Box Z^\mu - \frac{1}{2} a_2^{ZZ} H Z_{\mu\nu} Z^{\mu\nu}$$

$$- \frac{1}{2} a_3^{ZZ} H Z_{\mu\nu} \tilde{Z}^{\mu\nu} + a_1^{WW} m_W^2 H W_\mu^+ W^{-\mu}$$

$$- \frac{m_W^2}{(\Lambda_1^{WW})^2} \times \left(k_1^{WW} W_\mu^- \Box W^{+\mu} + k_2^{WW} W_\mu^+ \Box W^{-\mu} \right)$$

$$- \frac{1}{2} a_2^{WW} H W_{\mu\nu}^+ W^{-\mu\nu} - \frac{1}{2} a_3^{WW} H W_{\mu\nu}^+ \tilde{W}^{-\mu\nu} \qquad (A.2)$$

$$+ \frac{k_2^{Z\gamma}}{\Lambda_1^{Z\gamma}} m_Z^2 H Z_\mu \partial_\nu F^{\mu\nu} - a_2^{Z\gamma} H F^{\mu\nu} Z_{\mu\nu} - a_3^{Z\gamma} H F^{\mu\nu} \tilde{Z}_{\mu\nu}$$

$$- \frac{1}{2} a_2^{\gamma\gamma} H F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} a_3^{\gamma\gamma} H F_{\mu\nu} \tilde{F}^{\mu\nu} \,,$$

where H is the real Higgs field, Z_μ the Z boson field, $W_\mu^\pm$ the W$^\pm$ field, F_μ the γ^* field, $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$ the bosonic field strength and $\tilde{V}_{\mu\nu} = \frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} V^{\rho\sigma}$ the dual field strength and $\Box$ is the d'Alambert operator. This simplified effective Lagrangian takes already into account gauge invariance under the symmetries of the SM, in particular this leads to vanishing tree-level couplings between the Higgs and $Z\gamma$ or $\gamma\gamma$. The SM tree-level coupling shown in equation 2.1 is obtained by setting all coupling coefficients aside from a_1^{ZZ} and a_1^{WW} equal to 0. The contributions related to the a_2^{VV} coupling can arise both in the SM due to loop corrections to the tree-level coupling or from BSM contributions, as such they are treated as anomalous couplings. a_3^{VV} represents a CP-odd coupling and its experimental constrain is used to probe the CP nature of the Higgs boson.

These couplings have been investigated by the ATLAS and CMS experiments already during Run 1 [58, 61] in search for a CP-violating coupling. As an example, the search for CP violation in $H \to ZZ \to 4$ leptons[1] allowed to strongly constrain the relative contribution of the anomalous couplings. This is shown in Fig. A.1, in the form of the agreement of a test statistic based on the pure CP-even hypothesis with the one where the Higgs boson is hypothesized to have spin-parity J^P. The test statistic is defined based on the treatment described in [62], as the negative logarithmic ratio of the likelihood functions for the two pure CP hypothesis. These likelihood functions have been constructed respectively by setting all couplings but a_1^{ZZ} equal to 0 for the CP-even hypothesis, while for the other cases allo couplings but the one associated to a Higgs boson of spin-parity J^P have been placed to 0. The observed value of the test statistic

---

[1]Leptons is used here to indicate charged light leptons, namely muons and electrons.
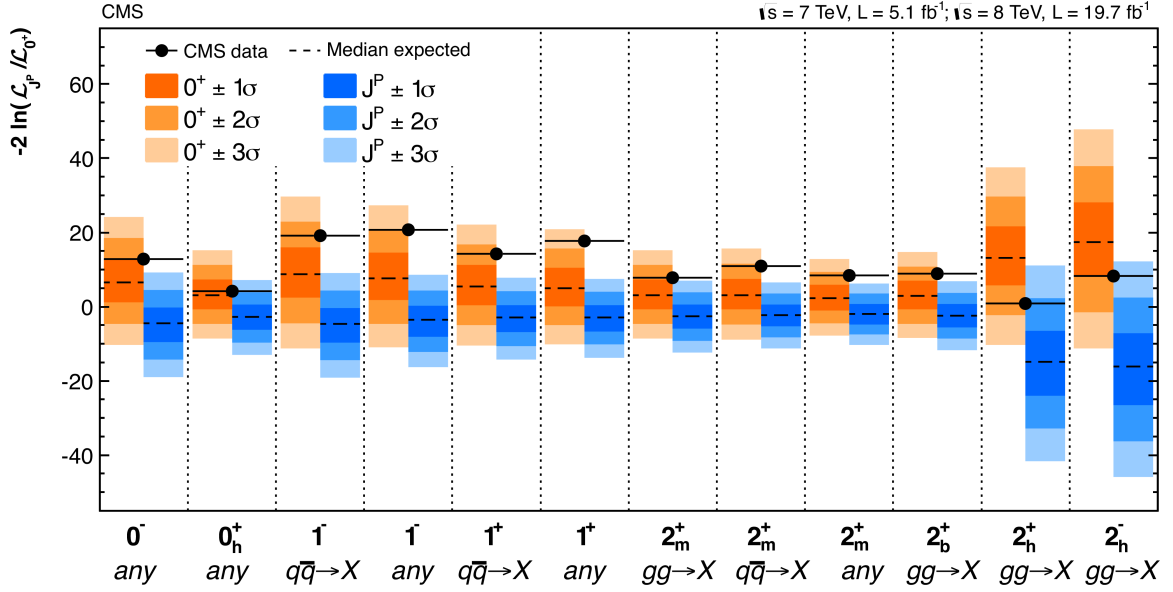
Figure A.1: Distribution of the test statistic $q = -2\ln\left(\mathcal{L}^{J^P}/\mathcal{L}^{0+}\right)$ of the anomalous HVV boson hypotheses (blue) tested against the SM Higgs boson hypothesis (red). The black points indicate the observed value obtained with Run 1 data from the CMS experiment [58].

(black points) is compared to distributions in samples of *pseudo experiments* obtained for the CP-even and anomalous spin-parity hypotheses with toy Monte Carlo simulations. Data shows a clear preference for the CP-even hypothesis. A pure CP-odd hypothesis ($J^P = 0^-$) is ruled out at 99% confidence level (CL). A similar level of exclusion is found for some hypotheses involving spin 2 Higgs bosons, while stronger exclusion limits are found for Higgs bosons of spin 1. The definition of couplings for Higgs boson of spin different from 0 can be found in [58].

As shown in Section 2.1, the effect of the anomalous coupling on the process cross-section is parametrized via the *effective fractional cross section*, defined as:

$$f_X^{VV} = \frac{\left|a_X^{VV}\right|^2 \sigma_X}{\left|a_1^{VV}\right|^2 \sigma_1 + \left|a_2^{VV}\right|^2 \sigma_2 + \left|a_3^{VV}\right|^2 \sigma_3 + \tilde{\sigma}_{\Lambda_1}/(\Lambda_1)^4 + \ldots} \ , \tag{A.3}$$

with σ_i representing the cross section of the $H \rightarrow VV$ process computed assuming $a_j^{VV} = 0 \ \forall j \neq i$, $\tilde{\sigma}_{\Lambda_1}$ is the effective cross section corresponding to $\Lambda_1 = 1$ TeV, and the denominator includes contribution of all anomalous couplings.

Subsequent analyses followed a more general approach, extending the measurement to other anomalous couplings and accounting for the interference between different couplings at the level of cross section calculation. Such mixing is represented as the argument

function for the ratio between higher order couplings and the SM one:

$$\phi_{a_i} = \arg\left(\frac{a_i}{a_1}\right) \ . \tag{A.4}$$

The latest CMS studies investigated the HVV couplings in the 4 lepton final state [63] and in the vector boson fusion production of a Higgs decaying into a pair of tau leptons [64]. The main features common to these analyses, which distinguish them from the study of the CP properties in the Yukawa coupling, is the use of an effective field theory approach: the anomalous couplings are introduced in the Lagrangian as additional terms, without focusing on the physical processes which could produce these couplings. Limits on the anomalous couplings are then estimated based on their effect on the overall cross section, represented by the effective fractional cross sections, and the kinematics of the Higgs boson and its accompanying jets. More precisely, the kinematics of the final state are used mainly in the study of a 4 lepton final state, which uses a multivariate discriminator based on the momenta, distance and direction of flight of all four reconstructed particles in order to distinguish between the different hypotheses regarding the presence of anomalous couplings. The kinematics of the initial state are instead used to study the Higgs production via vector boson fusion, and are also used to construct a multivariate discriminator. Other properties of the events, like the kinematics of the tau pair final state studied in [64], are not used for the estimation of the anomalous HVV couplings contribution, but only to identify events Higgs production and decay processes with respect to background processes which mimic their signatures in a detector. The combination of these analyses allowed to place strong constraints on the effective fractions and phases associated to anomalous HVV couplings. Fig. A.2 shows the limits placed on parameters of the form $f_{a_i}\cos(\phi_{a_i})$, with a value of 0 representing the SM prediction, i.e. with all anomalous couplings equal to 0. The values +1 and -1 represent instead models in which the SM coupling is negligible with respect to the anomalous ones, and as shown in the figures are all excluded with more than 99.7% CL.
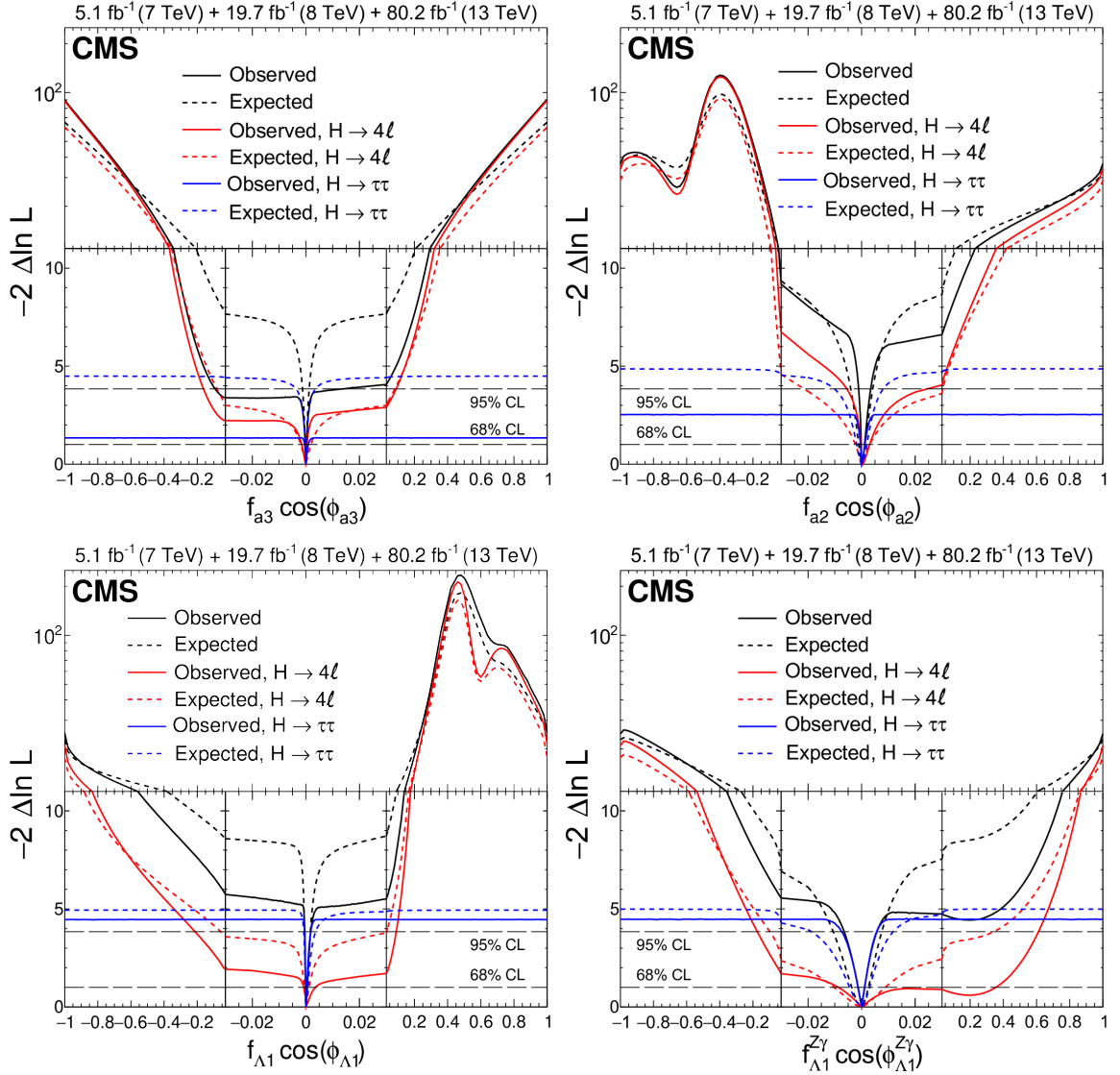
Figure A.2: Combination of limits on the anomalous HVV couplings obtained by studying the $H \to \tau\tau$ process [64] and $H \to 4$ leptons one [63]. The observed (solid) and expected (dashed) likelihood scans of $f_{a_3} \cos(\phi_{a_3})$ (top left), $f_{a_2} \cos(\phi_{a_2})$ (top right), $f_{\Lambda_1} \cos(\phi_{\Lambda_1})$ (bottom left), and $f_{\Lambda_1}^{Z\gamma} \cos\left(\phi_{\Lambda_1}^{Z\gamma}\right)$ (bottom right) are shown. The scales of the x and y axes are varied: the x axis is always in linear scale, but a zoom is applied in the range between -0.03 and +0.03, while y is shown linear scale in the lower part of the plot and in logarithmic scale for higher values of $-2\Delta \ln L$ [64].

Validation of the acoplanarity angle reconstruction

Section 2.2.3 provided a description of the methods used to reconstruct the acoplanarity angle on an event-by-event basis. The methods are formulated in order to be applied on the process of interest, i.e. the $H \to \tau\tau$, and other physical processes which produce similar signatures in the particle detector. For the purpose of studying the CP properties of the Yukawa coupling between the Higgs boson and tau leptons, the methods have to yield similar results in the Monte Carlo simulations for the various processes and in data recorded by the experiment. This validation cannot be performed in a phase space region enriched in signal-like events, as it requires to know the CP mixing angle, which is the physical parameter targeted by the measurement. To verify that the acoplanarity angle, in simulation, correctly models what observed in data a phase space region depleted in signal-like events must be used, where a modulation in φ_{CP} is also present. The Drell-Yan process, as discussed in Section 2.2.2, has a flat distribution with respect to φ_{CP}. In Eq. 2.31, reported here for convenience of the reader:

$$
\mathrm{d}\sigma_{DY}/\mathrm{d}\cos\left(\theta^+\right)\mathrm{d}\cos\left(\theta^-\right)\mathrm{d}\cos\left(\phi^+\right)\mathrm{d}\varphi_{CP}\mathrm{d}E^+\mathrm{d}E^- \propto \sum_{B_1,B_2=Z,\gamma} a(B_1,B_2)
$$

$$
\times \Bigg\{ V_\tau^{B_1} V_\tau^{B_2} \left[1 - \left(\cos\left(\theta^+\right)\cos\left(\theta^-\right) + \frac{1}{2}\sin\left(\theta^+\right)\sin\left(\theta^-\right)\cos\left(2\phi^+ - \varphi_{CP}\right) \right) \right]
$$

$$
+ A_\tau^{B_1} A_\tau^{B_2} \left[1 - \left(\cos\left(\theta^+\right)\cos\left(\theta^-\right) - \frac{1}{2}\sin\left(\theta^+\right)\sin\left(\theta^-\right)\cos\left(2\phi^+ - \varphi_{CP}\right) \right) \right] \tag{B.1}
$$

$$
+ (a^{B_1} V_\tau^{B_2} + V_\tau^{B_1} A_\tau^{B_2}) \left(\cos\left(\theta^+\right) - \cos\left(\theta^-\right) \right) \Bigg\},
$$

a modulation of the DY cross section on φ_{CP} is present, but is averaged by integrating over ϕ^+. If the cross section is calculated in different regions of ϕ^+ a modulation would be observed for the DY process, and can be used to validate the modelling of φ_{CP} in background simulations. A procedure was proposed in [70] to define different regions in ϕ^+ in order observe a non flat acoplanarity angle distribution. The method targets the **one prong** channels, but can be generalized to other decay channels using the notation

from Section 2.2.3. This allows to define an angle α_-:

$$\cos(\alpha_-) = \left| \frac{\hat{z} \times \vec{P}_-}{\left| \hat{z} \times \vec{P}_- \right|} \cdot \frac{\vec{R}_- \times \vec{P}_-}{\left| \vec{R}_- \times \vec{P}_- \right|} \right|, \tag{B.2}$$

with $\hat{z}$ unitary vector of the z axis, which for CMS is defined as the beam direction, and $\vec{P}_-$ and $\vec{R}_-$ are the P and R (see Table 2.1) vectors of the negative charged tau. By dividing the phase space into two regions based on $\alpha_- > \pi/4$ and $\alpha_- < \pi/4$ it is possible to observe a modulation in the DY cross section with respect to φ_{CP}. This was done as part of the measurement performed in the CMS experiment to measure the CP mixing angle.
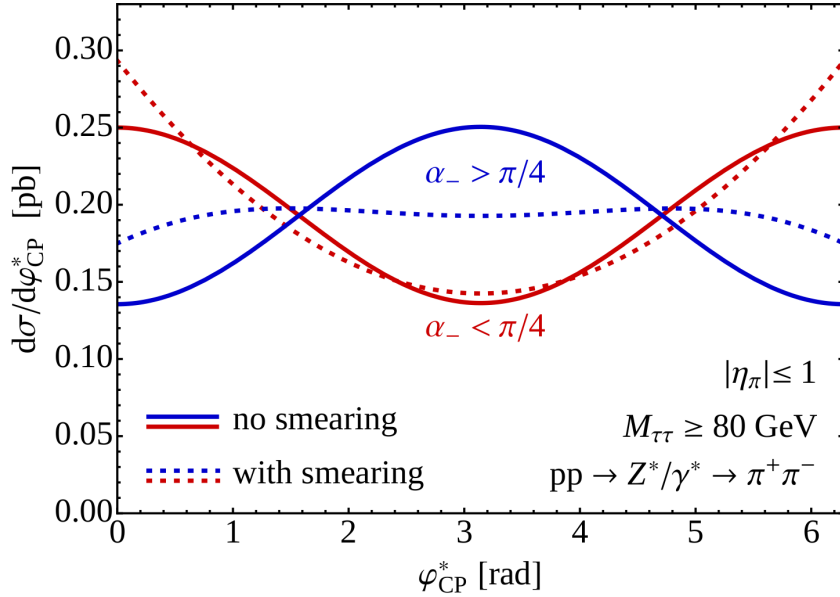
Figure B.1: Modulation of the DY cross section with respect to the acoplanarity angle, based on a categorization in α_- [70]. The plot was made explicitly regarding one prong decays, with the dashed lines representing the φ_{CP} modlulation for DY taking into account detector resolution effects, and thus the reconstruction level distribution in the $\tau_\mu \tau_\pi$ channel. The other decay channels present distributions more similar to the full lines, as the resolution on the PV has less impact in other decay channels.

As shown in Fig. B.2 and B.3, the agreement between data and simulation for the acoplanarity angle distribution split in α_- categories is good within uncertainties. This indicates that indeed the acoplanarity angle is well modelled in MC simulation for all decay channels.
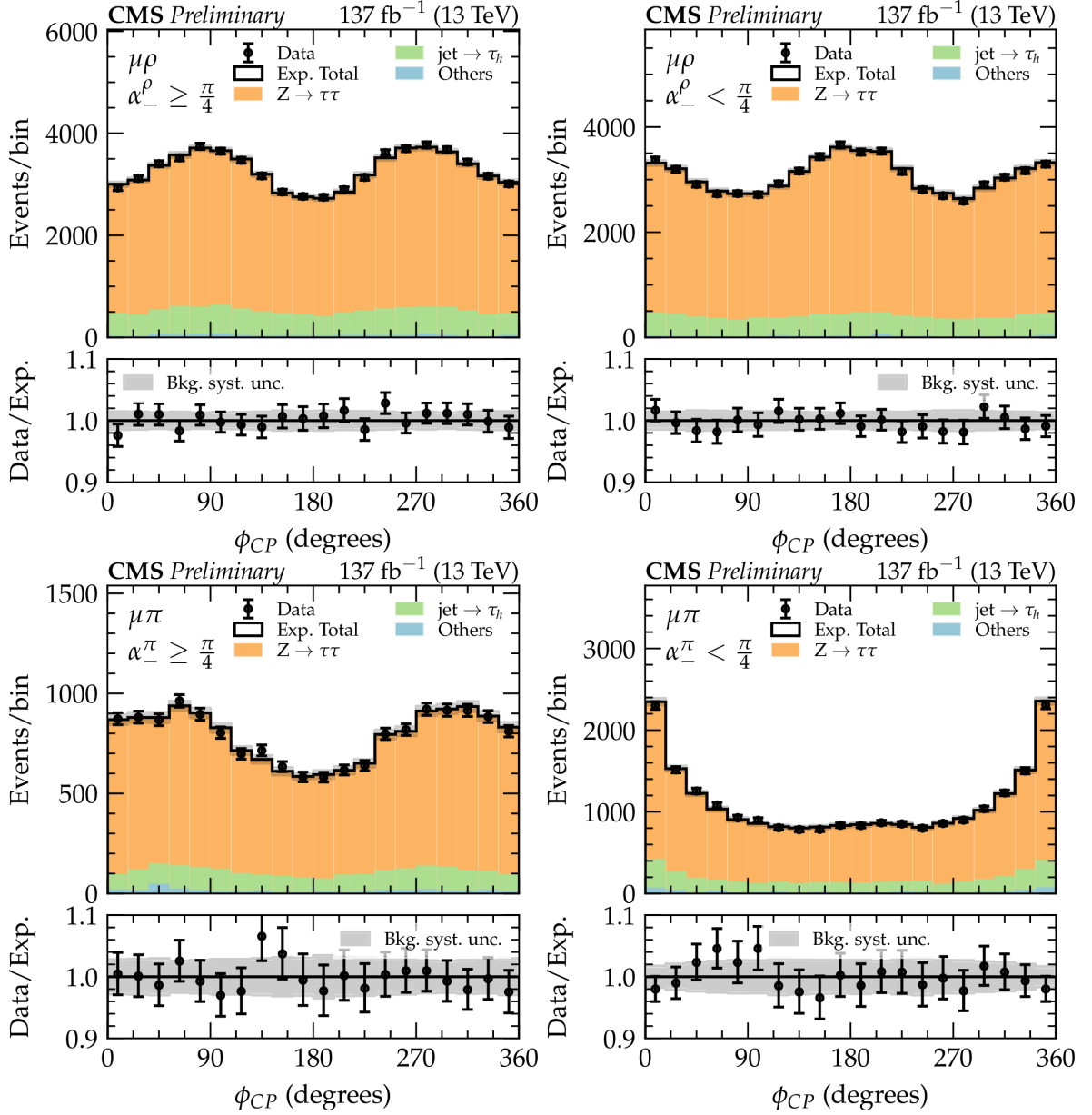
Figure B.2: Modulation of the DY cross section with respect to the acoplanarity angle measured in category of α_-. The top (bottom) row shows the $\tau_\mu \tau_\rho$ ($\tau_\mu \tau_\pi$) channel. On the left the categories $\alpha_- \geq \pi/4$ are shown, while on the right are the categories with $\alpha_- < \pi/4$.
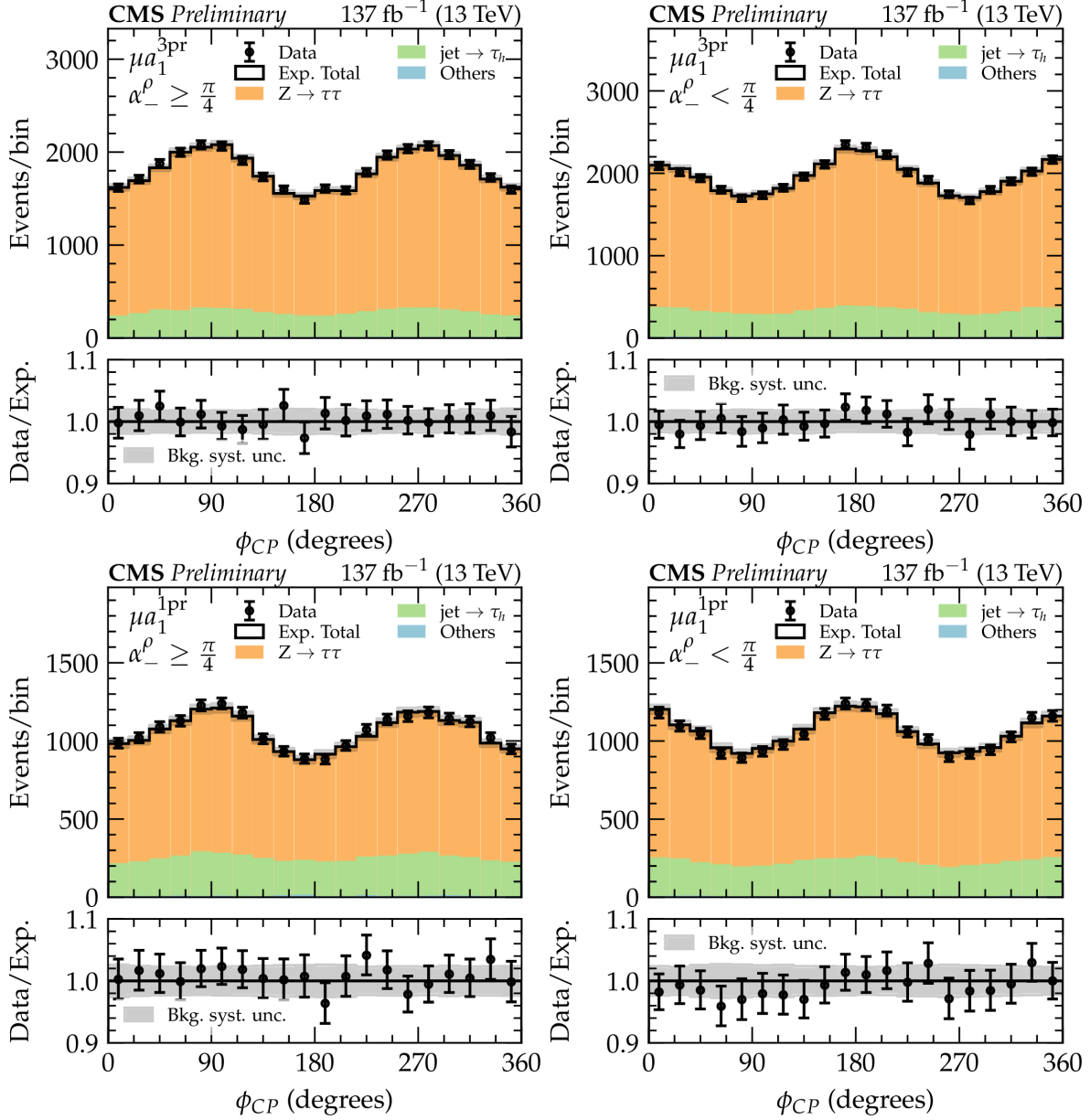
Figure B.3: Modulation of the DY cross section with respect to the acoplanarity angle measured in category of α_-. The top (bottom) row shows the $\tau_\mu \tau_{a_1^{3Pr}}$ ($\tau_\mu \tau_{a_1^{1Pr}}$) channel. On the left the categories $\alpha_- \geq \pi/4$ are shown, while on the right are the categories with $\alpha_- < \pi/4$.

Tau identification via cut-based and BDT-based algorithms

Contents

Before the deployment of the DeepTau ID [150,151] different discriminators were used to identify genuine τ_h against the major sources of contamination:

- jets coming from light quarks and gluons (see Section C.1);

- muons which can be misidentified a one prong decays (see Section C.2);

- electrons which can produce bremsstrahlung photons and be reconstructed as one prong + one π^0s decays (see Section C.3).

The methods used to reduce the corresponding misidentification rates are described in the following sections. The DeepTau ID is instead described in the main body of the thesis (see Section 4.4).

C.1   Discrimination against jets

The cut-based isolation sum discriminator is based on the idea that τ_h have generally well collimated decay products and are expected to be reconstructed as isolated from other particles in the detector. The sum of the transverse momenta of the particles found around the τ_h candidate, and not included among the tau decay products, is used to compute the tau isolation:

$$I_{\tau_h} = \sum p_T^{charged}(d_z < 0.2\text{cm}) + max\Big(0, \sum p_T^{\gamma} - \underbrace{\Delta\beta \sum p_T^{charged}(d_z > 0.2\text{cm})}_{\Delta\beta\text{-correction}}\Big) \ . \quad \text{(C.1)}$$

The first sum runs over the charged PF candidates reconstructed in a $\Delta R = 0.5$ isolation cone constructed around the τ_h direction of flight. To reduce the effect of PU contributions, and avoid rejecting genuine tau leptons produced in high multiplicity events, only PF candidates produced close to the PV from which the tau emerges, are used. The sum over the photons transverse momenta is corrected to account for possible bremsstrahlung radiation produced by charged particles emitted from PU vertices. This $\Delta\beta$-*correction* is operated by summing the transverse momenta of charged particles produced by PU vertices, i.e. with $d_z > 0.2$ cm, and reconstructed in a $\Delta R = 0.8$ cone. This sum is then scaled by the so-called $\Delta\beta$ factor, which was set to a value of 0.2. This value was obtained by taking the ratio of production rates between neutral and charged pions (0.5) and correct for the difference in cone sizes between the isolation cone and the one used to compute the PU correction: $0.5 \times (0.5/0.8)^2 \approx 0.195$.

With the dynamic strip reconstruction it is possible that a photon or electron is reconstructed outside the signal cone, that is the region of size $R_{sig} = (3.0\text{ GeV})/p_T$, resulting in a relative decrease of I_{τ_h} for $j \to \tau_h$ fakes. To reduce the misidentification rate a cut is placed on the energy fraction of the strip outside the signal cone:

$$p_T^{strip,outer} = \sum p_T^{e/\gamma}(\Delta R > R_{sig}) \; . \tag{C.2}$$

A cut $p_T^{strip,outer} < 0.1 \times p_T^{\tau_h}$ is used to decrease the misidentification rate from quark and gluon jets by 20%. Fig. C.1 shows the misidentification rate of $j \to \tau_h$ fakes as a function of the isolation cut-based discriminator using different values of $\Delta\beta$ and comparing the efficiency of the fixed strip reconstruction with the dynamic one. The multivariate against-jet discriminator combines the aforementioned tau isolation with several other variables in order to improve the tau identification efficiency and the rejection of quark and gluon jets. It is based on a Boosted Decision Tree (BDT) algorithm which uses multiple input variables to determine a probability that a given τ_h candidate is indeed a genuine tau. This method was developed already during Run 1, while the input variables were added progressively during the data-taking period. As of the end of Run 2 the input variables to the BDT are:

- the charged and neutral momenta sums defined in Eq. C.1;

- the tau decay mode identified by the HPS algorithm;

- the p_T, IP and IP$_{sig}$ of the leading hadron for the τ_h candidate;

- the distance $(|\vec{r}_{SV} - \vec{r}_{PV}|)$ and its significance $(|\vec{r}_{SV} - \vec{r}_{PV}|/\sigma_{\vec{r}_{SV}-\vec{r}_{PV}})$ between the reconstructed tau production (PV) and decay (SV) vertex, in cases when a SV was indeed reconstructed using the AVF [131] algorithm;

- τ lifetime, estimated by inverting Eq. 1.30 and replacing the average decay length with the aforementioned distance between SV and PV: $T_\tau = |\vec{r}_{SV} - \vec{r}_{PV}|/(c\gamma\beta)$;
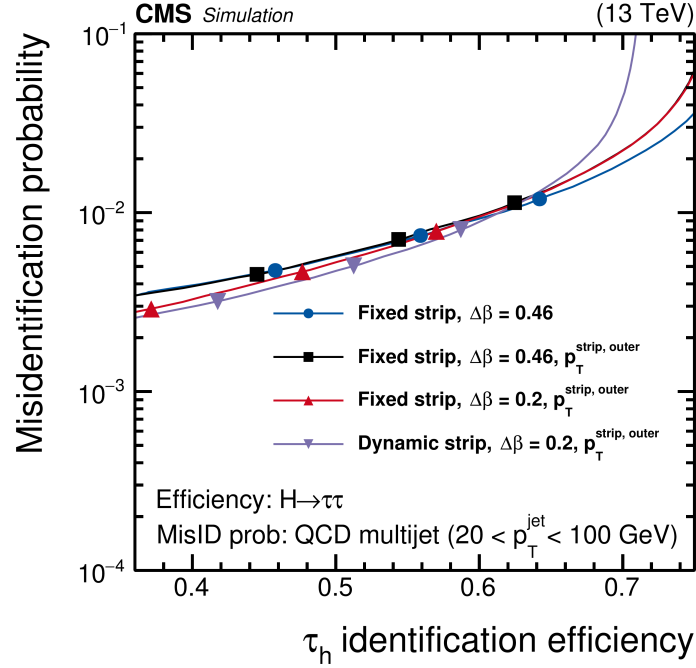
Figure C.1: Isolation sum against-jet discriminators [147]: the misidentification rate of quark and gluon jets, calculated on QCD multijet MC simulated events, is shown as a function of the efficiency in selecting genuine τ_h from a $H \to \tau\tau$ MC simulation.

- p_T weighted distances in η and φ between the strip constituents and the leading τ_h track;

- multiplicity of photons and electrons with $p_T > 0.5$ GeV.

The last two variables are grouped within the signal cone and the isolation cone and treated as separate inputs.

The BDT is trained using simulated samples of processes involving genuine τ_h such as decays of scalar and vector bosons both from SM (H, W, Z and γ^*) and BSM models (high mass scalar and pseudoscalar bosons predicted in MSSM, Z' and W'). The processes used in the training for the background of jets faking τ_h are instead W+jets, QCD multijet production and $t\bar{t}$. This method allowed to noticeably reduce the misidentification rate of $j \to \tau_h$ compared to the isolation sum-based discriminator. This is shown in fig. C.2 where the BDT-based discriminator (yellow line) is compared to the isolation sum-based ones used in Run 1 (green line), Run 2 (purple line) and the DeepTau discriminator (blue line), which is discussed later in this chapter.
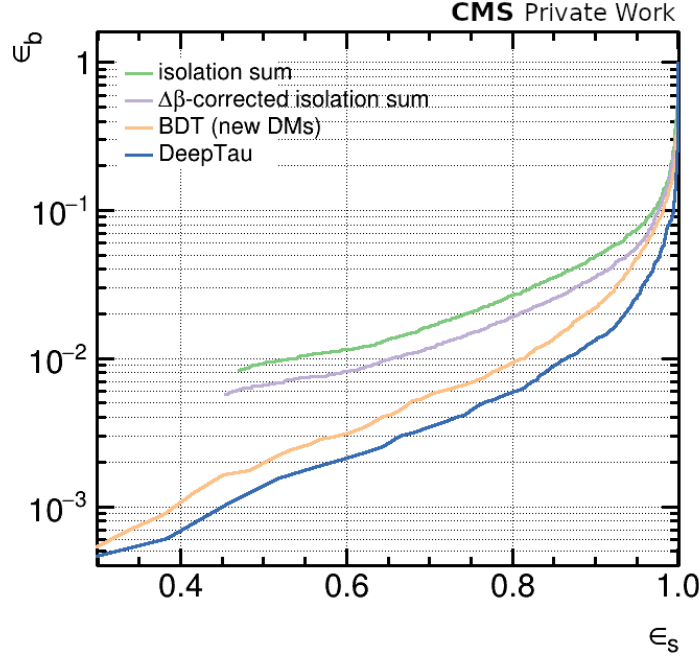
Figure C.2: Misidentification rate of $j \to \tau_h$ (ϵ_b) as a function of the efficiency in the selection of genuine τ_h (ϵ_s). The plot was made computing the misidentification rate and the tau identification efficiency on QCD multijet production and $H \to \tau\tau$ MC simulations respectively [200].

C.2   Discrimination against muons

Identification of hadronically decaying tau leptons from light leptons has historically been performed separately from the jet rejection, until the introduction of the DeepTau identification. This was motivated by the intrinsic differences of these misidentified objects compared to quark and gluon jets. Muons, in particular, can be reconstructed as τ_h candidates if they produce energy deposits in the calorimeters, mimicking a $h^\pm$ decay mode, a process which is much less probable than the $j \to \tau_h$ misidentification. The main reason to study this misidentification is that processes like the $Z \to \mu\mu$ have much larger cross sections compared to processes of interest in several analysis, like $H \to \tau\tau$ decays. Thus even if the misidentification probability is low, the effective number of events selected as $\tau_\mu\tau_h$ pairs is large enough to form a relevant background. The misidentification rate is reduced by vetoing events where a muon track found in the muon chambers can be extrapolated to the reconstructed τ_h candidate. More precisely, two working points are defined:

- *against-muon loose*: a τ_h candidate is rejected if at least two hits are found in the muon chambers leading to an extrapolated track which is within a cone of $\Delta R = 0.3$

from the τ_h candidate;

- *against-muon tight*: a τ_h fails this selection if it fails the loose WP or if a hit is present in the outermost muon stations contained in a cone of $\Delta R = 0.3$.

As previously mentioned this misidentification is relatively smaller compared to the $j \to \tau_h$ and is of the order of $\mathcal{O}(10^{-3})$ for a genuine tau identification efficiency larger than 99%.

C.3 Discrimination against electrons

Electron misidentification as τ_h falls between the jet and the muon misidentification as τ_h in terms of probability. The electromagnetic shower produced by an electron can occasionally reach HCAL producing a signature similar to a charged pion, while eventual bremsstrahlung photons can produce strip like signature. This means that electrons can fake mainly $h^\pm$ and $h^\pm\pi^0$ decay modes. The identification is performed using a BDT trained similarly to the one for the discrimination against jets. While the BDT has a similar structure, it is trained using different background simulated processes: $Z \to ee$, $W \to e\nu$ and the analogous decays for Z' and W'. The following input variables are used:

- the mass of the τ_h candidate;

- the p_T of the leading charged hadron;

- fraction of energy deposited in ECAL: $E_{ECAL}/(E_{ECAL} + E_{HCAL})$;

- ratios between energy deposited in ECAL and HCAL and the leading hadron momenta;

- number of photons per strip;

- the fraction of energy released as bremsstrahlung radiation, that is $F_{brem} = (p_{in} - p_{out})/p_{in}$, with p_{in} and p_{out} being the momentum of the leading GSFTrack (see Section 4.1.4), at the innermost and outermost tracker layer;

- the ratio between the energy collected in ECAL and the momentum of the leading track, $(E_e + \sum E_\gamma)/p_{in}$, with E_e energy of the electron cluster and E_γ total energy collected from photons;

- the ratio of energies of the bremsstrahlung photons measured in the ECAL and in the tracker, $\sum E_\gamma/(p_{in} - p_{out})$;

- the RMS of the p_T weighted distances between the strip constituents and the leading τ_h track;

- the energy fraction carried by photons.

The last three variables are kept as separate inputs for objects reconstructed inside and outside the τ_h signal cone.

Five working points are defined for this BDT-based classifier, ranging from very loose to very tight. Starting from the loosest WP, the BDT-classifier can reduce the rate of misidentified electrons to $\sim 5\%$ with an efficiency for genuine τ_h of $\sim 85\%$. Moving to tighter WPs both the misidentification rate and the genuine τ_h efficiency are reduced reaching respectively 10^{-3} and 70% for the tightest WP.

APPENDIX D

Additional figures for $H \to \tau_\mu \tau_h$

Contents

Chapter 6 describes the study of the CP properties of Higgs Yukawa coupling to tau leptons in the $\tau_\mu \tau_h$ final state. As discussed in Section 6.7, a neural network (NN) based classification was used to identify events involving $H \to \tau\tau$ decays from the dominant backgrounds. A total of 13 input variables are used as input for the NN, and can be found in Fig. 6.19 and 6.20 for data and background models in 2018. Sections D.1 and D.2 shown instead the corresponding distributions for 2016 and 2017 respectively. Event selection and background modeling are described in Sections 6.2, 6.5 and 6.6.

D.1   NN input variables in 2016
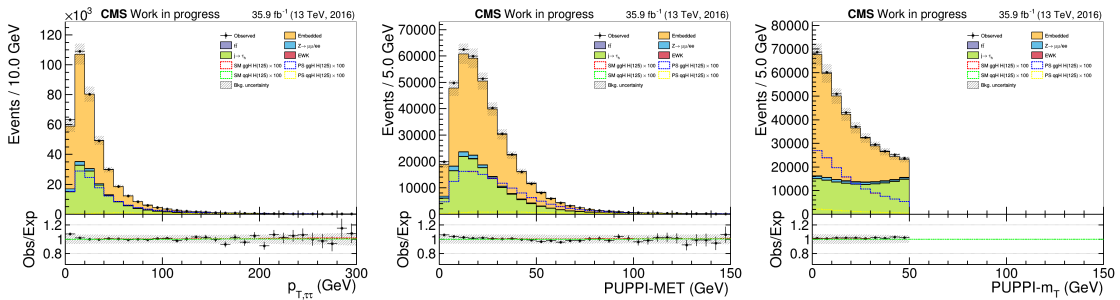
Figure D.1: Transverse momentum of the lepton pair (left), missing transverse energy (middle) and transverse mass (right) distributions using 2016 data and simulation.
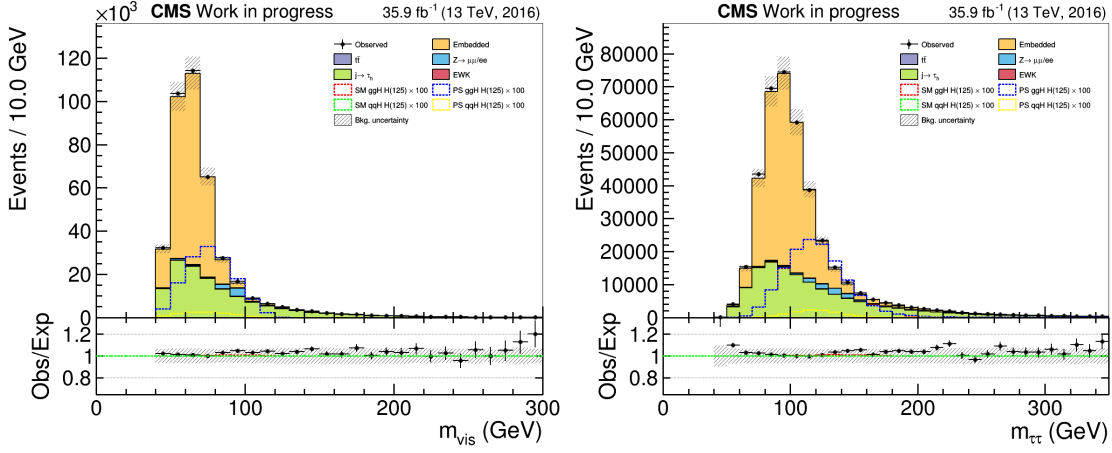
Figure D.2: Distribution of the invariant tau pair mass reconstructed using only the visible decay products (left) and with the SVFit algorithm (right) using 2016 data and simulation.
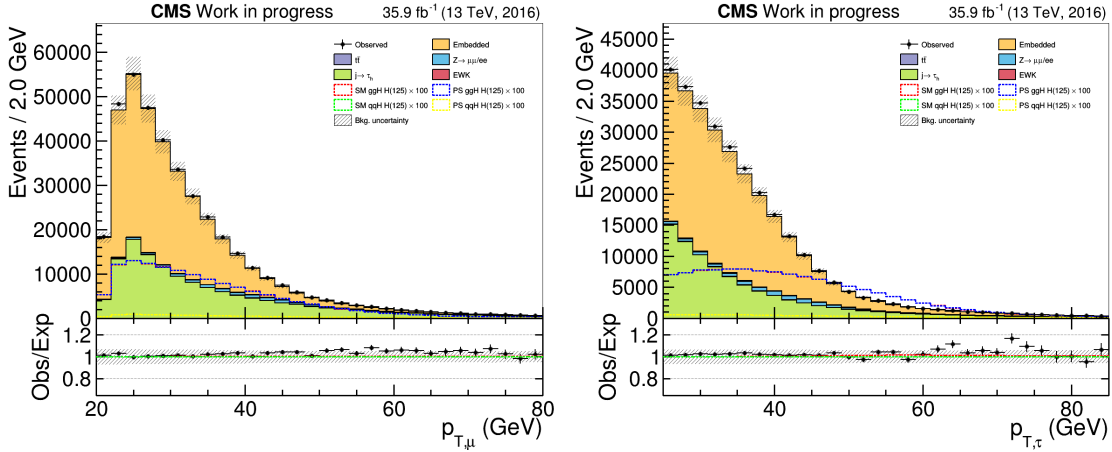
Figure D.3: Distribution of the muon (left) and tau (right) transverse momenta using 2016 data and simulation.
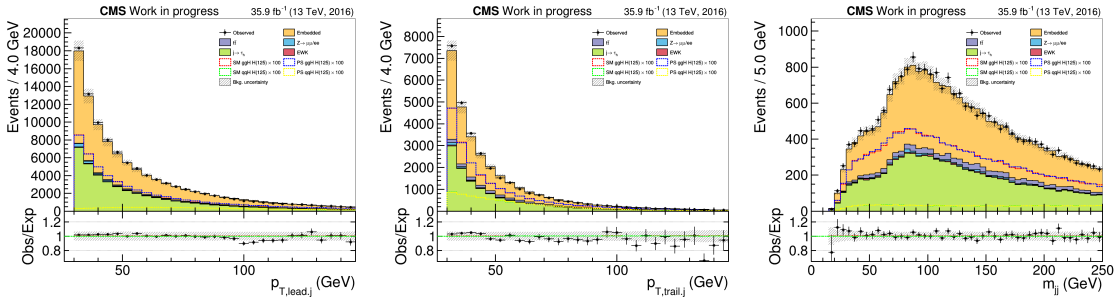
Figure D.4: Transverse momentum of leading (left) and subleading (middle) jets, and their invariant mass (right) distributions using 2016 data and simulation.
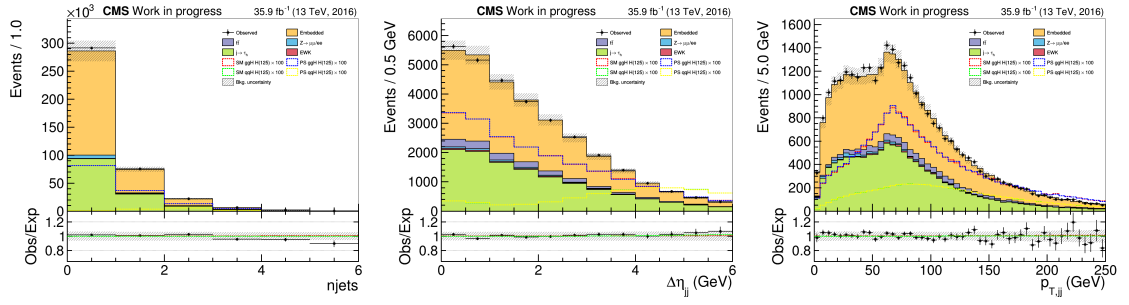
Figure D.5: Transverse momentum of leading (left) and subleading (middle) jets, and their invariant mass (right) distributions using 2016 data and simulation.

D.2 NN input variables in 2017
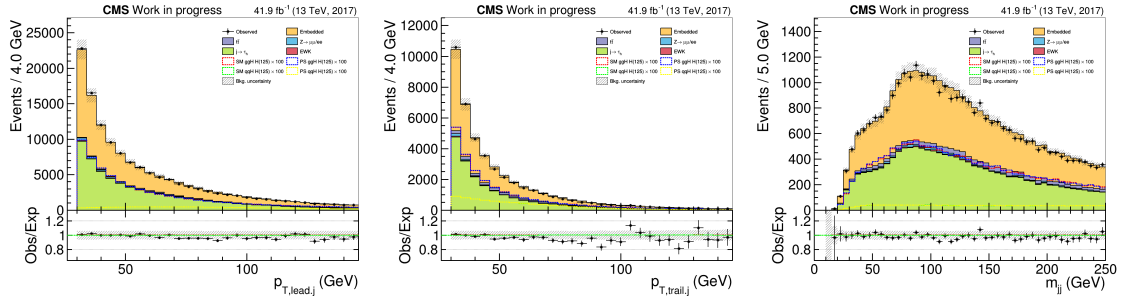
Figure D.6: Transverse momentum of leading (left) and subleading (middle) jets, and their invariant mass (right) distributions using 2017 data and simulation.
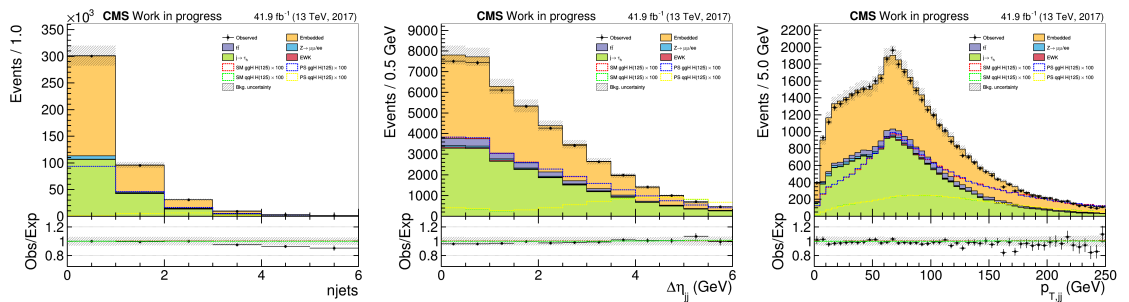
Figure D.7: Number of jets (left), eta separation between (middle) and transverse momentum of (right) the jets distributions using 2017 data and simulation.
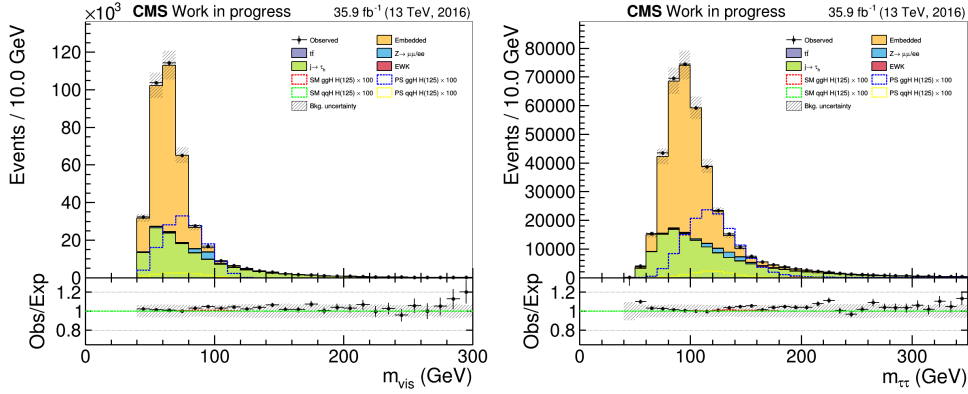
Figure D.8: Distribution of the invariant tau pair mass reconstructed using only the visible decay products (left) and with the SVFit algorithm (right) using 2017 data and simulation.
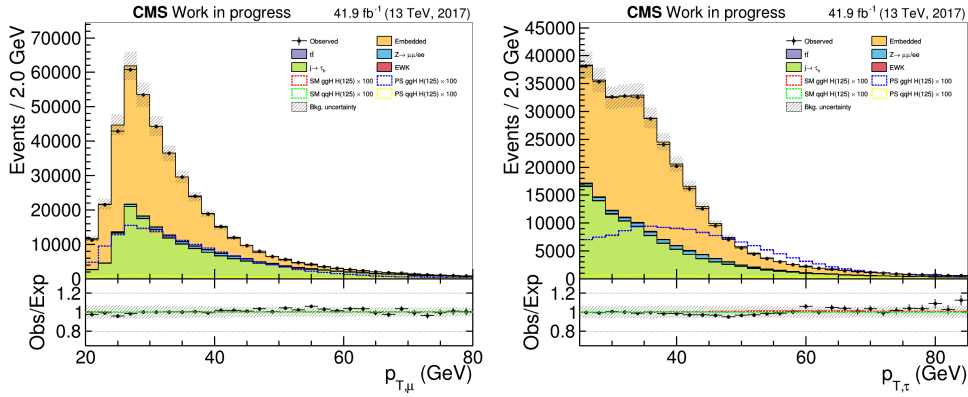
Figure D.9: Distribution of the muon (left) and tau (right) transverse momenta using 2017 data and simulation.
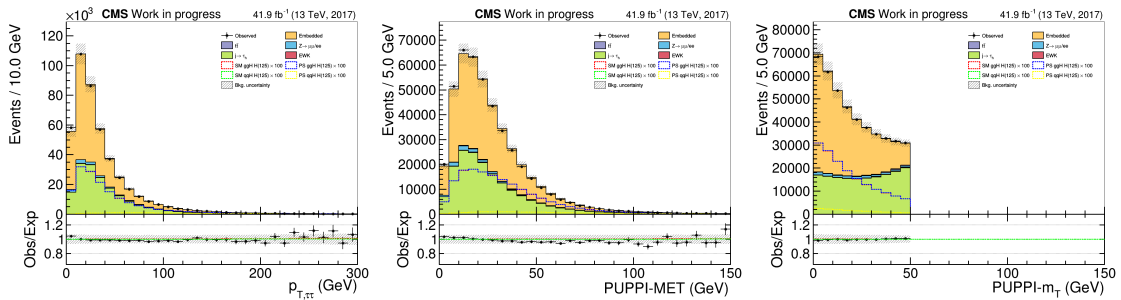
Figure D.10: Transverse momentum of the lepton pair (left), missing transverse energy (middle) and transverse mass (right) distributions using 2017 data and simulation.

Pulls and GOF tests for the combination of $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels

Contents

Several checks were performed as part of the analysis described in Chapters 6 and 7. As part of the Maximum Likelihood (ML) fit to determine the value of the CP mixing angle, the nuisance parameters were investigated in order to check that none was overconstrained or with a large pull on the measured parameters of interest. In Section E.1, the pulls of the most relevant nuisance parameters are shown when performing the ML fit on data. Since this analysis involved over 1000 nuisance parameters, only the 120 most relevant ones are shown.

Section E.2 shows instead the results of the goodness of fit (GOF) tests performed on the background categories to determine the agreement between the recorded data and the modeled backgrounds. The test performed were:

- a saturated model GOF test defined as in [183];

- a Kolmogorov-Smirnov GOF test described in [184].

All tests reported a good agreement between recorded data and signal and background models.

E.1   Pulls of the systematic model
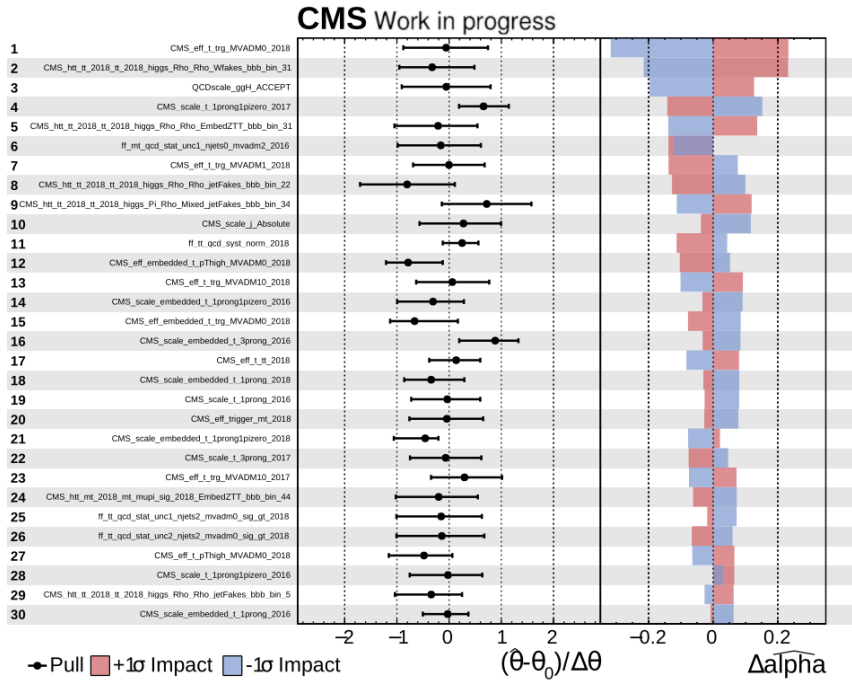
Figure E.1: Most relevant 30 nuisance parameters and associated pulls.
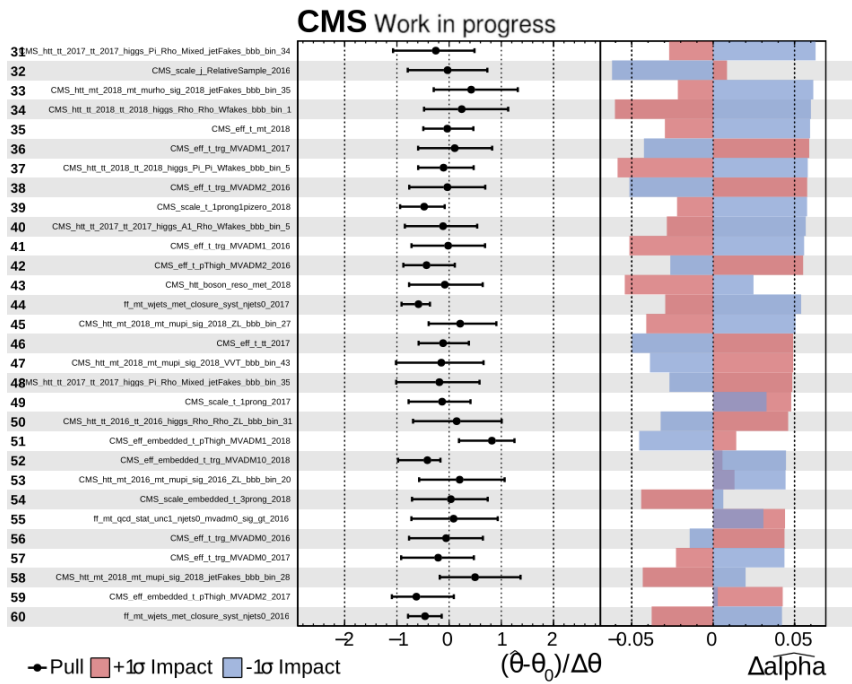
Figure E.2: Nuisance parameters 31-60 and associated pulls.
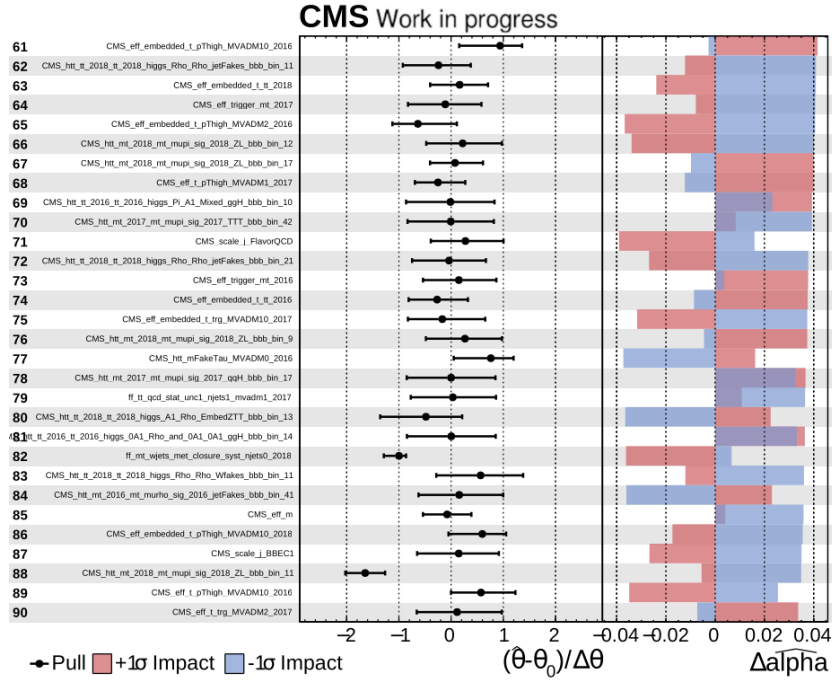
Figure E.3: Nuisance parameters 61-90 and associated pulls.
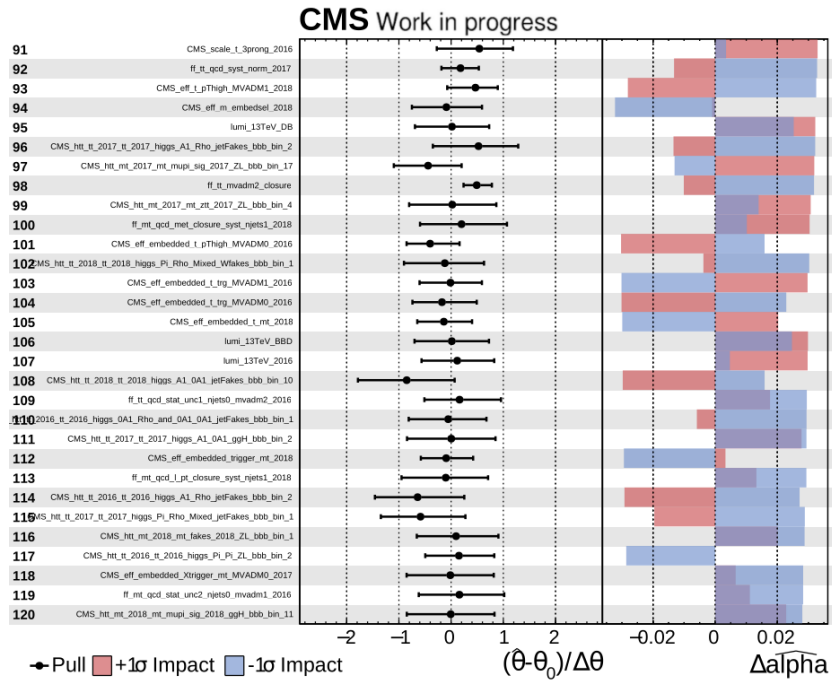
Figure E.4: Nuisance parameters 91-120 and associated pulls.

E.2   Goodness of fit tests for the combination of all categories

E.2.1   Saturated model test
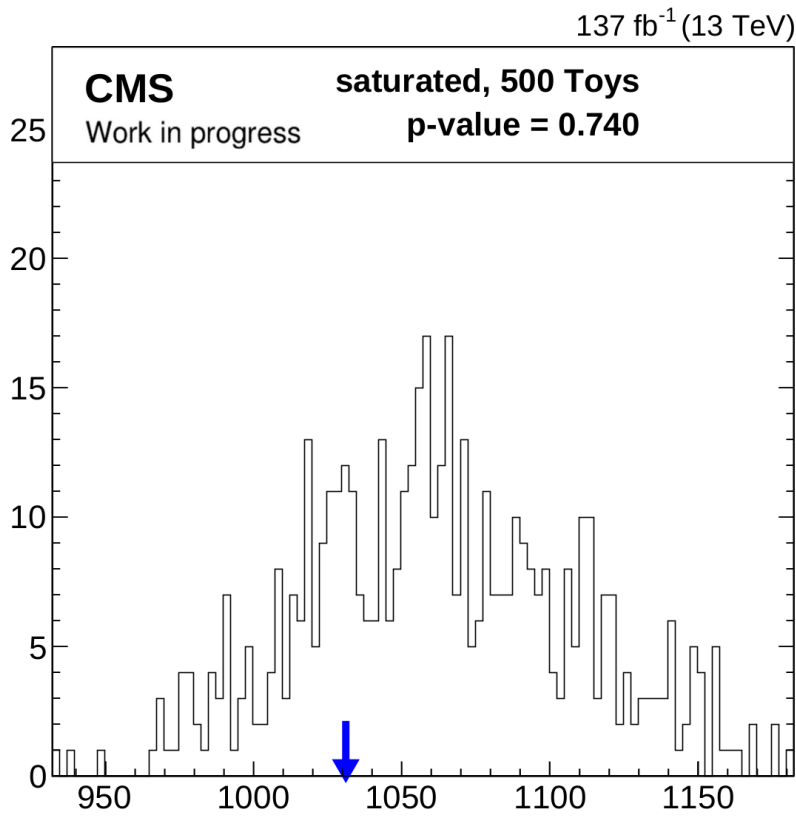
Figure E.5: Result of the saturated model GOF test performed on all categories.
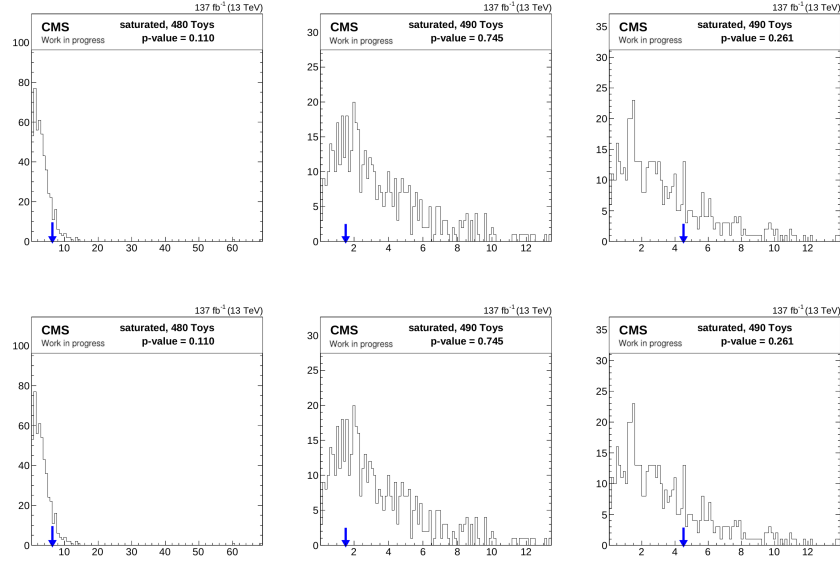
Figure E.6: Result of the saturated model GOF test performed on taus (top) and fakes (bottom) categories of the $\tau_\mu\tau_h$ channel.

E.2.2 Kolmogorov-Smirnov test
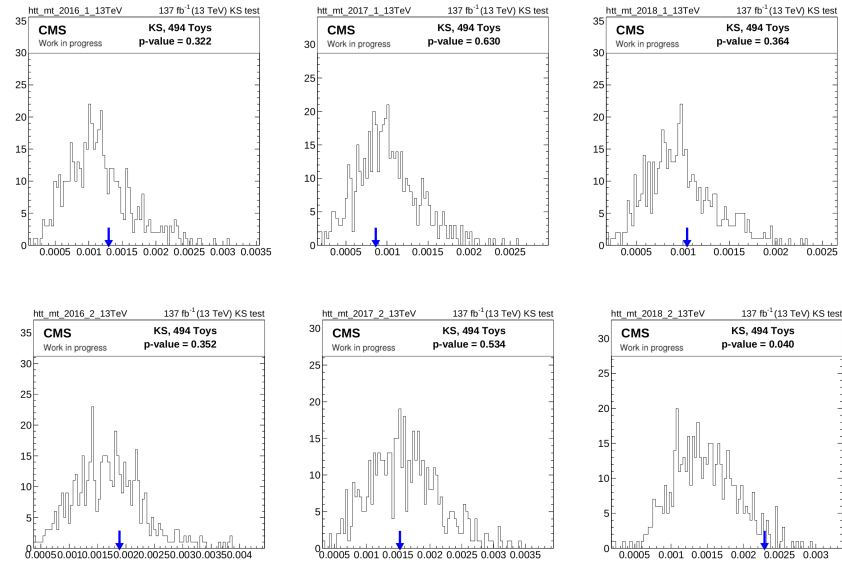
Figure E.7: Result of the Kolmogorov-Smirnov GOF test performed on taus (top) and fakes (bottom) categories of the $\tau_\mu\tau_h$ channel.

Notations and acronyms

- $\mathscr{O}(10^x$ in units of measure Y): order of magnitude for a certain quantity, generally expressed as a power of 10 followed by a unit of measure.

- φ: azymuthal angle.

- γ: photon.

- η: pseudorapidity.

- ρ: rho meson.

- $\mu(^\pm)$: can refer either to 1) a muon or 2) the signal strength of a process, usually specified as a subscript.

- $\tau(^\pm)$: tau lepton.

- $\varphi_{\tau\tau}$: CP mixing angle in Yukawa couplings between Higgs bosons and tau leptons.

- $\tau_{\pi/\rho/a_1^{3Pr}/a_1^{1Pr}}$: tau lepton decay to a charged pion $(\pi)/\pi + \pi^0/\pi + 2\,\pi^0/3\pi$.

- $\tau_{\mu/e}$: leptonically decaying tau lepton into a muon/an electron.

- φ_{CP}: acoplanarity angle for a tau lepton pair system.

- $\chi_{L/R}$: Left/right chirality symmetry.

- τ_h: hadronically decaying tau lepton.

- τ_l: leptonically decaying tau lepton.

- C_C: Color charge.

- T: Weak isospin.

- Y: Hypercharge.

- a_1^{1Pr}: a_1 meson decaying into one charged pion and 2 neutral pions.

- a_1^{3Pr}: a_1 meson decaying into 3 charged pions.

- $e(^{\pm})$: electron.

- $h^{\pm}$: charged hadron.

- j: jet.

- $l(^{\pm})$: light leptons, namely electron and muon.

- p_T: transverse momentum.

- ALICE: A Large Ion Collider Experiment.

- AOD: Analysis Object Data, dataset used by the CMS collaboration.

- APD: Avalanche photo-diode.

- ATLAS: A Toroidal LHC Apparatus.

- AVF: Adaptive Vertex Fitter.

- AVR: Adaptive Vertex Reconstruction.

- Adam: Adaptive Moment Estimation.

- BAU: Baryon Asymmetry in the Universe.

- BDT: Boosted Decision Tree.

- BPIX: Barrel Pixel detector.

- BR: Branching fraction.

- BS: Beam Spot.

- BSM: theory Beyond the Standard Model of particle physics.

- C: Charge conjugation symmetry.

- CERN: *Conseil Européen pour la Recherche Nucléaire* (European Council for Nuclear Research).

- CL: Confidence Level.

- CLEO: Cleopatra collaboration.

- CMS: Compact Muon Solenoid.

- CNN: Convolutional Neural Network.

- CP: Symmetry under charge conjugation and parity.

- CSC: Cathode Strip Chambers.

- DA: Deterministic Annealing.

- DLM: Dynamical Likelihood Method.

- DM: Decay Mode.

- DNN: Deep Neural Network.

- DR: Determination Region.

- DT: Drift Tube detector.

- DY: Drell-Yan production of pair of leptons (including tau leptons) via Z boson or virtual photon, possibly accompanied by jets.

- EB, EE, ES: ECAL Barrel, Endcap, or preshower.

- ECAL: Electromagnetic calorimeter.

- ES: Energy Scale.

- EW: Electroweak.

- EYETS: extended EYOTS.

- EYOTS: End of the year technical stop.

- FCC: Future Circular Collider.

- FES: Energy scale for misidentified objects.

- FPIX: Forward Pixel detector.

- FR: Fake Rate.

- FSR: Final State Radiation.

- GSF: Gaussian Sum Filter.

- HB, HE, HO, HF: HCAL Barrel, Endcap, Outer or Forward.

- HCAL: Hadronic calorimeter.

- HL-LHC: High-Luminosity Large Hadron Collider.

- HLT: High-level trigger.

- HPD: Hybrid Photo-detectors.

- HPS: Hadron-plus-strip.

- HVV: Higgs boson coupling to vector bosons.

- ID: Identification algorithm.

- ILC: Internationa Linear Collider.

- IP: Impact Parameter.

- ISR: Initial State Radiation.

- JEC: Jet Energy Correction.

- JES: Jet Energy Scale.

- KF: Kalman Filter algorithm.

- L1: level 1 trigger.

- LEP: Large Electron Positron Collider.

- LHC: Large Hadron Collider.

- LHCb: Large Hadron Collider beauty.

- LHCf: LHC forward.

- LINAC: Linear Accelerator.

- LO: Leading Order.

- LS: Long Shutdown of the LHC.

- MC: Monte Carlo.

- MET: Missing transverse energy.

- ML: depending on the context either 1) Machine Learning or 2) Maximum Likelihood.

- MM: Maximal mixing between CP-even and CP-odd Yukawa couplings ($\varphi_{\tau\tau} = \pi/4$).

- MSSM: Minimal Supersymmetric extension to the Standard Model.

- MVA-DM: MVA-based algorithm to determine the τ Decay Mode.

- MVA: Multivariate analysis technique.

- MoEDAL: Monopole and exotics detector at the LHC.

- N(N)LO: Next-to-(next-to-)leading order.

- NN: Neural Network.

- O(S)S: Opposite(Same) -Sign

- P1, P2, P5 and P8: collision points at the LHC.

- P: Parity symmetry.

- PCA: Point of Closest Approach.

- PF: Particle Flow.

- PIX: Pixel detector.

- PM: Photomultiplier.

- PMT: Photomultiplier tubes.

- POI: Parameter of interest.

- PS: can refer either to 1) Proton Synchrotron or 2) pseudoscalar decay of the Higgs boson ($\varphi_{\tau\tau} = \pi/2$).

- PSB: Proton Synchrotron Booster.

- PU: Pile-up.

- PUPPI: Pile-up per particle identification.

- PV: Primary (interaction) Vertex.

- QCD: Chromo Quantum-dynamics.

- RAW: raw data dataset collected by the CMS detector.

- RECO: reconstructed, can refer to either 1) a dataset type for the CMS collaboration or 2) reconstructed level for a MC simulated proton-proton collision.

- RP: Reference Point for a track.

- RPC: Resistive Plate Chambers.

- ReLU: Rectified Linear Unit.

- SC: Super Cluster.

- SF: Scale Factor.

- SLAC: Stanford Linear Accelerator Center.

- SM: most commonly used for Standard Model of particle physics, it refers in Eq. 6.25 to a CP-even Higgs decay ($\varphi_{\tau\tau} = 0$).

- SPS: Super Proton Synchrotron.

- SR: Signal Region.

- SUSY: Supersymmetry.

- SV: Secondary (interaction) Vertex.

- SVFit: SVFit algorithm for reconstructing the invariant mass of a tau lepton pair.

- SiPM: Silicon Photomultiplier.

- T: Time reversal symmetry.

- TEC: Tracker Endcaps.

- TIB: Tracker Inner Barrel.

- TID: Tracker Inner Disks.

- TOB: Tracker Outer Barrel.

- TOTEM: Total cross section, elastic scattering and diffraction dissociation measurement at the LHC.

- TP: Trainable Parameter (for a neural network).

- VBF: Vector Boson Fusion production mechanism for the Higgs boson.

- VH: Higgs-strahlung production mechanism.

- VPD: Vacuum photo-triode.

- VV: production of a pair of vector bosons.

- W($^\pm$)H: Higgs boson production accompanied by a W($^\pm$) boson.

- W+Jets: W boson production in association with jets.

- WLF: Wavelength shift fiber.

- WP: Working Point of an algorithm.

- ZEE: Drell-Yan production of pair of electrons via Z boson or virtual photon, possibly accompanied by jets.

- ZH: Higgs boson production accompanied by a Z boson.

- ZJ: Drell-Yan production of pair of leptons via Z boson or virtual photon, with a jet misidentified as hadronically decaying tau lepton.

- ZLL: Drell-Yan production of pair of electrons or muons via Z boson or virtual photon, possibly accompanied by jets.

- ZTT: Drell-Yan production of pair of τ leptons via Z boson or virtual photon, possibly accompanied by jets.

- bbH: b-associated production of the Higgs boson.

- e.g.: *exempli gratia* (for example).

- ggH: gluon-gluon fusion production mechanism for the Higgs boson.

- i.e.: *id est* (that is to say).

- lnN: log-normal uncertainty.

- miniAOD, nanoAOD: subsets of the AOD dataset for the CMS collaboration.

- nMSSM: next-to-Minimal Supersymmetric extension to the Standard Model.

- pp: proton-proton.

- prong: charged decay product of a tau lepton.

- strip: cluster of electrons and photons identified by the HPS algorithm.

- $t\bar{t}$: production of a top and anti-top pair.

- ttH: top-associated production of the Higgs boson.

Bibliography

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** (2012), no. 1, 1 – 29, `doi:10.1016/j.physletb.2012.08.020`.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (2012), no. 1, 30 – 61, `doi:10.1016/j.physletb.2012.08.021`.

[3] CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector", *Physics Letters B* **779** (2018) 283 – 316, `doi:10.1016/j.physletb.2018.02.004`.

[4] CMS Collaboration, "Observation of Higgs Boson Decay to Bottom Quarks", *Phys. Rev. Lett.* **121** (Sep, 2018) 121801, `doi:10.1103/PhysRevLett.121.121801`.

[5] CMS Collaboration, "Evidence for Higgs boson decay to a pair of muons", `arXiv:2009.04363`.

[6] D. J. Griffiths, "Introduction to elementary particles; 2nd rev. version". Physics textbook. Wiley, New York, NY, 2008.

[7] Y. Nagashima, "Elementary particle physics. Vol. 1: Quantum field theory and particles". 2010.

[8] P. Langacker, "The standard model and beyond; 2nd ed.". High energy physics, cosmology and gravitation. CRC Press, Jul, 2017.

[9] F. Mandl and G. Shaw, "Quantum field theory". John Wiley & Sons, 2010.

[10] D. Fehling, "The standard model of particle physics: A lunchbox's guide", 2008.

[11] J. Campbell, J. Huston, and F. Krauss, "The black book of quantum chromodynamics: a primer for the LHC era". Oxford University Press, Oxford, 2018.

[12] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, `doi:10.1103/PhysRevLett.13.321`.

[13] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133, `doi:10.1016/0031-9163(64)91136-9`.

[14] A. Zee, "Group Theory in a Nutshell for Physicists". Princeton University Press, USA, 2016.

[15] H. H. Denman, "Invariance and conservation laws in classical mechanics", *Journal of Mathematical Physics* **6** (1965), no. 11, 1611–1616.

[16] C. S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.* **105** (Feb, 1957) 1413–1415, `doi:10.1103/PhysRev.105.1413`.

[17] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π Decay of the K_2^0 Meson", *Phys. Rev. Lett.* **13** (Jul, 1964) 138–140, `doi:10.1103/PhysRevLett.13.138`.

[18] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture", *Phys. Lett. B* **47** (1973) 365–368, `doi:10.1016/0370-2693(73)90625-4`.

[19] H. D. Politzer, "Asymptotic Freedom: An Approach to Strong Interactions", *Phys. Rept.* **14** (1974) 129–180, `doi:10.1016/0370-1573(74)90014-3`.

[20] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. I", *Phys. Rev. D* **8** (Nov, 1973) 3633–3652, `doi:10.1103/PhysRevD.8.3633`.

[21] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215, `doi:10.1016/S0031-9163(64)92001-3`.

[22] F. Ynduráin, "Limits on the mass of the gluon", *Physics Letters B* **345** (1995), no. 4, 524–526, `doi:10.1016/0370-2693(94)01677-5`.

[23] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (Aug, 2018) 030001, `doi:10.1103/PhysRevD.98.030001`.

[24] Belle Collaboration, "Observation of a near-threshold enhancement in the e+e- $\rightarrow$ Lambda+(c) Lambda-(c) cross section using initial-state radiation", *Phys. Rev. Lett.* **101** (2008) 172001, `doi:10.1103/PhysRevLett.101.172001`, `arXiv:0807.4458`.

[25] Belle Collaboration, "Observation of a resonance-like structure in the $pi^{\pm}\psi'$ mass distribution in exclusive $B \rightarrow K\pi^{\pm}\psi'$ decays", *Phys. Rev. Lett.* **100** (2008) 142001, `doi:10.1103/PhysRevLett.100.142001`, `arXiv:0708.1790`.

[26] LHCb Collaboration, "Observation of structure in the J/ψ -pair mass spectrum", *Sci. Bull.* **65** (2020), no. 23, 1983–1993, `doi:10.1016/j.scib.2020.08.032`, `arXiv:2006.16957`.

[27] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, `doi:10.1103/PhysRevLett.115.072001`, `arXiv:1507.03414`.

[28] LHCb Collaboration, "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$", *Phys. Rev. Lett.* **122** (2019), no. 22, 222001, `doi:10.1103/PhysRevLett.122.222001`, `arXiv:1904.03947`.

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

[30] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377, `doi:10.1142/9789812795915_0034`.

[31] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, `doi:10.1103/PhysRevLett.19.1264`.

[32] A. Einstein, "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt", *Annalen der Physik* **322** (January, 1905) 132–148, `doi:10.1002/andp.19053220607`.

[33] P. A. M. Dirac, "Quantum theory of emission and absorption of radiation", *Proc. Roy. Soc. Lond. A* **114** (1927) 243, `doi:10.1098/rspa.1927.0039`.

[34] E. Fermi, "Sopra l'elettrodinamica quantistica", *Atti della Reale Accademia Nazionale dei Lincei* **9** (1929) 881–887, `doi:10.1098/rspa.1927.0039`.

[35] E. Fermi, "Quantum Theory of Radiation", *Rev. Mod. Phys.* **4** (1932), no. 1, 87–132, `doi:10.1103/RevModPhys.4.87`.

[36] J. R. Oppenheimer, "Note on the Theory of the Interaction of Field and Matter", *Phys. Rev.* **35** (Mar, 1930) 461–477, `doi:10.1103/PhysRev.35.461`.

[37] J. Lykken and M. Spiropulu, "The future of the Higgs boson", *Physics Today* **66** (2013), no. 12, 28–33, `doi:10.1063/PT.3.2212`.

[38] CMS Collaboration, "Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel", *Phys. Rev. Lett.* **125** (Aug, 2020) 061801, `doi:10.1103/PhysRevLett.125.061801`.

[39] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", `doi:10.23731/CYRM-2017-002`, `arXiv:1610.07922`.

[40] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel", *Physics Letters B* **805** (2020) 135425, `doi:10.1016/j.physletb.2020.135425`.

[41] CMS Collaboration, "Evidence for Higgs boson decay to a pair of muons", *JHEP* **01** (2021) 148, `doi:10.1007/JHEP01(2021)148`, `arXiv:2009.04363`.

[42] T. Nakano and K. Nishijima, "Charge Independence for V-particles", *Progress of Theoretical Physics* **10** (11, 1953) 581–582, `doi:10.1143/PTP.10.581`.

[43] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, `doi:10.1103/PhysRevLett.81.1562`, `arXiv:hep-ex/9807003`.

[44] "The Nobel Prize in Physics 2015". NobelPrize.org. Nobel Media AB 2021.

[45] P. Minkowski, "$\mu \to e\gamma$ at a Rate of One Out of 10^9 Muon Decays?", *Phys. Lett. B* **67** (1977) 421–428, `doi:10.1016/0370-2693(77)90435-X`.

[46] LHCb Collaboration, "Test of lepton universality in beauty-quark decays", `arXiv:2103.11769`.

[47] Y. Bai and J. Berger, "Muon $g - 2$ in Lepton Portal Dark Matter", `arXiv:2104.03301`.

[48] Y.-S. Tsai, "Decay Correlations of Heavy Leptons in $e^+ + e^- \to l^+ + l^-$", *Phys. Rev. D* **4** (Nov, 1971) 2821–2837, `doi:10.1103/PhysRevD.4.2821`.

[49] M. L. Perl et al., "Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation", *Phys. Rev. Lett.* **35** (Dec, 1975) 1489–1492, `doi:10.1103/PhysRevLett.35.1489`.

[50] V. Cherepanov, "Measurement of the polarization of tau–leptons produced in Z decays at CMS and determination of the effective weak mixing angle". PhD thesis, Aachen, Tech. Hochsch., 2016.

[51] CLEO Collaboration, "Hadronic structure in the decay tau- → tau-neutrino pi-pi0 pi0 and the sign of the tau-neutrino helicity", *Phys. Rev. D* **61** (2000) 012002, `doi:10.1103/PhysRevD.61.012002`, `arXiv:hep-ex/9902022`.

[52] BaBar Collaboration, "Measurement of the tau lepton lifetime with BaBar",
Nucl. Phys. B Proc. Suppl. **144** (2005) 105–112,
`doi:10.1016/j.nuclphysbps.2005.02.014`.

[53] The Belle Collaboration Collaboration, "Measurement of the τ Lepton Mass and
an Upper Limit on the Mass Difference between τ^+ and τ^-", *Phys. Rev. Lett.* **99**
(Jul, 2007) 011801, `doi:10.1103/PhysRevLett.99.011801`.

[54] A. Soni and R. M. Xu, "Probing CP violation via Higgs boson decays to four
leptons", *Phys. Rev. D* **48** (Dec, 1993) 5259–5263,
`doi:10.1103/PhysRevD.48.5259`.

[55] S. Bolognesi et al., "Spin and parity of a single-produced resonance at the LHC",
Phys. Rev. D **86** (Nov, 2012) 095031, `doi:10.1103/PhysRevD.86.095031`.

[56] Y. Gao et al., "Spin determination of single-produced resonances at hadron
colliders", *Phys. Rev. D* **81** (Apr, 2010) 075022,
`doi:10.1103/PhysRevD.81.075022`.

[57] S. Choi, D. Miller, M. Mühlleitner, and P. Zerwas, "Identifying the Higgs spin and
parity in decays to Z pairs", *Physics Letters B* **553** (2003), no. 1, 61 – 71,
`doi:10.1016/S0370-2693(02)03191-X`.

[58] CMS Collaboration, "Measurement of the properties of a Higgs boson in the
four-lepton final state", *Phys. Rev. D* **89** (Dec, 2013) 092007. 45 p,
`doi:10.1103/PhysRevD.89.092007`.

[59] ATLAS Collaboration, "Evidence for the spin-0 nature of the Higgs boson using
ATLAS data", *Phys. Lett. B* **726** (2013) 120–144,
`doi:10.1016/j.physletb.2013.08.026`, `arXiv:1307.1432`.

[60] CMS Collaboration, "Constraints on the spin-parity and anomalous HVV
couplings of the Higgs boson in proton collisions at 7 and 8 TeV", *Phys. Rev. D*
92 (Jul, 2015) 012004, `doi:10.1103/PhysRevD.92.012004`.

[61] CMS Collaboration, "Study of the Mass and Spin-Parity of the Higgs Boson
Candidate via Its Decays to Z Boson Pairs", *Phys. Rev. Lett.* **110** (Feb, 2013)
081803, `doi:10.1103/PhysRevLett.110.081803`.

[62] L. Lista, "Hypothesis Tests", pp. 175–204. Springer International Publishing,
Cham, 2017. `doi:10.1007/978-3-319-62840-0_9`.

[63] CMS Collaboration, "Measurements of the Higgs boson width and anomalous
HVV couplings from on-shell and off-shell production in the four-lepton final

state", *Phys. Rev. D* **99** (2019), no. 11, 112003,
`doi:10.1103/PhysRevD.99.112003`, `arXiv:1901.00174`.

[64] CMS Collaboration, "Constraints on anomalous *HVV* couplings from the production of Higgs bosons decaying to τ lepton pairs", *Phys. Rev. D* **100** (2019), no. 11, 112002, `doi:10.1103/PhysRevD.100.112002`, `arXiv:1903.06973`.

[65] CMS Collaboration, "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state", `arXiv:2104.12152`.

[66] V. D. Barger et al., "Higgs bosons: Intermediate mass range at e+ e- colliders", *Phys. Rev. D* **49** (1994) 79–90, `doi:10.1103/PhysRevD.49.79`, `arXiv:hep-ph/9306270`.

[67] K. Desch, A. Imhof, Z. Was, and M. Worek, "Probing the CP nature of the Higgs boson at linear colliders with τ spin correlations; the case of mixed scalar–pseudoscalar couplings", *Physics Letters B* **579** (2004), no. 1, 157–164, `doi:10.1016/j.physletb.2003.10.074`.

[68] M. Davier, L. Duflot, F. Le Diberder, and A. Rouge, "The Optimal method for the measurement of tau polarization", *Phys. Lett. B* **306** (1993) 411–417, `doi:10.1016/0370-2693(93)90101-M`.

[69] J. H. Kuhn and E. Mirkes, "Structure functions in tau decays", *Z. Phys. C* **56** (1992) 661–672, `doi:10.1007/BF01474741`. [Erratum: Z.Phys.C 67, 364 (1995)].

[70] S. Berge, W. Bernreuther, and S. Kirchner, "Determination of the Higgs CP-mixing angle in the tau decay channels at the LHC including the Drell–Yan background", *Eur. Phys. J. C* **74** (2014), no. 11, 3164, `doi:10.1140/epjc/s10052-014-3164-0`, `arXiv:1408.0798`.

[71] D. Jeans and G. W. Wilson, "Measuring the *CP* state of tau lepton pairs from Higgs decay at the ILC", *Phys. Rev. D* **98** (Jul, 2018) 013007, `doi:10.1103/PhysRevD.98.013007`.

[72] Z. Was and M. Worek, "Transverse spin effects in H/A $\to \tau^+\tau^-$: $\tau^\pm \to \nu X^\pm$, Monte Carlo approach", *Acta Phys. Polon. B* **33** (2002) 1875–1884, `arXiv:hep-ph/0202007`.

[73] N. Davidson et al., "Universal interface of TAUOLA: Technical and physics documentation", *Computer Physics Communications* **183** (2012), no. 3, 821–843, `doi:10.1016/j.cpc.2011.12.009`.

[74] G. R. Bower, T. Pierzchala, Z. Was, and M. Worek, "Measuring the Higgs boson's parity using tau → rho nu", *Phys. Lett. B* **543** (2002) 227–234, `doi:10.1016/S0370-2693(02)02445-0`, `arXiv:hep-ph/0204292`.

[75] S. M. Berman and M. Jacob, "Systematics of Angular and Polarization Distributions in Three-Body Decays", *Phys. Rev.* **139** (Aug, 1965) B1023–B1038, `doi:10.1103/PhysRev.139.B1023`.

[76] Z. Was and J. Zaremba, "Study of variants for Monte Carlo generators of $\tau \to 3\pi\nu$ decays", *Eur. Phys. J. C* **75** (2015), no. 11, 566, `doi:10.1140/epjc/s10052-015-3780-3`, `arXiv:1508.06424`. [Erratum: Eur.Phys.J.C 76, 465 (2016)].

[77] Z. Czyczula, T. Przedzinski, and Z. Was, "TauSpinner Program for Studies on Spin Effect in tau Production at the LHC", *Eur. Phys. J. C* **72** (2012) 1988, `doi:10.1140/epjc/s10052-012-1988-z`, `arXiv:1201.0117`.

[78] W. Bernreuther, O. Nachtmann, and P. Overmann, "The CP violating electric and weak dipole moments of the tau lepton from threshold to 500-GeV", *Phys. Rev. D* **48** (1993) 78–88, `doi:10.1103/PhysRevD.48.78`.

[79] E. Wilson and J. Gibbs, "Vector Analysis: A Text-book for the Use of Students of Mathematics & Physics: Founded Upon the Lectures of J. W. Gibbs". Yale bicentennial pub. Scribner, 1901.

[80] R. Józefowicz, E. Richter-Was, and Z. Was, "Potential for optimizing the Higgs boson CP measurement in $H \to \tau\tau$ decays at the LHC including machine learning techniques", *Phys. Rev. D* **94** (2016), no. 9, 093001, `doi:10.1103/PhysRevD.94.093001`, `arXiv:1608.02609`.

[81] S. Berge and W. Bernreuther, "Determining the CP parity of Higgs bosons at the LHC in the tau to 1-prong decay channels", *Phys. Lett. B* **671** (2009) 470–476, `doi:10.1016/j.physletb.2008.12.065`, `arXiv:0812.1910`.

[82] V. Cherepanov and A. Zotz, "Kinematic reconstruction of $Z/H \to \tau\tau$ decay in proton-proton collisions", `arXiv:1805.06988`.

[83] S. Jadach, J. H. Kühn, and Z. Was, "TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons", *Computer Physics Communications* **64** (1991), no. 2, 275–299, `doi:10.1016/0010-4655(91)90038-M`.

[84] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[85] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (aug, 2008) S08001–S08001, `doi:10.1088/1748-0221/3/08/s08001`.

[86] S. Myers and E. Picasso, "The Design, construction and commissioning of the CERN Large Electron Positron collider", *Contemp. Phys.* **31** (1990) 387–403, `doi:10.1080/00107519008213789`.

[87] R. Allen, "The Higgs Bridge", *Physica Scripta* **89** (11, 2013) `doi:10.1088/0031-8949/89/01/018001`.

[88] "New Linac + 'old' Booster = many protons. Nouveau Linac + «Booster» = multiplication des protons", *CERN Bulletin* (1978) 1–2.

[89] K. H. Reich, "The CERN Proton Synchrotron Booster", *IEEE Trans. Nucl. Sci.* **16** (1969) 959–961, `doi:10.1109/TNS.1969.4325414`.

[90] D. Cundy and S. Gilardoni, "The Proton Synchrotron (PS): At the Core of the CERN Accelerators", *Adv. Ser. Direct. High Energy Phys.* **27** (2017) 39–85. 47 p, `doi:10.1142/9789814749145_0003`.

[91] J. B. Adams, "The CERN 400 GeV Proton Synchrotron (CERN SPS)", in *10th International Conference on High-Energy Accelerators*. 1977.

[92] F. Marcastel, "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN", (Oct, 2013). General Photo.

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

[94] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, `doi:10.1088/1748-0221/3/08/S08005`.

[95] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, `doi:10.1088/1748-0221/3/08/S08002`.

[96] TOTEM Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007, `doi:10.1088/1748-0221/3/08/S08007`.

[97] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006, `doi:10.1088/1748-0221/3/08/S08006`.

[98] J. L. Pinfold, "The MoEDAL Experiment at the LHC", in *13th ICATPP Conference on Astroparticle, Particle, Space Physics and Detectors for Physics Applications*. 2012. `doi:10.1142/9789814405072_0076`.

[99] J.-L. Caron, "Cross section of LHC dipole. Dipole LHC: coupe transversale.", (May, 1998). AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[100] "2010–2013: the LHC's first long run", *CERN Courier* **53** (Sep, 2013) 25–28.

[101] LHC Press Release, "CERN sets date for first attempt at 7 TeV collisions in the LHC".

[102] A. V. Praag, "During the event of LHC first physics on March 30th 2010, physicists from both ATLAS and CMS collaboration gather in the atrium of building 40 at CERN to celebrate.", (Mar, 2010).

[103] HL-LHC official website.
`https://hilumilhc.web.cern.ch/content/hl-lhc-project`.

[104] R. Bruce et al., "Review of LHC Run 2 Machine Configurations", in *9th LHC Operations Evian Workshop*, pp. 187–197. Geneva, Switzerland, 2019.

[105] CMS Public Luminosity Results twiki. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots`.

[106] CMS Collaboration and T. Mc Cauley, "Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill", (Nov, 2016). CMS Collection.

[107] CMS Collaboration, "Technical Proposal for the Phase-II Upgrade of the CMS Detector", technical report, Geneva, Jun, 2015. Upgrade Project Leader Deputies: Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota).

[108] CMS Collaboration, CMS Warsaw, K. Pozniak, G. Kasprowicz, R. Romaniuk and W. Zabolotny, "A New Boson with a Mass of 125 GeV Observed with the CMS Experiment at the Large Hadron Collider", *Science* **338** (12, 2012) 1569–1575.

[109] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade", Technical Report CERN-LHCC-2012-016. CMS-TDR-11, Sep, 2012.

[110] CMS Public Tracker Material Budget twiki. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots`.

[111] CMS Public Tracker Detector Performance Results twiki.
`https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK`.

[112] CMS Collaboration, "Overview of the CMS Detector Performance at LHC Run 2", *Universe* **5** (2019), no. 1, 18, `doi:10.3390/universe5010018`.

[113] L. Caminada, "Performance and operation of the CMS Phase 1 pixel detector", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **936** (2019) 688–690, `doi:10.1016/j.nima.2018.10.084`. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[114] P. L. Connor, "Inclusive b Jet Production in Proton-Proton Collisions". Springer Theses. Springer, 2019.

[115] V. Blobel and C. Kleinwort, "A New method for the high precision alignment of track detectors", in *Conference on Advanced Statistical Techniques in Particle Physics*. June, 2002. `arXiv:hep-ex/0208021`.

[116] V. Karimaki, T. Lampen, and F. Schilling, "The HIP algorithm for track based alignment and its application to the CMS pixel detector", Technical Report CERN-CMS-NOTE-2006-018, Jan, 2006.

[117] E. Auffray, "Overview of the 63000 PWO Barrel Crystals for CMS_ECAL Production", *IEEE Trans. Nucl. Sci.* **55** (2008), no. 3, 1314–1320, `doi:10.1109/TNS.2007.913935`.

[118] CMS Collaboration, "The CMS ECAL performance with examples", *Journal of Instrumentation* **9** (feb, 2014) C02008–C02008, `doi:10.1088/1748-0221/9/02/c02008`.

[119] F. Cavallari, "Performance of calorimeters at the LHC", *Journal of Physics: Conference Series* **293** (apr, 2011) 012001, `doi:10.1088/1742-6596/293/1/012001`.

[120] CMS Collaboration, "Performance of CMS hadron calorimeter timing and synchronization using test beam, cosmic ray, and LHC beam data. Performance of CMS Hadron Calorimeter Timing and Synchronization using Cosmic Ray and LHC Beam Data", *JINST* **5** (Nov, 2009) T03013. 31 p, `doi:10.1088/1748-0221/5/03/T03013`.

[121] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 06, P06015, `doi:10.1088/1748-0221/13/06/P06015`, `arXiv:1804.04528`.

[122] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020), no. 10, P10017, `doi:10.1088/1748-0221/15/10/P10017`, `arXiv:2006.10165`.

[123] L. R. F. Castillo, "The ATLAS and CMS detectors", in *The Search and Discovery of the Higgs Boson*, 2053-2571, pp. 4–1 to 4–8. Morgan & Claypool Publishers, 2015. `doi:10.1088/978-1-6817-4078-2ch4`.

[124] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *Journal of Instrumentation* **12** (Oct, 2017) P10003–P10003, `doi:10.1088/1748-0221/12/10/p10003`.

[125] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (Oct, 2014) P10009–P10009, `doi:10.1088/1748-0221/9/10/p10009`.

[126] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME–Journal of Basic Engineering* **82** (1960), no. Series D, 35–45.

[127] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987), no. 2, 444 – 450, `doi:10.1016/0168-9002(87)90887-4`.

[128] T. Miao, N. Leioatts, H. Wenzel, and F. Yumiceva, "Beam Position Determination using Tracks", Technical Report CMS-NOTE-2007-021, CERN, Geneva, Aug, 2007.

[129] F. Brivio, "Beam spot resolution in CMS during Run2018A". Private communication.

[130] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proceedings of the IEEE* **86** (1998), no. 11, 2210–2239, `doi:10.1109/5.726788`.

[131] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar, 2007.

[132] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *Journal of Instrumentation* **7** (Oct, 2012) P10002–P10002, `doi:10.1088/1748-0221/7/10/p10002`.

[133] A. Sirunyan et al., "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV", *Journal of Instrumentation* **13** (Jun, 2018) P06015–P06015, `doi:10.1088/1748-0221/13/06/p06015`.

[134] CMS Collaboration, "Muon identification and isolation efficiencies with 2017 and 2018 data", Technical Report CMS-DP-2018-042, CERN-CMS-DP-2018-042, Jul, 2018.

[135] Kyeongpil Lee on behalf of the CMS Collaboration, "Muon performance with the CMS detector in Run2 of LHC", *PoS* **ICHEP2018** (2019) 690, `doi:10.22323/1.340.0690`.

[136] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", `arXiv:2012.06888`.

[137] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (jul, 2005) N9–N20, `doi:10.1088/0954-3899/31/9/n01`.

[138] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-ktjet clustering algorithm", *Journal of High Energy Physics* **2008** (apr, 2008) 063–063, `doi:10.1088/1126-6708/2008/04/063`.

[139] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[140] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *Journal of Instrumentation* **13** (May, 2018) P05011–P05011, `doi:10.1088/1748-0221/13/05/p05011`.

[141] D. Guest et al., "Jet flavor classification in high-energy physics with deep neural networks", *Phys. Rev. D* **94** (Dec, 2016) 112002, `doi:10.1103/PhysRevD.94.112002`.

[142] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *Journal of Instrumentation* **14** (jul, 2019) P07004–P07004, `doi:10.1088/1748-0221/14/07/p07004`.

[143] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", *Journal of High Energy Physics* **2014** (Oct, 2014) 59, `doi:10.1007/JHEP10(2014)059`.

[144] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *Journal of Instrumentation* **15** (sep, 2020) P09018–P09018, `doi:10.1088/1748-0221/15/09/p09018`.

[145] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *Journal of Instrumentation* **12** (feb, 2017) P02014–P02014, `doi:10.1088/1748-0221/12/02/p02014`.

[146] CMS Collaboration, "Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2", Technical Report CMS-PAS-TAU-16-002, CERN, Geneva, 2016.

[147] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and $\nu\tau$ in pp collisions at $\sqrt{s}$=13 TeV", *Journal of Instrumentation* **13** (Oct, 2018) P10005–P10005, `doi:10.1088/1748-0221/13/10/p10005`.

[148] I. Neutelings, "Extension to the Tikz latex package". Private communication.

[149] CMS Collaboration, "Tau Identification Performance in 2017 Data at $\sqrt{s} = 13$ TeV", Technical Report CMS-DP-2018-026, Jun, 2018.

[150] CMS Collaboration, "Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons", Technical Report CMS-DP-2019-033, Oct, 2019.

[151] Andrea Cardini on behalf of the CMS Collaboration, "Tau identification exploiting deep learning techniques", in *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)*, volume 390, p. 723. 2021. `doi:10.22323/1.390.0723`.

[152] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks", *CoRR* **abs/1511.08458** (2015) `arXiv:1511.08458`.

[153] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016. `http://www.deeplearningbook.org`.

[154] T. Dozat, "Incorporating Nesterov Momentum into Adam", 2016.

[155] Combine tool website. `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit`.

[156] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011.

[157] F. James and M. Roos, "MINUIT: a system for function minimization and analysis of the parameter errors and corrections", *Comput. Phys. Commun.* **10** (Jul, 1975) 343–367. 38 p.

[158] CMS Collaboration, "Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-HIG-20-006, CERN, Geneva, 2020.

[159] CMS Collaboration, "Measurement of Higgs boson production in the decay channel with a pair of τ leptons", Technical Report CMS-PAS-HIG-19-010, CERN, Geneva, 2020.

[160] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, "Reconstruction of the Higgs mass in H$\to \tau\tau$ Events by Dynamical Likelihood techniques", *Journal of Physics: Conference Series* **513** (jun, 2014) 022035, `doi:10.1088/1742-6596/513/2/022035`.

[161] K. Kondo, "Dynamical likelihood method for reconstruction of events with missing momentum. I. Method and toy models", *Journal of the Physical Society of Japan* **57** (1988), no. 12, 4126–4140.

[162] K. Kondo, "Dynamical likelihood method for reconstruction of events with missing momentum. II. Mass spectra for $2\to 2$ processes", *Journal of the Physical Society of Japan* **60** (1991), no. 3, 836–844.

[163] H. A. Bethe, "Molière's Theory of Multiple Scattering", *Phys. Rev.* **89** (Mar, 1953) 1256–1266, `doi:10.1103/PhysRev.89.1256`.

[164] V. Cherepanov and W. Lohmann, "Methods for a measurement of τ polarization asymmetry in the decay $Z \to \tau\tau$ at LHC and determination of the effective weak mixing angle", `arXiv:1805.10552`.

[165] CMS Collaboration, "Identification of hadronic tau decay channels using multivariate analysis (MVA decay mode)", Technical Report CMS-DP-2020-041, Jul, 2020.

[166] T. Chen and C. Guestrin, "XGBoost", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Aug, 2016) `doi:10.1145/2939672.2939785`.

[167] C. Oleari, "The POWHEG-BOX", *Nucl. Phys. B Proc. Suppl.* **205-206** (2010) 36–41, `doi:10.1016/j.nuclphysbps.2010.08.016`, `arXiv:1007.3893`.

[168] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, `doi:10.1016/j.cpc.2008.01.036`, `arXiv:0710.3820`.

[169] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, `doi:10.1007/JHEP06(2011)128`, `arXiv:1106.0522`.

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

[171] A. Schälicke et al., "An event generator for particle production in high-energy collisions", *Progress in Particle and Nuclear Physics* **53** (2004), no. 1, 329–338, `doi:doi.org/10.1016/j.ppnp.2004.02.031`. Heavy Ion Reaction from Nuclear to Quark Matter.

[172] T. Przedziński, E. Richter-Was, and Z. Was, "TauSpinner: a tool for simulating CP effects in $H \to \tau\tau$ decays at LHC", *The European Physical Journal C* **74** (Nov, 2014) 3177, `doi:10.1140/epjc/s10052-014-3177-8`.

[173] CMS Collaboration, "Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques", *Eur. Phys. J. C* **78** (2018), no. 9, 708, `doi:10.1140/epjc/s10052-018-6146-9`, `arXiv:1801.03535`.

[174] CMS Collaboration, "Measurement of Higgs boson production and decay to the $\tau\tau$ final state", Technical Report CMS-PAS-HIG-18-032, 2019.

[175] CMS Collaboration, "An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data", *JINST* **14** (2019), no. 06, P06032, `doi:10.1088/1748-0221/14/06/P06032`, `arXiv:1903.01216`.

[176] CMS Collaboration, "Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **09** (2018) 007, `doi:10.1007/JHEP09(2018)007`, `arXiv:1803.06553`.

[177] CMS Public Cross Section Measurements twiki. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined`.

[178] CMS Collaboration, "Measurement of differential cross sections for Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **78** (2018), no. 11, 965, `doi:10.1140/epjc/s10052-018-6373-0`, `arXiv:1804.05252`.

[179] Mareike Meyer on behalf of the CMS collaboration, "CMS H(125) fermionic decay results". Talk presented at "Higgs Hunting 2019", Orsay-Paris, on 29/07/2019.

[180] N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research* **15** (2014), no. 56, 1929–1958.

[181] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", `arXiv:1412.6980`.

[182] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, `doi:10.1140/epjc/s10052-011-1554-0`, `arXiv:1007.1727`. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[183] R. Cousins, "Generalization of Chisquare Goodness-of Fit Test for Binned Data Using Saturated Models , with Application to Histograms", 2013.

[184] W. T. Eadie et al., "Statistical methods in experimental physics; 1st ed.". North-Holland, Amsterdam, 1971.

[185] Daniel Winterbottom on behalf of the CMS Collaboration, "Higgs boson measurements in final states with taus at CMS", in *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)*, volume 390, p. 723. 2021. `doi:10.22323/1.390.0090`.

[186] S. F. King, M. Muhlleitner, R. Nevzorov, and K. Walz, "Exploring the CP-violating NMSSM: EDM Constraints and Phenomenology", *Nucl. Phys. B* **901** (2015) 526–555, `doi:10.1016/j.nuclphysb.2015.11.003`, `arXiv:1508.03255`.

[187] S.-F. Ge, G. Li, P. Pasquini, and M. J. Ramsey-Musolf, "CP-violating Higgs Di-tau Decays: Baryogenesis and Higgs Factories", `arXiv:2012.13922`.

[188] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and antimatter in the universe", *New Journal of Physics* **14** (sep, 2012) 095012, `doi:10.1088/1367-2630/14/9/095012`.

[189] A. D. Sakharov, "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35, `doi:10.1070/PU1991v034n05ABEH002497`.

[190] ATLAS Collaboration, "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the

ATLAS detector", *Phys. Rev. D* **99** (2019) 072001,
`doi:10.1103/PhysRevD.99.072001`, `arXiv:1811.08856`.

[191] CMS Collaboration, "Projected Performance of an Upgraded CMS Detector at the
LHC and HL-LHC: Contribution to the Snowmass Process", in *Community
Summer Study 2013: Snowmass on the Mississippi*. Jul, 2013. `arXiv:1307.7135`.

[192] "Projections for measurements of Higgs boson signal strengths and coupling
parameters with the ATLAS detector at a HL-LHC", Technical Report
ATL-PHYS-PUB-2014-016, Oct, 2014.

[193] X. Chen and Y. Wu, "Probing the CP-Violation effects in the $h\tau\tau$ coupling at the
LHC", *Phys. Lett. B* **790** (2019) 332–338,
`doi:10.1016/j.physletb.2019.01.038`, `arXiv:1708.02882`.

[194] CEPC Study Group Collaboration, "CEPC Conceptual Design Report: Volume 2
- Physics & Detector", `arXiv:1811.10545`.

[195] FCC Collaboration, "FCC-ee: The Lepton Collider: Future Circular Collider
Conceptual Design Report Volume 2", *Eur. Phys. J. ST* **228** (2019), no. 2,
261–623, `doi:10.1140/epjst/e2019-900045-4`.

[196] "The International Linear Collider Technical Design Report - Volume 2: Physics",
`arXiv:1306.6352`.

[197] R. P. Feynman, "What Is Science", *The Physics Teacher* **7** (1969), no. 6,
313–320, `doi:10.1119/1.2351388`.

[198] CMS Collaboration, "Sensitivity projections for Higgs boson properties
measurements at the HL-LHC", Technical Report CMS-PAS-FTR-18-011, 2018.

[199] ATLAS Collaboration, "Evidence for Higgs boson decays to a low-mass dilepton
system and a photon in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector",
`arXiv:2103.10322`.

[200] J. Steggeman, E. Asilar, A. Cardini, and F. Zhang, "Tau short exercise for
CMSvDAS 2020". Private communication.

Acknowledgments