Analysis of Higgs bosons utilizing hadronic tau decays at the CMS experiment

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der RWTH Aachen University zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Physiker

Bastian Kargoll

aus Werne

Berichter: Univ.-Prof. Dr. rer. nat. Achim Stahl
apl.-Prof. Dr. Wolfgang Lohmann

Tag der mündlichen Prüfung: 21.11.2016

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Abstract

For many years the Higgs boson was the last undiscovered piece of the successful standard model of particle physics. It was discovered in 2012 by the ATLAS and CMS collaborations in decays to $\gamma\gamma$, ZZ, and WW. This thesis describes a search for the Higgs boson in its decay into two τ leptons. This decay is of particular interest to probe the Yukawa coupling of the Higgs boson to fermions. CMS data of the year 2012 are used, corresponding to an integrated luminosity of $19.7\,\mathrm{fb}^{-1}$ of proton-proton collisions collected at a center-of-mass energy of 8 TeV. Triggers have been developed, implemented, and maintained to collect events with two hadronically decaying tau leptons in the final state. Their performance is described. Additionally, this thesis investigates Higgs boson events with a muon and a hadronic tau decay in the final state. If the tau lepton decays into three charged hadrons and a neutrino, the reconstruction of the tau decay vertex is used to suppress background and to reconstruct the kinematic properties of the event. Prospects of such methods in the context of Higgs analyses are discussed.

Das Higgs-Boson war für viele Jahre der letzte unentdeckte Baustein des erfolgreichen Standardmodells der Teilchenphysik. Im Jahr 2012 wurde es von den ATLAS und CMS Kollaborationen in Zerfällen in $\gamma\gamma$, ZZ sowie WW entdeckt. Die vorliegende Arbeit beschreibt eine Suche nach dem Higgs-Boson in seinem Zerfall in zwei Tau-Leptonen. Dieser Zerfallskanal erlaubt die Untersuchung der Yukawa-Kopplung des Higgs-Bosons an Fermionen. Dazu werden CMS-Daten des Jahres 2012 verwendet, entsprechend einer integrierten Luminosität von $19,7\,\mathrm{fb}^{-1}$ von Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 8 TeV. Trigger zur Aufzeichnung von Ereignissen mit zwei hadronisch zerfallenden Tau-Leptonen wurden entwickelt, implementiert und gewartet. Verschiedene Parameter zur Beurteilung ihrer Leistungsstärke werden beschrieben. Zusätzlich werden in dieser Arbeit Higgs-Boson-Ereignisse mit einem Myon und einem hadronisch zerfallenden Tau im Endzustand untersucht. Sofern das Tau-Lepton in drei geladene Hadronen und ein Neutrino zerfällt, wird der Tau-Zerfallsvertex zur Unterdrückung von Untergründen sowie zur Rekonstruktion der Ereignis-Kinematik verwendet. Die Erfolgsaussichten derartiger Methoden im Kontext von Higgs-Analysen werden diskutiert.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The science of Particle Physics is probing some of the most extreme environments that are in reach today. Trying to extend our knowledge about the fundamental principles of nature, it is probing the smallest known objects, called particles. Using the highest energy densities (thus its alternative name high energy physics), these smallest objects are produced and then studied by the largest machines ever built for science. Thousands of people are collaborating to enhance our understanding of this smallest scale.

Over the last decades, the standard model of particle physics (SM) has proven to be an accurate description of elementary particles. It describes the smallest matter constituents as fermions with spin $1/2$, which are further separated into quarks and leptons. While the former are taking part in the strong interaction, and thus build the nuclei of atoms, the latter are mainly interacting via the electroweak interaction. These interactions, also named forces, are mediated by gauge bosons with spin 1, which build the second big group of particles within the standard model: four electroweak bosons and eight strong bosons. All of these particles have been proven to exist, and their properties studied in detail.

The last missing piece of the standard model was the so-called Higgs boson. This particle was predicted by Peter Higgs in 1964 [1]. The search for the Higgs boson was one of the main reasons to build the Large Hadron Collider (LHC) and its two general purpose detectors, ATLAS and the Compact Muon Solenoid (CMS). After decades of planning, construction and commissioning of these huge machines, both ATLAS and CMS collaborations announced the discovery of a new particle on July 4, 2012 [2–4]. Although the properties of this object that could already be measured show good agreement with the SM prediction, the final determination of its true nature will need much more data and even more scrutinized analyses.

This thesis will describe the analysis of the Higgs boson at the CMS experiment in its decay into two τ leptons. This specific channel is experimentally challenging, but offers access to important characteristics of the boson. A direct coupling of the Higgs boson to τ leptons would give insight in the mass generation of fermions, which is different than for bosons. Additionally, the subsequent decay of the τ leptons allows access to the spin and CP properties of the decaying resonance.

This document consists of four major parts. Chapter 1 gives an overview of the theoretical and experimental foundation of this work. Event selection that happens simultaneously to the collection of the data is called triggering. Triggers with two τ leptons in the final state thus are the first step in the search for $\mathrm{H} \to \tau\tau$ decays and are discussed in Chapter 2. A search for $\mathrm{H} \to \tau\tau$ decays in the CMS data is described in Chapter 3, with a focus on the $\mu\tau_\mathrm{h}$ channel. The fourth Chapter will then present ideas on how the analysis strategy could be improved.

1.1. Theoretical background

1.1.1. Standard model of particle physics

The standard model (SM) of particle physics describes the fundamental constituents of matter as elementary particles as well as their interactions. It allows to describe the electromagnetic, weak, and strong forces with a single Lagrangian. The SM is often parametrized into two parts: The electroweak sector offers a description of the unification of quantum electrodynamics (QED) and the weak force, while quantum chromodynamics (QCD) models the strong force as interaction between colored objects.

Fermions with a spin of $1/2$ can carry charges and are thereby categorized into leptons and quarks. The gauge bosons with spin 1 are the quanta of the corresponding force fields and transmit the forces between particles. The Higgs boson has spin 0 and belongs to neither of these groups. It is the excitation of the Higgs field, which will be discussed in more detail in Section 1.1.3. Table 1.1 and Table 1.2 give an overview of the elementary particles in the standard model.

The SM is a renormalized gauge field theory, represented by a gauge group of the form $SU(3)_C \times SU(2)_L \times U(1)_Y$. The corresponding charges are named color, weak isospin, and weak hypercharge, respectively. Particle masses are not part of the Lagrangian at this stage. They are introduced by spontaneous symmetry breaking (SSB) of the electroweak gauge group $SU(2)_L \times U(1)_Y$ as described by the Englert-Brout-Higgs (EBH) mechanism [1, 5–9]. The $SU(2)_L$ and $U(1)_Y$ groups mix and give rise to the photon γ, the Z, and the $W^{\pm}$ bosons as gauge bosons of the electromagnetic and weak force, respectively. The remaining $SU(3)_C$ gauge group describes the QCD as force coupling to three charges (named colors), carried by eight gluons g.

The first particle discovered was the electron in 1897 [10], followed by the photon in 1923 [11]. In the 1960s, the theory of electroweak interactions unified the electromagnetic and the weak force [12–14]. The theoretical description of the SM can be considered complete since the strong interaction was incorporated into the same formalism in the early 1970s [15–19]. The experimental proof took much longer: Only since 2012 all predicted elementary particles have been discovered. Figure 1.1 shows that the Higgs boson was the toughest challenge for the experimentalists and was proven to exist, about 40 years after its proclamation, by the ATLAS and CMS experiments [2–4].

Although the standard model is well tested experimentally, it cannot answer all questions of fundamental physics. Some observations, like the non-zero masses of neutrinos [20], are not described in the SM. Cosmological data show that the energy content of the universe is dominated by Dark Matter and Dark Energy [21], which require further extensions of the theory. Nonetheless, up to today no fundamental object beyond the scope of the standard model could be discovered.

1.1.2. QCD in proton collisions

QCD plays a vital role in the hadron collisions at the LHC. Protons are no fundamental particles. Their substructure in first order consists of three *valence quarks*, namely two u and one d quark. These are bound by the strong force and define the proton quantum numbers.

A closer investigation of the substructure reveals a more complex composition. Additionally to the valence quarks, the proton contains gluons and *sea quarks*. The former mediate

Table 1.1: The elementary fermions of the standard model and their respective charges. Leptons and quarks are ordered in three families, which differ only in mass, and appear in left-handed doublets (L) and right-handed singlets (R) of the weak isospin T. In this table the eigenstates of the weak interaction are listed, which are not identical to the mass eigenstates. The weak hypercharge Y is related to the electrical charge Q and T_z by the relation $Q = T_z + \frac{1}{2}Y$. The color charge is denoted c, where c is either red, green, or blue.

Quarks						
1	2	3	$Q[e]$	T_z	Y	color
$\begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix}_L$	$\begin{pmatrix} \mathrm{c} \\ \mathrm{s} \end{pmatrix}_L$	$\begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix}_L$	$+2/3$ $-1/3$	$+1/2$ $-1/2$	$+1/3$ $+1/3$	c c
$\mathrm{u_R}$	$\mathrm{c_R}$	$\mathrm{t_R}$	$+2/3$	0	$+4/3$	c
$\mathrm{d_R}$	$\mathrm{s_R}$	$\mathrm{b_R}$	$-1/3$	0	$-2/3$	c
Leptons						
1	2	3	$Q[e]$	T_z	Y	color
$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	0 -1	$+1/2$ $-1/2$	-1 -1	– –
e_R^-	μ_R^-	τ_R^-	-1	0	-2	–

Table 1.2: The elementary bosons of the standard model after spontaneous symmetry breaking. Gauge bosons with spin 1 mediate the fundamental forces, while the Higgs boson with spin 0 is connected to the generation of particle masses. There are eight gluons g in the SM, corresponding to eight independent linear combinations of red, green, and blue.

Gauge Bosons	m[GeV]	$Q[e]$	color	mediated force
γ	0	0	–	electromagnetic
W	80.4	± 1	–	weak
Z	91.2	0	–	weak
g	0	0	$c\bar{c}$	strong
Higgs Boson	m[GeV]	$Q[e]$	color	
H	125.1	0	–	

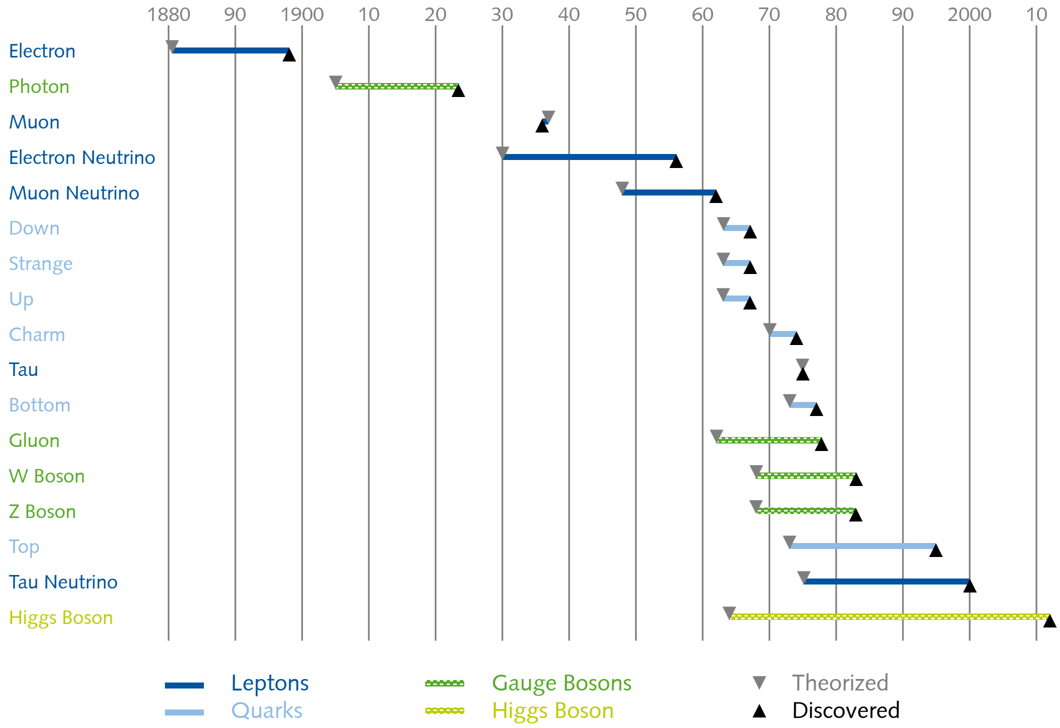


Figure 1.1: The elementary particles of the standard model and their years of theoretical description and experimental discovery.

the strong force and thus connect the individual constituents. Sea quarks can be described as virtual quark-antiquark pairs occurring within the proton. The term *parton* denotes an individual constituent of the proton, i.e. a gluon or quark. Within the proton, all partons are *asymptotically free*, meaning that they can be described as free particles as long as they remain inside of the proton radius.

At high energies, as they appear at the LHC, a collision of protons must be described on parton level. The initial state of hard interactions is therefore qq', qg, or gg. While the center-of-mass energy of the proton pair is denoted $\sqrt{s}$, the corresponding value of the colliding partons is $\sqrt{\hat{s}}$. Its value differs for each collision, as it depends on the momenta of the colliding partons. In contrast to lepton-lepton colliders, the center-of-mass frame at LHC is unknown a priori.

The momentum of a given parton depends on its type. It is described by the parton distribution function (PDF). For each type of parton, the PDF describes the probability density as a function of the momentum fraction x in relation to the full proton momentum and the energy scale Q^2. Figure 1.2 illustrates the PDFs for valence quarks, gluons, and the sum of all sea quarks. As expected, each of the three valence quarks is most likely to carry one third of the proton momentum. A factor of two between the probabilities of u and d valence quarks corresponds to the defining structure of the proton. At lower momentum fractions x, the valence quarks become rare and sea quarks and gluons dominate. For SM processes with typical energies $\sqrt{\hat{s}} = \mathcal{O}(100\,\text{GeV})$, gluons dominate at the LHC.

The term *QCD interaction* is used to emphasize that the initial state at the LHC always consists of strongly interacting partons. This has consequences for all hard scattering processes, including resonant production of weak or Higgs bosons.

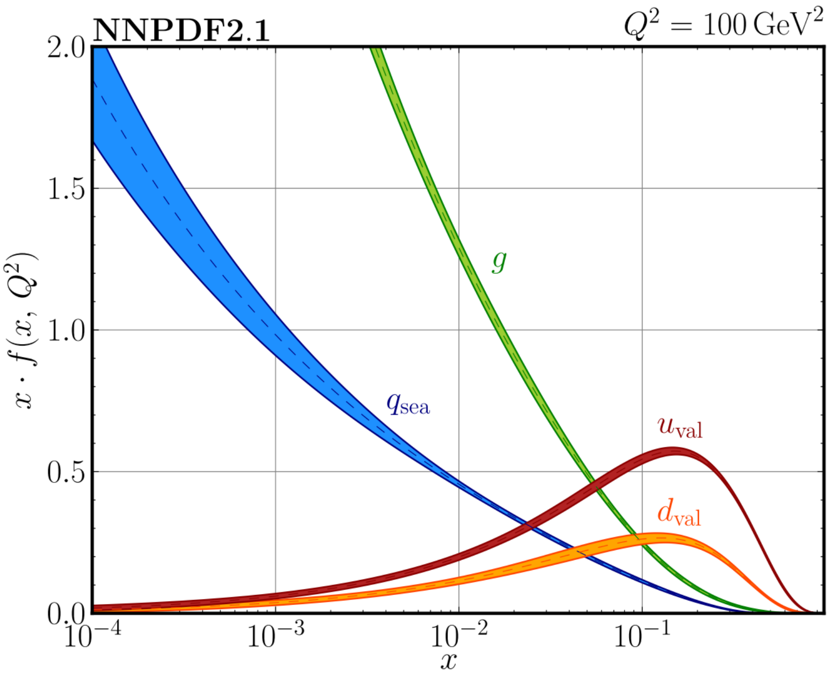


Figure 1.2: Proton PDF at $Q^2 = 100\,\mathrm{GeV}^2$ as a function of the momentum fraction x. Shown are the PDFs of the valence quarks $\mathrm{u_{val}}$ and $\mathrm{d_{val}}$, gluons g, and the sum of the sea quarks q_{sea}. The figure was adopted from Ref. [22] and visualizes the PDF set provided by the NNPDF collaboration [23].

1.1.3. The EBH mechanism and the Higgs boson

1.1.3.1. Particle masses in the standard model

In order to keep the standard model renormalisable, its Lagrangian in its unbroken form may not contain particle masses. Mass terms for both fermions and bosons violate the fundamental principle of gauge invariance. Contrary to this requirement, massive particles are observed in nature. Additionally, massive vector bosons would lead to large scattering amplitudes and violate the unitary requirement.

The EBH mechanism introduces a fundamental field ϕ, called Higgs field. This complex scalar $SU(2)$ doublet reads

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_1 + i\phi_2 \end{pmatrix} , \tag{1.1}$$

with real components ϕ_i, $i = 1 \ldots 4$. The simplest choice for the potential $V(\phi)$ that solves the aforementioned problems is of the form

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 , \tag{1.2}$$

with a real, positive parameter λ, and $\mu^2 = -\left|\mu^2\right| < 0$. This potential has a shape as depicted in Figure 1.3. Its nontrivial minima ϕ_{min} are given by

$$\frac{\mathrm{d}V}{\mathrm{d}\phi^\dagger\phi} = \mu^2 + 2\lambda\phi^\dagger\phi = 0$$

$$\Rightarrow |\phi_{min}| = \sqrt{\frac{-\mu^2}{2\lambda}} . \tag{1.3}$$

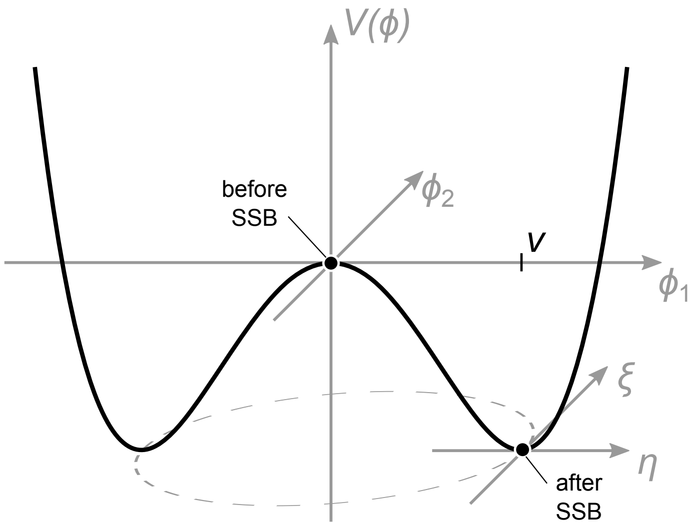


Figure 1.3: Higgs potential $V(\phi)$. Only two of the four real field components ϕ_i are shown, the components ϕ_3 and ϕ_4 are omitted. A rotational symmetry is implied. After spontaneous symmetry breaking (SSB), an arbitrary solution is chosen out of a set of degenerated minima located at the vacuum expectation value v (a normalization factor $\sqrt{2}$ is omitted for simplicity). An expansion around this point is performed, introducing new coordinates ξ in direction of the phase symmetry and η perpendicular to it.

As the vacuum is defined by having minimal energy, the concrete vacuum solution is given by an arbitrary but fixed point on a circle with radius ϕ_{min}. The gauge symmetry remains preserved by the potential, but the chosen vacuum solution is not symmetric. It can be shown that this spontaneous symmetry breaking gives a renormalizable theory [24].

One can arbitrarily choose

$$\begin{aligned} \phi_2 = \phi_3 = \phi_4 = 0 \\ \phi_1 = \phi_{min} \end{aligned} \tag{1.4}$$

to be the vacuum solution. New coordinates are introduced, η pointing in radial direction and ξ along the phase symmetry direction. This allows to express a fluctuation around the minimum as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \eta + i\xi \end{pmatrix} , \tag{1.5}$$

introducing the vacuum expectation value

$$v = \sqrt{2}\phi_{min} = \sqrt{\frac{-\mu^2}{\lambda}} . \tag{1.6}$$

To analyze the system, we will begin by ignoring the ξ direction for now, as the potential is flat along the rotational symmetry. The potential defined in Equation (1.2) can be expressed using the new coordinate η and Equation (1.6) as

$$\begin{aligned} V(\phi) &= \frac{\mu^2}{2} (v+\eta)^2 + \frac{\lambda}{4} (v+\eta)^4 \\ &= -\frac{\lambda}{4} v^4 + \lambda v^2 \eta^2 + \lambda v \eta^3 + \frac{\lambda}{4} \eta^4 . \end{aligned} \tag{1.7}$$

The constant term $-\frac{\lambda}{4}v^4$ gives the non-zero energy value of the vacuum. More importantly, $\lambda v^2\eta^2$ can be identified as a mass term in the Lagrangian, which is of the shape $\frac{1}{2}m_\eta\eta^2$. Thus, we have introduced a new scalar particle, which is identified as the Higgs boson and whose mass is given by

$$m_{\mathrm{H}} = \sqrt{2\lambda}v \,. \tag{1.8}$$

The scalar field η defined in Equation (1.5) corresponds to the field of the Higgs boson. Higher order terms in Equation (1.7) describe trilinear and quartic self-couplings of the Higgs boson. In the gauge of the SM the vacuum expectation value is given as

$$v = (\sqrt{2}G_F)^{-1/2} \approx 246\,\mathrm{GeV}\,, \tag{1.9}$$

where G_F is the well-measured Fermi constant. The values of λ and μ, and therefore also the Higgs mass m_H, remain free parameters.

According to the Goldstone theorem, spontaneous symmetry breaking always gives rise to massless Nambu-Goldstone bosons [25, 26], which correspond to a fluctuation of the field in direction of the symmetry. We denoted one of this directions above as ξ. In total we expect three Nambu-Goldstone bosons, one for each remaining degree of freedom. In the context of a local gauge theory, these bosons couple to the gauge fields and mix with the longitudinal modes of the electroweak gauge bosons. Accordingly, the weak bosons $\mathrm{W}^\pm$ and Z obtain mass, while the fourth boson γ remains massless.

The Higgs field also allows to introduce fermion masses into the SM Lagrangian. For a fermion field ψ the Dirac mass term is

$$m\bar{\psi}\psi = m\left(\bar{\psi}_{\mathrm{L}}\psi_{\mathrm{R}} + \bar{\psi}_{\mathrm{R}}\psi_{\mathrm{L}}\right)\,. \tag{1.10}$$

The standard model describes fermions as left-handed doublets ψ_{L} and right-handed singlets ψ_{R} (see Table 1.1). A Dirac mass term is therefore not allowed in the SM Lagrangian, as it violates local gauge symmetry: It describes an interaction that changes chirality from left-handed to right-handed, or vice versa. The interaction of a fermion field and the scalar Higgs field via gauge invariant Yukawa-type operators produces terms $m\bar{\psi}\psi$ if the vacuum expectation value is not vanishing [13]. As was shown above, this is true after spontaneous symmetry breaking. The masses of the SM fermions remain free parameters.

1.1.3.2. Properties of the Higgs boson

The Higgs boson is different from any other particle of the standard model, as it has zero spin. Its eigenvalue in CP-symmetry operations is $+1$. The Higgs boson does neither carry electric nor color charge and does not partake in the associated interactions. Nonetheless, it carries weak isospin and hence interacts with heavy gauge bosons and fermions.

It is important to note that the coupling strengths of the Higgs boson to the standard model particles are fixed by the theory described in Section 1.1.3.1. They are only dependent on the particle mass, resulting in stronger couplings to heavier objects. This is true for all of its couplings, i.e. trilinear and quartic couplings to weak bosons V, Yukawa couplings to fermions f, and even the trilinear and quartic Higgs self-couplings [27, p. 163],

$$g_{\mathrm{H}f\bar{f}} = \frac{m_f}{v}\,, \qquad g_{\mathrm{H}VV} = \frac{2m_V^2}{v}\,, \qquad g_{\mathrm{HH}VV} = \frac{2m_V^2}{v^2}\,, \tag{1.11}$$

$$g_{\mathrm{HHH}} = \frac{3m_{\mathrm{H}}^2}{v}\,, \qquad g_{\mathrm{HHHH}} = \frac{3m_{\mathrm{H}}^2}{v^2}\,. \tag{1.12}$$

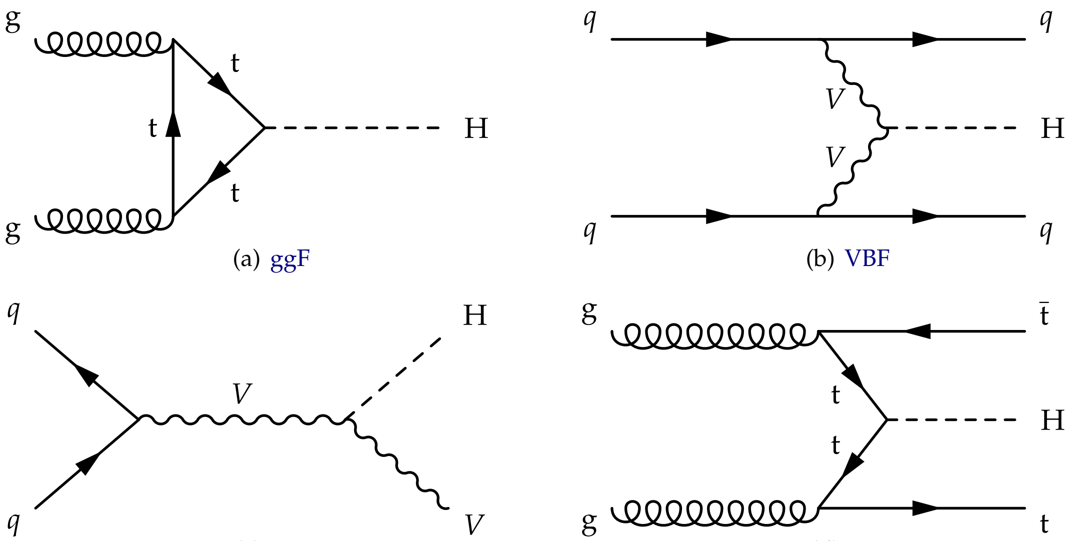


Figure 1.4: Relevant processes to produce Higgs bosons at the LHC. Shown are the Feynman diagrams in leading order for gluon fusion (ggF), vector boson fusion (VBF), Higgs strahlung (VH), and associated production with top quarks (ttH). In the graphs, q denotes any quark, and V denotes either one of the weak bosons Z^0 and $W^{\pm}$.

As the mass m_{H} is not predicted within the standard model, the first systematic surveys of the Higgs boson had to cover a huge mass range leading to very versatile phenomenology, and initially also preferred light Higgs masses $\mathcal{O}(\,\mathrm{MeV})$ [28]. Nonetheless, there are theoretical limitations on m_{H} obtained from requiring unitarity of the scattering amplitudes and a pertubative description of the Higgs self-couplings. A consideration of the naturalness of the standard model implies additional limits, as neither fine-tuning of independent parameters should be necessary in physics theories, nor should the ground state of the vacuum be unstable. A discussion of these arguments can be found in Ref. [29, Sec. 1.4]. While some of these limits are independent of the values of SM parameters, others have become more and more stringent as the parameters of the SM have been determined precisely. In 2010, before the discovery of the Higgs boson, only masses in the range of

$$130\,\mathrm{GeV} \lesssim m_{\mathrm{H}} \lesssim 180\,\mathrm{GeV}$$

were in agreement with the validity of the SM up to the Planck scale [30]. Masses as small as 115 GeV were allowed if a metastable vacuum with a lifetime greater than the age of the universe was accepted.

The phenomenology of a Higgs boson at the LHC has been extensively studied. A complete summary is given in Refs. [32–34]. At a pp collider, such as LHC, four Higgs production processes have to be considered. Their Feynman graphs are shown in Figure 1.4. Additional processes either have cross sections that are too small to be observed or have too large background.

The production cross sections for pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ are depicted in Figure 1.5 as a function of the Higgs boson mass. A detailed discussion of the uncertainties on the Higgs boson production cross section calculations can be found in Ref. [34]. The uncertainties on the theory predictions are typically in the order of 10% and are shown as uncertainty bands in the figure.

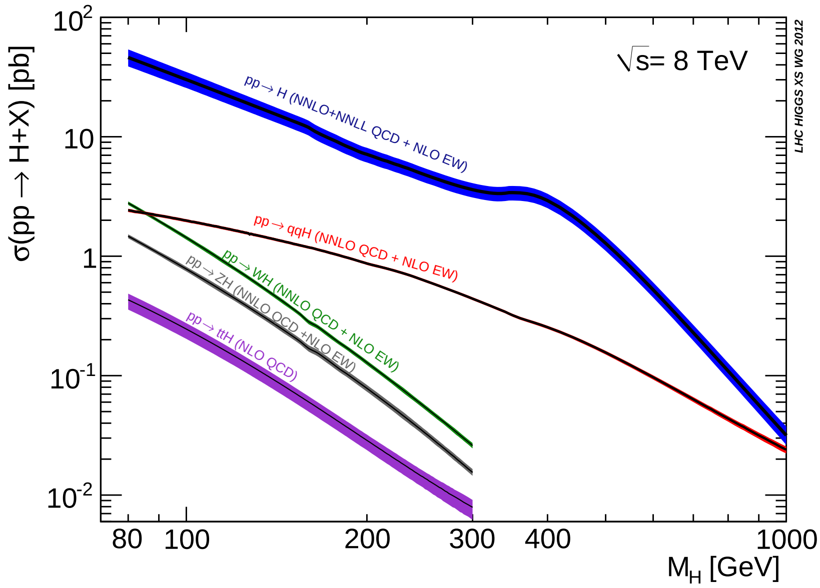


Figure 1.5: Cross sections of the important Higgs boson production processes for $\sqrt{s} =$ 8 TeV pp collisions as a function of the Higgs boson mass [31]. The shown processes from top to bottom are ggF, VBF, VH (split into WH and ZH), and ttH.

In gluon fusion (ggF) processes, Figure 1.4(a), two gluons produce an H via a fermion loop. As the Higgs couples to the mass of fermions, the particle in the loop is almost exclusively the top quark. The production cross section for this process [34],

$$\sigma_{\text{ggF}} = 19.3\,\text{pb} \qquad (m_{\text{H}} = 125.0\,\text{GeV}) , \tag{1.13}$$

is at least one order of magnitude larger than for the other processes. The main challenge is caused by the absence of additional objects in the final state, which means that only the Higgs decay products can be used to disentangle the Higgs signal from background processes. Additionally, the Higgs boson produced via gluon fusion in leading order is expected to have small transverse momentum. Nonetheless, a sizable transverse momentum can be caused by initial state radiation (ISR), when a gluon or top radiates another gluon or quark. In these cases, the boost of the Higgs can be utilized, and the additional quark or gluon jet in the final state might be used for the event selection. The most important irreducible background in the $\text{H} \to \tau\tau$ analyses, the Drell-Yan (DY) processes $qq \to \text{Z}/\gamma^* \to \tau\tau$, have quarks in the initial state, which radiate gluons less likely.

Higgs production via vector boson fusion (VBF) is characterized by two additional quark jets in the final state. Figure 1.4(b) shows that these quarks are scattered out of the proton and thus carry a high momentum in beam direction. In the detector this can be observed as two back-to-back jets close to the beam axis with very high energy. The Higgs boson can carry sizable transverse momentum. For many Higgs analyses, especially those with complex final states such as $\text{H} \to \tau\tau$, this production channel is the most promising. The cross section of vector boson fusion [34],

$$\sigma_{\text{VBF}} = 1.58\,\text{pb} \qquad (m_{\text{H}} = 125.0\,\text{GeV}) , \tag{1.14}$$

still allows for enough events to be analyzed, given an efficient event selection.

Higgs strahlung (VH) denotes production of a weak boson W or Z, which then radiates a Higgs boson. The name Higgs strahlung originates from the similarity of the Feynman graph, Figure 1.4(c), with bremsstrahlung. It has a cross section of [34]

$$\sigma_{\mathrm{VH}} = 0.705\,\mathrm{pb}\ (\mathrm{WH}) + 0.415\,\mathrm{pb}\ (\mathrm{ZH}) \qquad (m_{\mathrm{H}} = 125.0\,\mathrm{GeV})\ . \tag{1.15}$$

This process is sometimes also denoted as associated production (AP), although AP could include ttH, depending on the author. The vector boson in the final state of VH production can offer a good handle for background suppression, especially if it decays leptonically.

Finally, associated production with top quarks (ttH) is a rare process in proton collisions [34],

$$\sigma_{\mathrm{ttH}} = 0.129\,\mathrm{pb} \qquad (m_{\mathrm{H}} = 125.0\,\mathrm{GeV})\ . \tag{1.16}$$

With three heavy objects in the final state, these events can be identified by the large multiplicity of jets and leptons. As both the Higgs boson and top quarks decay via various channels, a large number of final states in the detector has to be considered. In both the VH and ttH processes, the Higgs boson is expected to obtain transverse momentum from the recoil against the heavy objects. Due to its small cross section and the complex final state, associated production with top quarks is not considered in this thesis.

Figure 1.6 shows the cross sections of ggF, WH, and VBF Higgs boson production in comparison with other standard model processes. Compared to the total pp cross section σ_{tot}, Higgs boson production is a rare process. Only one Higgs boson is produced in approximately 10^{10} pp collisions. The individual production mechanisms show different dependency on $\sqrt{s}$, making VBF more important at LHC than at Tevatron. With higher collider energy, the Higgs production cross sections become bigger, meaning that less luminosity is necessary to observe the same amount of Higgs bosons. Most of the important background processes for Higgs searches, such as Z production or non-resonant production of b quarks, show a smaller rise of the cross section as a function of energy, which makes Higgs studies simpler at higher energies. One crucial exception is top-quark production, which is an important background to various Higgs decay channels due to the possible leptons in the final state.

The standard model Higgs boson decays via various channels, according to the couplings given in Equation (1.11). The branching fractions are dependent on the Higgs mass, as can be seen in Figure 1.7. At high values of m_{H}, decays into the heavy gauge bosons Z and W dominate, and only $\mathcal{O}(10\%)$ of the Higgs bosons decay into fermions, namely exclusively into top quarks. For the WW, ZZ, and $\mathrm{t\bar{t}}$ channels, resonance effects at the sum of the final state particle masses are visible.

At lower masses, $m_{\mathrm{H}} \lesssim 200\,\mathrm{GeV}$, the phenomenology of Higgs decays is more versatile. It is thus necessary to perform Higgs searches in a wide range of final states, and to use sophisticated statistical methods to extract a clean Higgs signal. However, this allows to compare measured branching fractions in different decay channels, and therefore to test the predictions by the SM. The important, calculated branching fractions $\mathcal{B}$ for a Higgs boson of mass $m_{\mathrm{H}} = 125.0\,\mathrm{GeV}$ are given as [34]

$$\begin{aligned}
\mathcal{B}(\mathrm{H} \to \mathrm{b\bar{b}}) &= (5.77 \pm 0.18) \times 10^{-1} & \mathcal{B}(\mathrm{H} \to \mathrm{c\bar{c}}) &= (2.91 \pm 0.36) \times 10^{-2} \\
\mathcal{B}(\mathrm{H} \to \mathrm{WW}) &= (2.15 \pm 0.09) \times 10^{-1} & \mathcal{B}(\mathrm{H} \to \mathrm{ZZ}) &= (2.64 \pm 0.11) \times 10^{-2} \\
\mathcal{B}(\mathrm{H} \to \tau\tau) &= (6.32 \pm 0.35) \times 10^{-2} & \mathcal{B}(\mathrm{H} \to \mu\mu) &= (2.19 \pm 0.13) \times 10^{-4} \\
\mathcal{B}(\mathrm{H} \to \mathrm{gg}) &= (8.57 \pm 0.87) \times 10^{-2} & \mathcal{B}(\mathrm{H} \to \gamma\gamma) &= (2.28 \pm 0.11) \times 10^{-3}\ .
\end{aligned} \tag{1.17}$$

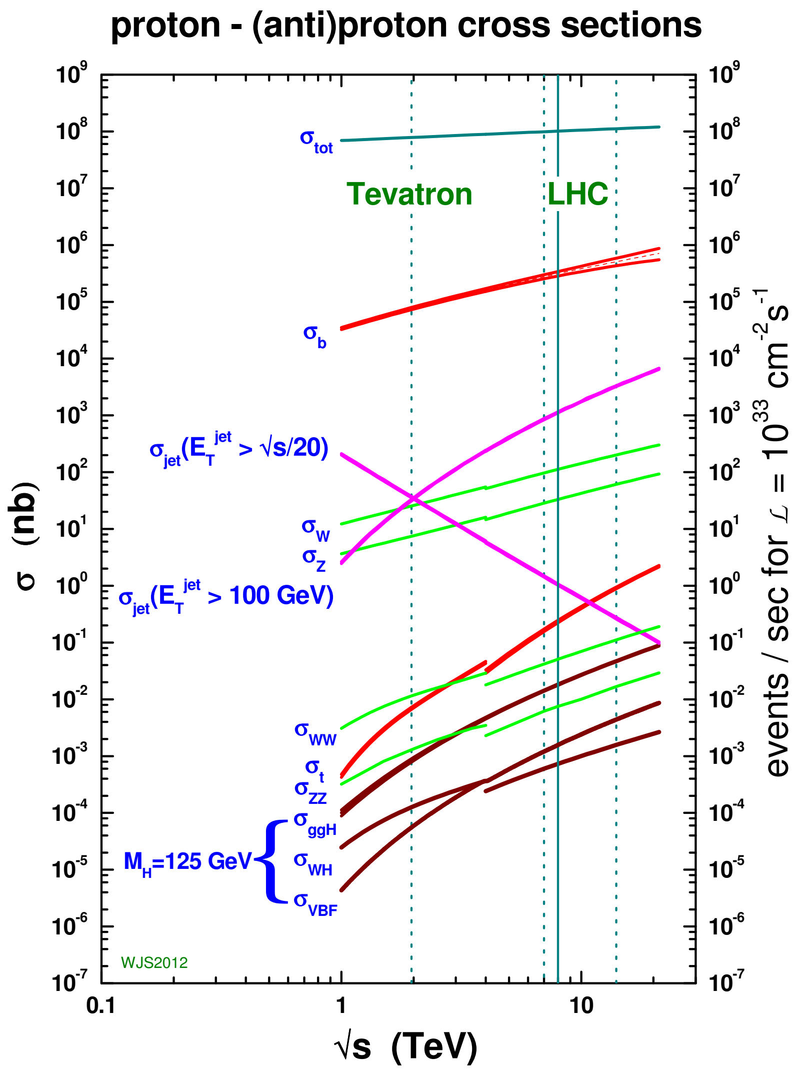


Figure 1.6: Cross sections of important SM processes as a function of the center-of-mass energy of a pp or p$\bar{\text{p}}$ collider [35]. The vertical lines show $\sqrt{s}$ of Tevatron and the LHC, with the 2012 LHC center-of-mass energy of 8 TeV as a solid line. The cross section curves show $\sigma_{\text{p}\bar{\text{p}}}$ for $\sqrt{s} < 4$ TeV and σ_{pp} above this value, exhibiting a discontinuity at 4 TeV.

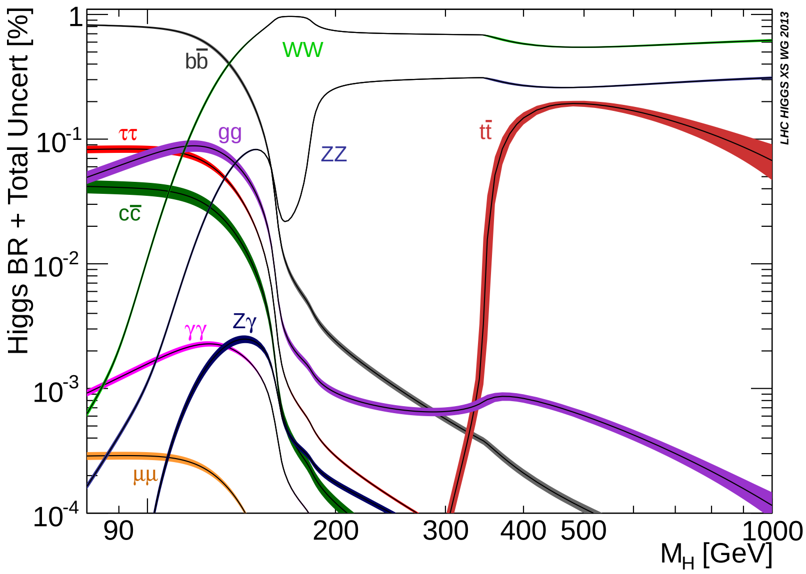


Figure 1.7: Branching fractions of Higgs decays into various final states as a function of the Higgs boson mass [31].

The VV channels WW and ZZ show non-negligible branching ratios well below the kinematic threshold defined by their masses. In these cases, one of the vector bosons is produced off-shell (often denoted V^*), while the other is expected to have nominal mass. If the bosons decay into leptons, $Z \to \ell\ell$ or $W \to \ell\nu_\ell$ with $\ell = e, \mu$, background from QCD multijet events is highly reduced. For this reason, $H \to ZZ \to 4\ell$ is often labeled the "golden channel", as it has a clean signature and small background. The good lepton energy resolution at CMS allows to search for a narrow resonance on a broad background spectrum. As W and Z are unstable, analysis of the angular correlations of their decay products allows access to Higgs boson properties like spin and parity.

The decay into top quarks does not occur since $t\bar{t}$ pairs are too heavy. Hence $H \to b\bar{b}$ is the predominant decay channel in the low mass range. Separating bottom quarks from light-quark jets is possible at CMS, but remains challenging (see Section 1.3.4). Since the background from $b\bar{b}$ dijet events is huge, searches for $H \to b\bar{b}$ seem only feasible with additional objects in the final state, and therefore only in association with the VBF, VH, and ttH production mechanisms. The other $q\bar{q}$ decay channel with non-negligible branching ratio is the decay into charm quarks. While the complications described for bottom quarks are even more severe for charm quarks, the branching ratio is approximately a factor of 10 smaller. Therefore, $H \to c\bar{c}$ is not in reach within the first years of LHC operation. The same argument holds for Higgs decays into gluons.

Although photons are massless, the Higgs boson can decay into $\gamma\gamma$ via top-quark or W-boson loops. These decays are rare, but the channel is of big importance for Higgs searches. Photons do not decay and can be reconstructed in the electromagnetic calorimeter (ECAL) with high accuracy. As diphoton events are rare in pp collisions, the small branching ratio is sufficient to observe a clear, narrow peak on a broad background.

Finally, the leptonic channels of interest for Higgs searches at the LHC are $H \to \tau\tau$ and $H \to \mu\mu$. While electrons have too small mass and thus couplings to the SM Higgs boson, decays into muons are sufficiently frequent. Given its clean signature, $H \to \mu\mu$ could be observed at LHC.

The decay of Higgs bosons into tau leptons is of importance for multiple reasons. With a relatively high branching ratio, $\mathcal{B}(\mathrm{H} \to \tau\tau) = \mathcal{O}(10\,\%)$, a sufficient amount of such events is expected in all production mechanisms. Thus, the channel can provide direct evidence for the Yukawa couplings of the Higgs boson to fermions. Even more, as taus are the heaviest leptons they are predestined to test the mass generation mechanism in the leptonic sector, which could be different than the one of quarks. Although the reconstruction of hadronically decaying taus is challenging (see Section 1.1.4 and Section 1.3.6), the kinematics of the decay products provide information on Higgs properties like spin, parity, and the CP eigenvalue.

1.1.3.3. Searches for the Higgs boson

While the EBH mechanism provides an elegant solution to the problems described in Section 1.1.3.1 caused by massive standard model particles, it is challenging to proof it experimentally. The scalar Higgs field is impossible to observe by itself. The prediction of an additional massive boson accompanying this field seems not to have caused much interest at first, as it did not impact the problems that ought to be solved. Thus, only Peter Higgs mentioned explicitly the necessary existence of additional "waves whose quanta have (bare) mass" [1]. Today, this particular quantum is called the Higgs boson H. Its importance, offering a prediction that could be tested by experiment, was noted only later, when the standard model of particle physics became more and more popular.

Direct searches for the Higgs boson have been performed by various experiments. Before the LHC, results obtained at two facilities provided the best limits. Higgs bosons with a mass $m_{\mathrm{H}} < 114.4\,\mathrm{GeV}$ were excluded at 95% confidence limit (CL) by the four experiments at the Large Electron-Positron Collider (LEP), using $\mathrm{e}^+\mathrm{e}^-$ collisions at center-of-mass energies $189\,\mathrm{GeV} < \sqrt{s} < 209\,\mathrm{GeV}$ [36]. Complementary, the detectors CDF and D0 at the Tevatron $\mathrm{p\bar{p}}$ collider searched for Higgs bosons in the mass range of $90-200\,\mathrm{GeV}$. Their combination excludes a standard model Higgs boson at 95% CL from 90 to 109 GeV as well as from 149 to 182 GeV [37].

Since the discovery of the top quark in 1995 [38, 39] all but one parameters of the standard model were measured by experiments. This allowed to predict the last missing value m_{H} by fitting electroweak precision measurements, including higher order processes and assuming that the standard model is valid and complete. Among others, the Gfitter Group has performed such a fit [40]. Excluding information from direct Higgs searches from the fit, a most probable value $m_{\mathrm{H}} = 91^{+30}_{-23}\,\mathrm{GeV}$ was found. Figure 1.8 shows the $\Delta\chi^2$ distribution as obtained by this fit. In combination with the direct searches, only a small mass window above 114.4 GeV remained open, if the standard model holds.

This final hideout was eventually investigated by the LHC experiments ATLAS and CMS. Both collaborations claimed observation of a new boson in 2012, presenting analyses with significances of 5.9 and 5.8 standard deviations, respectively [2–4]. Since then more pp collision data have been collected in the LHC Run-I, amounting to an integrated luminosity of approximately $5\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 7\,\mathrm{TeV}$ and $20\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 8\,\mathrm{TeV}$ for each experiment. Extensive studies of the new resonance have been performed using this data, all of which are in agreement with the hypothesis of a standard model Higgs boson. An overview of these studies can be found in Ref. [41], which also lists references to publications of interest.

The current state of understanding on the Higgs boson discovered in 2012, as of the writing

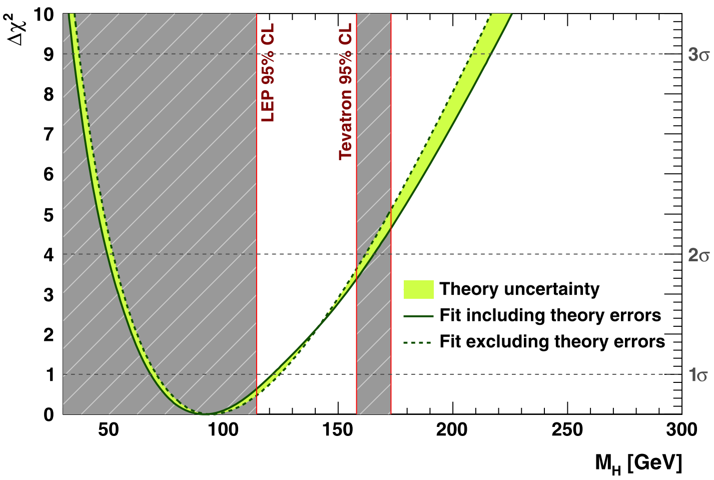


Figure 1.8: Determination of m_{H} from electroweak precision measurements, excluding information from direct Higgs searches, as performed by the Gfitter group [40]. Shown is $\Delta\chi^2$ as a function of the Higgs-boson mass, giving 91 GeV as most probable value. The gray areas show exclusion regions from direct Higgs searches.

of this thesis, can be best described by pointing out some important results. The resonance significance is found to be close to or above 5σ in the $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$, and $\mathrm{H} \to \mathrm{WW}$ decay channels individually, for both ATLAS and CMS [41]. Couplings to fermions, and thus the concept of Yukawa couplings, are evident not only by the top-quark loops in $\mathrm{H} \to \gamma\gamma$, but are found directly in $\mathrm{H} \to \tau\tau$ and $\mathrm{H} \to \mathrm{bb}$ decays [42, 43].

The combined measurement of the mass [44], using the full data sets and various decay channels from both experiments, is

$$m_{\mathrm{H}} = 125.09 \pm 0.21(\mathrm{stat}) \pm 0.11(\mathrm{syst})\,\mathrm{GeV}\,. \tag{1.18}$$

No inconsistency of the mass values obtained from different decay channels is observed.

The signal strength is commonly defined as the ratio of the measured cross section and its SM prediction,

$$\mu = \frac{\sigma_{\mathrm{meas}}}{\sigma_{\mathrm{SM}}}\,. \tag{1.19}$$

A global fit to the combined ATLAS and CMS data from LHC Run-I [41] results in

$$\mu = 1.09^{+0.11}_{-0.10}\,, \tag{1.20}$$

which is compatible with the expectation of $\mu = 1$. The same statement holds for the individual signal strengths in all investigated production and decay channels. It should be noted that the systematic component of the given uncertainty on μ in Equation (1.20) is already larger than the statistical uncertainty, and itself is dominated by theoretical uncertainties on the Higgs boson production cross sections in the SM.

One important prediction of the EBH mechanism is the coupling of the Higgs boson H(125) to the masses of the standard model particles. Equation (1.11) shows that the coupling strength is a linear function of the mass for Yukawa couplings, and goes quadratic with the mass for weak bosons. Figure 1.9 shows the corresponding measurements from CMS. The coupling strength to vector bosons is transformed to allow a representation of the results in

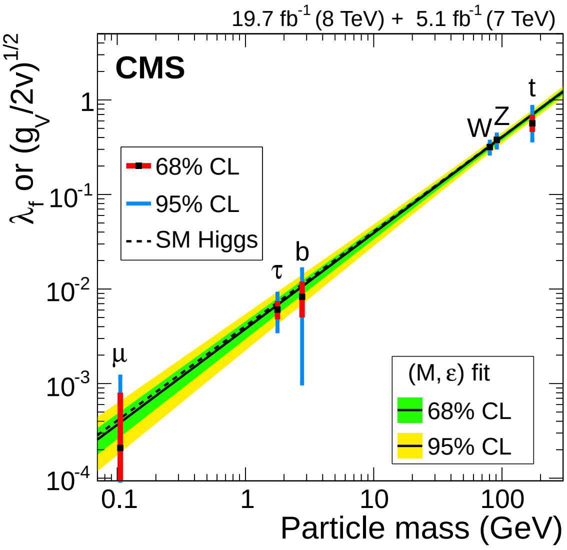


Figure 1.9: Measured couplings of H(125) to standard model particles as a function of the particle mass [45]. The variables λ_f and g_V in the figure correspond to $g_{\mathrm{H}f\bar{f}}$ and $g_{\mathrm{H}VV}$ in Equation (1.11), respectively. To show Yukawa and weak boson couplings in one plot, the latter are modified as $\sqrt{g_V/2v}$ to account for the different scaling of the couplings with mass. The SM expectation (dashed line) and a fit to the data (solid line) show good agreement. A preliminary version of a similar figure combining the results from ATLAS and CMS can be found in Ref. [41].

one plot. A fit to the mass dependence is in good agreement with the SM prediction. The mass range already covers four orders of magnitude, which strongly supports mass generation by the Higgs field. Only the light SM particles might not obtain their mass via the scalar H(125) field, as their coupling to the Higgs boson could not be tested yet.

Finally, all measurements of spin-parity of H(125) are compatible with the SM expectation of $J^P = 0^+$ [46, 47]. The observed boson decays into a pair of photons. According to the *Landau-Yang theorem* [48, 49], it therefore cannot be of spin 1. The assumption of spin 2 is excluded by both experiments, and a pure pseudoscalar hypothesis is reported to be strongly disfavored.

Given the very good understanding of the new resonance H(125), it is safe to claim that ATLAS and CMS have discovered *a Higgs boson* as predicted by the EBH mechanism. Some important characteristics of the SM Higgs boson remain yet to be tested, others have to be studied in more detail. Examples include Higgs self-couplings and its *CP* nature, respectively.

Although the search for a Higgs boson was eventually successful, it does not necessarily end here. Many beyond the standard model (BSM) theories predict additional Higgs bosons with similar or even identical properties. One prominent example are the so-called two-Higgs-doublet models (2HDMs), which expand the standard model by adding a second Higgs doublet. This gives rise to four additional degrees of freedom, resulting in a total of five Higgs bosons. Two neutral *CP*-even states ($\mathrm{h}^0$and H^0) are distinguished only by their mass, one boson (A^0) is *CP*-odd, and two more carry electric charge ($\mathrm{H}^\pm$). The search for such new resonances will persist for the whole running period of the LHC.

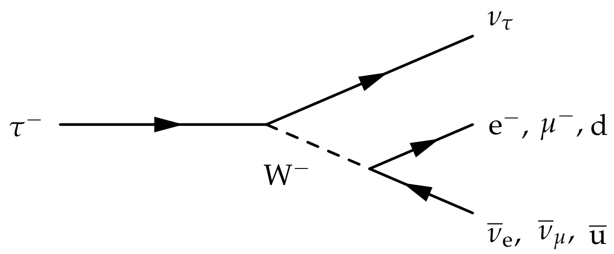


Figure 1.10: Feynman graph for the decay of a tau lepton with negative charge. The W decays in $e^-\overline{\nu}_e$, $\mu^-\overline{\nu}_\mu$, or $d\overline{u}$. Occurrences of s quarks in the final state are rare.

1.1.4. Tau leptons

The tau is the charged lepton in the third family of the standard model of particle physics. It was discovered in 1975 by Martin Perl et al. in e^+e^- collisions at SLAC [50]. With a mass of $m_\tau = 1.777\,\text{GeV}$ it is the heaviest of the three charged leptons. As its mass is bigger than some meson masses, the tau lepton can decay hadronically. This leads to a very different decay phenomenology than for muons, which can only decay leptonically, i.e. into electrons and neutrinos.

Figure 1.10 shows the Feynman diagram of a τ^- decay. For leptonic decays, $\tau^- \to e^-\overline{\nu}_e\nu_\tau$ or $\mu^-\overline{\nu}_\mu\nu_\tau$, the final state contains two neutrinos. In these cases, the emerging electron or muon is the only object that can be observed by particle detectors. Given the tau's short lifetime $c\tau(\tau) = 87.03\,\mu\text{m}$, the decay vertex is expected to be near the τ production vertex. An observation of the decay is the only handle to separate leptonic tau decays from direct e or μ production (often called *prompt leptons*). The decay can be seen either as kink in the reconstructed track or as offset of the track from the production vertex.

In case of a hadronic decay, only one neutrino occurs in the final state. The emerging quarks in these decays are subject to the confinement of QCD. They build hadrons, mostly pions and kaons, via intermediate resonance states. The final state then typically consists of one or three charged plus sometimes additional neutral mesons. A higher multiplicity of charged mesons is possible, but rare. In this thesis, τ_h denotes the hadronic decay products of a tau decay, $\tau \to \tau_h\nu_\tau$. Hadronic tau decays are often categorized by their number of *prongs*, which is the number of charged hadrons in the final state. As the final state particles are mostly pions, the symbols $\tau_{1\pi}$ and $\tau_{3\pi}$ are used to denote 1-prong and 3-prong tau decays, respectively[1].

Table 1.3 gives an overview of the most important tau decay modes. Many more modes have been measured, but are omitted here as they play no role in CMS analyses. For the hadronic decays, the intermediate resonance is also listed.

Both leptonic and hadronic tau decays play a vital role in analyses searching for Higgs boson decays. Electrons and muons can be easily identified and measured with high accuracy by CMS (see Section 1.3.3) and therefore offer a good handle to suppress the large QCD background. The disadvantage of leptonic tau decays is the relatively small branching ratio as well as their similarity to prompt leptons. Quark and gluon jets are difficult to distinguish from hadronic tau decays (see Section 1.3.6). However, the high branching ratio of 64.8% allows for a higher number of signal events after the selection. For these reasons, for most analyses of $\tau\tau$ final states both leptonic and hadronic tau decays have to be considered.

[1] In this thesis, the term *1-prong* denotes only hadronic tau decays.

Table 1.3: Decay modes of tau leptons [27]. Listed are the branching ratios $\mathcal{B}$ as well as the main intermediate resonance for hadronic decays. For simplification the symbol π for pions is used in this table, although the listed branching ratios include small fractions of decays into kaons. Various resonances contribute to the 4π final state, which is why no main resonance is listed.

Decay Mode	Resonance	$\mathcal{B}/\%$	
Leptonic Modes			35.2
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8	
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4	
Hadronic Modes			64.8
$\tau^- \to \pi^- \nu_\tau$		11.5	
$\tau^- \to \pi^- \pi^0 \nu_\tau$	$\rho(770)$	26.0	
$\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5	
$\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$	$a_1(1260)$	9.8	
$\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$		4.8	
other hadronic modes		3.2	

1.2. Experimental setup

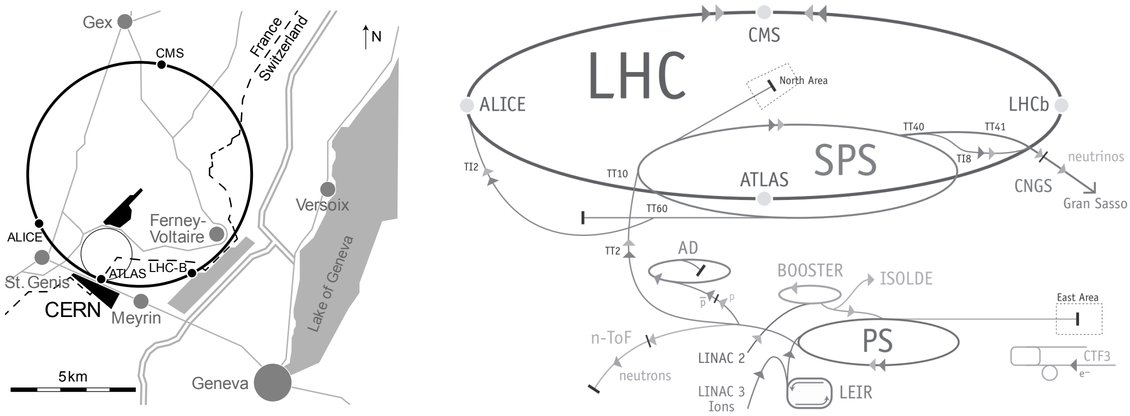
1.2.1. The LHC

The Large Hadron Collider (LHC) [51] is a particle accelerator and collider operated by the European Organization for Nuclear Research (CERN) near Geneva. Its goal is to provide proton-proton, proton-ion or ion-ion collisions at four interaction points, which are observed by experiments. In pp mode the LHC is designed to provide collisions at a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$. This provides access to an energy that could not be studied in laboratory experiments before.

One of the main assignments of CERN is to provide and operate large scale experiments. To investigate the current questions of high energy physics (HEP), the complexity and thus the cost of the experimental setup are increasing. This often exceeds the capabilities of individual institutes or even countries. CERN as an international collaboration provides the political basis as well as the infrastructure for these endeavors. The main site of the laboratory is located in Meyrin, next to Geneva in Switzerland, close to the border to France (see Figure 1.11(a)).

The LHC is the latest of several accelerators operated by CERN. A sketch of the facilities is shown in Figure 1.11(b). Many of the older machines are now used as part of the LHC accelerator chain. Each constituent of the chain increases the particle energy by a specific amount. The particles are then transferred to the next chain link, where they are accelerated further. The last link is the LHC, which is provided with 450 GeV protons by the Super Proton Synchrotron (SPS). To provide head-on collisions, two beams are injected into the LHC, running clockwise and counter-clockwise, respectively. Within approximately twenty minutes the protons are accelerated up to a maximum energy of 7 TeV. At peak energy the particles are then stored for several hours and brought into collision at four collision points.

The LHC is hosted in a circular tunnel with a length of 26.7 km, built between 50 m and 175 m below the surface. This tunnel was previously used for LEP. The particles are accelerated in



(a) Location [52]. (b) Accelerators and experiments [53].

Figure 1.11: Location of CERN and the LHC at the Franco-Swiss border (left) and a schematic view of the CERN accelerators and experiments (right). For the operation of the LHC in pp mode, the pre-accelerators Linac 2, PSB (Booster), PS, and SPS are used.

the LHC using eight superconducting radio frequency (RF) cavities. Each of these provides 2 MV at 400 MHz, corresponding to an accelerating field of 5.5 MV/m. Using RF fields implies that the particle beam is bunched. In pp mode, a maximum of 2808 bunches are stored in each beam, separated by gaps of 25 ns. Each bunch consists of up to 1.15×10^{11} protons.

A total of 1232 dipole magnets, each of a length of 14.3 m, provide magnetic fields of 8.3 T. To reach such high fields, superconducting NbTi cables are employed. Additional magnetic fields of higher order provide corrections for the particle trajectories and thus keep the beam focused. These fields are provided either by individual magnets, or are embedded in the cold mass of the main dipoles or quadrupoles. At each of the four interaction points the transverse size of the particle beam is minimized by a dedicated array of magnets to about 16 μm.

The LHC design parameters allow for a luminosity of up to $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Luminosity is a measure of the amount of possible collisions provided by a particle collider. It allows to calculate the rate of events of a certain collision process as

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \sigma \cdot \mathcal{L} \,. \tag{1.21}$$

Here, the cross-section σ is a physical quantity, whereas $\mathcal{L}$ is defined by the machine parameters. It can be given as

$$\mathcal{L} = \frac{n_\mathrm{b} \cdot N_\mathrm{b}^2 \cdot f_\mathrm{rev}}{A_\mathrm{T}^\mathrm{eff}} \,. \tag{1.22}$$

In this formula, n_b is the number of bunches per beam, and N_b the number of protons in each bunch. The revolution frequency of the protons is denoted f_rev. The effective transverse area of the beam $A_\mathrm{T}^\mathrm{eff}$ can be given as $A_\mathrm{T}^\mathrm{eff} = 4\pi\sigma_x\sigma_y$ if the beam profile is Gaussian with widths $\sigma_{x,y}$ in x and y direction, respectively, and if the crossing angle is zero.

The four major LHC experiments are each located at one of the four interaction points: ALICE, ATLAS, CMS, and LHCb. The main purpose of A Large Ion Collider Experiment (ALICE) is to study the quark-gluon plasma produced in heavy-ion collisions. High energy densities in such collisions correspond to a huge number of particles in the final state, which is one driving aspect of the detector design of ALICE. The Large Hadron Collider beauty (LHCb) is an asymmetric detector. Its focus lies on analyses of B mesons, which are often boosted along the beam direction. The main goal of LHCb is to study CP violation associated to b quarks.

ATLAS and CMS are multi-purpose detectors. Their objective is to cover a wide range of experimental tasks, including detailed studies of the standard model, the search for and investigation of a Higgs boson, and exploration of the wide field of BSM theories. At the same time they must also look for surprising new phenomena that have not been predicted by theories. Both detectors thus follow a similar strategy: Covering a large solid angle with several layers of particle detectors allows to detect all particles of the SM, with the exception of neutrinos. The main differences between ATLAS and CMS are the different magnetic field configurations as well as the location of the calorimeters. These are installed outside and inside the magnetic coil for ATLAS and CMS, respectively. Furthermore, the individual detector layers are implemented using different technologies. The CMS detector is described in more detail in Section 1.2.2.

1.2.1.1. LHC operation

The first beam was injected into the LHC in 2008. In 2010 the LHC provided first collisions of protons at a center-of-mass energy of 7 TeV. This year was used both for commissioning the machine with dedicated studies and also for collecting physics data [54]. At the end of 2010, the first lead-ion collisions were recorded. As heavy-ion collisions play no role in this thesis, they are not considered further in this section.

Since 2011, the highest priority of the LHC is to provide collisions at the four interaction points. The instantaneous luminosity $\mathcal{L}$ steadily increased. After reducing the bunch spacing from 75 ns to 50 ns, the number of stored bunches reached its maximum of 1380 in August 2011 [54]. The peak instantaneous stable luminosity in 2011 was $3.54 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [55].

For 2012, the beam energy was increased to 4 TeV, corresponding to $\sqrt{s} = 8\,\mathrm{TeV}$. The bunch spacing remained at 50 ns, allowing for only half as many bunches as the design value. Nonetheless, with a maximal value of $\mathcal{L} = 7.67 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ the design luminosity of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ was almost reached.

For physics analyses, the integrated luminosity

$$L = \int \mathcal{L}\,\mathrm{d}t \tag{1.23}$$

is essential, as it allows to measure cross sections according to Equation (1.21). Figure 1.12 shows the integrated luminosity during the year 2012. In total $23.30\,\mathrm{fb}^{-1}$ have been delivered by the LHC to CMS, of which $21.79\,\mathrm{fb}^{-1}$ have been recorded by the experiment. Depending on the analysis, an additional certification of the collected data is performed by the CMS collaboration. For $19.7\,\mathrm{fb}^{-1}$ all detector components have been fully functional. This is the data set used for most physics analyses, including the one presented in this thesis.

The large instantaneous luminosity of the LHC has a drawback for the experiments. In each bunch crossing, multiple proton-proton interactions occur. Most of these collisions are parton interactions with a relatively small center-of-mass energy $\hat{s}$ and are therefore not of interest. The detector has to disentangle these so-called pileup (PU) interactions from the interaction of interest, which is often called the hard interaction. The distribution of the number

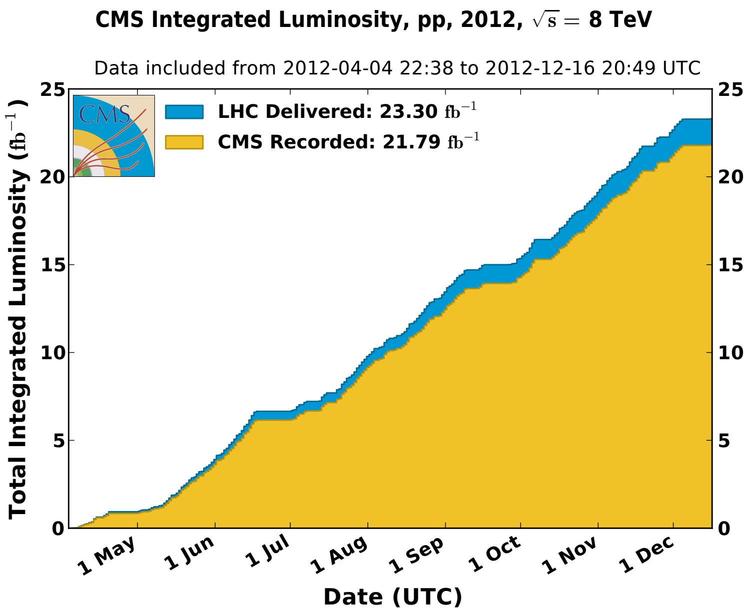


Figure 1.12: Integrated luminosity delivered by the LHC and recorded by CMS during the year 2012 [55].

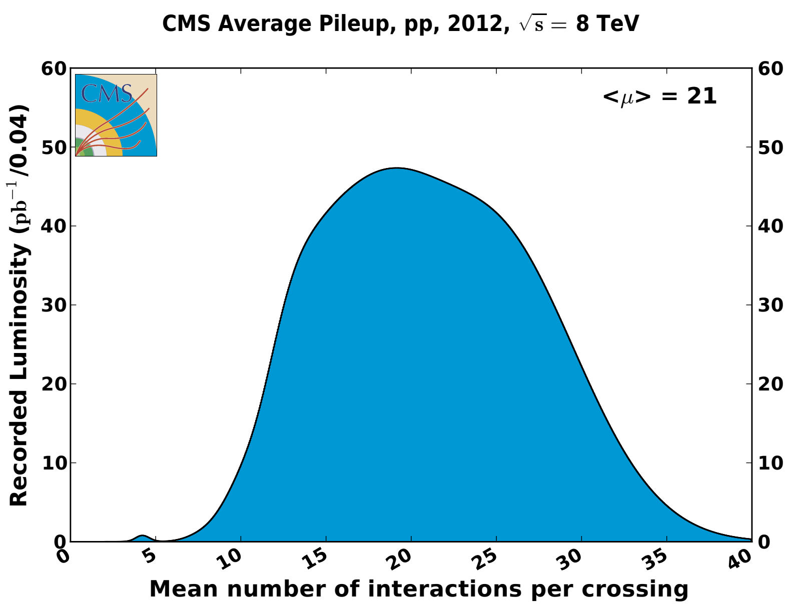


Figure 1.13: Distribution of the number of interactions per bunch crossing (pileup) in the 2012 CMS data set [55].

of interactions per bunch crossing is shown in Figure 1.13. It is calculated using Poisson statistics from the distribution of the instantaneous luminosity in 2012.

The period from 2010 to 2012 is denoted LHC Run-I. After a shutdown period of two years, Run-II started in 2015 with a center-of-mass energy of 13 TeV. In this year, the bunch spacing was reduced to the design value of 25 ns. While this allows for higher luminosities, it also makes it more difficult to associate the detector signals to the correct bunch crossing. This effect, called out-of-time pileup (OOT-PU), had no big impact on physics analyses during Run-I, but will become more important in Run-II. The data taken during Run-II will not be considered in this thesis.

1.2.2. The CMS detector

The Compact Muon Solenoid (CMS) is a general purpose particle detector recording proton-proton, lead-proton and lead-lead collisions provided by the LHC. The chosen acronym for the apparatus identifies its main features: First, the design is compact, with only a minimum amount of empty space. Second, muons are reconstructed with highest precision, although other final state objects are not neglected. Finally, a superconducting solenoid provides a strong magnetic field of 3.8 T. A detailed description of the detector can be found in Ref. [56]. This section will provide an overview of its most important features.

The general design of CMS resembles a barrel, closed by an endcap on each side. The LHC beam line defines the axis of its rotational symmetry, the nominal collision point is located in the center of the detector. Different sub-detectors are mounted in layers around this center. As most of these components resemble a barrel-and-endcaps shape themselves, the performance of sub-detectors and therefore particle identification and reconstruction algorithms is often given separately for the barrel and endcap regions.

In order to measure the momentum of charged particles, their tracks are bent by the magnetic field. A strong magnet is thus mandatory for a multi-purpose detector. At CMS, the superconducting solenoid has an inner diameter of 6 m and a length of 12.5 m. The return yoke consists of 10 000 t of iron that are distributed along five wheels in the barrel region, and three disks at each endcap.

Figure 1.14 shows a cutaway view of the CMS apparatus as it was during LHC Run-I. Going from the collision point outwards, the first detector layer directly outside of the beam tube is a silicon pixel tracker, followed by a silicon strip tracker. The calorimetry system is also located within the free bore of the solenoid, minimizing the energy loss of particles before the calorimeters. It consists of a lead tungstate crystal ECAL, followed by a brass and scintillator hadronic calorimeter (HCAL). Additional forward calorimeters are installed to reduce energy escaping close to the beam axis. Outside of the solenoid, gas-ionization detectors are embedded into the return yoke to measure muons.

1.2.2.1. The CMS coordinate system

CMS uses a coordinate system centered at the nominal collision point. The x-axis points horizontally towards the center of the LHC, the y-axis vertically upward. To obtain a right-handed system, the z-axis must point along the beam that follows the tunnel in the counter-clockwise direction. The location of detector components is best described using cylindrical coordinates. Here, φ denotes the azimuthal angle and is measured from the x-axis in the x-y plane. The radial coordinate r is defined in the same plane.

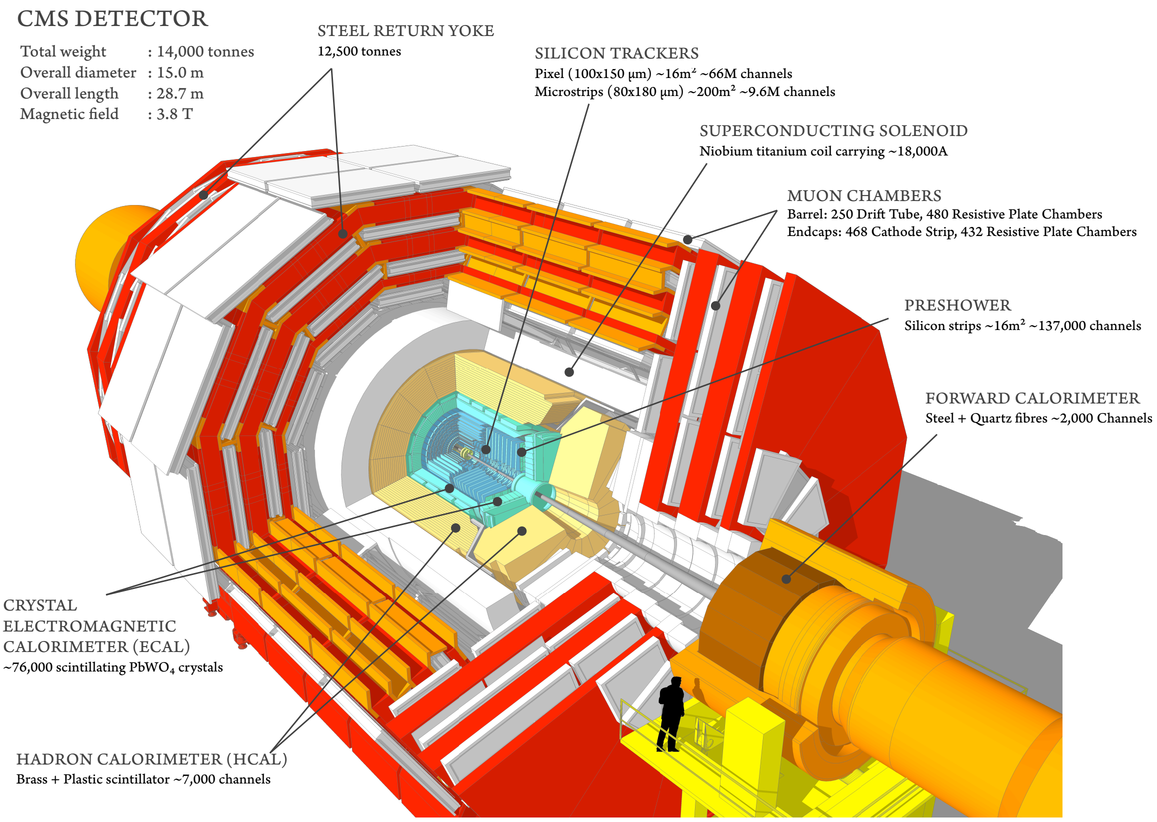
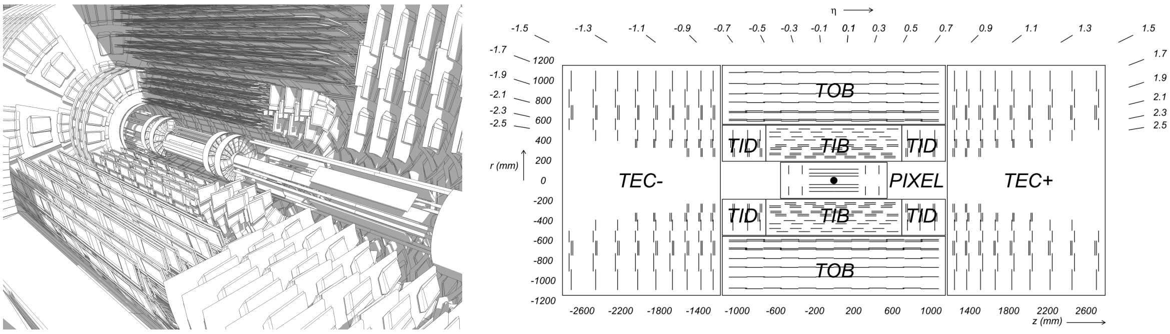


Figure 1.14: Cutaway view of the CMS detector [57]. The picture shows the configuration without the upgrades installed for LHC Run-II.



(a) Cutaway view [57].   (b) Cross section view [56].

Figure 1.15: The tracking system of CMS. The pixel tracker is located at the center and consists of three barrel layers and two endcaps. The components of the silicon strip tracker are labeled in the cross section view (right).

Particle trajectories are commonly described using spherical coordinates. The polar angle θ is measured from the z-axis, the azimuthal angle φ remains as defined before. Instead of θ, the pseudorapidity η is often used. It is defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) . \tag{1.24}$$

Accordingly, an angular measure between two particles is defined as

$$\Delta R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2} . \tag{1.25}$$

Physical quantities that are calculated only in the transverse plane, i.e. neglecting the z component, are denoted with a T in the index. Typical examples for this are the transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ and the corresponding energy E_{T}.

1.2.2.2. Tracking system

The innermost layer of CMS is the silicon tracking system shown in Figure 1.15. Charged particles traversing the silicon deposit energy, which is used to reconstruct the particle track. As multiple interactions are expected per bunch crossing, one main task of the tracker is to reconstruct the multiple interaction points, and to associate each track to one of these primary vertices (PVs). This is the most important measure to mitigate the impact of PU on physics results. Additionally, the tracking system allows to reconstruct decay vertices of unstable particles, called secondary vertices (SVs). The second important task of the tracker is to measure the bending of the charged particles caused by the magnetic field. The bending strength is used to determine the transverse momentum p_{T} of charged particles.

The design of the tracker system is constraint by a few key aspects. To minimize the amount of particles not reconstructed by the tracker, a large solid angle needs to be covered. As shown in Figure 1.15(b), the tracker covers a pseudorapidity range $|\eta| < 2.5$. At design luminosity, 1000 particles hit the tracker per bunch crossing [56], i.e. 40 million times per second. Therefore it is crucial to have small detector elements, which allow to keep the occupancy per element reasonably low. Additionally, the high particle flux causes radiation damage to the inner detectors, which needs to be minimized and corrected for. The design needs to ensure the functionality of the tracking system for an expected lifetime of 10 years. Finally, to allow for accurate spatial resolution in both the tracker and the calorimeters, the amount of material employed in the tracker has to be minimized.

The tracker consists of two parts. Closest to the interaction point, directly outside of the LHC beam tube, is the silicon pixel tracker. Each pixel gives a two-dimensional information on the track position to reduce the amount of ambiguities. Outside of the pixel detector resides the silicon strip tracker. Each silicon strip reaches a similar resolution only in the φ coordinate. Given the much larger size of the strip tracker compared with the pixel tracker, this design allows for a reasonable price while achieving a good resolution for both vertex reconstruction and momentum measurements.

The pixel tracker consists of three cylindrical layers, located at radii of 4.4 cm, 7.3 cm, and 10.2 cm, each being 53 cm long [56]. Additionally, on either side two endcap discs are located at $z = \pm 34.5$ cm and ± 46.5 cm. The setup is designed such that the pixel tracker measures three space points with high precision for each charged particle trajectory, for most of the covered η range. Each of the 66 million pixels has a size of $100 \times 150\,\mu\text{m}^2$, which allows for precise track reconstruction in both r-φ and z directions. This is especially important for the reconstruction of secondary vertices with low track multiplicity, as they occur in tau decays. An analog readout of each pixel allows to make use of a spread of the deposited charge over multiple pixels, reducing the spatial resolution to $15-20\,\mu$m in both directions.

The strip tracker consists of two separate barrel as well as two separate endcap components, which are labelled in Figure 1.15(b). The tracker inner barrel (TIB) consists of four layers of silicon strip modules, extending to a radius of 55 cm. The high η range is covered by the tracker inner disks (TIDs), consisting of three disks at each side of the detector. The strips are aligned with the beam axis in the barrel region, and radial in the disks. The TIB/TID system provides up to four r-φ or z-φ measurements for each trajectory with a single point resolution of $23-35\,\mu$m. Outside of this system resides the tracker outer barrel (TOB) with an outer radius of 116 cm. It provides up to six additional r-φ measurements with single point resolution of $35-53\,\mu$m. Finally, the tracker endcaps (TECs) are located at $124\,\text{cm} < |z| < 282\,\text{cm}$. Each endcap consists of nine disks and thus provides up to nine additional measurements.

As strips allow to measure only one coordinate accurately, some of the detector modules are double-sided. This means that a second silicon strip module is mounted on the back of a given module, tilted by an angle of 100 mrad. The second coordinate can then be measured with a single point resolution as good as $230\,\mu$m. Double-sided detector modules are used in all four components of the strip tracker, and are identified in Figure 1.15(b) as a double line.

The silicon strip tracker provides at least nine hits for all trajectories within $|\eta| < 2.4$. Four of these measurements provide information also on the coordinate parallel to the strip orientation. If fewer measurements are accepted, the tracker covers a range up to $|\eta| < 2.5$. Depending on the pseudorapidity, the material budget of the full CMS tracker system corresponds to $0.4-1.8$ radiation lengths. The biggest budget is reached in the transition region between barrel and endcaps at $|\eta| \approx 1.4$ [56].

The performance of the tracking system is best described by an investigation of reconstructed tracks and vertices. It is therefore described after a summary of the corresponding algorithms in Section 1.3.1.

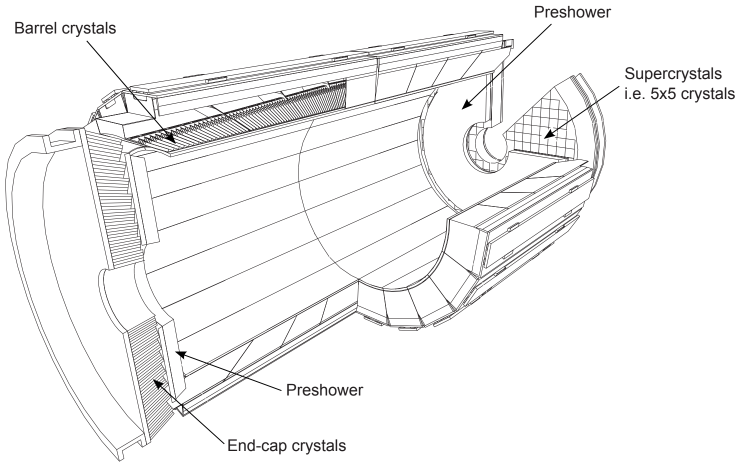


Figure 1.16: Cutaway view of the electromagnetic calorimeter of the CMS experiment [58, modified].

1.2.2.3. Calorimetry system

The calorimetry system is located outside of the tracker. Its task is to measure the energy of all particles produced in the pp collision, except for muons and neutrinos. To achieve this, the particles are stopped in dense material and the deposited energy is measured. One of the distinguishing features of CMS is the location of the calorimeters inside of the magnet coil (see Figure 1.14), which minimizes energy loss in non-instrumented material.

QCD interactions are likely to give rise to final state particles close to the beam line. Therefore, a coverage of the calorimeters up to high η values is crucial. At the same time this hermetic design allows for an accurate measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$, which depends on the reconstruction of (at least almost) all final state objects. In the central region up to $|\eta| = 3.0$, a separate electromagnetic and hadronic calorimeter is employed. A specific hadron forward (HF) calorimeter extends the coverage down to $|\eta| = 5.2$.

Electromagnetic calorimeter The ECAL employs 75 848 lead tungstate ($PbWO_4$) crystals to both induce electromagnetic showers and measure the deposited energy. Figure 1.16 shows a sketch of the ECAL. In the barrel region $|\eta| < 1.479$, the crystals have a cross section of 0.0174×0.0174 in φ-η [56]. Accordingly, 360×170 crystals are implemented along the φ and η directions, respectively. Their length of 230 mm corresponds to 25.8 radiation lengths X_0. The endcaps cover $1.479 < |\eta| < 3.0$. They consist of crystals with a fixed size of $28.62 \times 28.62\,\mathrm{mm}^2$ and a length of $24.7X_0$. The scintillating light is detected using avalanche photodiodes in the barrel region and vacuum phototriodes in the endcaps.

An additional preshower detector is installed at $1.653 < |\eta| < 2.6$. It helps to identify π^0, to discriminate electrons against minimum ionizing particles, and improves the position resolution. The preshower is a sampling calorimeter, built of layers of lead radiators followed by silicon strip detectors.

The energy resolution of the ECAL barrel was determined in an electron test beam [58, 59] and can be parametrized as

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + \left(\frac{12\%}{E/\mathrm{GeV}}\right)^2 + (0.3\%)^2 \ . \tag{1.26}$$

Measurements in 7 TeV data have been performed using electrons from $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ decays, yielding an energy resolution better than 2% for $|\eta| < 0.8$, and 2% to 5% elsewhere [58]. The energy resolution for photons with $E_\mathrm{T} \approx 60\,\mathrm{GeV}$, as are expected from H decays, is determined to lie between 1.1% and 5%.

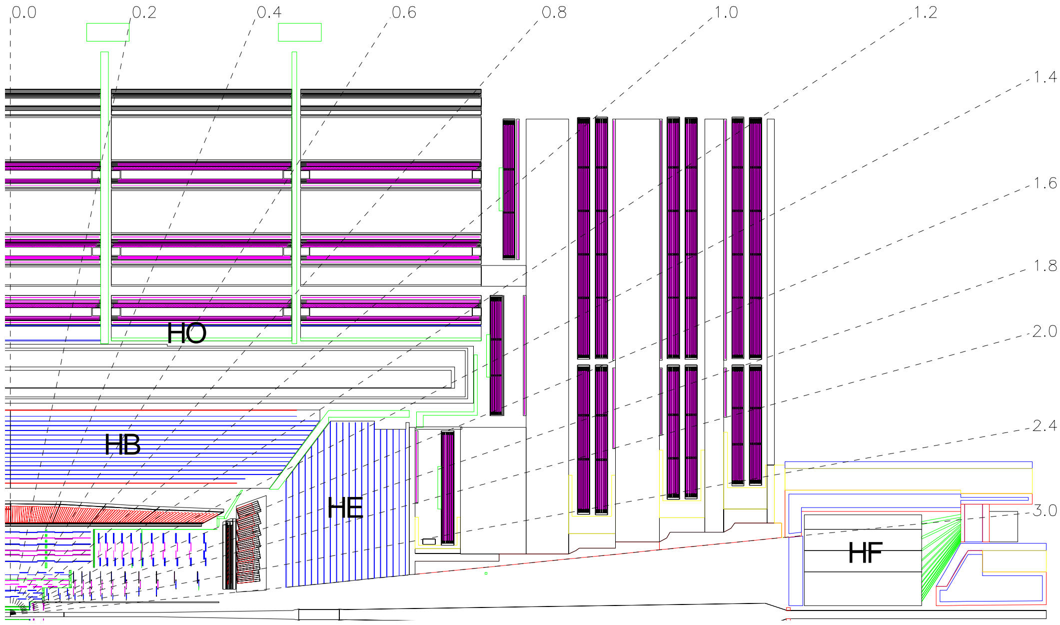
Hadronic calorimeter To measure the energy of hadrons in each event is not only important for the reconstruction of hadronic jets, but is also a crucial requirement for an accurate determination of the missing transverse energy (see Section 1.3.5). At CMS, the HCAL consists of four parts, which are labelled in Figure 1.17(a).

Hadron barrel (HB) ($|\eta| < 1.3$) and hadron endcap (HE) ($1.3 < |\eta| < 3.0$) make up the main part of the hadron calorimeter system and are located within the solenoid [56]. Both are built as sampling calorimeters, consisting of layers of absorber material to induce particle showers, followed by active material. Given the large magnetic field, brass is used as non-magnetic absorber material. The inner- and outermost layers of HB are made of steel to improve the structural integrity. Plastic scintillator tiles of 3.7 mm thickness are used as active material. The scintillating light is collected via wavelength-shifting fibers and measured by hybrid photodiodes. In HB, 14 layers of scintillator are employed, plus 2 additional layers in front of the first and behind the last absorber layer. HE consists of 18 layers. The size of the scintillator tiles defines the granularity of the detector. In the barrel it is given by $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$. The endcaps provide the same size up to $|\eta| < 1.6$. At higher $|\eta|$ values, the granularity is reduced to 0.17×0.17.

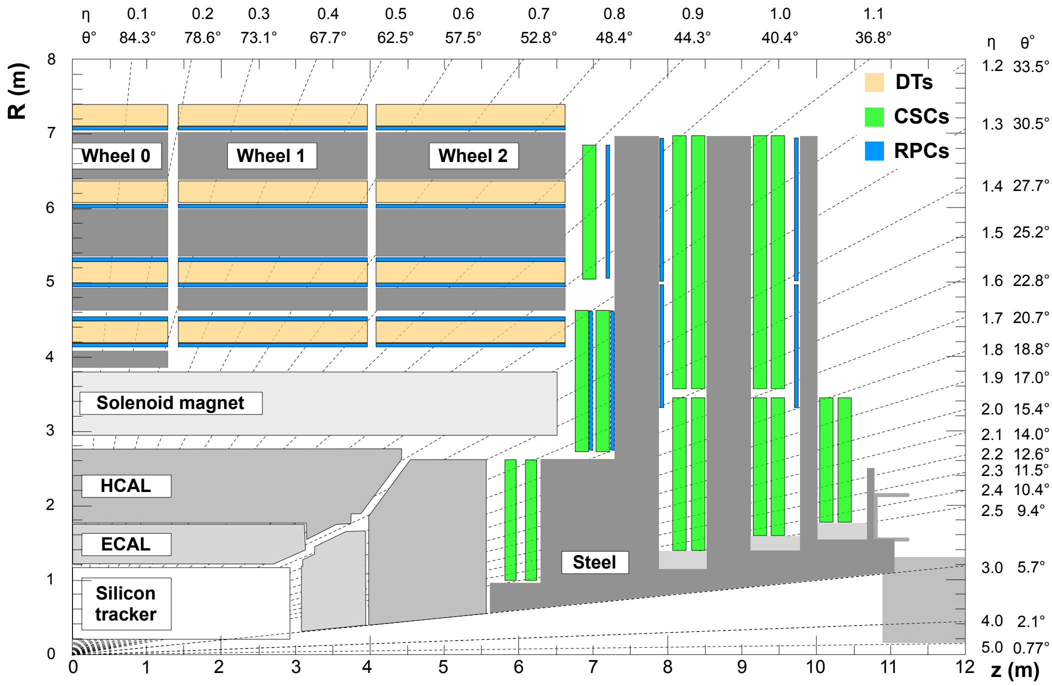
The thickness of HB and HE amounts to 5.8 to 10.6 interaction lengths, depending on $|\eta|$. The material of the ECAL corresponds to approximately one interaction length and has to be added to these numbers. This is not enough to fully contain hadronic showers. Therefore, the hadron outer (HO) is implemented outside of the magnetic coil. It is intended to act as a tail catcher of hadronic showers, and allows to discriminate muons from hadronic showers leaking into the muon chambers. HO consists of one or two layers of plastic scintillator with a granularity identical to HB. During LHC Run-I HO has not been used for event reconstruction.

Finally, HF extends the pseudorapidity range to $|\eta| < 5.0$. The large particle flux close to the beam-line enforces the use of radiation-hard technology. As absorber, HF implements steel. Quartz fibers are used as active material. Particles passing them generate Cherenkov light, which is captured by photomultiplier tubes. The fibers are aligned with the beam line and are bundled to form towers of $\Delta\eta \times \Delta\varphi = 0.175 \times 0.175$. The HF is crucial to measure forward jets produced in VBF Higgs production. It is also used to measure the instantaneous luminosity online, i.e. during data taking [61].

The energy resolution of the HCAL has been measured in a test beam of charged pions [62]. In the barrel, a stochastic resolution of $100\%/\sqrt{E}$ was obtained. A later measurement using pions, kaons, protons, and antiprotons, provided by the H2 beam-line at CERN, estimated the energy resolution of HB in combination with the electromagnetic barrel calorimeter [63]. Above energies of 5 GeV, a stochastic term of 84.7% was achieved. The constant term was measured to be 7.4%.



(a)



(b)

Figure 1.17: Cross section views of one quarter of the CMS detector. Both figures show similar information, but highlight either (a) the HCAL or (b) the muon system. Both images are modified; the originals are taken from (a) Ref. [56] and (b) Ref. [60].

1.2.2.4. Muon system

In pp collisions, muons are an important handle to identify interesting processes and separate them from the huge QCD multijet background. At CMS, a muon system making use of three different gaseous detector technologies is implemented within the magnet's iron return yoke. Its main objectives are to identify muons, measure their transverse momentum, and allow for triggering. Figure 1.17(b) shows the location of these detectors within CMS. In the barrel region, a magnetic field of about 2 T is contained in the yoke's iron, while the field in between is small. The endcaps reside in strong and non-uniform magnetic fields, and a significantly bigger occupancy is expected. This demands the use of more robust detector technologies in the endcap than in the barrel.

Drift tube (DT) chambers are used to cover the range $|\eta| < 1.2$ [56]. They are designed as simple, rectangular gas cells with a cross section of $13 \times 42\,\mathrm{mm}^2$. The anode wire in the middle has a typical length of 2.4 m. These chambers are arranged in stations, which consist of eight layers of chambers to measure r-φ, and four layers to measure the z coordinate. As can be seen in Figure 1.17(b), four of these stations are interspersed with the yoke. The outermost station does not allow for a z measurement. The stations are arranged in five wheels, which cover the full barrel range.

In each endcap, i.e. for $0.9 < |\eta| < 2.4$, four stations of cathode strip chambers (CSCs) are deployed. Each station consists of CSCs with six layers, where a layer is defined as a gas-filled gap between planes. Each plane carries cathode strips, and anode wires are strained in the middle between two planes. The CSCs are of trapezoidal shape with a length of up to 3.4 m. While the cathode strips are oriented radially and allow to measure φ, the wires are installed in azimuthal direction and give access to the r or η coordinate.

The third detector type used in the muon system are resistive plate chambers (RPCs). They are installed both in the barrel and endcap region and cover $|\eta| < 1.6$. RPCs are fast and thus allow for triggering on muons above a given p_T threshold. Six layers of RPCs are installed in the barrel, and three layers in the endcaps. Each RPC consists of two gas-filled gaps, each between two resistive plates. Readout strips between these two chambers sample the image charges induced on the resistive plates. The chambers are operated in avalanche mode to allow operation at high rates.

The performance of the individual muon subsystems was measured in 7 TeV data taken in 2010 [60]. The resolution of a single DT chamber is measured to be $80-120\,\mu$m, the CSCs reach $40-150\,\mu$m. As expected, RPCs have a worse spatial resolution of $0.8-1.2$ cm. A reconstruction efficiency of $95-98\%$ is reported for hits and track segments of genuine muons traversing the muon chambers. The time resolution, which is crucial to associate each hit to the correct bunch crossing, is better than 3 ns in each system. For the trigger system, CSCs and RPCs provide track segments with an efficiency of 96%, of which 99.5% are associated to the correct bunch crossing.

If muons are reconstructed offline using only information from the muon system, often called standalone muons, their p_T resolution is below 10% in the barrel region, and typically below 20% in the endcaps [64]. The momentum resolution in the muon system is limited by multiple scattering for muons up to 200 GeV, in the endcaps also by the complicated magnetic field [60]. Therefore, for muons used in physics analyses, inclusion of information obtained by the inner tracking system is crucial (see Section 1.3.3).

1.2.2.5. Luminosity

The value of the integrated luminosity, L, is a crucial quantity for many physics analyses. Following Equation (1.21), it allows to determine the cross section of a given process by counting the number of events, N, as

$$\sigma = \frac{N}{L} \,. \tag{1.27}$$

Equation (1.21) also describes the event rates that are expected to be produced within CMS, and which have to be handled by the trigger system (see Section 1.2.2.6). Therefore, the monitoring of the instantaneous luminosity $\mathcal{L}$ is not only important for the operation of the LHC machine, but also essential for CMS data taking.

To monitor the value of $\mathcal{L}$, the HF calorimeter is used [56, 61]. Each collision at the interaction point produces scattered particles that are observed by HF (see Section 1.2.2.3). Monitoring the occupancy of this detector allows to extract the current value of $\mathcal{L}$. The absolute calibration of this method is obtained by Van der Meer scans [61]. Alternatively, the instantaneous luminosity can be measured by a beam conditions monitor (BCM), called BCM1F [65], and since LHC Run-II also by the pixel luminosity telescope (PLT). These detectors use single-crystal diamonds or silicon pixel sensors, respectively, located close to the beam line [56].

The value of the integrated luminosity L is obtained either using HF, as explained above, or by exploiting the occupancy of the pixel tracker (see Section 1.2.2.2). The latter system has proven to be very stable over time and has an occupancy of below 0.1% at design luminosity [66]. This allows for a precise measurement of $\mathcal{L}$ by counting the clusters in the pixel tracker in a given time interval. As for the HF method, the absolute normalization is obtained from Van der Meer scans. Usage of Equation (1.23) finally allows to obtain the integrated luminosity L. The details of this method are given in Ref. [66].

While both measurements, using HF and the pixel tracker, give similar results, the pixel cluster counting method has smaller uncertainties and is recommended by the CMS collaboration. The uncertainty on the measured integrated luminosity is reported as 2.5% (syst.) + 0.5% (stat.) [66].

1.2.2.6. Trigger

The CMS detector produces data with a size of about 1 MB per event. At a collision rate of 40 MHz, this corresponds to a data output of approximately 40 TB/s. Such an amount of raw data cannot be handled by the computing farms responsible for the event reconstruction. At the same time, as Figure 1.6 shows, most of the events are not of interest for the CMS physics program. For these reasons, a trigger system is employed to discard the majority of the events before the data is permanently stored. This is often referred to as *online* selection, whereas all later steps are labeled *offline*.

CMS uses a two-layered trigger system. The first layer, called Level-1 (L1), reduces the event rate to below 100 kHz [67]. It is based on programmable, custom-built electronics. The system has to trigger a readout command within 4 μs, during which the full data of each sub-detector is buffered in pipelines on the front-end electronics. Of this time, only 1 μs is available for the trigger decision process itself, while the rest is used for signal transportation, etc. The L1 system uses information from the muon chambers and the calorimeters to reconstruct the so-called L1 objects: muons, electrons, photons, taus and jets [68]. Additionally, the calorimeter information is evaluated to estimate total and missing E_{T} sums as well as total and missing E_T sums contained in jets (H_T). This information is then combined and used by up to 128 trigger *algorithms*, each capable of triggering the readout of the event data. The algorithms can apply simple kinematic cuts on the individual L1 objects, e.g. on E_T, p_T, or η, but can also apply requirements on object correlations, e.g. having two objects close to each other.

As the instantaneous luminosity changes drastically over longer periods of data-taking, the menu of trigger algorithms is adopted when necessary. Within each LHC fill the value of $\mathcal{L}$ decreases, corresponding to smaller trigger rates. To ensure full efficiency of data-taking, i.e. to store as much data as possible, the chosen prescale column can be changed within a fill. A trigger is prescaled by scaling down the rate by an integer value, and therefore discarding a fraction of events. Each prescale column defines prescale integers for all algorithms, including the possibility to switch off algorithms. As prescales can have serious impact on physics analyses, this procedure is performed with caution.

The second layer of the trigger system is the high level trigger (HLT). Here, software-based algorithms have access to the full detector data and can perform more complex and thus more precise calculations. The reconstruction algorithms used at HLT are similar to the offline algorithms, with the main difference being the time constraint: At 100 kHz input rate, each event has a time budget of 175 ms [68]. This value is obtained using a computer farm that offered approximately 13 000 CPU cores at the end of 2012. Each HLT trigger *path* uses one or more L1 algorithm as seed, and selects events on multiple levels. At each level, the event can be discarded, before the next, generally more complicated analysis step is performed. An event is kept if at least one HLT trigger path selects it. The HLT reduces the event rate to an average value of 400 Hz, which is stored offline.

In 2012, the amount of data available for physics analyses was not limited by the trigger system nor by limitations on storing data on tape, but by the offline event reconstruction. It is performed on a computer farm at CERN generally within 48 hours after data taking. To allow for more data to be stored, additional 300−350 Hz of data were *parked*. This means that the data was stored to tape, but reconstructed only later, for example during longer shut-down periods of the LHC.

As part of this thesis, contributions to the design, optimization, and maintenance of L1 and HLT triggers for $\mathrm{H} \rightarrow \tau\tau$ searches have been made. Their performance have been measured as part of these tasks, and at the same time as input for physics analyses. The triggers for hadronic tau decays are thus explained in more detail in Chapter 2. A detailed description of the trigger system, including studies of its performance during LHC Run-I, is going to be published in Ref. [68]. Some of these studies are part of this thesis and are also presented in Chapter 2.

1.3. Event reconstruction and simulation

The data that is read out from the detectors and saved on tape is raw. This means that it consists of the signals describing the detector response to the physics event, but does not correspond to the physics objects themselves. Once stored, each event is analyzed to reconstruct the objects that most likely induced the measured signals. These fully reconstructed particles are then used as input for the final analyses. The event reconstruction is therefore a crucial part of the CMS physics program. It has to provide a reliable identification of particles, which is often quantified by the identification efficiency and the fake rate. The latter describes the chance of a different physics object to be misidentified as the object of interest. At the same time, the object's properties like momentum, energy, direction, and charge should be deduced with high precision. Here, the performance is often measured by its bias and resolution.

In general, event reconstruction starts with low-level entities, such as tracks or energy clusters in the calorimeters. These are then used to build more complex structures, and finally to construct candidates corresponding to physics particles. At CMS, most analyses make use of an approach that tries to connect the individual detector components at a very early stage of the reconstruction chain. This algorithm, called particle flow, is described in Section 1.3.2. The following sections give an overview of the necessary steps to reconstruct the full event.

Studies of events that have been simulated using a Monte Carlo (MC) method are crucial in various stages of the CMS physics program. They allow to study individual detector components even before the detector is built. Effects in the reconstruction of different particles can be investigated individually, without the need to understand their interdependence. Knowledge of the "true" physics process allows to disentangle signal from background events. And, most importantly, simulated events allow to compare predictions from theory with reality, i.e. the collision events recorded at CMS. The reconstruction algorithms used for real and simulated events are identical. More information on event simulation is given in Section 1.3.7.

Throughout this thesis, the term *data* is used to describe collision events that have been recorded by the CMS detector. Simulated events are not labeled data.

1.3.1. Tracks and vertices

Charged particles emerging from a collision are bent by the magnetic field. Their trajectory is therefore given by a helix that originates at the collision vertex. *Tracking* describes the process that reconstructs this helix. Knowledge of the helix parameters allows to derive the transverse momentum p_T of the particle, as the bending radius depends on p_T and the magnetic field strength.

The reconstruction of tracks and vertices at CMS is described in Ref. [69]. Tracking is based on an adaption of the Kalman filter [70–73]. Starting from a seed track, consisting of only a few hits close to the interaction point, the trajectory is extrapolated outwards. If a hit consistent with the track is found, it is associated to the track and the track parameters are updated. Once all hits that originated from the charged particle are known, the track is refitted twice, inside-out and outside-in. The final track parameters are set to the weighted average of these two results. This ensures that all hits are used to determine the track parameters at any given point. The tracks have to fulfill quality criteria and are otherwise discarded. Finally, the used hits are removed from the hit collection, and the procedure is repeated. By relaxing the necessary quality criteria for the seeds and the tracks with each iteration, it is ensured that high quality tracks originating from the hard interaction are reconstructed with best accuracy. At the same time the fake rate is reduced, as more complicated tracks are built with reduced combinatorics, i.e. less hits to consider.

Although electrons are not of direct interest for this thesis, it should be noted that the aforementioned tracking algorithm needs to be adapted for electrons. The non-Gaussian energy loss due to bremsstrahlung cannot be described by a conventional Kalman filter. The Gaussian-sum filter (GSF) technique [74] models these losses by a mixture of multiple Gaussians, and is thus used for electrons. Track reconstruction for all other particles rely on the method described above.

Primary vertices are reconstructed by clustering tracks that are likely to originate from the same interaction and using these tracks to fit the vertex position [69]. Tracks are clustered using a deterministic annealing (DA) algorithm [75]. The luminous region of pp collisions at the center of CMS has a typical length of 5−6 cm in z direction, while its transverse size is only 20−30 μm. Therefore, the z coordinate of the track's closest approach to the current beam spot is examined for clustering. In the DA approach, each track can be associated to multiple clusters (i.e. vertices) with a probability between 0 and 1. Given the large number of PU interactions (see Figure 1.13), this technique allows to separate vertices without a need for a fixed cut on the distance. All clusters with two or more tracks are fitted to estimate the vertex position. The fit also returns a covariance matrix and fit quality parameters. In this thesis, the vertex with the highest value of

$$\sum_{\text{tracks}} p_{\mathrm{T}}^2 \tag{1.28}$$

is identified as the signal vertex, whereas all other vertices are considered as PU.

The performance of the tracking system has been studied extensively in both simulation and data. Detailed information is available in Ref. [69]. The track reconstruction efficiency for isolated muons with $p_{\mathrm{T}} < 100\,\mathrm{GeV}$ is larger than 99% while retaining a negligible fake rate. For charged hadrons, such as $\pi^{\pm}$, nuclear interactions degrade the tracking performance. Pions with $p_{\mathrm{T}} = 1\,\mathrm{GeV}$ are reconstructed with at least 90% efficiency, depending on η. Below $\approx 700\,\mathrm{MeV}$, the efficiency drops significantly. The track fake rate is estimated to be below 3% for pions with fixed momenta of $p_{\mathrm{T}} = 1$ and $10\,\mathrm{GeV}$. A resolution of approximately 100 μm and 25 μm is reached for these pions in the transverse impact parameter, respectively. The p_{T} resolution is found to be $< 1\%$ in the central region and $< 3\%$ in the endcaps.

If three or more tracks are used to reconstruct a PV, the reconstruction efficiency is very close to 1. The spatial resolutions are strongly dependent on the number of tracks associated to the vertex, and the p_{T} of these tracks. In 7 TeV data, the resolutions are found to be as low as 10 μm or 12 μm for the transverse or longitudinal coordinates, respectively, for interesting pp collision vertices [69].

1.3.2. Particle flow

Most analyses performed by the CMS collaboration rely on an integral approach called particle flow (PF) [76–79]. Its goal is to combine information from all sub-detectors in order to reconstruct individual stable particles, namely electrons, muons, photons, charged hadrons, and neutral hadrons. These *PF candidates* are then used to build jets, hadronic tau decays, $E_{\mathrm{T}}^{\mathrm{miss}}$, and other relevant quantities. The CMS detector fulfills the necessary preconditions for this approach. The high granularity of the ECAL allows to disentangle energy deposited by a charged particle from close-by photons. Even more importantly, the strong magnetic field and the high-resolution tracker allow to reconstruct tracks with high precision even at small transverse momenta. The PF algorithm ensures that the excellent momentum resolution of tracks is propagated to more complex objects.

The particle-flow event reconstruction uses tracks in the inner tracker, muon tracks, and calorimeter clusters as starting points. These components are linked if they are consistent with each other, i.e. match in space and energy. Each of these blocks is then independently

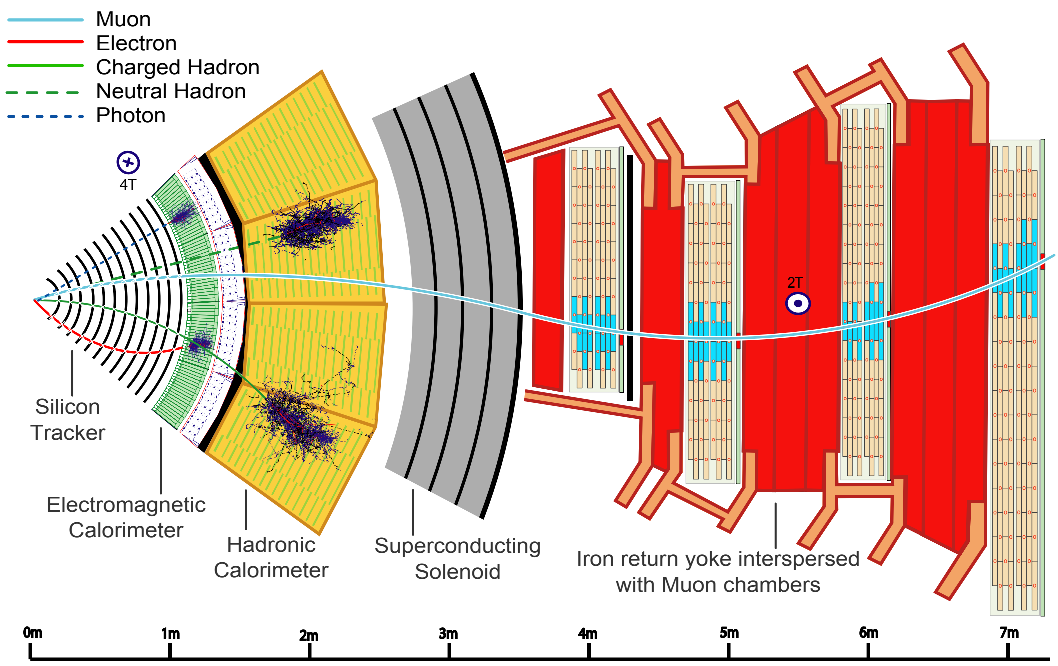


Figure 1.18: A slice of the CMS detector in the r-φ plane [81, modified]. Starting from the interaction point on the left, different stable particles are depicted as they traverse the detector layers. Charged particles are reconstructed as tracks in the silicon tracker and are drawn with solid lines. Neutral particles (dashed lines) are not bent by the magnetic field. Depending on the dominant interaction, all particles except muons shower either in the electromagnetic or the hadronic calorimeter.

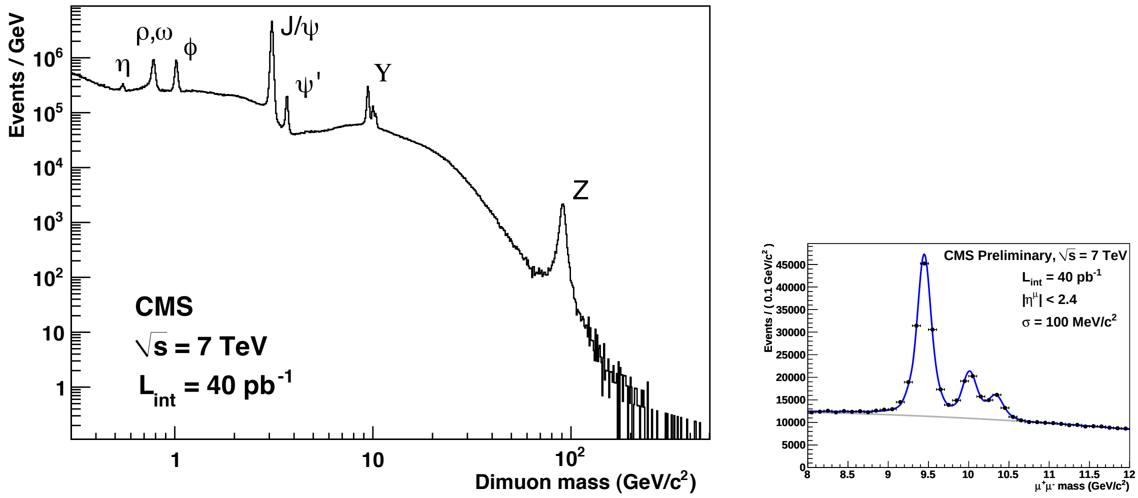
analyzed and a particle hypothesis from the list above is created. Multiple particle candidates can be created from a block, if one particle alone is not sufficient to explain the detector response. As an example, a track can be used to estimate the energy deposit of a particle in the ECAL. In case of significantly larger ECAL signals, the remaining energy can be assigned to a photon or neutral hadron.

Performance studies of the PF algorithm are summarized in Ref. [80]. PF reconstructs above 95% of a jet's energy, while calorimeter jets reconstruct only up to 80%. This means that significantly smaller corrections have to be applied, and the energy and angular resolution of jets are improved up to a factor 3. The hadronic tau reconstruction at CMS, which is described in Section 1.3.6, relies on PF, which allows to reconstruct the tau decay channel. The relative $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is improved by almost a factor 2 [76] regardless of its true value.

1.3.3. Muons

According to Bethe's formula [27, Chapter 32], the energy loss of muons with typical LHC energies is small while they traverse the CMS detector. Therefore they are not stopped within the detector, as is shown in Figure 1.18. Nonetheless, muons deposit energy in the active material of the sub-detectors and thus give rise to distinct signatures. Below a few GeV, the bending of the magnetic field is so strong that muons do not reach the muon chambers. Above this threshold, muon reconstruction is based on track segments both in the inner tracker and in the muon system.

To achieve the best possible efficiency, muon reconstruction can begin either with standalone



(a) (b)

Figure 1.19: Spectrum of the invariant mass of dimuon events collected in 2010 [64]. Using a loose double-muon trigger, CMS is capable to observe resonances over a wide energy range. The zoom (b) shows that the three $\Upsilon(n\mathrm{S})$ peaks can be clearly resolved.

muon tracks (outside-in approach) or with tracker tracks (inside-out approach) [64]. In the outside-in approach, standalone muon tracks are extrapolated to the inner tracker. If an inner track can be matched to the muon, a Kalman-filter fit is employed to reconstruct the muon kinematics. Muons reconstructed this way are labeled *global muons*. While the muon chamber signals are crucial for muon identification for all muon momenta, the momentum resolution of global muons is better than the track-based resolution only for high momenta $p_\mathrm{T} \gtrsim 200\,\mathrm{GeV}$.

Muons reconstructed by the inside-out approach are labeled *tracker muons*. Here, at least one short muon track has to match the inner track. As muons do not have to pass all layers of the muon system, tracker muons are more efficient than global muons for $p_\mathrm{T} \lesssim 5\,\mathrm{GeV}$. If the same tracker track is used by both a global and a tracker muon, these two are merged into one muon candidate.

Many physics analyses rely on a robust selection of high-quality muons. For this purpose, a *tight muon* selection is defined. Muons with this quality flag must be reconstructed as global muon, with a track fit that includes at least one muon-chamber hit and has a normalized χ^2 smaller than 10. At the same time, the muon has to be reconstructed as tracker muon with hits in at least two muon stations, more than 10 hits in the inner tracker including a minimum of 1 pixel hit, and a transverse impact parameter $|d_{xy}| < 2\,\mathrm{mm}$. The longitudinal impact parameter $|d_z|$ has to be smaller than 5 mm. Both impact parameters are calculated with respect to the leading primary vertex (see Section 1.3.1). Tight muons are reconstructed in $\mathrm{Z} \rightarrow \mu\mu$ decays ($p_\mathrm{T} > 20\,\mathrm{GeV}$) with an efficiency of about 96%. About 3% of global muons are lost in the tight muon selection, which at the same time reduces the amount of muons that are faked by hadronic punch-throughs from 5.4% to 0.2% [64].

Muons are reconstructed with high accuracy at CMS. Figure 1.19 shows a dimuon mass spectrum recorded with a low-threshold double-muon trigger in 2010. Resonances are precisely reconstructed over a wide mass range. The good mass resolution of $\leq 100\,\mathrm{MeV}$ allows to separate even the $\Upsilon(1\mathrm{S})$, $\Upsilon(2\mathrm{S})$, and $\Upsilon(3\mathrm{S})$ resonances. Individual tight muons in the range

$20 < p_T < 100$ GeV, as is typical for Z or H decays, are reconstructed with a p_T resolution of about 2% [64].

1.3.4. Jets

Most inelastic pp interactions give rise to partons (quarks or gluons) in the final state. These carry color charge and are subject to QCD confinement. Therefore they fragment and hadronize into multiple stable and unstable hadrons. This collimated spray of particles is called a jet. It often also contains other particles like photons or leptons, which originate from radiation processes or hadron decays.

As partons are not accessible directly by the detector, jets can be seen as the link between theoretical description and experimental observation of partons in the final state. Jet fragmentation models provide the theoretical description and allow to estimate the parton kinematics from the observed jet parameters. Particle detectors can detect the stable particles within a jet. It is important to note that neither "parton" nor "jet" are unambiguously defined in the context of QCD. A proper definition is essential though, if observations are to be compared with theoretical predictions. The concepts and approaches feasible for the LHC are reviewed in Ref. [82].

Jet reconstruction at CMS is commonly based on PF objects. Most analyses make use of the *anti-*k_t algorithm [83]. In this approach, objects are sequentially combined if a specific distance measure is small. It is infrared and collinear safe, meaning that neither radiation of soft particles nor the collinear splitting of a constituent has impact on the number of jets or their parameters. The analyses documented in this theses use jets reconstructed by the anti-k_t algorithm with a distance parameter $R = 0.5$. The algorithm is executed using the *FastJet* package [84].

To obtain an estimate of the parton energy from the reconstructed jet, jet energy corrections (JECs) have to be applied. CMS uses a factorized approach, where each factor is intended to correct for a different effect [85]. First, an offset term subtracts energy that does not originate from the hard scattering process, but is caused by effects like pileup or electronic noise. Second, scale factors dependent on η and φ are applied. These are derived from MC and ensure that the energy response is flat over the whole detector. After this step, the corrected energy of a jet should correspond to the original parton's energy. A third factor is applied only on data events and removes residual differences in the jet response between data and MC. Additional corrections are provided by the CMS collaboration, but are not mandatory and not used in this thesis. The values of the correction factors used for analyses of 2012 data are given in Ref. [86]. It also shows the uncertainties introduced by the individual stages of JEC. While they depend on η and φ, the total uncertainty on the JECs lies between 1 and 5%. The p_T resolution of PF jets has been estimated in 7 TeV data as 5−15% [85].

Energy from PU interactions is statistically distributed over the whole detector. Thus, it does not only affect the measurement of jet kinematics, but also gives rise to additional jets that do not correspond to a parton from the hard interaction. These so called pileup jets can fake parton jets, especially when multiple objects from different vertices overlap and merge into one single jet. Identification and removal of these jets is crucial for analyses that categorize events by the number of jets, search for VBF signatures, or rely on PU insensitive E_T^{miss}. Most Higgs searches rely on these features, including the $H \to \tau\tau$ search presented in Chapter 3.

An algorithm to separate parton jets from PU jets has been developed and used in all CMS

Higgs searches. This so-called PU jet ID [87] implements a multivariate analysis (MVA). In general, MVAs use machine learning techniques to analyze a set of input variables. Their result is most often a single discriminant that combines the discrimination power of these input variables. In case of the PU jet ID, the MVA investigates variables related to the vertices of charged PF candidates as well as the shape of the jet. The former exploits that PU jets mostly consist of multiple overlapping jets, and tracks therefore point to more than one vertex. The latter includes the neutral energy, which is not uniformly distributed within a PU jet. The MVA is trained on Z+jets MC events independently in four $|\eta|$ regions. For central jets ($|\eta| < 2.5$), over 85% of PU jets are removed, while the efficiency for quark and gluon jets is above 99%. In the forward region ($3.0 < |\eta| < 5.0$) a PU removal of 40% is achieved at a signal jet efficiency of $\approx 80\%$. These values are obtained for jets with $30\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$ [87].

Jets originating from b quarks are identified using the combined secondary vertex (CSV) b-tag algorithm [88]. It exploits the long lifetime of B mesons by combining information from a possible SV, track-based lifetime information, and further properties like the number of tracks in the jet. Working points are provided, allowing to select a combination of efficiency and light-quark fake rates suitable for the given analysis.

1.3.5. Missing energy

Missing transverse momentum is defined as the negative vector sum of all reconstructed particles, projected on the transverse plane,

$$\vec{E}_\text{T}^\text{miss} := \vec{p}_\text{T}^\text{miss} = -\sum_i \vec{p}_{\text{T},i}\,. \tag{1.29}$$

Its absolute value is denoted E_T^miss or $\not{E}_\text{T}$ and referred to as missing transverse energy (MET). Neglecting mismeasurements, E_T^miss corresponds to the energy carried away by objects invisible to the detector, such as neutrinos.

In the particle flow approach, the sum in Equation (1.29) runs over all PF candidates. This also means that all uncertainties in the reconstruction of individual objects in the event are accumulated. Accordingly, E_T^miss is very sensitive to experimental effects like PU, jet energy resolution, or acceptance limits in the individual detectors. Different corrections are provided to mitigate the impact of these effects, most importantly the jet energy corrections and pileup. They are described and studied in Ref. [89].

In events with small hadronic activity and typical PU for 2012 data, the biggest contribution to the E_T^miss resolution originates from PU. The analysis described in Chapter 3 therefore uses an algorithm called *MVA* E_T^{miss} to reduce this impact [89, Chap. 7]. It utilizes the pileup jet ID described in Section 1.3.4. Events with well reconstructed vector bosons (e.g. $Z \to \mu\mu$) are investigated. Defining the momenta of the vector boson as $\vec{q}$ and the hadronic recoil against this boson as $\vec{u}$, momentum conservation in the transverse plane requires

$$\vec{q}_\text{T} + \vec{u}_\text{T} + \vec{E}_\text{T}^\text{miss} = 0\,. \tag{1.30}$$

In the presence of PU the resolution of $\vec{u}_\text{T}$ is diluted, which directly worsens the E_T^miss resolution. For this reason corrections are applied to both the direction and the magnitude of $\vec{u}_\text{T}$, obtained by training of two individual boosted decision trees (BDTs). The variables used are different combinations of PF candidates, depending on their charge, their association to the selected PV or to a PU vertex, and on the PU jet ID of their associated jet.

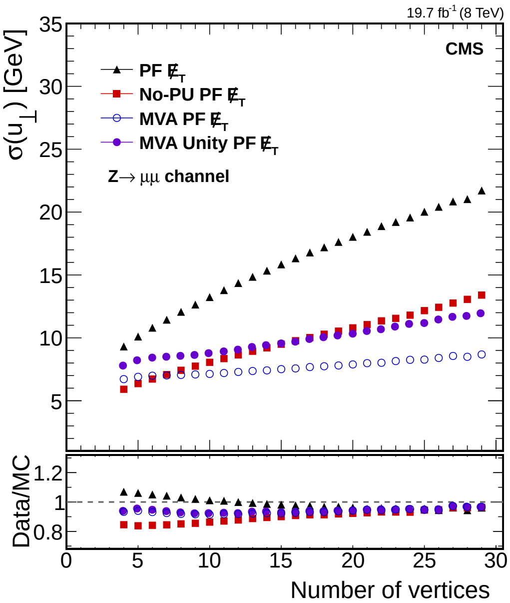


Figure 1.20: Resolution of the perpendicular component of the hadronic recoil $\vec{u}$ as function of the number of vertices in the event, as measured in $Z \rightarrow \mu\mu$ events [89]. The default PF $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown as black triangles, the MVA $E_{\mathrm{T}}^{\mathrm{miss}}$ that is described in the text as blue open circles.

The impact of this regression technique is shown in Figure 1.20. The symbol $u_{\perp}$ describes the component of $\vec{u}_{\mathrm{T}}$ that is perpendicular to the vector-boson direction, and is therefore mostly independent of a correct measurement of the hadronic recoil momentum. For a typical value of 20 PU vertices, the resolution $\sigma(u_{\perp})$ is halved with respect to the default PF $E_{\mathrm{T}}^{\mathrm{miss}}$. Also, the dependence on the amount of PU is drastically reduced. This improves the sensitivity of the search for $H \rightarrow \tau\tau$, which is described in Chapter 3 and strongly depends on $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, by about 20%.

1.3.6. Tau decays

As described in Section 1.1.4, tau leptons decay either into leptons or hadrons. As neutrinos cannot be observed in collider experiments, a separation of leptonic tau decays from prompt leptons is challenging. In principle it is possible by observation of a kink in the measured track. Given the tau's short lifetime, in LHC collisions taus decay well within 1 cm of the interaction point and thus within the beam pipe. Accordingly, such a kink cannot be observed directly at CMS. It is possible nonetheless to resolve impact parameters between the lepton track and the vertex of the primary interaction. Another handle to distinguish prompt leptons from leptonic tau decays is the two-neutrino system, which is only evident by missing energy. As such missing energy can have various sources, including all kind of mis-measurements, a separation of leptonic tau decays and prompt leptons remains challenging at the LHC. For these reasons leptonic τ decays are not explicitly reconstructed at CMS. They are treated identically to prompt electrons and muons.

Hadronic tau decays make up 64.8% of all decays. The hadrons in the final state make it difficult to differentiate τ_{h} from quark or gluon jets, which have a similar signature in the detector. The hadronization of quarks and gluons originating directly from the hard interaction gives rise to jets that typically have a higher hadron multiplicity than tau decays. Addition-

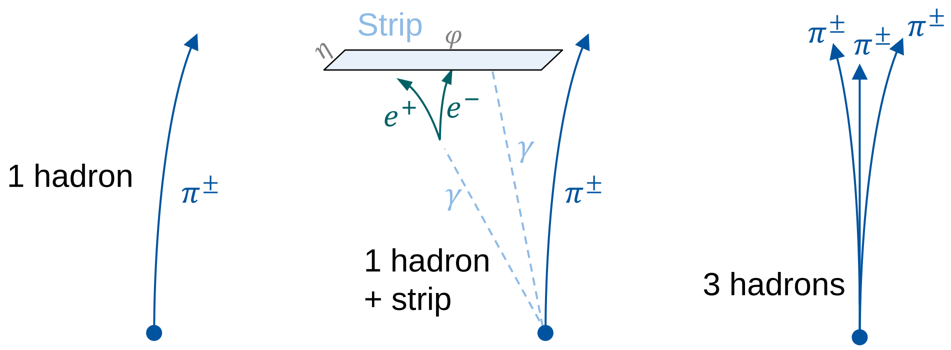


Figure 1.21: Sketch of the hadronic decay modes reconstructed by the HPS algorithm. Photons and e^+e^- pairs emerge from π^0 decays and are clustered in strips in the η-φ plane. An additional decay mode consisting of one $h^\pm$ and two strips is not shown.

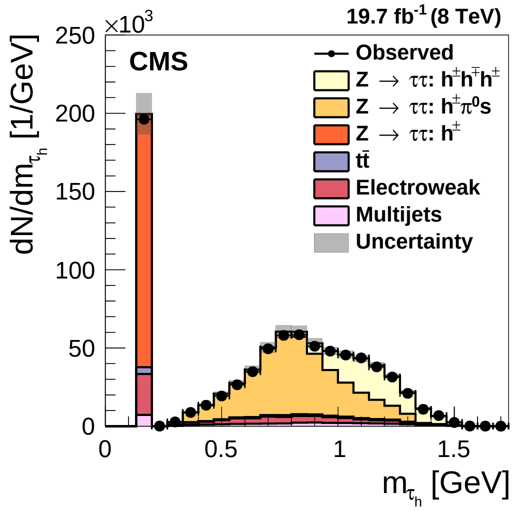
ally these jets show a wider angular spread of the final state objects. These two items, particle multiplicity and cone size, are the main handles to separate tau decays from quark or gluon jets. While $\tau_{1\pi}$ are expected to have a non-zero impact parameter with respect to the primary vertex, a higher multiplicity of charged tracks in the final state allows to reconstruct the tau decay vertex. For $\tau_{3\pi}$ this offers an additional handle to suppress quark and gluon jets.

Reconstruction of hadronic tau decays at CMS relies on the precise reconstruction of individual particles by the PF algorithm. In a bottom-up approach, individual PF candidates are aggregated to form one τ_h candidate. This allows not only to distinguish them from quark or gluon jets, but also to reconstruct the τ decay mode and analyze them independently. The method used is called hadrons plus strips (HPS) algorithm. A detailed description as well as performance studies have been published in Ref. [90]. This section will summarize the aspects important to this thesis.

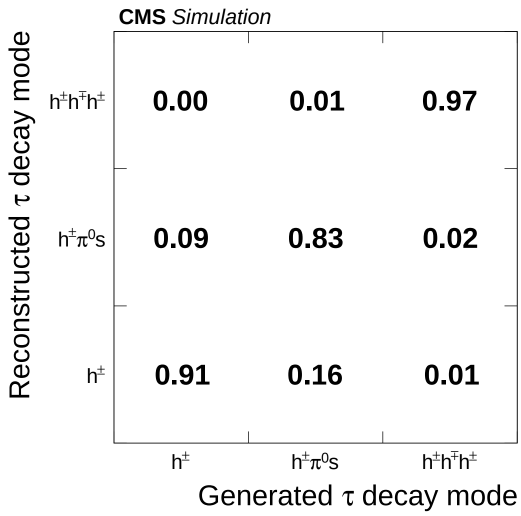
The reconstruction starts by identifying charged hadron candidates ($h^\pm$) with certain quality criteria within a jet. Neutral pions, which can occur in the tau decay, decay further into photons, $\pi^0 \rightarrow \gamma\gamma$. With a high probability, these photons convert to e^+e^- pairs within the CMS tracking system. For this reason γ, e^+, and e^- candidates are clustered into *strips* with a size of 0.05×0.20 in $\eta \times \varphi$. The larger extension in φ is motivated by the bending of the electrons and positrons in the magnetic field. A strip is identified as π^0 candidate if its transverse momentum is above 2.5 GeV.

From these constituents τ_h candidates are built. Figure 1.21 shows the decay modes considered by the HPS algorithm: $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$, and $h^\pm h^\pm h^\pm$. This corresponds to 88% of all hadronic tau decays, or 57% of all tau decays (see Table 1.3). As separation of $h^\pm\pi^0$ and $h^\pm\pi^0\pi^0$ is challenging, these decay modes are often combined. The chosen category has to be compatible with a tau decay and therefore must fulfill additional quality criteria. The invariant mass of the constituents has to match the intermediate hadronic resonance. For the 3-prong decay, the tracks have to originate from the same vertex, and have a combined charge of ± 1. For the $h^\pm$ mode, no additional quality criteria are applied. If more than one decay mode hypothesis remains, the one leading to the highest $p_T(\tau_h)$ is chosen.

Figure 1.22 shows the performance of the decay mode identification. The invariant mass of the hadrons in the final state of a hadronic tau decay is labeled τ_h mass. Its spectrum (Figure 1.22(a)) shows a good agreement between data and MC. The different masses and



(a) Invariant mass of the τ_h constituents.



(b) Migration between generated and reconstructed τ_h decay modes.

Figure 1.22: Performance of the decay mode reconstruction by the HPS algorithm. (a) The invariant mass spectrum is corrected for small differences in the τ_h energy scale between data and MC. The individual decay modes show shapes as expected by the intermediate resonances. (b) shows the migration between generated and reconstructed decay modes. Both figures are taken from Ref. [90], which also explains the applied event selection.

widths of the respective intermediate resonances are clearly visible. For the $h^\pm$ decay mode the τ_h mass is equal to the $\pi^\pm$ mass. The migration matrix in Figure 1.22(b) shows that for about 90% of taus the correct decay channel is identified. While the number of prongs can be reconstructed with high accuracy, identification of π^0 is more challenging.

Once the τ_h candidate is reconstructed, discriminators against quark and gluon jets, electrons, and muons are computed. Tau decay products are collimated at typical LHC energies, and tau leptons in signal events are expected to appear isolated from other particles. For these reasons, isolation variables are the main handle to suppress background from quark and gluon jets. A simple measure for this purpose is the scalar p_T sum of all charged particles and photons within an isolation cone of $\Delta R = 0.5$ around the τ_h axis. Objects that have been used in the τ_h reconstruction, or whose tracks are incompatible with the signal vertex, are not considered in this sum. While this reduces the impact of PU by charged particles, energy depositions of neutral PU particles are removed on a statistical basis. The isolation variable can thus be given as

$$I_\tau = \sum_{\substack{\text{charged} \\ \Delta z < 0.2\,\text{cm}}} p_T + \max\left\{0, \sum_{\gamma} p_T - \Delta\beta\right\} \tag{1.31}$$

$$\text{with} \quad \Delta\beta = 0.46 \sum_{\substack{\text{charged} \\ \Delta z > 0.2\,\text{cm}}} p_T\,. \tag{1.32}$$

Compatibility with the signal vertex is tested by the longitudinal distance Δz between the vertex and the track. The factor 0.46 in the $\Delta\beta$ correction describes the energy ratio of neutral to charged particles introduced by PU interactions.

A more sophisticated discriminator against quark and gluon jets combines the isolation variable with other information in an MVA approach. It includes the tau decay mode, the transverse impact parameter and its significance, as well as the τ flight length as given by the difference between the primary and secondary vertex, if available. This discriminator was not available for analyses of the LHC Run-I data set. The impact of flight length variables for the τ_h identification is discussed in Section 4.1.1.

Electrons and muons can fake taus in the $h^{\pm}$ decay mode, or in the $h^{\pm}\pi^{0}$ decay mode if a photon is radiated. Discrimination against electrons is based on multiple variables sensitive to the energy distributions in the ECAL, bremsstrahlung along the track, and particle multiplicity in the showering process. They are combined in a BDT. As discrimination of τ_h against muons is less challenging, a robust, cutoff-based approach is used in this thesis. A loose working point is defined by requiring that there is no muon track segment ($\geq$ 2 stations) within $\Delta R = 0.5$ around the τ_h, and the sum of the deposited energies in the ECAL and HCAL is less than 20% of the momentum of the tau's leading track. To pass the tight working point, hits in the two outermost muon stations within $\Delta R = 0.3$ around the τ_h direction must not be present, additionally to the requirements of the loose working point.

The performance of the HPS algorithm has been studied both in MC and data [90]. Requiring the isolation variable I_τ to be below 2.0 GeV (1.0 GeV), an efficiency to identify τ leptons of 49% (41%) has been measured in $Z/\gamma^* \to \tau\tau$ MC events. The corresponding misidentification rate from jets is below 1% (0.6%). The loose working point of the MVA variable against electrons has an efficiency for taus of 91%, while 0.4% of electrons from $Z/\gamma^* \to ee$ events are misidentified as τ_h. Muon discrimination at the tight working point is 99% efficient for taus, and removes all but 8×10^{-4} muons from $Z/\gamma^* \to \mu\mu$ events. Measurements in data confirm the MC predictions.

For $p_T(\tau_h) < 500$ GeV, the energy resolution of τ_h is about 15%. The τ_h energy scale, which is defined as the reconstructed energy over the true energy of the visible tau decay products, is crucial for accurate reconstruction of resonances decaying into taus. It has been observed to be $\neq 1$ and is therefore corrected. A correction factor for the $h^{\pm}\pi^{0}$and $h^{\pm}h^{\pm}h^{\pm}$ decay modes is obtained by investigation of the m_{τ_h} distribution, i.e. the invariant mass of the τ_h constituents. Altering the energy scale modifies m_{τ_h}, and the energy scale is set to the value that leads to the best data-MC agreement. Obviously this method is not feasible for $\tau \to h^{\pm}\nu_\tau$ decays. An alternative method uses the invariant mass of $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events for the same purpose, which also allows to obtain an energy scale correction for 1-prong decays.

1.3.7. Monte Carlo simulation

Simulation of a pp collision starts by defining the hard interaction process. This is done by external MC simulation tools, which use different methods to model the matrix element and therefore the differential cross section of the physics process. Accuracy can be increased by including higher orders in the pertubative description of the process. In this thesis, MC samples are used that have been generated with PYTHIA 6.4 [91], POWHEG 1.0 [92–95], and MADGRAPH 5.1 [96]. The two latter generators are interfaced with PYTHIA to model parton shower and fragmentation processes. Additional ISR and final state radiation (FSR) can be added to the event, as well as objects originating from other parton-parton collisions in the same pp interaction, called *underlying event*. The Z2* tune in PYTHIA is used for this purpose in the 8 TeV samples [97].

Additional proton-proton collisions are simulated with PYTHIA and added to the event, to account for the presence of PU. As the PU distribution depends on the LHC running parameters and thus is not known before the data has been collected, MC samples are produced with a predefined PU distribution. By comparison of the simulated distribution with the one of the collision data (see Figure 1.13), a weight distribution can be calculated. Each MC event is then reweighted by the ratio of real and simulated events with the same amount of PU.

Decays of tau leptons are simulated by TAUOLA [98], which is interfaced to all used event generators. By including most known decay channels and results from precision measurements, it allows to model tau decays with high accuracy. The kinematic distributions of the tau decay products are improved by consideration of the tau polarization.

The output of the previous steps is a list of particles and their four-momenta. These are now further simulated by GEANT4 [99]. Weak decays of semi-stable particles are modeled by this program as well as the interaction of all particles with the detector material. The resulting energy depositions are subject to a simulation of the detectors electronics. The detector response obtained this way is then subject to the event reconstruction described in the previous sections. The software used for this is identical for both simulated events and data.

Chapter 2

Trigger for $\mathrm{H} \rightarrow \tau\tau$ analyses

Investigation of CMS collision data starts while the data is recorded. The event selection performed by the trigger system (see Section 1.2.2.6) discards events irrevocably. It is therefore crucial to define a trigger configuration that suits the intended physics analyses before data is recorded. This mandatory online event selection must not be too restrictive, as it should not eliminate interesting collisions. At the same time it has to be tight enough to discard most background events, which in turn allows to record more interesting collisions with different event signatures while keeping the event rate within manageable limits. The main task of trigger development at CMS is to find a compromise between these aspects.

The trigger configuration is based on a set of selection chains, each of which is specialized for one or more physics analyses or for monitoring of detector components. In the L1 trigger system these chains are called *algorithms*. At HLT, the equivalent is labeled *path*. Both terms have previously been introduced in Section 1.2.2.6. Each HLT path is seeded by one or more L1 algorithms, meaning that it is only executed if the event passed that algorithm. In that respect, L1 algorithms can be considered as part of the HLT path. The term *trigger menu* comprises a specific set of deployed trigger paths and the settings for the trigger system at a given time.

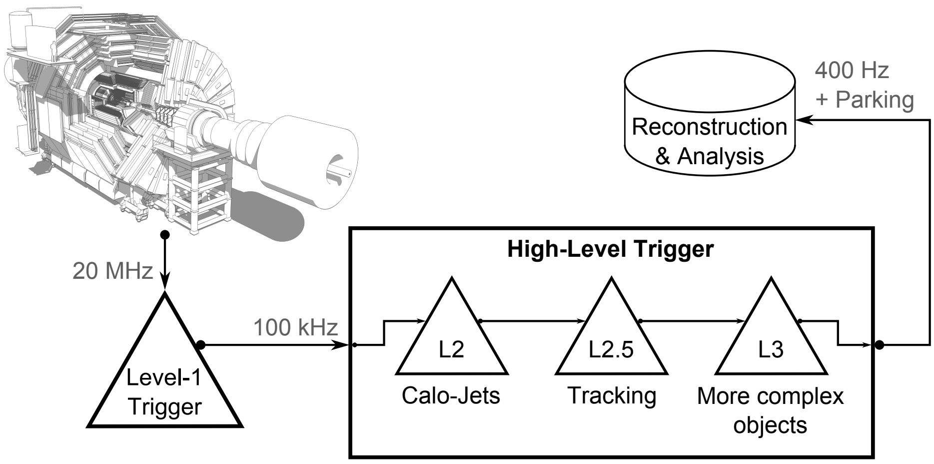


Figure 2.1: Trigger sequence at CMS. The information collected by the detector components is first analyzed by the L1 system. The HLT executes several layers of increasing complexity; shown here are the three layers of τ_{h} triggers as an example. Events passing both trigger systems are stored, reconstructed, and made available for analysis. Drawing of the CMS detector from Ref. [57].

Figure 2.1 shows the trigger sequence that is executed for each event, showing details of a τ_h trigger path as an example. The flexibility of the high level trigger allows to begin with a coarse characterization of the event, followed by more and more complex investigation. Each of these levels (L2, L2.5, and L3) reconstructs parts of the event, which are then analyzed. Execution of the path is stopped at any level if the event does not meet the imposed requirements. This ensures that the full computing time is spent only on events with a signature resembling the desired signal. For the same purpose, the intermediate reconstruction steps can be shared by multiple trigger paths. An event is finally discarded if it fails all paths. All selected events are stored on tape for offline reconstruction and analysis.

Many trigger paths exploit more than one final state particle. This allows to drastically reduce backgrounds while keeping the selection on each individual particle loose. This is especially important for complex final state objects like τ_h. The selection chain for one final state object in the trigger is denoted by *leg*. Paths that use two or more different legs are called *cross triggers*. The trigger menu contains both single object triggers, which can be used for a broad variety of analyses, and cross triggers, which are often optimized for specific analyses.

During data-taking the trigger system is permanently monitored. This ensures that sudden changes, for example in the rate of some algorithms, are observed quickly and appropriate actions can be taken. The data quality monitoring (DQM) includes information from both L1 and HLT. It ensures that the data are examined shortly after they are collected. DQM typically investigates individual LHC runs, i.e. data taken within several hours. Finally, data validation is performed on larger data samples. It may include complex analysis methods and enough signal events to obtain statistically significant information even for rare processes. Should problems emerge during validation, the data might be recovered by adaptions of the reconstruction software. Additionally, the experiment can be reconfigured immediately to ensure that only a minimal amount of data is affected.

On a longer timescale, the trigger performance is analyzed and optimized with respect to specific analyses. As small improvements in the trigger configuration can have sizable impact on the physics reach, trigger paths often have to be adapted. For these reasons, most analyses use data collected with multiple trigger menus.

Monte Carlo simulation also includes trigger information. The L1 hardware is emulated after GEANT4 simulation. The software-based HLT decision is obtained as for data events. As slight differences in the trigger configuration used in data and MC cannot be avoided, MC events are often reweighted such that the trigger performance corresponds to the one in data. This involves performance measurements of trigger paths both in data and MC. Knowledge of the trigger efficiencies as a function of p_T and η of the signal particles is the minimal requirement.

As part of this thesis, the τ_h reconstruction at the L1 trigger has been examined (Section 2.1.1). Taus at HLT are described in Section 2.1.2. They have been monitored by DQM and validation tools, which were developed and maintained. A measurement of the trigger efficiency in 2012 data (Section 2.2.1) was performed. The performance of tau triggers in 2012 is summarized in Section 2.2.2. The necessity of a $\tau_\mathrm{h}\tau_\mathrm{h}$ path as part of the data parking campaign is described in Section 2.3. This path was developed, proposed, and monitored in the scope of this thesis.

2.1. Structure of ditau triggers

2.1.1. Tau leptons at Level-1

Separation of hadronic tau decays from quark and gluon jets is challenging because of the similar signatures. Identification of hadronically decaying tau leptons at the Level-1 trigger is based solely on calorimeter data. Information from the tracker is not available, as its large amount of data cannot be handled on the available time scale. Tracking information is the key component of offline τ_h identification (see Section 1.3.6). Additionally, the L1 hardware does not allow to access calorimeter data with full granularity, but enforces the usage of merged readout channels. For these reasons, separation of τ_h from quark or gluon jets can only be rudimentary at L1.

The hardware used in the L1 system that is relevant to τ_h identification is presented in detail in Appendix A. It defines boundaries for the event reconstruction and therefore the tau identification. Hence, a summary is given here. Starting from the ECAL and HCAL front-end electronics, the data is first used to build trigger primitives by combining data from multiple readout channels into a single object called *trigger tower*. These are forwarded to the regional calorimeter trigger (RCT), where a tau veto bit is set and electrons and photons are identified. The RCT also combines trigger towers into *trigger regions* and forwards its result to the global calorimeter trigger (GCT). Here, jet and tau candidates are built from the trigger regions, total and missing energy sums are calculated, and the identified L1 objects are sorted. As the L1 trigger decision has to be determined within microseconds, all of these steps are performed simultaneously in multiple regions of the calorimeter system, while ensuring that objects hitting the region borders are reconstructed with the same efficiency. The individual trigger algorithms, and therefore the final trigger decision, is computed in the global trigger (GT). It is also responsible to provoke the read-out from the data buffers and establishes the connection between the L1 system and the data acquisition (DAQ).

The granularity of the calorimeter data that is imposed by the hardware is illustrated in Figure 2.2. The smallest accessible object is a trigger tower. It consists of 5×5 crystals in the ECAL barrel. In the ECAL endcaps and HCAL it has a similar size, while trigger towers in HF are larger. A trigger region aggregates 4×4 trigger towers and has a size of $\Delta\eta \times \Delta\varphi = 0.348 \times 0.348$. As described later, an L1 jet consists of 3×3 trigger regions. These large proportions make it impossible to disentangle particles close to each other. To minimize bandwidth, properties of objects are stored in as few bits as possible, allowing only for specific values of energies and directions.

L1 jets are reconstructed using a sliding window technique. A jet is found if the central region in a window of 3×3 trigger regions has a higher E_T than the eight outer regions. To suppress noise, the central region E_T has to be above an adjustable threshold. The jet energy is the sum of the E_T values of the nine regions, its position is defined as the center of the window.

The L1 configuration defines three types of jets. *Forward jets* have $|\eta| > 3.0$. The remaining jets are labeled *central jets* or *taus*, depending on the outcome of the tau identification. For each type, the four objects with the highest E_T are made available to the GT and can be used to calculate trigger algorithms. Jet energy corrections are applied dependent on η.

Two requirements sensible to the collimation of the deposited energy are implemented to define L1 tau jets. A tau veto bit is calculated for each trigger region. It is set if the energy

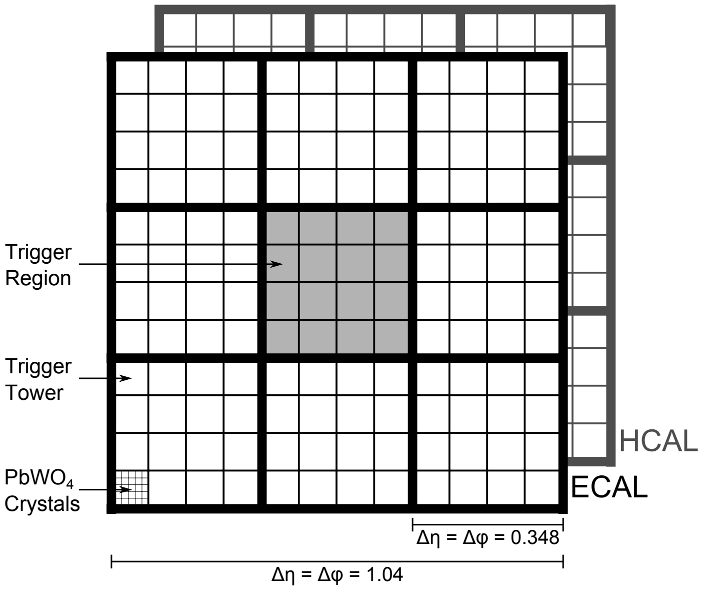
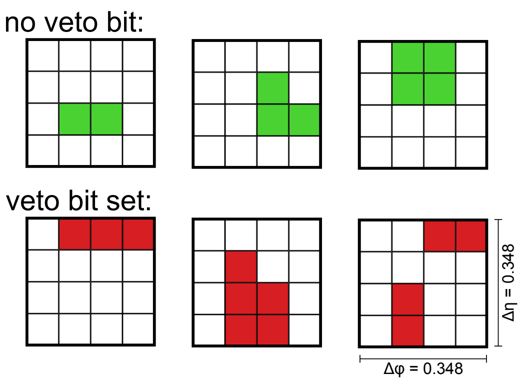
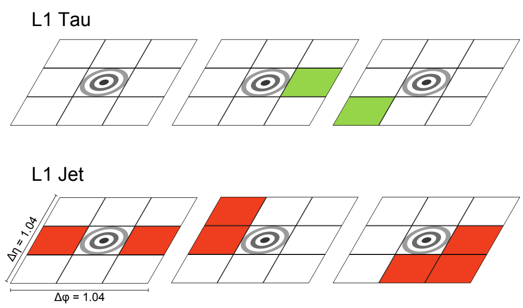


Figure 2.2: Granularity of the calorimeter data used by the L1 trigger. A trigger tower consists of 5×5 crystals in the ECAL barrel, and a structure of similar size in the endcaps. Each trigger region combines 4×4 trigger towers from both ECAL and HCAL. L1 jet finding uses windows of 3×3 trigger regions.



(a) The tau veto bit investigates the substructure of a trigger region.



(b) The tau isolation requirement is defined in a 3×3 window of trigger regions.

Figure 2.3: Isolation criteria used to discriminate taus from jets at L1. (a) The tau veto bit in each trigger region is set if the energy is not contained within 2×2 trigger towers. (b) An isolation requirement for taus is implemented by allowing only one out of the eight surrounding trigger regions to contain significant energy.

deposit within the region is spread over more than 2×2 trigger towers (see Figure 2.3(a)). If the veto bit is set in any of the nine trigger regions that form the L1 tau object, the tau hypothesis is discarded. To investigate the candidate on a larger scale, a second isolation variable is used. Seven out of the eight non-central regions must contain only small energy deposits $E_T < 2\,\text{GeV}$. Candidates that fail either of the isolation requirements are considered L1 central jets. They are expected to correspond to quarks or gluons and not to hadronic tau decays.

A discussion of the performance of L1 τ_h triggers is given in Section 2.2.2.

2.1.2. Tau leptons at HLT

Leptonic tau decays are treated as prompt electrons or muons by the HLT and are therefore not discussed in this section. Hadronic tau decays are reconstructed explicitly by HLT. Similar to the offline τ_h reconstruction, the main task is to separate quark and gluon jets from τ_h. In contrast to L1, the full calorimeter granularity is available as well as data from the tracker. Ideally, τ_h reconstructed by HLT are identical to their offline counterparts. However, this cannot be realized as the computing time per event is limited in the trigger system. Finding the best trade-off between these aspects is the main task when defining HLT paths.

The chosen approach for τ_h reconstruction implements three stages that are executed successively (see Figure 2.1). They can be considered as independent trigger levels since each of them reconstructs and selects hadronic tau decays. In case an event does not pass the selection at a given level, the succeeding stages are not executed. As the complexity of τ_h reconstruction is increased with every level, this reduces the overall computing time. For the same reason most of the event reconstruction at HLT differs from the offline reconstruction in various extents. This includes track and vertex reconstruction and the PF algorithm. Details on these can be found in Ref. [68].

The first step in HLT is called L2 and uses calorimeter jets, i.e. jets using only calorimeter data as input. Jet reconstruction is performed using an iterative cone algorithm [82] with a cone radius of 0.2. The jets are seeded by both L1 tau and central jets, which ensures that the same objects (i.e. the same area of the calorimeter) are used for the L1 and L2 decision. The reconstruction based on L1 seeds also reduces the necessary computing time at L2. The event passes L2 if such a calorimeter jet exists above a p_T threshold. L2 jets are expected to have a worse p_T resolution than objects that were reconstructed including tracking information. The purpose of this trigger level is to reduce the event rate as much as possible before the CPU-intensive tracking is executed.

The L2 stage does not allow to disentangle τ_h from quark or gluon jets. L2.5 is the first trigger level to engage in this task. Here, a track-based isolation criterion is applied on the τ_h candidates selected in L2. Data from the pixel tracker is used to reconstruct tracks in the vicinity of L2 jets. Vertices are reconstructed from these pixel tracks. Tracks that do not originate from the signal vertex are discarded. L2 τ_h are considered to be isolated if no track with $p_T > 1.2\,\text{GeV}$ is found in a hollow isolation cone of $0.2 < \Delta R < 0.4$ around the tau candidate. Signal tracks corresponding to the hadronic decay products of the τ_h are expected to lie within $\Delta R < 0.2$. Compared to the full tracking, the regional reconstruction of pixel tracks reduces the computing time by more than a factor of 10.

The last stage, called L3, reconstructs the event using the particle flow approach. This includes full tracking. In contrast to the complex HPS algorithm, the online τ_h reconstruction uses a simple *fixed cone* tau algorithm. This technique defines an inner signal cone that is expected to contain all τ_h decay products and an outer isolation annulus. The τ_h kinematics are estimated using the objects within the signal cone. Particles that originate from the same vertex and fall within the isolation annulus are not expected in genuine τ decays, and are therefore used to define an isolation criterion. At HLT, only tracking information is used to calculate the isolation. Loose and medium working points are defined. The former is mainly used in paths that include an electron or muon, which allow for a clean selection. It is using a signal cone of $\Delta R < 0.18$ and an isolation cone of $0.18 < \Delta R < 0.45$. The τ_h candidate is considered isolated if no track with $p_\mathrm{T} > 1.5\,\mathrm{GeV}$ is found in the isolation annulus. For the medium working point, the signal cone radius is reduced to $\Delta R < 0.15$, and the inner boundary of the isolation annulus is adjusted accordingly. A candidate fails the medium isolation if a track with $p_\mathrm{T} > 1.0\,\mathrm{GeV}$ is found within the isolation annulus.

2.1.3. Tau triggers used in 2012

Triggers for hadronic tau decays are a crucial part of several analyses using the 2012 data. The search for $\mathrm{H} \to \tau\tau$ decays was a driving force for the definition of the trigger menu. For each of the $\tau\tau$ decay channels of this search, at least one dedicated path must be defined that accommodates the fundamental requirements: optimal signal efficiency, a rate as small as possible corresponding to a good background rejection, thrifty use of the computing resources, and robustness against changes in the running conditions.

An overview of the main trigger paths that have been used during 2012 to collect hadronic tau decays is given in Table 2.1. Each listed path has been used in different searches and analyses, one of which is given as an example in the table. As the LHC conditions changed, the trigger paths have been altered. In particular, paths are adapted to higher instantaneous luminosities by applying tighter selection criteria. The given cut values on p_T, $|\eta|$, and the isolation correspond to the menu used at the end of 2012 at the highest luminosity, and thus have been looser before. All of the listed trigger paths were not prescaled during 2012, except for the $\tau_{1\pi}E_\mathrm{T}^\mathrm{miss}$ path, which is designed for the search for charged Higgs bosons.

The $\mu\tau_\mathrm{h}$ and $\mathrm{e}\tau_\mathrm{h}$ paths investigate the hadronic tau decay last, i.e. after the lighter lepton. On L1, only the muon (electron) is selected, applying a p_T cut of 14 GeV (18 GeV). The HLT part starts with the reconstruction of the muon or electron as described in Ref. [68]. This includes application of isolation requirements on these objects. Compared to the single muon or single electron trigger, which do not implement a second trigger leg, the p_T threshold is lower. While the lowest, unprescaled single muon trigger available in the 2012 trigger menu applies a p_T threshold of 24 GeV, requiring an additional τ_h allows to reduce this cut value to 18 GeV. Reconstruction of hadronic tau decays starts only after a good muon (electron) was found. Only a very loose cut of $p_\mathrm{T} > 5\,\mathrm{GeV}$ is applied on the L2 calorimeter jets. In cross triggers that exploit τ_h in addition to an electron, muon, or $E_\mathrm{T}^\mathrm{miss}$, the L2.5 regional pixel isolation is not used. This allows for a higher τ_h selection efficiency. The rate reduction introduced by L2.5 is not crucial in these paths, as it is provided by the other leg. At L3, a loose isolation is applied to the τ_h candidate. The signal tracks as well as the tracks used for the isolation must be compatible with the vertex that is associated with the light lepton. The trigger accepts the event if a τ_h candidate with these properties and $p_\mathrm{T} > 20\,\mathrm{GeV}$, $|\eta| < 2.5$ is found.

A good performance of the τ_h identification at trigger level is most crucial in channels without a light lepton in the final state. These include the paths $\tau_\mathrm{h}\tau_\mathrm{h}$ + jet, $\tau_{1\pi}\tau_{1\pi}$, and $\tau_\mathrm{h}\tau_\mathrm{h}$ listed

Table 2.1: Trigger paths that use τ_h objects applied during 2012 data taking. Only the main paths that have been used for published physics results are listed in the table. The listed cut values correspond to the last trigger level and can be different at earlier levels. For some triggers the cut values have changed during 2012. In these cases, only the tightest cuts are given, corresponding to the trigger menu running at highest luminosities. The triggers were not prescaled, unless stated otherwise. For triggers with two identical objects in the final state, the *First Leg* column gives the requirements for these two objects, while the *Second Leg* defines a possible third object.

Final State	First Leg				Second Leg				Analysis	Comment
	Type	p_T / GeV	$\|\eta\|$	Iso.	Type	p_T / GeV	$\|\eta\|$	Iso.	Example	
$\mu\tau_\mathrm{h}$	μ	> 18	< 2.1	yes	τ_h	> 20	2.5	Loose	$\mathrm{H} \to \tau\tau \to \mu\tau_\mathrm{h}$	
$\mathrm{e}\tau_\mathrm{h}$	e	> 22	< 2.1	yes	τ_h	> 20	2.5	Loose	$\mathrm{H} \to \tau\tau \to \mathrm{e}\tau_\mathrm{h}$	
$\tau_\mathrm{h}\tau_\mathrm{h} + \mathrm{jet}$	τ_h	> 30	< 2.1	Medium	Jet	> 30	3.0		$\mathrm{H} \to \tau\tau \to \tau_\mathrm{h}\tau_\mathrm{h}$	jet in L2 only
$\tau_{1\pi}\tau_{1\pi}$	$\tau_{1\pi}$	> 35	< 2.1	Medium	–				$\mathrm{H} \to \tau\tau \to \tau_\mathrm{h}\tau_\mathrm{h}$	only 1-prong decays
$\tau_\mathrm{h}\tau_\mathrm{h}$	τ_h	> 35	< 2.1	Medium	–				$\mathrm{H} \to \tau\tau \to \tau_\mathrm{h}\tau_\mathrm{h}$	parking
$\tau_{1\pi}E_\mathrm{T}^\mathrm{miss}$	τ_h	> 35	< 2.5	Loose	$E_\mathrm{T}^\mathrm{miss}$	> 70			$\mathrm{H}^\pm \to \tau\nu_\tau \to \tau_{1\pi}$	only 1-prong decays, prescaled

in Table 2.1. Accordingly, τ_h reconstruction is performed on all trigger levels: L1, L2, L2.5, and L3. All fully hadronic tau triggers require two identified L1 taus with $p_\mathrm{T} > 44\,\mathrm{GeV}$ and $|\eta| < 2.1$. Anticipating the results presented in Figure 2.6(a), it must be noted here that for L1 taus the transverse momentum is significantly overestimated. A cut of $p_\mathrm{T} > 44\,\mathrm{GeV}$ thus suits the L3 cut at 30 or 35 GeV. In consideration of an inefficiency of L1 tau identification at higher p_T values (see Figure 2.4), L1 central jets are included for $p_\mathrm{T} > 64\,\mathrm{GeV}$. The calorimeter jets on L2 and L2.5 are required to fulfill the same nominal cut values on p_T and $|\eta|$ as on L3, which are listed in Table 2.1. To further reduce the QCD background contribution, the detailed event description provided by the PF algorithm is exploited at L3. The medium isolation criterion is applied. A p_T cut on the leading charged PF candidate in the signal cone was reduced from $p_\mathrm{T} > 5\,\mathrm{GeV}$ to 1 GeV during 2012. The track multiplicity in the signal cone is limited to 4, reducing the background rate by more than 40%. The L3 τ_h candidates must coincide with the L1 taus within $\Delta R < 0.5$. Finally, both τ_h may not overlap within $\Delta R < 0.5$, but must be associated with vertices that are close to each other, requiring $\Delta z < 0.2\,\mathrm{cm}$.

Despite the additional quality requirements applied for fully hadronic ditau paths, the trigger rate is still large. To collect the important data nonetheless, different strategies are applied. The trigger rate of the $\tau_\mathrm{h}\tau_\mathrm{h}$ + jet and $\tau_{1\pi}\tau_{1\pi}$ paths is reduced by application of additional requirements. In case of the $\tau_\mathrm{h}\tau_\mathrm{h}$ + jet path, an additional quark or gluon jet with $p_\mathrm{T} > 30\,\mathrm{GeV}$ is expected in the final state. Usage of such a trigger reduces the main backgrounds of $\mathrm{H} \to \tau_\mathrm{h}\tau_\mathrm{h}$ searches, namely QCD multijet and DY, more than the signal. This is motivated by the Higgs production mechanisms: Gluons in the initial state of ggF Higgs processes are more likely to radiate quarks or gluons than the quarks in the initial state of DY processes. Higgs bosons produced by VBF are also accompanied by high energy jets. A $\tau_\mathrm{h}\tau_\mathrm{h}$ + jet path does not allow for an inclusive study of all $\mathrm{H} \to \tau_\mathrm{h}\tau_\mathrm{h}$ decays since its selection efficiency is highly dependent on the H production process. The technical implementation of the additional jet requirement considers only Level-2 calorimeter jets. This means that the energy resolution at the cut-off is not optimal, but no additional CPU time is consumed for the reconstruction of PF jets.

A second non-inclusive trigger path is intended to collect events where both taus decay into 1-prong final states. This reduces the background rate from QCD multijet events, as quark and gluon jets mostly show a bigger multiplicity of constituents. The usage of a fixed cone τ_h algorithm at HLT does not allow to reconstruct the tau decay mode, as is done by HPS. Therefore, the 1-prong requirement is implemented by requiring the number of charged PF candidates within the signal cone to be less or equal to 2. This allows for one additional track within the signal cone, which can be produced e.g. by pileup.

The initial plan of the CMS Collaboration was to run without an inclusive ditau path, but rely on the aforementioned triggers. With an expected trigger rate around 50 Hz the inclusive $\tau_\mathrm{h}\tau_\mathrm{h}$ path alone would consume 12% of the full HLT rate, and therefore drastically impede other analyses. If the applied cuts were tightened, the decrease of the signal phase space would have been too severe. Given the importance of the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel in H and Z decays, finding a solution to this problem was one of the first goals within the scope of this thesis. A first approach was to optimize the L1 tau identification. The achieved improvements were too small due to the technical limitations of the L1 trigger and thus not implemented. Data parking offered a solution. Although parked data was not available for analysis in 2012, the events could be investigated eventually. The development, proposal, maintenance, and monitoring of the inclusive $\tau_\mathrm{h}\tau_\mathrm{h}$ parking trigger was part of the PhD project described in this thesis. Additional information on the parking trigger is given in Section 2.3.

2.2. Trigger performance

The performance of trigger paths is characterized by a set of criteria. Most importantly, the path needs to collect as many signal events as possible. This is described by the trigger efficiency. It is commonly determined for each trigger leg individually as a function of p_T and η. A method to measure the trigger efficiency of $\tau_h\tau_h$ paths in data was derived and is described in Section 2.2.1. It was used in the first CMS $H \rightarrow \tau\tau$ search that included the $\tau_h\tau_h$ channel [100].

Other important characteristics include the trigger rate, which is limited by the trigger system and also reflects the background contribution to the specific path. The CPU time consumption needs to be controlled. This is especially important with respect to an increasing number of pileup interactions, which complicates the event reconstruction. All of these characteristics are regularly checked during data-taking.

2.2.1. Measurement of trigger efficiencies in data

In this thesis, the method used to determine the efficiency of a trigger path in data is based on a tag-and-probe procedure. It exploits $Z \rightarrow \mu\tau_h$ events. The μ represents the tag: It is used to trigger the event and is the main handle to suppress backgrounds in the offline selection. The hadronic tau decay is being probed. After offline selection, which is as close as possible to the selection used in the physics analyses, it is tested whether the offline reconstructed τ_h is found by the trigger. This defines the efficiency ε of one individual τ_h trigger leg as the likelihood of an offline selected τ_h to be identified accordingly by the trigger. A different definition of efficiency, which compares triggered τ_h to genuine tau leptons in the final state of the hard interaction, is not accessible in data, as the genuine final state is not known. Nonetheless, the efficiency defined here is suitable for trigger performance studies and, most importantly, for the correction of trigger effects impacting a physics result. The efficiency of a trigger path, consisting of one or more trigger legs, can be obtained by multiplication of the efficiencies of the individual legs. The method used in this section aims to characterize the fully hadronic trigger paths, $\tau_h\tau_h$.

In order to measure trigger efficiencies in data, one has to use an independent data set. All data that are available for analysis have already been selected by the trigger system. The tag-and-probe technique allows to obtain an unbiased sample by restricting the trigger path that is used for the event selection to the tagging leg. The events used for the presented study have been triggered by a single-muon path. The design of the trigger system ensures that the single-muon data set includes, but is not restricted to, events that contain an additional τ_h. The τ_h leg is therefore unbiased by the applied trigger selection.

As will be explained later, the trigger decision is emulated by rerunning the HLT software offline. Therefore, the required input data needs to include the full set of unprocessed detector data, called RAW. This is where HLT starts from. The common data format for offline data analysis, RECO, is limited to reconstructed quantities like particle candidates and their properties. The RAW data corresponds to approximately 400 kB per event when stored on disk, while RECO makes up 1200 kB. A smaller event size is favorable since it allows for faster data processing. As access to RAW data is rarely needed in physics analyses, the RAW information is not included in RECO. To measure τ_h trigger efficiencies, a data set is used that combines RAW and RECO data. It is produced specifically for this purpose. To limit its size, the data is loosely preselected to be consistent with a $\mu\tau_h$ final state.

The event selection starts with the choice of one specific single-muon trigger path: one HLT muon with $p_{\mathrm{T}} > 24\,\mathrm{GeV}$ and $|\eta| < 2.1$ must be reconstructed. This is reflected in the offline selection by a selection of exactly one tight muon (see Section 1.3.3) with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.1$. The cuts on the impact parameters are tightened to $|d_{xy}| < 0.45\,\mathrm{mm}$ and $|d_z| < 1\,\mathrm{mm}$. A $\Delta\beta$-corrected isolation is calculated, similar to the τ_{h} isolation given in Equation (1.31). The muon isolation uses a cone size of $\Delta R = 0.4$ and a $\Delta\beta$ correction factor of 0.5. The isolation variable may not exceed 10% of $p_{\mathrm{T}}(\mu)$. This muon selection is reasonably tight to suppress background from QCD events. It does not bias the selection of τ_{h} and therefore neither the trigger efficiency measurement.

To obtain a trigger efficiency with respect to the τ_{h} candidates used in the final analysis, the τ_{h} selection is close to the one described in Ref. [100]. However, the different final states in the efficiency measurement ($\mu\tau_{\mathrm{h}}$) and the analysis itself ($\tau_{\mathrm{h}}\tau_{\mathrm{h}}$) demand a few changes in the τ_{h} selection to account for the different background compositions. The implemented selection requires at least one HPS tau that is considered isolated by the MVA based discriminator at the medium working point. A medium electron rejection is applied as well as a tight muon rejection. The latter is important to suppress background from $\mathrm{Z} \rightarrow \mu\mu$ events. The muon and tau candidates must not overlap within $\Delta\mathrm{R} = 0.5$. As the efficiency is measured as a function of the kinematic properties p_{T} and η, cuts on these values are only applied when appropriate. The used kinematic cuts are $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ and $|\eta| < 2.1$. If more than one τ_{h} candidate that fulfills the aforementioned requirements exists, the candidate with the highest value of p_{T} is used for the trigger efficiency measurement.

The μ and τ_{h} candidates are supposed to originate from a Z boson decay. Accordingly, they must have opposite charge and may not be closer than $\Delta R = 0.5$ to each other. Given the similarity of $\mathrm{Z} \rightarrow \mu\tau_{\mathrm{h}}$ signatures with $\mathrm{W} + j \rightarrow \mu + j$ events, where the jet j is misidentified as τ_{h}, an additional cut to reduce this background is applied. The exploited quantity is the transverse mass m_{T}, corresponding to the invariant mass of an object calculated using only the x and y components of momenta. The transverse mass of two objects can thus be calculated universally as

$$\begin{aligned} m_{\mathrm{T}} &:= \sqrt{E_{\mathrm{T}}^2 - p_{\mathrm{T}}^2} \\ &= \sqrt{m_1^2 + m_2^2 + 2\left(E_{\mathrm{T},1}E_{\mathrm{T},2} - \vec{p}_{\mathrm{T},1}\vec{p}_{\mathrm{T},2}\right)}\,. \end{aligned} \tag{2.1}$$

The transverse energy E_{T} was defined in Section 1.2.2.1. Symbols without index 1 or 2 denote properties of the sum of the two four-vectors. If the particles can be considered massless, which is often valid for typical LHC energies, this reduces to

$$\begin{aligned} m_{\mathrm{T}} &= \sqrt{p_{\mathrm{T}}^2 - p_x^2 - p_y^2} \\ &= \sqrt{2p_{\mathrm{T},1}p_{\mathrm{T},2}\left(1 - \cos\Delta\varphi\right)}\,. \end{aligned} \tag{2.2}$$

To suppress the W background, the transverse mass of the W candidate is calculated using the μ and $E_{\mathrm{T}}^{\mathrm{miss}}$ four-momenta. Neither E_{T} nor m_{T} correspond to physical quantities. Nonetheless, m_{T} is a suitable approximation to the W invariant mass, which cannot be reconstructed directly as the z component of the neutrino momentum is not accessible. The W background is expected to show large values in this quantity. In Z and also H events, the neutrinos originate from the τ decays. Due to the boost into the laboratory frame, they are aligned with the μ momentum. This is reflected in a small value of m_{T}. A cut $m_{\mathrm{T}} < 40\,\mathrm{GeV}$ is applied to the event selection.

Table 2.2: Versions of the $\tau_h\tau_h$ + jet trigger path in the 2012 data set.

	Name	Luminosity [fb^{-1}]	[%]	Changes
1	2012A	0.8	4	
2	2012B	4.5	23	p_T threshold; isolation track quality
3	2012C part 1	1.8	9	leading track p_T; isolation area
4	2012C part 2	5.2	26	global tag
5	2012D part 1	4.2	21	global tag; track $\sigma(p_T)/p_T$ cut
6	2012D part 2	3.2	16	global tag

After all cuts, about 600 000 events are selected from the 2012 data set. The trigger decision is recalculated for these events by running the HLT software again. Contrary to the online HLT, events are not discarded if they fail the trigger selection. Instead, a positive or negative outcome of the event filters at the individual trigger layers is stored. This allows to measure the efficiency of each trigger layer independently of the outcome of the previous layers.

While the general layout of the $\tau_h\tau_h$ trigger paths remained the same throughout 2012, details of the implementation have changed. This is discussed in this work using the $\tau_h\tau_h$ + jet path as example. Six different trigger versions have been deployed. At the beginning of 2012, the p_T threshold on the τ_h objects was set to 25 GeV. The leading track in the τ_h signal cone was required to be above 5 GeV, and tracks in the isolation cone were ignored if they did not meet loose quality criteria. For a second version, the τ_h threshold was raised to 30 GeV and the isolation track quality criteria were tightened. Later the cut on p_T of the leading track was loosened to 1 GeV and simultaneously the size of the isolation area was adapted. Version 4 is identical to the previous version, but general settings like correction factors have changed. This can have an impact on the trigger efficiency. These general settings are read from a database when data is collected and reconstructed. A distinct collection of settings is called *global tag*. The next change includes an update of the global tag as well. Also, a cut on the uncertainty $\sigma(p_T)/p_T$ of PF tracks was relaxed, as it introduced an inefficiency for high-p_T taus ($p_T \gtrsim 150$ GeV). The last update is again connected to the global tag. An overview of the trigger versions and the amount of data collected with each of them is given in Table 2.2.

Although the selected data is independent of the $\tau_h\tau_h$ trigger version, the HLT decision of each event is reproduced using the corresponding trigger menu, i.e. the version that was deployed when the event was recorded. Especially changes of the global tag can have concealed, although mostly minor, impacts on the trigger reconstruction and therefore introduce correlations between trigger paths. This is diminished by using the matching trigger menu. At the same time, the method ensures to use an appropriate amount of events for each data period, resulting in an adequate statistical accuracy. Nevertheless, the ratio of the number of selected $Z \to \mu\tau_h$ events and the number of events triggered by the $\tau_h\tau_h$ path is not necessarily constant over the different data periods. For example, it can change if the trigger efficiency of the used single-muon trigger changes. Therefore, the τ_h trigger efficiency is measured for each data period independently. To obtain results that are valid for the full 2012 data set, these have to be combined. The weights used for combination are the shares of each period of the total luminosity, as given in Table 2.2.

The modular structure of the HLT software allows to access the reconstructed trigger objects at the individual trigger levels L1, L2, L2.5, and L3. For each event, it is tested if the selected

HPS tau can be matched to a trigger object. The offline and trigger τ_h candidates are considered to originate from the same genuine tau decay if they are close to each other, $\Delta\mathrm{R} < 0.3$ for L2, L2.5, and L3. Level-1 tau candidates have a worse angular resolution. Their matching cone size is increased to $\Delta\mathrm{R} < 0.5$, accordingly. If a trigger tau candidate is found within the matching cone, the event is counted positively: the trigger has worked as expected. Otherwise, the trigger selection was not successful. The full trigger selection is successful only if the tau candidate is reconstructed and identified at all four trigger levels.

Two of the cuts applied at L3 have to be considered separately. First, a matching of L3 taus to L1 objects is implemented in the trigger path. It is removed for the efficiency measurement of L3 tau candidates, as it introduces a dependence of the L3 decision on the L1 trigger. This dependence would inhibit a determination of the standalone L3 efficiency. The removal of this requirement does not impact the measurement of the total trigger efficiency, as a matching between L1 and L3 objects is intrinsically introduced by the matching of both objects to the same offline τ_h candidate. Second, the $\tau_\mathrm{h}\tau_\mathrm{h}$ trigger paths apply a cut on the distance in z between the vertices of the two L3 tau candidates. This introduces a correlation between the two τ_h legs. As there is only one τ_h present in the selected events for this study, the effect of this cut cannot be determined. To estimate its impact, a similar distance between the vertices of the L3 tau and the offline tau is computed. It is shown to have no significant impact on the final trigger efficiency measurement.

A possible problem of the described method might arise from the QCD background in the selected data sample. Although the event selection is applying tight cuts, a sizable fraction of the offline τ_h candidates originates from quark and gluon jets, which cannot be fully suppressed. To quantify the impact of this contribution, the cut on m_T is inverted in the event selection, $m_\mathrm{T} > 40\,\mathrm{GeV}$. Given the high cross section of W + jets events, this sample is highly dominated by jets that are misidentified as τ_h decay. The trigger efficiency is measured identically as explained above. It is found to be equal to the efficiency obtained using the regular event selection within the statistical uncertainty. Two aspects have to be considered to explain this observation. First, the trigger efficiency is measured with respect to offline selected τ_h candidates, which do not necessarily correspond to genuine tau decays. Second, τ_h signatures from genuine tau decays and from quark and gluon jets that passed the event selection are very similar; this is the reason why the jets could not be rejected in the event selection. A combination of these arguments explains that the trigger efficiency of genuine tau decays and jets misidentified as HPS taus do not differ. The obtained trigger efficiency from this method can therefore be applied to samples of genuine tau decays or a mixture of genuine and misidentified decay signatures, as it commonly appears in data analyses.

2.2.2. Performance of ditau triggers in 2012 data

The constraints at L1 do not allow for relative isolation variables, i.e. variables that normalize the deposited energy around the object to the object's energy. These are commonly used in offline analyses. The use of absolute isolation criteria causes an inefficiency for L1 tau reconstruction at high energies. This is driven by the hadronic shower profile of final state pions, which gets broader for higher energetic pions. The energy leaking into surrounding trigger regions can then be big enough to fail the isolation requirement. The effect is thus more pronounced for 3-prong than for 1-prong tau decays.

Figure 2.4 shows the efficiency to reconstruct a given τ_h in the Level-1 trigger as a function of the offline reconstructed p_T. Using a threshold of $p_\mathrm{T} > 44\,\mathrm{GeV}$ on L1 taus, the efficiency rises

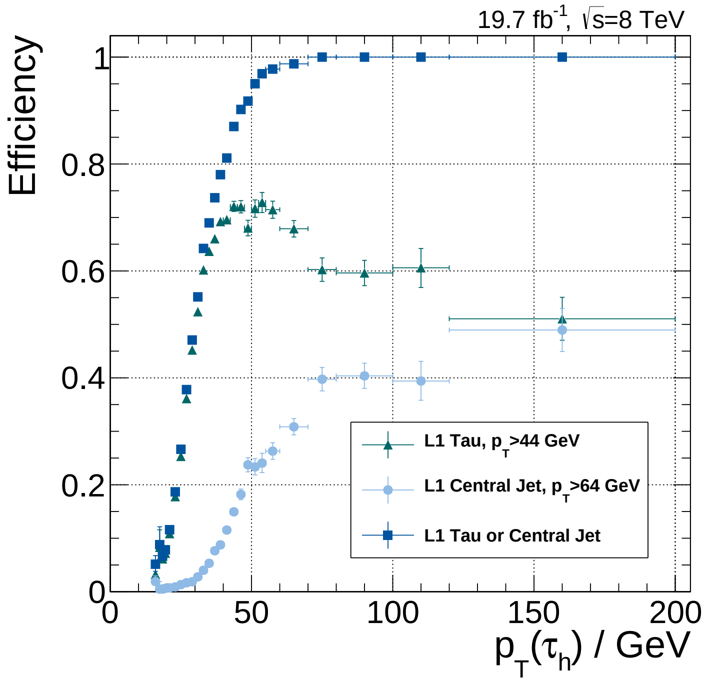


Figure 2.4: Efficiency of τ_h at Level-1 as a function of the offline reconstructed $p_T(\tau_h)$. The τ_h candidates are identified either as L1 tau (triangles) or as L1 central jets (circles). They are accepted by the trigger if the reconstructed L1 object has a p_T above 44 GeV or 64 GeV, respectively. To obtain a flat plateau, the two channels are combined (squares).

around the threshold value before dropping significantly at higher energies. The inefficiency can be fully recovered by including L1 central jets. Choosing a higher threshold of $p_T >$ 64 GeV for L1 central jets allows to reach a plateau in the τ_h efficiency at 100%, while rejecting most of the non-τ_h jets at lower energies. L1 jets with low p_T are expected to be highly dominated by quark and gluon jets. Rejecting them is a necessity to keep the rate of the trigger path within its boundaries.

Figure 2.5 shows the L1 trigger rate for an algorithm selecting two L1 taus with $|\eta| < 2.17$ above a given p_T threshold (triangles). The rate of L1 central jets with $|\eta| < 3.0$ above a given p_T threshold (circles) is significantly higher over the full p_T range, showing that most jets at CMS are non-isolated and thus likely originate from quarks and gluons. To obtain the trigger rates presented in Figure 2.5, a sample of events is used that was collected without trigger requirements, except that any activity must be observed within the CMS detector. This sample is therefore unbiased. The trigger rate can be deduced by testing how often two L1 taus (or central jets) have been reconstructed by the L1 trigger system. The shown rates are normalized to an instantaneous luminosity $\mathcal{L} = 5 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, a typical value for 2012 data taking. The normalization is calculated under the assumption of a linear dependence between the rate and the luminosity, according to Equation (1.21). Effects like pileup can change this dependence and make it non-linear. As these effects are difficult to predict, the rate needs to be controlled regularly during data taking.

The rate of the double-tau selection drops significantly faster with p_T than the rate of the double-jet selection. This corresponds to the aforementioned observation that the absolute isolation requirement for L1 taus is not suitable for high-p_T objects. It should be noted here that the events collected by these L1 algorithms are almost exclusively QCD multi-jet events

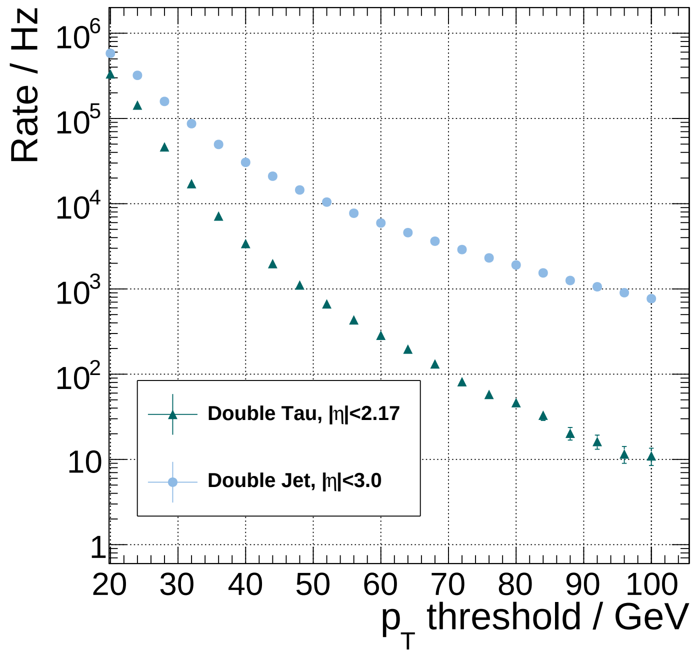
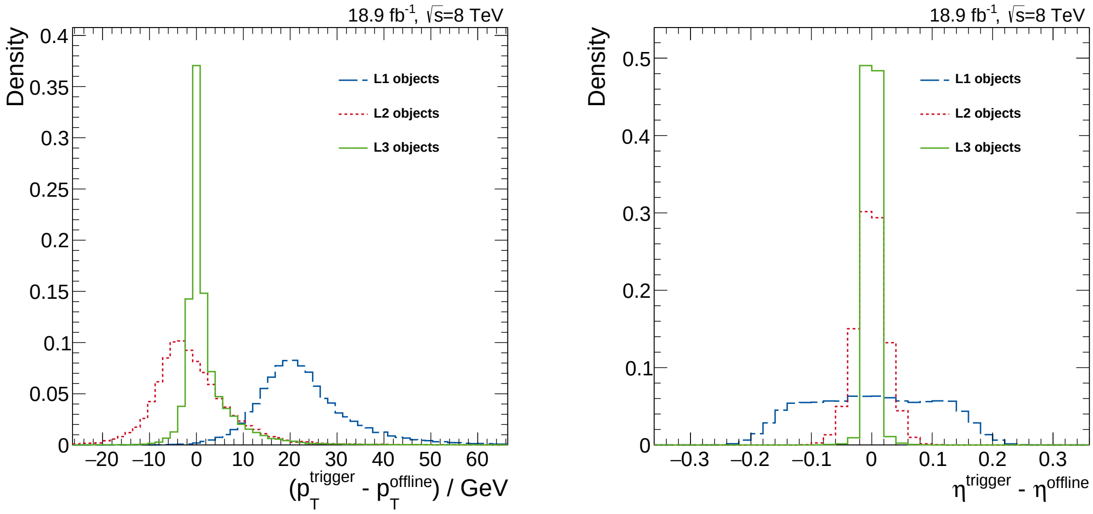


Figure 2.5: Level-1 trigger rates as a function of the p_T cut value on the trigger objects. The rate is normalized to an instantaneous luminosity $\mathcal{L} = 5 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Shown are the two algorithms that are used as seed for the $\tau_\mathrm{h}\tau_\mathrm{h}$ HLT paths: two L1 tau objects with $|\eta| < 2.17$ (triangles), or two L1 central jet objects with $|\eta| < 3.0$ (circles). The implemented thresholds in the L1 trigger were 44 GeV for L1 taus and 64 GeV for L1 central jets.



(a) p_T resolution. (b) η resolution.

Figure 2.6: Resolution (a) in p_T and (b) in η of τ_h trigger objects at different trigger levels. The resolution is defined as the difference of the quantity reconstructed at trigger level and reconstructed offline.

and contain only few genuine taus. It can be seen that the coarse isolation requirement is met by a sizable fraction of QCD jets. Nonetheless, Figure 2.5 shows that L1 tau identification can reduce the rate and thus the background contribution by a factor 2 (for $p_T \gtrsim 20$ GeV) up to a factor of 100 (for $p_T \gtrsim 100$ GeV). The L1 thresholds used during the 2012 data taking period correspond to a combined rate of 6.6 kHz, of which 2.0 kHz are introduced by the double-tau part. An algorithm without L1 identification, selecting two jets with $p_T > 44$ GeV, would give a rate of 23 kHz and thus would not be feasible. L1 tau identification is therefore a crucial ingredient for $\tau_h\tau_h$ analyses.

The purpose of the trigger system is to identify objects and to reconstruct their properties. To quantify the performance of τ_h reconstruction, resolution curves are examined. For a given quantity ζ we define its resolution as the difference of the quantity reconstructed at trigger level and the same quantity reconstructed offline, $\zeta^{\text{trigger}} - \zeta^{\text{offline}}$. Resolution curves for p_T and η are presented in Figure 2.6 for the different trigger levels L1, L2, and L3. A curve for L2.5 objects is omitted, as it is similar to the L2 curve; L2.5 uses pixel tracks for τ_h identification, but does not alter its reconstructed properties.

As expected, the τ_h reconstruction becomes more precise for the higher trigger levels, which include more detector information and use more complex reconstruction techniques. The p_T resolution of L1 tau objects (Figure 2.6(a), long dashes) is particularly broad, with a root mean square (RMS) of 10.9 GeV. Its mean is located at 23.2 GeV, meaning that the L1 tau energy is significantly overestimated. The energy measurement of the calorimeters is subject to a complex calibration routine which differs for each type of particle in the offline event reconstruction. At L1 only an imprecise jet energy correction is applied. It is measured using QCD dijet events and is thus not optimal for genuine tau decays. Nonetheless, it is applied to both L1 central jets and L1 tau jets, causing the observed energy mismeasurement. While the broad resolution has an impact on physics analyses, the shift of the energy values is not critical. Instead of its nominal value, the underlying physical L1 p_T threshold has to match the HLT thresholds. For the $\tau_h\tau_h$ paths the L1 p_T threshold is set to 44 GeV and thus, taking the shift into account, reasonably well below the HLT cut values of 30 or 35 GeV.

The calorimeter jets at L2 (short dashes) have a slightly better momentum resolution with an RMS of 9.1 GeV and are centered around the nominal value. It implements a more sophisticated jet reconstruction algorithm but is still limited by the sole usage of calorimeter information. Inclusion of the tracker and combination of the information using the PF algorithm causes a significant improvement of the resolution at L3 (solid line). The resolution distribution is centered around zero and has a prominent narrow peak. This peak corresponds to tau candidates for which the L3 PF tau reconstruction and the offline HPS tau method construct the same PF candidates. The RMS value of 0.2 GeV is explained by differences in the HLT and offline PF algorithms. A second component of the distribution is described by a Gaussian with a width of about 2.0 GeV, centered around zero. It originates from events where the fixed cone algorithm used at L3 and the offline HPS taus have different, mainly neutral, constituents. This occurs mainly for τ decays into $\pi^{\pm}\pi^{0}$, as collecting all π^{0} decay products is challenging. Additionally, it includes effects from parton jets misidentified as tau decays. The latter is more prominent in 3-prong decays, which have a signature more similar to quarks or gluons. Finally, the fixed cone used at HLT might include additional high-p_{T} objects that originate from pileup or the underlying event. These are excluded from the tau decay by the HPS algorithm, as they are not consistent with the decay model. As a consequence a tail to higher energies is apparent.

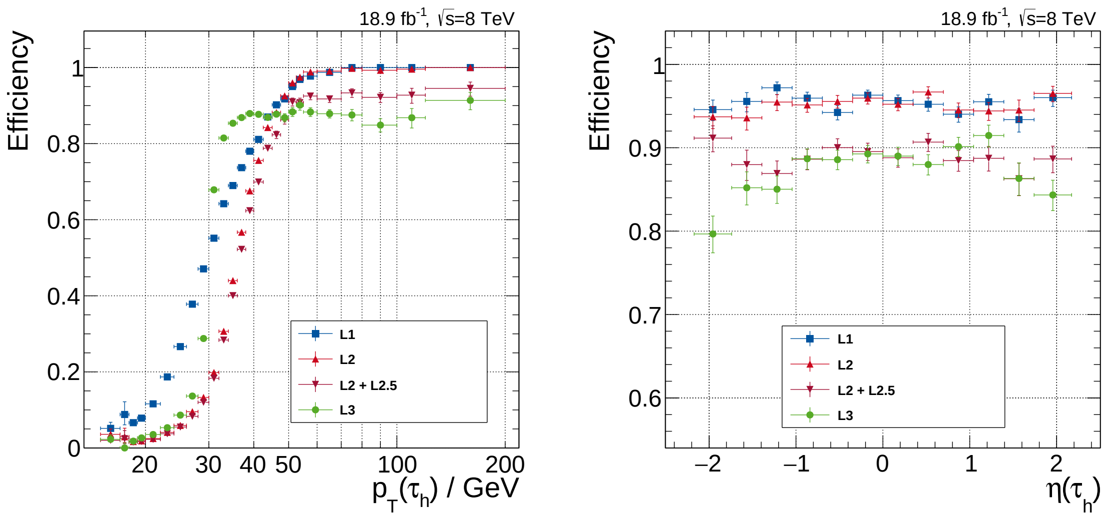
The angular resolution of trigger objects along the η coordinate is presented in Figure 2.6(b). The coordinates of L1 taus are set to the center of the associated trigger region. Accordingly, the angular resolution is described by a uniform distribution around 0 with a width of $\Delta\eta = 0.348$. If taus hit a trigger tower that is located at the border of two trigger regions, it is assigned to either of the two neighboring regions. A transition region of the size of a trigger tower is the result. The η resolution at L2 is significantly better than at L1, as the L2 jet reconstruction does allow continuous η values. The RMS is extracted as 0.026. Usage of the full set of information at L3 improves this value to 0.007. The direction measurements obtained from HPS taus and at the final level of HLT are practically identical.

An important input for $\mathrm{H} \to \tau\tau$ analyses is the trigger efficiency as function of p_{T}. It is often called *turn-on curve*. If the trigger efficiency is dependent on η, turn-on curves are provided in appropriate η bins. The final turn-on curve is a combination of the trigger efficiencies at the individual trigger levels. At each level the efficiency curve is given by a convolution of the resolution with a step function

$$s(p_{\mathrm{T}}) = \begin{cases} 0: & p_{\mathrm{T}} < p_{\mathrm{T}}^{\mathrm{thresh}} \\ \hat{\varepsilon}: & p_{\mathrm{T}} \geq p_{\mathrm{T}}^{\mathrm{thresh}} \end{cases} . \tag{2.3}$$

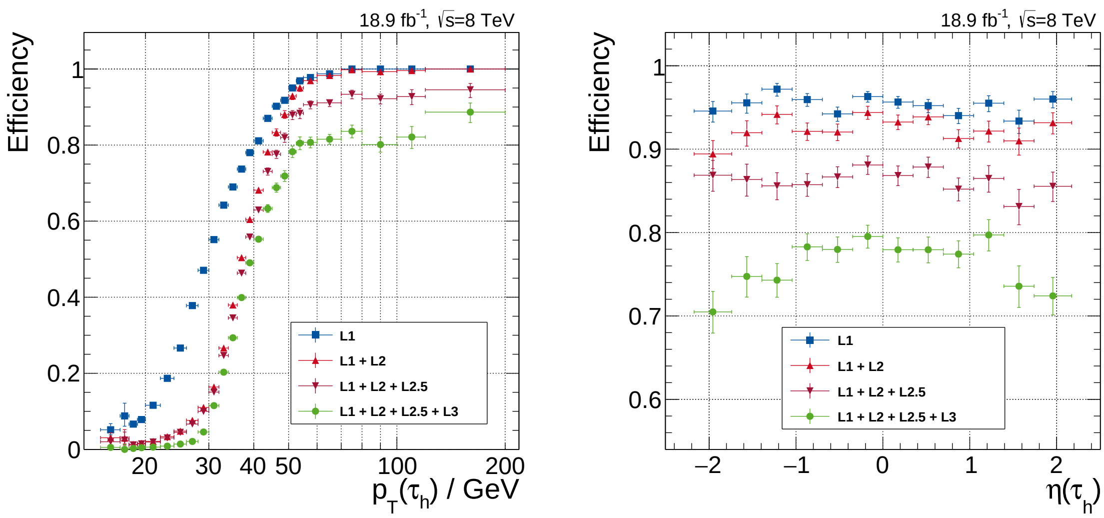
Its parameter $p_{\mathrm{T}}^{\mathrm{thresh}}$ denotes the nominal threshold value and the plateau efficiency $\hat{\varepsilon}$ gives the efficiency for objects sufficiently far above the threshold. In reality $\hat{\varepsilon}$ can itself be a function of p_{T}, if the identification of the trigger objects depends on their kinematic properties. In ideal cases the plateau efficiency is 100% and the resolution is described by a Gaussian; the efficiency curve is then given by an error function.

The measured efficiency curves for each of the four trigger levels are presented in Figure 2.7(a). They are independent from each other, meaning that the efficiency values of a given level do not depend on the trigger decision at other levels. While the efficiency of L1 taus raises only slowly with p_{T}, the edge of the distribution gets sharper for the later levels. This is in agreement with the previously discussed resolution curves. The plateau efficiency



(a) Efficiency as function of p_T. (b) Efficiency as function of η.

Figure 2.7: Efficiency of τ_h objects reconstructed at individual trigger levels. The shown efficiency values for each level are independent of the trigger decision at the other levels.



(a) Efficiency as function of p_T. (b) Efficiency as function of η.

Figure 2.8: Efficiency of τ_h objects after each level of the trigger chain. The shown efficiency values include all trigger selections that have been applied before the given level.

of the calorimeter-based trigger levels L1 and L2 is close to 100%. Inclusion of the tracking detectors is crucial to separate genuine tau decays from parton jets. The selections that are applied for this purpose also remove about 10% of genuine taus. At the nominal threshold value $p_{\mathrm{T}}^{\mathrm{thresh}} = 30\,\mathrm{GeV}$ the L3 curve has a very sharp edge, corresponding to the central narrow component of the resolution curve. In the final turn-on curve this sharp edge will not be visible, as it is convoluted with the L1 efficiency. A sharp turn-on is preferable, because it is easier to model.

While a sharp turn-on given the current implementation can be achieved by increasing the nominal L3 threshold value, this would cut too deeply into the phase space of the important physics processes $\mathrm{Z} \rightarrow \tau_{\mathrm{h}}\tau_{\mathrm{h}}$ and $\mathrm{H} \rightarrow \tau_{\mathrm{h}}\tau_{\mathrm{h}}$. The chosen threshold values are a compromise between rate and efficiency. To improve this situation, the most important component is the L1 trigger system. By inclusion of tracking information into the L1 decision, as is planned for the Phase-II upgrade of the CMS detector [101], the accessible phase space could be significantly increased.

The efficiencies show only a small dependency on the pseudorapidity for all trigger levels. The data presented in Figure 2.7(b) was obtained for hadronic tau decays with $p_{\mathrm{T}} > 45\,\mathrm{GeV}$. A significant deviation from a flat distribution over η is observed only for level-3 objects. Nonetheless, the efficiency for all individual trigger levels is at or above 80% over the whole η range.

The final trigger efficiency for a given offline τ_{h} candidate is the aggregate of the individual level efficiencies. Figure 2.8 shows the efficiency after each level as a function of p_{T} and η. While the slope of the final turn-on curve combining all trigger levels (circles in Figure 2.8(a)) is determined by the L1 efficiency, the plateau efficiency value is defined by the two last levels. For objects with $p_{\mathrm{T}} > 45\,\mathrm{GeV}$, the final trigger efficiency is about 78% in the barrel and about 73% in the endcap region (Figure 2.8(b)).

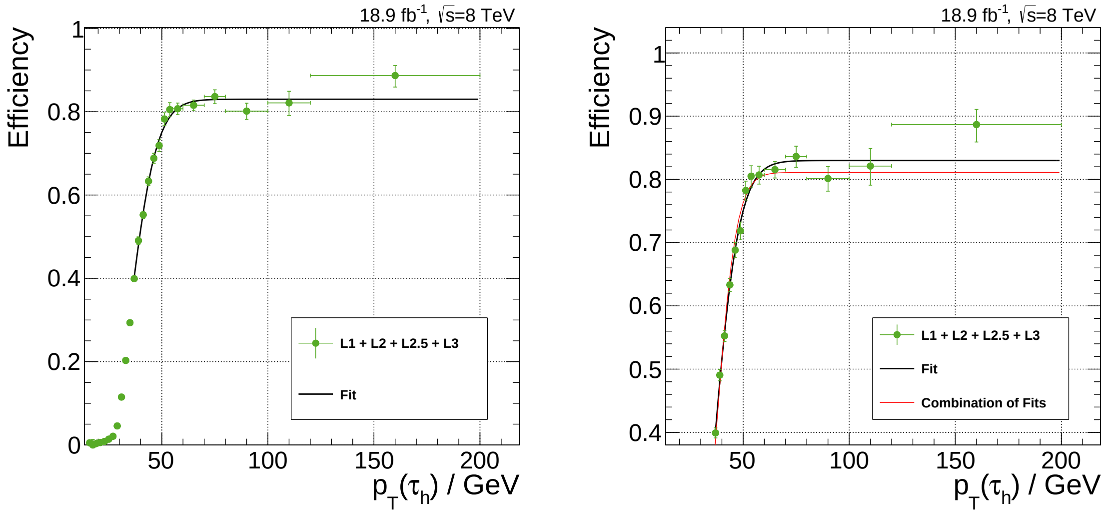
To account for the trigger efficiency in $\mathrm{H} \rightarrow \tau_{\mathrm{h}}\tau_{\mathrm{h}}$ analyses, the values obtained in each bin could be used directly. The better alternative is to fit an appropriate function to the data points. This reduces the impact of statistical fluctuations in the trigger efficiency measurement to the final result. One possible approximation of the functional form is

$$\varepsilon = \frac{1}{2}\hat{\varepsilon}\left(\mathrm{erf}\left(\frac{p_{\mathrm{T}} - \mu}{2\sigma\sqrt{p_{\mathrm{T}}}}\right) + 1\right) , \tag{2.4}$$

where the error function is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \mathrm{d}t . \tag{2.5}$$

It is different from the cumulative distribution of a Gaussian by a factor $1/\sqrt{2p_{\mathrm{T}}}$ in the argument of the error function, which results in a less sharp transition between the raising part and the plateau region of the turn-on curve. As the final turn-on is defined by multiple trigger levels, each with non-Gaussian components in the underlying p_{T} resolution curves, it is reasonable that an adaption of the functional form is necessary. The parametrization given in Equation (2.4) is found to model the upper part of the turn-on curve appropriately. The lower part, i.e. for p_{T} values reasonably far below the threshold value, the form of the turn-on curve is not of importance for physics analyses, as the events are discarded anyway. Other functional forms have been studied, but did not improve the final result. Especially in the plateau region the fitting result is limited by the statistical uncertainties of the trigger



(a) Fit to the combined efficiency bins.

(b) Comparison of fit methods.

Figure 2.9: Fit of a modified error function to the τ_h trigger efficiency as a function of p_T. (a) The result of a fit to the combined efficiency bins, i.e. the efficiency values of each data period have been combined before in each individual bin. (b) A comparison of the aforementioned method (black curve) with an alternative approach, where a fit is performed individually in each data period. The shown red curve is then computed as the weighted average of the individual fits.

efficiency measurement, since the number of events entering the fit is limited. Inclusion of more degrees of freedom into the fit is therefore not appropriate.

Figure 2.9 shows the results of the fit. Fitting has been performed using a binned likelihood method. Usage of an unbinned likelihood gives similar results. The data points shown in the figures are identical to the points representing the full trigger efficiency (circles) in Figure 2.8(a). They are obtained by combining the efficiency values of the six trigger periods (see Table 2.2) for each bin individually. A fit is then applied to these data points. The result is presented in Figure 2.9(a). The analysis documented in Ref. [100] uses an alternative approach: An individual fit is performed for each trigger period. The analytic representations are then added up, again using the luminosities as weights. Figure 2.9(b) compares these two methods. Although the underlying data is identical, the difference in the plateau efficiency between the two methods is about 2%. Differences like these that occur when using different parametrizations, binning, fitting ranges, or event selections have been considered in the systematic uncertainty. For each τ_h leg, an uncertainty of 4.5% is assigned for the trigger efficiency measurement.

2.3. Parking trigger for ditau events

The $\tau_h\tau_h$ parking trigger is identical to the other fully hadronic paths, but omitting the last selection step, i.e. the additional jet or 1-prong requirement. Its purpose is to allow for inclusive measurements of $\tau_h\tau_h$ final states that are biased as little as possible by irrevocable trigger selections. Inclusion of such a path with reasonable low p_T thresholds was not acceptable at the beginning of 2012, as it would have restricted other important triggers. The average event rate that could be collected by CMS was limited to 300 to 350 Hz by the prompt event reconstruction at the computing center. The bandwidth of the detector readout itself allowed to store more events on tape. Accordingly, an additional 300 to 350 Hz of data were *parked*: the raw data was stored, but the event reconstruction was not executed right away. The data could therefore not be used for analyses during 2012. During the LHC shutdown that followed the 2012 running period, the computing resources were not needed for the prompt reconstruction and could be used to reconstruct the parked data set. Thus, data parking offered a good opportunity to recover a significant amount of data that otherwise would have been lost. Data parking was enabled at the beginning of the CMS running period 2012B (see Table 2.2).

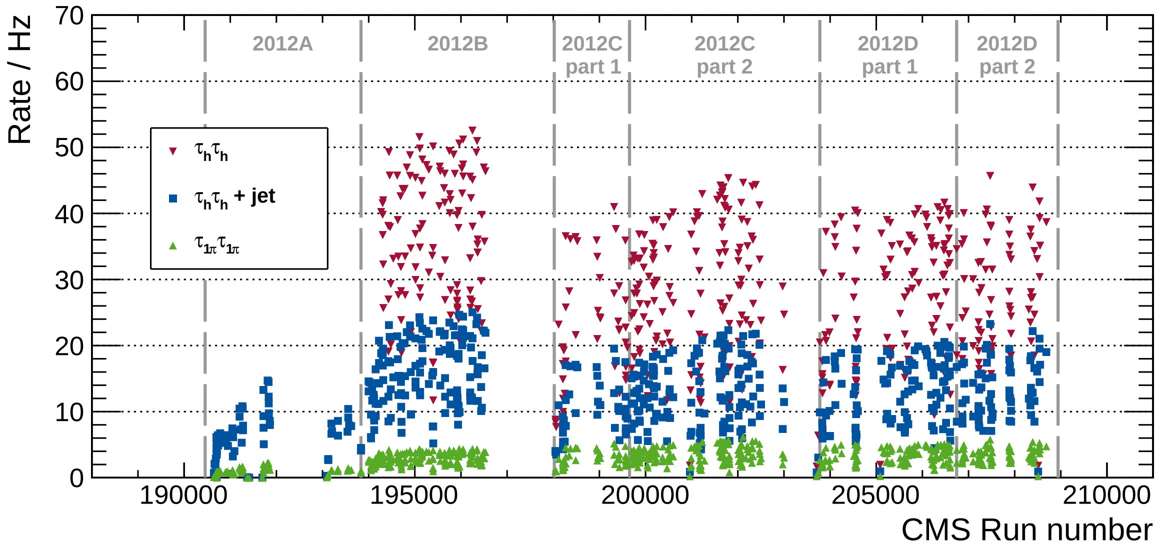
The trigger software needs to handle the supplementary trigger paths. As the aforementioned non-inclusive paths are already part of the menu, the $\tau_h\tau_h$ parking trigger does not consume much additional CPU time. The structure of HLT ensures that code is not run twice. While the total CPU time of the inclusive path at $\mathcal{L} = 1 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ is quite large, the time consumed exclusively by this path is negligible. Per event, 9 ms is spent on the path in total, of which only about 10 μs does not overlap with the $\tau_{1\pi}\tau_{1\pi}$ path. Thus, the trigger rate remains the only critical item and needs to be kept within limits even in the parking data set.

Figure 2.10(a) shows the trigger rate of the fully hadronic ditau trigger paths over the year 2012. The rates are averaged within each CMS *Run*, a period of data collection that typically lasts a few hours. Within a given Run, the rate is expected to decrease as the LHC luminosity decreases. The parking trigger reached a maximum averaged rate of 52.7 Hz, meaning that the according number of events were stored on tape. The impact of the additional selection steps for the $\tau_h\tau_h + \mathrm{jet}$ and $\tau_{1\pi}\tau_{1\pi}$ is evident. While the jet requirement reduces the rate to approximately half, the 1-prong selection reduces it by almost 90%. For this reason the pure trigger rate of the parking trigger, i.e. events that are not selected by any other trigger path, is in the order of 20 Hz. An additional event yield of this magnitude is acceptable given the possible physics reach. Adaption of the τ_h isolation area at the beginning of the trigger period 2012C reduced the rate significantly. It was crucial to allow execution of the parking trigger path over the whole year.

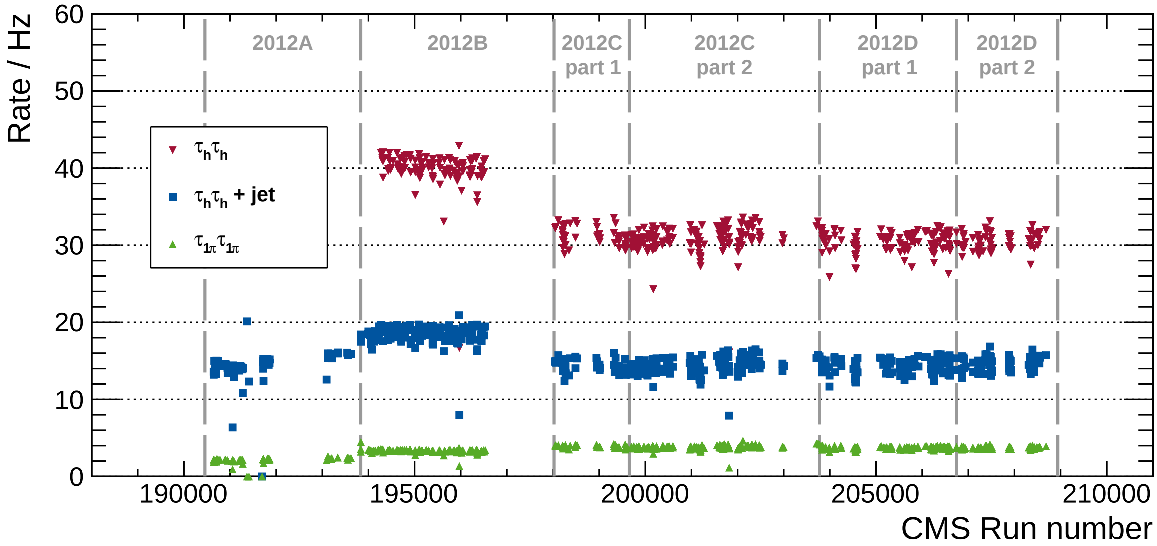
Trigger rates are expected to be proportional to the instantaneous luminosity. As the value of $\mathcal{L}$ fluctuates from Run to Run, the rate fluctuates accordingly. To study the performance of the trigger paths independently from the LHC conditions, the rates can be normalized to a specific value of $\mathcal{L}$. This is done in Figure 2.10(b) for $\mathcal{L} = 5 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. As expected, the normalized trigger rates show only small fluctuations and do not change significantly over time. The linear dependence on $\mathcal{L}$ also means that there is only marginal dependence on effects like pileup or changes in the detector performance.

To ensure good quality of the recorded data, it is necessary to perform tests shortly after it is recorded. This allows to identify problems and to adjust or repair the trigger system quickly, so that the remaining data taking is not affected. For parking triggers, this procedure is restricted as the data is not reconstructed right away. To ensure good data quality, a fraction of the events collected by parking triggers is reconstructed and thus available for offline studies. During 2012, regular checks have been performed on these events. It was shown that the path does collect 3-prong tau decays that were missing in the non-parking data set. Investigations of the kinematic distributions and the trigger efficiency showed no unexpected behaviour.

Ref. [102] describes the legacy analysis of $\mathrm{H} \rightarrow \tau\tau$ decays during LHC Run-I. It includes the additional events collected by the parking trigger. This increases the event yield of Higgs



(a) Trigger rate as collected by CMS.



(b) Trigger rate normalized to $\mathcal{L} = 5 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

Figure 2.10: Trigger rates of fully hadronic ditau paths during 2012. Shown are the average rates within a given CMS Run. The trigger periods as described in Table 2.2 are marked; as the triggers have changed between the trigger periods, changes in the rate are expected at the period borders. (a) The trigger rate collected by the CMS detector is dependent on the average value of $\mathcal{L}$ within a Run and fluctuates accordingly. (b) By normalization of the trigger rate to $\mathcal{L} = 5 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, this dependency is removed.

bosons produced in the VBF final state by about 10%. In combination with other developments with respect to previous CMS $\mathrm{H} \rightarrow \tau\tau$ results, the sensitivity of the fully hadronic decay channel was significantly improved. Despite the challenges that are associated with triggering and reconstruction of τ_{h}, the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ channel reaches a sensitivity similar to the $\mathrm{e}\tau_{\mathrm{h}}$ channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. It is inferior only to the $\mu\tau_{\mathrm{h}}$ channel, which can rely on the good muon identification and reconstruction.

Chapter 3

Search for $\mathrm{H} \rightarrow \tau\tau$ decays at CMS

The search for Higgs boson decays into pairs of τ leptons has been an important task of the CMS collaboration since the beginning of data collection. Several publications are concerned with this search within the data collected during the LHC Run-I. All of them follow a similar analysis strategy and differ mainly in the amount of analyzed data as well as improvements in individual analysis components. The legacy analysis published in Ref. [102] represents the final iteration of this process for the Run-I data. An overview of the analysis of the 8 TeV data set is given in this chapter with a focus on the $\mu\tau_{\mathrm{h}}$ final state.

3.1. Overview of the analysis

A search for $\mathrm{H} \rightarrow \tau\tau$ decays has to comprise a variety of final state signatures, which depend on the tau decay modes. CMS follows an integrated analysis approach: Despite the variety of final states, the event reconstruction and selection remains identical as far as possible. Nonetheless, a classification of events based on the initial and final state of the underlying $\mathrm{H} \rightarrow \tau\tau$ process is employed. It allows to exploit particularities in the event signatures of specific processes and thus leads to a better significance. Both aspects are integrated in a global fit of all individual analysis channels.

Events with a Higgs boson produced via ggF or VBF processes contain two leptons in the final state. Associated production with a weak boson can give rise to additional leptons; this is exploited in Ref. [102], but is neglected in this work as the sensitivity in these channels is small. Depending on the decay of the two tau leptons, six different *channels* are analyzed: $\mu\tau_{\mathrm{h}}$, $\mathrm{e}\tau_{\mathrm{h}}$, $\mathrm{e}\mu$, $\mu\mu$, ee, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. These leptons as well as event properties like $E_{\mathrm{T}}^{\mathrm{miss}}$ are used in a baseline event selection. Within each channel, all events that pass the baseline selection are classified into *categories*. This is based firstly on the number and properties of jets that occur in addition to the leptons in the event, and secondly on kinematic properties of the signal leptons.

The invariant mass of the ditau system, $m_{\tau\tau}$, is used as final discriminant, as it allows to separate Higgs decays from the dominant DY background. The distribution of $m_{\tau\tau}$ is modeled for signal and background contributions in each category individually. To allow for a correlation of systematic uncertainties, the applied methods are similar as far as possible for all categories. With these inputs, the final global fit can constrain nuisance parameters related to the signal and backgrounds and extract a possible signal contribution.

3.2. The $\mu\tau_\mathrm{h}$ channel

3.2.1. Baseline selection

The first layer of event selection is the trigger, which was described in detail in Chapter 2. Events of the $\mu\tau_\mathrm{h}$ channel are triggered by the corresponding trigger path listed in Table 2.1.

A muon with $p_\mathrm{T} > 20\,\mathrm{GeV}$ and within $|\eta| < 2.1$ that meets the tight muon criteria (see Section 1.3.3) is the starting point for the offline selection. The relative, $\Delta\beta$-corrected isolation I_μ has to be below 10% and the muon track needs to be compatible with the selected primary vertex within $|d_z| < 2\,\mathrm{mm}$ and $|d_{xy}| < 0.45\,\mathrm{mm}$. To ensure that the candidates reconstructed by HLT and offline correspond to the same particle, their angular distance is constraint to $\Delta R < 0.5$.

Hadronically decaying taus are reconstructed by the HPS algorithm (see Section 1.3.6). A threshold of 20 GeV is applied on the reconstructed p_T, and the τ_h must have $|\eta| < 2.3$. Separation against quark and gluon jets is achieved by an isolation requirement of $I_\tau < 1.5\,\mathrm{GeV}$, where I_τ is the $\Delta\beta$-corrected isolation defined in Equation (1.31). Discrimination of τ_h against electrons is performed using the loose working point. To suppress background from DY dimuon events, the tight discriminator against muons is applied. Similar to the muon, the τ_h must be reconstructed within $\Delta R < 0.5$ of the corresponding HLT object. Additionally, it must not overlap with the selected muon within $\Delta R < 0.5$.

Events are kept only if they have at least one selected muon and one selected τ_h. If more than one lepton of a given kind is selected, the one with the highest p_T value is chosen. To further suppress $\mathrm{Z} \to \mu\mu$ decays, events are rejected if they contain two loosely selected muons of opposite charge with $\Delta R > 0.15$. These muons do not have to fulfill the tight muon identification, but must be reconstructed as both tracker and global muon. The isolation requirement is loosened to 30% and the kinematic constraints are set to $p_\mathrm{T} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$.

VV ($V = \mathrm{Z},\mathrm{W}$) events and processes involving top quarks can give rise to more than two leptons in the final state. They are reduced by a tri-lepton veto. If an additional tight muon of any charge is found that meets $p_\mathrm{T} > 10\,\mathrm{GeV}$, $|\eta| < 2.4$, and relative isolation below 30%, the event is discarded. The same holds for any electron with similar selection criteria except for $|\eta| < 2.5$. An MVA-based electron identification is used.

$\mathrm{W} + \mathrm{jets}$ events are an important source of background to $\mu\tau_\mathrm{h}$ final states. If the W decays into a muon and the radiated quark or gluon is misidentified as τ_h, the m_T variable introduced in Equation (2.2) offers a way of separation from genuine $\tau_\mu\tau_\mathrm{h}$ events. The transverse mass is calculated from the reconstructed MVA $E_\mathrm{T}^\mathrm{miss}$ and the selected muon. While m_T is large for $\mathrm{W} + \mathrm{jets}$ events, $\tau_\mu\tau_\mathrm{h}$ final states are expected to show smaller $E_\mathrm{T}^\mathrm{miss}$ values aligned with the visible τ decay products, which leads to a small transverse mass. Figure 3.1 shows the distribution of the $m_\mathrm{T}(\mu, E_\mathrm{T}^\mathrm{miss})$ variable after all other selection cuts have been applied. Only events with $m_\mathrm{T}(\mu, E_\mathrm{T}^\mathrm{miss}) < 30\,\mathrm{GeV}$ are selected. This reduces the background from $\mathrm{W} + \mathrm{jets} \to \mu + \mathrm{jets}$ events by 93%.

Top-quark production gives rise to events with multiple genuine leptons in the final state. Therefore, the lepton selections are not efficient to suppress background from top quarks. As these decays are accompanied by b quarks, a veto against b jets is applied. The medium working point of the CSV algorithm (see Section 1.3.4) is employed to identify b jets. If such a jet is found with $p_\mathrm{T} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$, the event is rejected.

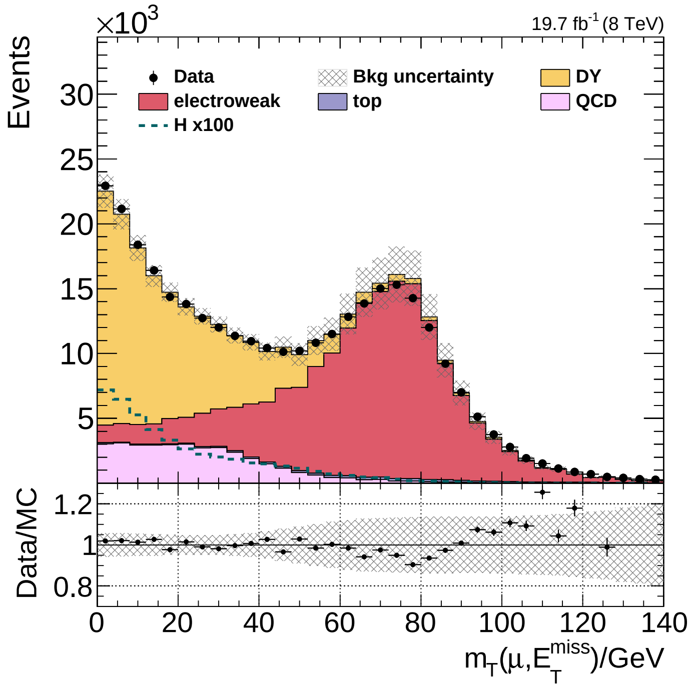


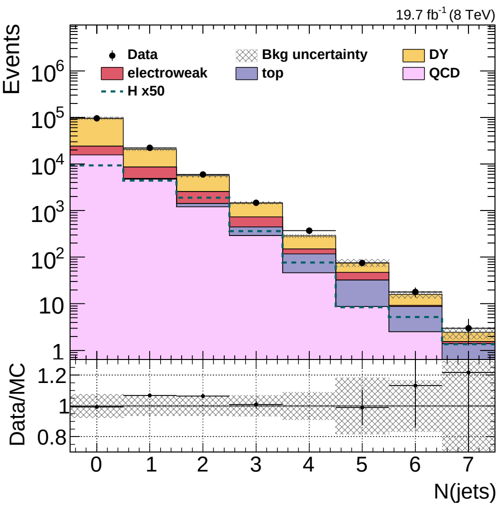
Figure 3.1: Distribution of the transverse mass $m_T(\mu, E_T^{miss})$. All cuts of the $\mu\tau_h$ baseline selection have been applied, except for the cut on m_T itself.

3.2.2. Event categories

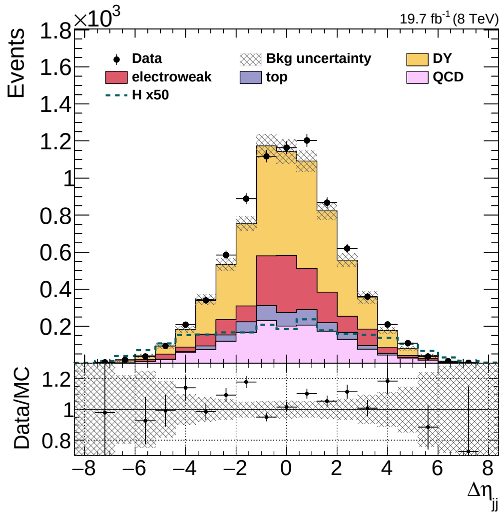
The baseline event selection is focused on the two signal leptons, i.e. the Higgs boson decay products. It is an inclusive selection as it neither distinguishes between different initial states, nor does it separate events with high sensitivity from ones with low sensitivity. The Higgs boson production mechanisms, whose Feynman diagrams were given in Figure 1.4, give rise to additional event signatures that were described in Section 1.1.3.2. These additional signatures are mainly parton jets that originate from the initial state. Other kinematic properties of the event can be used to suppress background and thus obtain larger sensitivities. At this stage of the analysis, events are not discarded, but merely sorted into categories that differ in signal and background contributions and thus in sensitivity.

The categorization of events is primarily based on the number and properties of jets that accompany the signal leptons. These jets are selected using the loose working point of the PU jet ID introduced in Section 1.3.4 and a threshold of $p_T > 30\,\text{GeV}$. A wide acceptance in η is important to collect jets from VBF production, which are expected to occur close to the LHC beam. Accordingly, jets with $|\eta| < 4.7$ are accepted. Jets must not overlap with the signal leptons within a cone of radius $\Delta R = 0.5$.

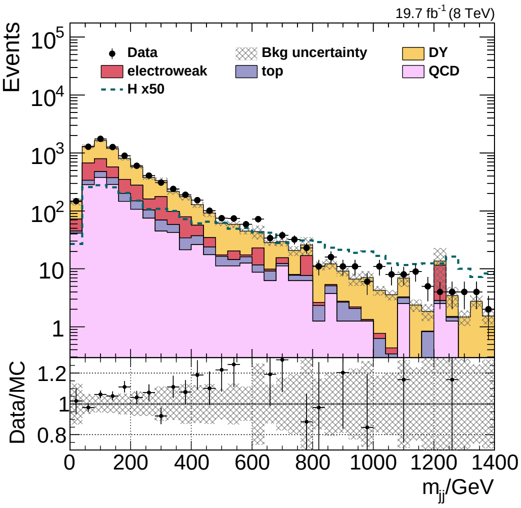
Distributions of variables used in the event categorization are presented in Figure 3.2. The shown background predictions are obtained as outlined in Section 3.2.3. The Higgs boson signal is depicted as dashed curve; it combines Higgs boson production via the ggF and VBF mechanisms. The number of jets in the event (Figure 3.2(a)) is utilized first. While DY events are mainly found without additional jets, the bins with one or more jets show a larger signal contribution. Correlations of the two p_T-leading jets are investigated to identify VBF signatures. This includes the η gap shown in Figure 3.2(b), where jets from VBF Higgs boson production are expected to be close to the beam line, and thus show large $|\Delta\eta_{jj}|$. For



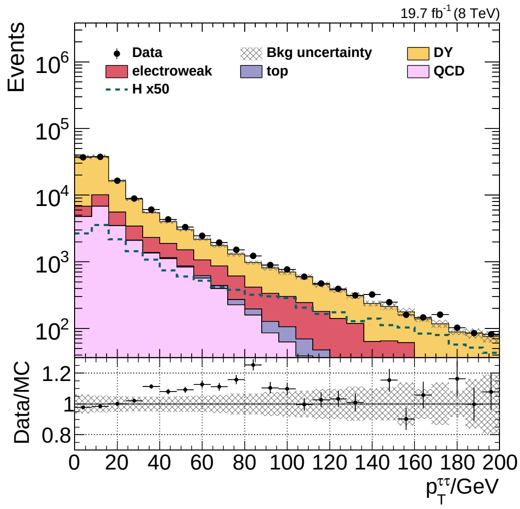
(a) Number of jets.



(b) Gap between jets $\Delta\eta_{jj}$.



(c) Invariant mass m_{jj} of the dijet system.



(d) H candidate transverse momentum $p_{\mathrm{T}}^{\tau\tau}$.

Figure 3.2: Variables used to categorize events that pass the baseline event selection. Backgrounds are estimated using the methods described in Section 3.2.3.

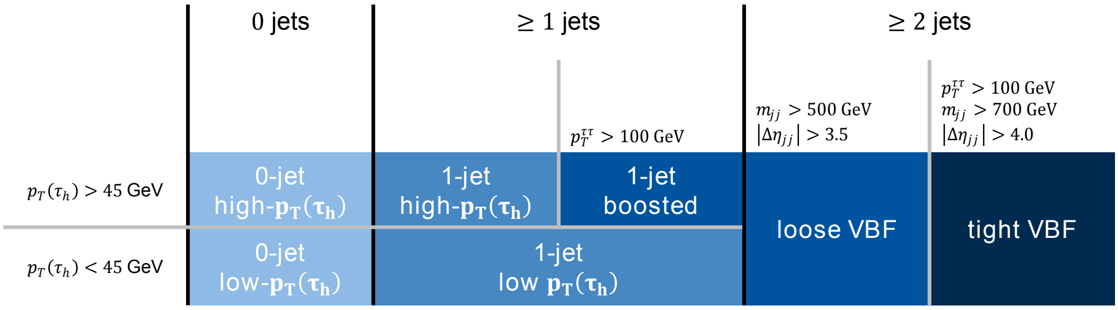


Figure 3.3: Event categories in the $\mu\tau_h$ channel. Categories with higher signal significance have a darker shade.

the same reason, the invariant mass m_{jj} of the dijet system (Figure 3.2(c)) is large for VBF Higgs events. The transverse momentum $p_T^{\tau\tau}$ of the Higgs boson candidate is shown in Figure 3.2(d). It is obtained from the vectorial sum of the momenta of the two signal leptons and $\vec{E}_T^{miss}$ and is sensitive to the fact that ggF Higgs boson production is more likely to give rise to additional jets from ISR than Z boson production. Accordingly, Higgs bosons show a harder p_T spectrum than Z bosons. More importantly, a boosted Higgs boson decays into two taus with similar directions in the laboratory frame. Accordingly, the sum of the neutrino momenta is larger. As larger values of E_T^{miss} can be reconstructed more accurately, events with boosted Higgs bosons allow for a better ditau mass resolution (see Section 3.3).

An overview of the seven event categories in the $\mu\tau_h$ channel is shown in Figure 3.3. VBF Higgs boson production is targeted by 2-jet categories. A *VBF tag* is assigned if two or more jets are found that fulfill a set of requirements. The η gap between the two p_T-leading jets must be large, $|\Delta\eta_{jj}| > 3.5$. If there are more than two jets, no other jet is allowed within the η gap between the leading jets. As the two selected jets are assumed to originate from quarks that have been scattered from a beam proton, their invariant mass must be large. A cut at $m_{jj} > 500\,\text{GeV}$ is applied. These requirements match the particular features of the VBF Higgs boson production (see Section 1.1.3.2) and efficiently reduce background from DY processes.

A further partition of the VBF category is introduced to obtain a category with high signal significance. Events with $|\Delta\eta_{jj}| > 4.0$, $m_{jj} > 700\,\text{GeV}$, and $p_T^{\tau\tau} > 100\,\text{GeV}$ are assigned to a *tight VBF* category. Events that fail this selection, but have a VBF tag assigned, are collected in a *loose VBF* category. While the VBF categories contain only a small amount of signal events, the categorization suppresses most of the background. This results in the best signal significance among the categories.

Event categories with one jet mainly aim at ggF Higgs boson production. The additional jet originates from initial state radiation. This suppresses Drell-Yan background as quarks in the initial state, compared to gluons, are less likely to radiate partons. A given event is assigned to a 1-jet category if it contains one or more jets and fails the VBF tag. The *1-jet low-$p_T(\tau_h)$* category is chosen if the selected τ_h candidate has $p_T < 45\,\text{GeV}$, while events with $p_T(\tau_h) \geq 45\,\text{GeV}$ are sorted into the *1-jet high-$p_T(\tau_h)$* category. The distribution of the τ_h transverse momentum is presented in Figure 3.5(a). A Higgs boson with larger mass than the Z boson gives rise to leptons with larger $p_T(\tau_h)$. Simultaneously, misidentified leptons are more likely to have smaller p_T. Consequently, the 1-jet high-$p_T(\tau_h)$ category is enriched with Higgs boson events. Events with $p_T^{\tau\tau} > 100\,\text{GeV}$ are split from the 1-jet high-$p_T(\tau_h)$ category to build a separate *1-jet boosted* category. It is the category with best significance towards ggF Higgs boson production.

Events with no jets are sorted into a *0-jet low-*$p_{\mathrm{T}}(\tau_{\mathrm{h}})$ and a *0-jet high-*$p_{\mathrm{T}}(\tau_{\mathrm{h}})$ category. The separation value is set to $p_{\mathrm{T}}(\tau_{\mathrm{h}}) = 45\,\mathrm{GeV}$. Although most of the Higgs bosons produced at the LHC are not accompanied by additional jets, the 0-jet categories offer only negligible sensitivity for Higgs searches, because the irreducible DY background dominates the Higgs production. Simultaneously, the background from misidentified quark and gluon jets is substantial.

Each category has distinct contributions from the ggF and VBF signals, but also from the different background processes. The global event fit as last step of the analysis exploits this information (see Section 3.4). The 0-jet categories play an essential role in the $\mathrm{H} \to \tau\tau$ search since various sources of systematic uncertainties can be constrained by fitting the background to the observed data in the 0-jet channels. As most systematic uncertainties are correlated between categories and channels, this improves the signal deduction in the high-sensitivity categories.

3.2.3. Signal samples and background estimation

Monte-Carlo samples

A description of MC generation at CMS has been given in Section 1.3.7. For this analysis, signal samples are used to estimate the yield and shape of events with $\mathrm{H} \to \tau\tau$ decays that have been produced via ggF or VBF processes. The MC event generation for this purpose is performed using POWHEG. A correction to the p_{T} spectrum of the produced Higgs bosons is applied. The reweighted spectrum complies with the prediction by a next-to-next-to-leading order (NNLO) calculation [103]. As the Higgs-boson mass was not known a priori, samples with masses between 110 and 145 GeV are analyzed. The difference in m_{H} between two consecutive samples is 5 GeV.

In all channels the estimation of the most important backgrounds is based on data. Subordinate processes are simulated by MC techniques. The $\mu\tau_{\mathrm{h}}$ channel presented in this thesis uses samples produced by MADGRAPH for Z + jets, W + jets, and VV production. Events generated by POWHEG are used to estimate backgrounds with top quarks, namely $\mathrm{t\bar{t}}$ and single-top-quark processes.

Drell-Yan

A proper modeling of $\mathrm{Z} \to \tau\tau$ processes is crucial for $\mathrm{H} \to \tau\tau$ searches. The signal is expected to manifest as an excess on the high-mass flank of this irreducible background. Therefore, the $m_{\tau\tau}$ shape of this contribution must be well understood. Also important are knowledge of the DY yield and well-populated background templates with small statistical uncertainties in the signal region. All of this is achieved by usage of so-called *embedded* events [104]. This technique is based on a sample of selected $\mathrm{Z} \to \mu\mu$ events from data. Given their clean signature, such events can be effectively selected with minimal background contribution. The PF muons are removed from the event and replaced by taus with identical kinematic properties. As for the MC samples, the tau decays are subsequently simulated using TAUOLA. The PF candidates reconstructed from the simulated tau decay products are added to the event and the object reconstruction is recomputed. This means that jets, $E_{\mathrm{T}}^{\mathrm{miss}}$, τ_{h} candidates, and lepton isolation are determined from a data event with embedded simulated tau decays.

Normalization of the embedding sample is performed on all events that pass the baseline selection excluding the m_T cut. The yield N_{tot} of these events is normalized to the observed yield in $Z \to \mu\mu$ events. To obtain the embedding yield N_{cat} in a given category, the efficiency ε_{m_T+cat} of embedding events to pass the m_T cut and the category selection is applied, $N_{cat} = N_{tot} \cdot \varepsilon_{m_T+cat}$. The transverse mass m_T and the categorization variables are modeled better in the embedding sample than in MC events. As ε_{m_T+cat} is determined from the embedding sample, inaccuracies in the MC modeling do not affect the yield estimate.

Usage of the embedding sample significantly reduces the systematic uncertainties stemming from the dominant background. Since most of the detector response originates from data, effects like pileup and also detector alignment, noise, and response do not have to be modeled. The uncertainties connected to the jet energy scale, E_T^{miss}, and luminosity can be neglected for the embedding sample.

Background from $Z \to \mu\mu$ and, less importantly, $Z \to ee$ is modeled using MC simulation.

W + jets

The shape of the background originating from W + jets is modeled by MC. To limit the impact of statistical fluctuations within the simulated sample, some of the selection requirements are relaxed in the categories with small event yields, i.e. 1-jet boosted, loose VBF, and tight VBF. To select W + jets events from the MC sample, only in these categories the two signal leptons may have the same charge. Additionally, the requirements on τ_h isolation, m_{jj}, $|\Delta\eta_{jj}|$, and the central jet veto are relaxed.

The yield of this background process is obtained from data. For this purpose a sideband region in data is defined by $m_T > 70\,\text{GeV}$ ($60 < m_T < 120\,\text{GeV}$ for the VBF categories). As shown in Figure 3.1, this sideband is dominated by W + jets background. To account for contamination from other backgrounds, distributions of DY, top quark and diboson processes within the sideband are estimated using MC samples and subtracted from data. The yield obtained in the sideband in data is extrapolated to the signal region, $m_T < 30\,\text{GeV}$, using the ratio of MC events in the signal and sideband regions. Similarly to the shape estimate, the event selection is relaxed in the VBF categories for the determination of the extrapolation factor.

QCD multijet

QCD multijet background cannot be accurately modeled by MC. The determination of its shape and event yields relies on data.

To obtain the shape, a sideband in data is defined by selection of lepton candidates with identical charge, instead of opposite charge in the default event selection, and an inverted relative muon isolation, $20\% < I_\mu < 50\%$. Alteration of these cuts does not affect the kinematic distributions of QCD multijet events. The sideband is highly dominated by QCD multijet processes. Contributions from other processes are negligible. To obtain a smooth shape in the categories with small event yields, event selection cuts are relaxed. In the VBF categories jet-related cuts are loosened, while the tau isolation requirement is softened in the 1-jet boosted category.

The QCD multijet event yield is estimated by multiplication of the event yield obtained in a sample with same-sign lepton candidates with a scale factor $f_{OS/SS}$. In contrast to the shape estimate, the muon isolation is not inverted. As a small contribution from other backgrounds in this same-sign sample is expected, the yields of these processes are subtracted. This includes DY, top quark and diboson processes, which are estimated from MC, and W + jets events as predicted by the corresponding method described above. As the charges of μ or τ_h candidates that originate from misidentified parton jets are not correlated with the muon isolation, $f_{\text{OS/SS}}$ is obtained in the inverted relative muon isolation sideband $20\% < I_\mu < 50\%$.

For the 1-jet boosted, loose VBF, and tight VBF categories the event yields in the sideband regions are small. This leads to large uncertainties on the predicted yield of QCD multijet events in the signal region. Accordingly, an alternative method is deployed in these categories. It is based on the QCD multijet yield after the baseline selection, which is obtained by the method described before. This yield is multiplied with the efficiency of QCD multijet events to pass the category selection. To obtain the efficiency, sidebands are defined by alteration of the charge requirement, the τ_h isolation and the μ isolation.

Quality of background prediction

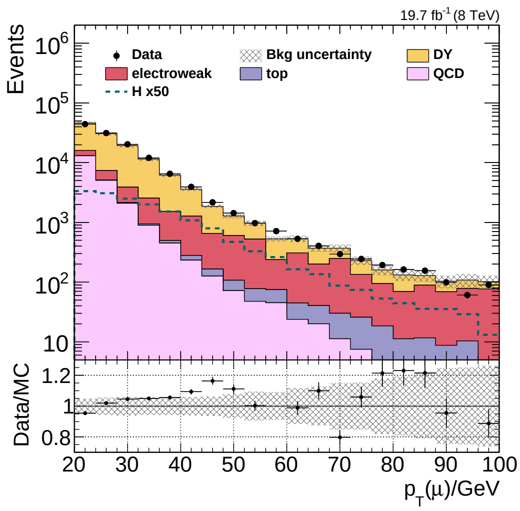
Figure 3.4 and Figure 3.5 show kinematic distributions of the selected μ and τ_h candidates, respectively. Only the baseline event selection has been applied and the background distributions are obtained by the methods explained above. The background predictions agree with the data reasonably well. Differences up to 20% are observed in the shape of the DY background modeled by the embedding technique at medium and large transverse momenta as well as at large $\eta(\mu)$. The origin of these effects is attributed to a set of corrections that have to be applied to the embedding sample. These corrections include weights that account for the influences of the underlying $\text{Z} \to \mu\mu$ event selection, trigger efficiencies, tau polarization, and a bias introduced by repeating the tau decay simulation until the tau passes the event selection. The set of corrections used in this work is not identical to the analysis presented in Ref. [102].

Systematic uncertainties in the background estimation are represented by shaded bands in the figures. A discussion of these uncertainties is given in Section 3.4. Additionally, the uncertainty bands include statistical uncertainties on the bin content as well as overall yield uncertainties on each of the background components. Shape uncertainties as well as yield uncertainties that differ for each category are not included in the figures. For illustration, the expected distribution of a SM Higgs boson with a mass of 125 GeV, scaled by a factor 50, is shown as a dashed curve.

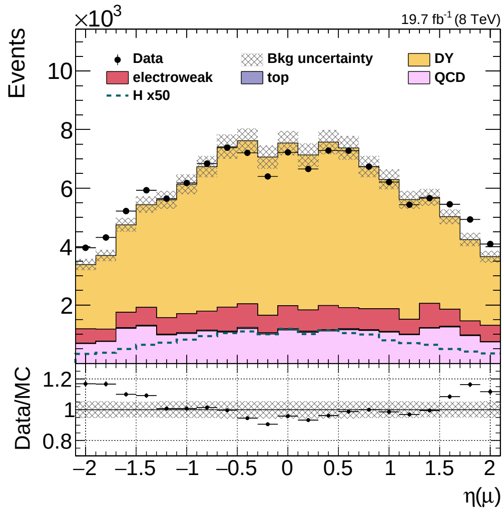
As the background is estimated independently in each category, the quality of the prediction can vary. The distributions presented in Figures 3.1-3.7 are obtained from events passing the baseline selection. They have not been subject to event categorization, but rather represent the aggregate of the individual categories used in the global fit. As described in Section 3.4.1, the global fit can account for disagreements between data and predicted backgrounds. Especially the 0-jet categories, which show large event yields but small sensitivity, are exploited to improve the quality of the background estimation.

3.3. Reconstruction of the invariant ditau mass

To separate Higgs boson and DY events, the invariant mass of the ditau system is the decisive variable. A simple estimator is the visible mass $m_{\text{vis}}(\tau_\text{h}, \mu)$, which is constructed as the invariant mass of the selected μ and τ_h candidates. The visible mass shows good agreement between data and predicted background, as can be seen in Figure 3.6. However, it neglects the neutrinos from both tau decays. Accordingly, the visible mass underestimates the true

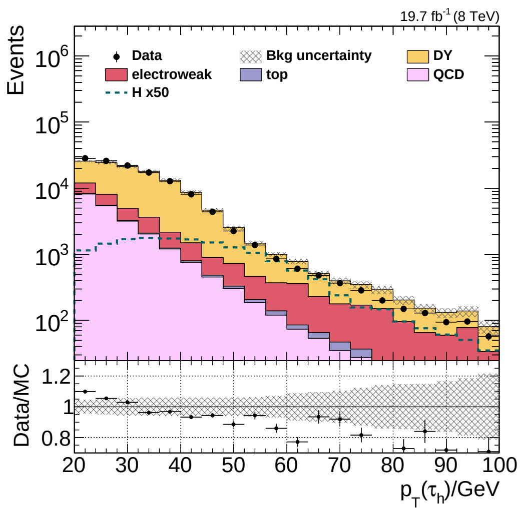


(a) Transverse momentum.

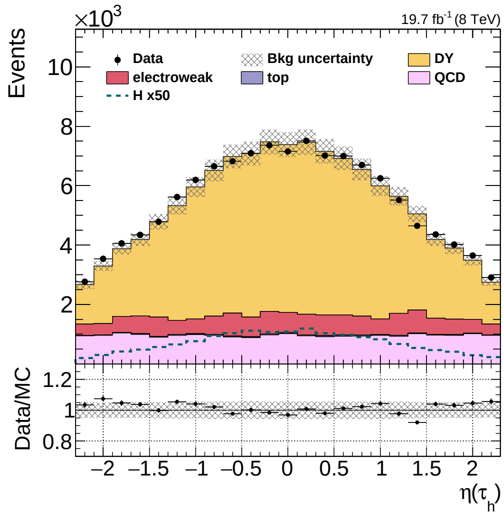


(b) Pseudorapidity.

Figure 3.4: Transverse momentum (a) and pseudorapidity (b) of the selected muon candidates for events passing the baseline event selection.



(a) Transverse momentum.



(b) Pseudorapidity.

Figure 3.5: Transverse momentum (a) and pseudorapidity (b) of the selected τ_h candidates for events passing the baseline event selection.

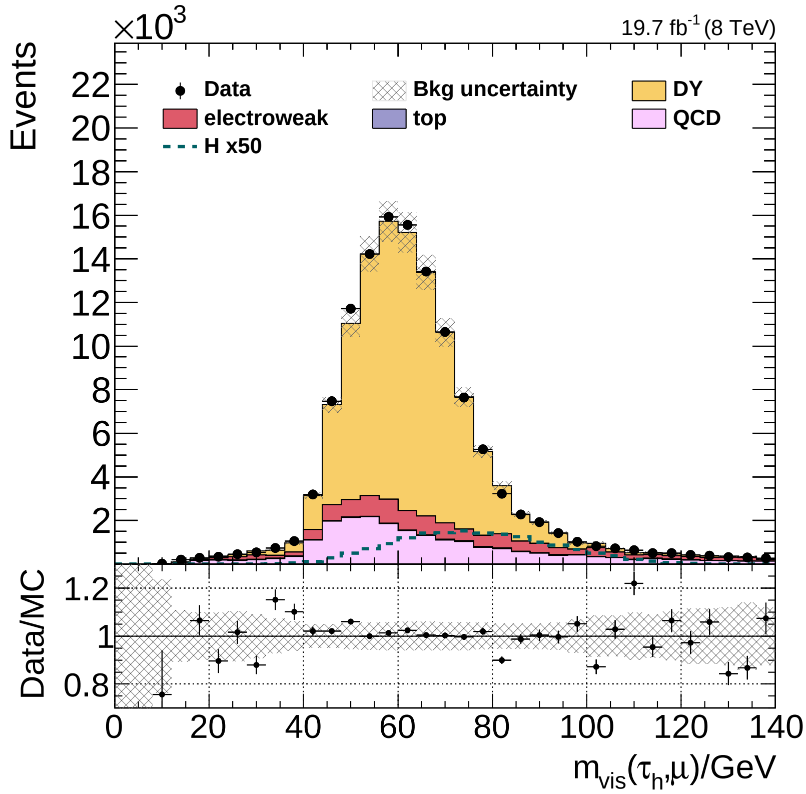
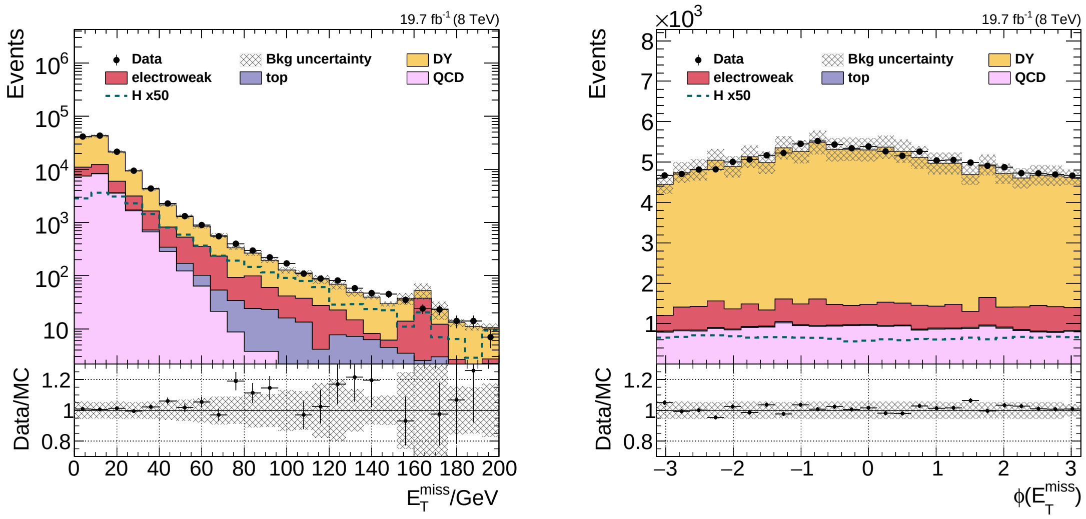


Figure 3.6: Visible mass of the selected $\mu\tau_\mathrm{h}$ system for events passing the baseline event selection.



(a) Absolute value. (b) Azimuthal angle.

Figure 3.7: Absolute value (a) and azimuthal angle (b) of the reconstructed missing transverse momentum for events passing the baseline event selection.

value of $m_{\tau\tau}$. For the same reason, the mass resolution is limited.

As the momenta of the neutrinos produced in the τ decays cannot be measured directly, the full invariant ditau mass $m_{\tau\tau}$ is not conveniently accessible. This can be compensated by use of the SVFIT technique. It reconstructs $m_{\tau\tau}$ by use of a likelihood method that incorporates information from the tau decay model and the measured directional $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ [102]. Neutrinos from tau decays are the only source of genuine $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ for events with Higgs bosons that were produced via the ggF or VBF mechanisms as well as for the main background from DY processes. The two measured components of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, i.e. its amplitude and azimuthal angle, therefore provide two constraints on the ditau system and its invariant mass. Figure 3.7 shows the distributions of the absolute value $E_{\mathrm{T}}^{\mathrm{miss}}$ as well as the direction of the corresponding momentum as reconstructed by the MVA $E_{\mathrm{T}}^{\mathrm{miss}}$ technique described in Section 1.3.5.

Tau leptons decay into a visible part, which is reconstructed either as light lepton or as τ_{h}, and an invisible part, which contributes to $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. While the four-momentum of the visible component is measured, the invisible component gives rise to three unconstrained parameters; the fourth parameter is fixed by the known tau mass. The free parameters are chosen as the energy fraction x of the tau energy that is carried by the visible component, the azimuthal angle φ of the tau lepton in the laboratory frame, and the invariant mass $m_{\nu\nu}$ of the two-neutrino system in leptonic tau decays. The set of parameters that describe a tau decay is therefore given as $(\vec{p}_i, \vec{a}_i)$, with the visible momentum $\vec{p}_i$ of tau i and the unknown parameters

$$\vec{a}_i = (x_i, \varphi_i, m_{\nu\nu,i}) \qquad \text{with } i = 1, 2\,. \tag{3.1}$$

The energy fraction parameter x can take values between 0 and 1 for leptonic tau decays. The mass of the neutrino system is constrained by $0 \leq m_{\nu\nu} \leq m_\tau\sqrt{1-x}$. For hadronic decay modes the boundaries are given as $m_{\tau_{\mathrm{h}}}^2/m_\tau^2 \leq x \leq 1$, where $m_{\tau_{\mathrm{h}}}$ is the invariant mass of the visible tau decay products, i.e. of the reconstructed τ_{h} candidate. The mass of the neutrino system is zero, as only one neutrino is produced. Accordingly, a hadronic tau decay is described by two parameters.

Overall, the combination of two tau decays in the event with two constraints from $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ results in an underconstrained system. A likelihood method is used to find the most probable value of $m_{\tau\tau}$. The likelihood function $\mathcal{L}$ is constructed as a product of three terms,

$$\mathcal{L} = \mathcal{L}_{\tau_1}(\vec{p}_1, \vec{a}_1) \cdot \mathcal{L}_{\tau_2}(\vec{p}_2, \vec{a}_2) \cdot \mathcal{L}_{\nu}(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{a}_1, \vec{a}_2) \ . \tag{3.2}$$

The first two terms describe the physical compatibility of the assumed parameter set $(\vec{p}, \vec{a})$ with a tau decay for both hemispheres. The form of this term depends on the tau decay mode. For leptonic tau decays, the matrix element for unpolarized τ decays is used [105],

$$\mathcal{L}_{\tau_\ell} = \frac{\mathrm{d}\Gamma}{\mathrm{d}x\mathrm{d}\varphi\mathrm{d}m_{\nu\nu}} \propto \frac{m_{\nu\nu}}{4m_\tau^2}\left[(m_\tau^2 + 2m_{\nu\nu}^2)(m_\tau^2 - 2m_{\nu\nu}^2)\right] \ . \tag{3.3}$$

Hadronic tau decays are described using the phase space of a two-body decay,

$$\mathcal{L}_{\tau_\mathrm{h}} = \frac{\mathrm{d}\Gamma}{\mathrm{d}x\mathrm{d}\varphi} \propto \frac{1}{1 - m_{\tau_\mathrm{h}}^2/m_\tau^2} \ . \tag{3.4}$$

This neglects differences between the individual hadronic decay channels introduced by the intermediate resonances. The likelihood of the measured $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ to originate from a ditau system with parameters $(\vec{a}_1, \vec{a}_2)$ is described by

$$\mathcal{L}_{\nu}(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}) = \frac{1}{2\pi\sqrt{|\mathbf{V}|}} \exp\left[-\frac{1}{2}(\vec{E}_{\mathrm{T}}^{\mathrm{miss}} - \Sigma\vec{p}_\nu)^\intercal \, \mathbf{V}^{-1} \, (\vec{E}_{\mathrm{T}}^{\mathrm{miss}} - \Sigma\vec{p}_\nu)\right] , \tag{3.5}$$

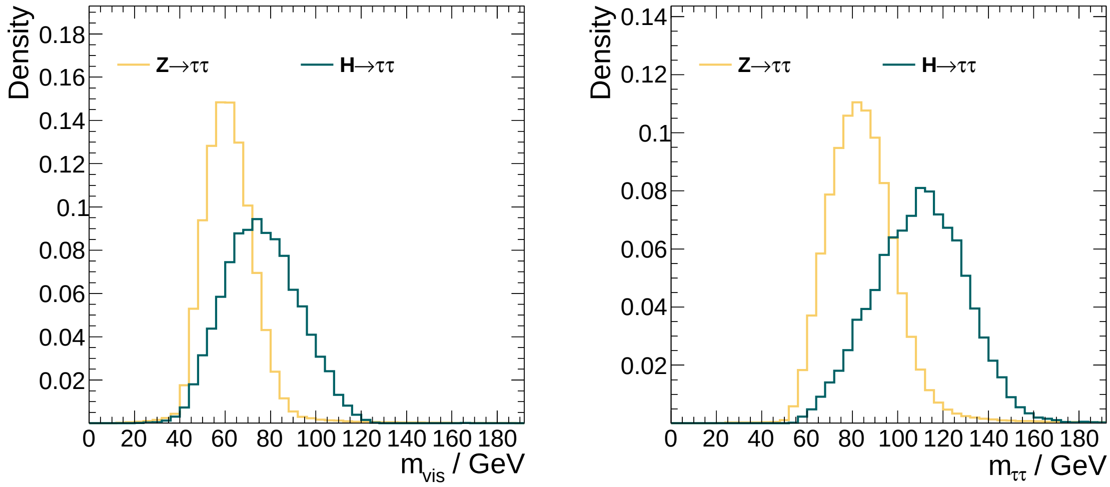
where $\Sigma\vec{p}_\nu$ denotes the momentum sum of the neutrino systems for the given parameter set. The covariance matrix $\mathbf{V}$ of the measured missing transverse energy is determined on an event-by-event basis by propagating the uncertainties of the objects used to construct $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$.

The reconstructed value of $m_{\tau\tau}$ is the mass hypothesis $\tilde{m}_{\tau\tau}$ that maximizes the probability

$$P(\tilde{m}_{\tau\tau}) = \int \mathrm{d}\vec{a}_1 \mathrm{d}\vec{a}_2 \, \mathcal{L}(\vec{p}_1, \vec{p}_2, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{a}_1, \vec{a}_2) \, \delta\left[\tilde{m}_{\tau\tau} - m_{\tau\tau}(\vec{p}_1, \vec{p}_2, \vec{a}_1, \vec{a}_2)\right] \ . \tag{3.6}$$

The integration is performed over all physically allowed values of x, φ, and $m_{\nu\nu}$. While $\mathcal{L}$ itself does not depend on x, the dependency is introduced by the integration boundaries. The angles φ are constraint by $\mathcal{L}_\nu$. The mass of the neutrino system appears in the matrix element used in $\mathcal{L}_{\tau_\ell}$.

The mass resolution achieved by SVFIT depends on the ditau decay channel. Ref. [102] gives values of about 10% in $\tau_\mathrm{h}\tau_\mathrm{h}$, 15% in $\ell\tau_\mathrm{h}$, and 20% in $\ell\ell$ channels. The difference is caused by the amount of neutrinos in the tau decays. Additionally, the resolution varies within a few percent between the categories of the $\mathrm{H} \to \tau\tau$ search. Especially in categories with boosted Higgs bosons, where $p_{\mathrm{T}}^{\tau\tau}$ is large, a better $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution is achieved, which directly translates to the mass resolution. Figure 3.8 shows the shape distributions of the reconstructed mass after the baseline $\mu\tau_\mathrm{h}$ selection presented in Section 3.2.1. The SVFIT mass $m_{\tau\tau}$ shown in Figure 3.8(b) achieves a resolution of 18% for both $\mathrm{Z} \to \tau\tau$ and $\mathrm{H} \to \tau\tau$ events. The mean of the distributions underestimate the true mass values by about 10%. This is a crucial improvement compared to the visible mass with a resolution of about 21%. The values of m_{vis} underestimate the true mass value by 31% for DY events and 39% for Higgs boson events. Separation of the signal from the dominant irreducible background is significantly better using the SVFIT mass $m_{\tau\tau}$.



(a) Visible mass. (b) SVFIT mass.

Figure 3.8: Shape of the reconstructed mass for $\mathrm{H} \to \tau\tau$ and $\mathrm{Z} \to \tau\tau$ events that passed the baseline $\mu\tau_\mathrm{h}$ event selection. The mass is calculated using (a) the visible mass and (b) the SVFIT technique.

3.4. Statistical inference

3.4.1. Global fit

To extract a signal from the selected data, a global fit is deployed. It can be considered as a template fit using $m_{\tau\tau}$ as discriminating variable. Accordingly, it uses the $m_{\tau\tau}$ distribution of the selected data as input as well as the yield and shapes of all individual background and signal components. To benefit from the event categories, the data as well as the templates are constructed for each category individually. The fit is then performed simultaneously for all categories in all channels.

In each category, systematic uncertainties are estimated and implemented as nuisance parameters in the fit. Uncertainties in the yield of background components are implemented as normalization nuisance parameters with a log-normal probability density. As $m_{\tau\tau}$ distributions are exploited by the fit, uncertainties affecting the shape in this variable are also considered. They are modeled assuming a Gaussian probability density. While the nuisance parameters are treated individually for each category, they can be fully or partially correlated between categories and even channels. This feature allows to use the combined information from all categories and all channels to constrain important uncertainties. Most importantly, low-sensitivity categories like the 0-jet categories in the $\mu\tau_\mathrm{h}$ channel are used to constrain the shape and normalization of DY background.

The fitting technique is a binned maximum likelihood fit. The likelihood combines all $m_{\tau\tau}$ bins from all categories. For each bin, the likelihood term is constructed by comparing data with the sum of estimated backgrounds and a possible Higgs boson signal. The signal component is predicted by a SM Higgs boson sample. Samples with different values of the Higgs boson mass are used, and the fit is performed once for each mass value. The signal is weighted by the signal strength modifier μ that was introduced in Equation (1.19). It is a free parameter in the fit and acts as parameter of interest.

The results presented in this thesis solely rely on the $\mu\tau_\mathrm{h}$ channel using the 8 TeV data set. Accordingly, the global fit is performed using only the categories of this channel. This means that correlations between channels cannot be considered. Even with identical inputs, the results will therefore differ from the results obtained in the full analysis.

3.4.2. Systematic uncertainties

The analysis presented in this thesis uses systematic uncertainties as described in Ref. [102]. They are summarized in this section. The global fit treats systematic uncertainties as nuisance parameters. That means that the fit is allowed to vary the value of the parameter within the allowed range. Various elements like event yields, kinematic distributions, and background models differ between categories. Accordingly, the sizes of systematic uncertainties can also differ. In these cases a range is given.

Several nuisance parameters affect the signal prediction. An uncertainty of 2.6% in the measurement of the integrated luminosity [66] is applied to all signal samples. PDF uncertainties are estimated to amount to 10% for ggF and 4% for VBF Higgs production. The renormalization and factorization scales in the signal simulation are varied, which impacts the signal yield by 8−31% for the ggF and 1−3% for the VBF sample, depending on the category. The model of the underlying event and parton showering introduces a rate uncertainty of 1−5%. For ggF events that enter one of the VBF categories, this uncertainty increases to 12%.

As the event selection is based on identification of the signal leptons, several uncertainties related to these objects are considered. The selection efficiency for muons accounts to a 2% rate uncertainty for all signal and background processes with a genuine muon in the final state. This number includes effects caused by the trigger and the muon ID. As the reconstruction and identification of τ_h is more complex, the corresponding uncertainty for hadronic tau decays is larger; it is the dominating uncertainty in the $\mu\tau_\mathrm{h}$ channel. A rate uncertainty of 8% is applied to signal, $\mathrm{Z} \to \tau\tau$, $\mathrm{t\bar{t}}$, and VV events and is correlated between all categories. An additional uncertainty of 1−3% is applied that is correlated between the low-p_T categories and, separately, between the high-p_T categories.

The τ_h energy scale is determined in $\mathrm{Z} \to \tau_\mu\tau_\mathrm{h}$ events by fitting the invariant mass of the τ_h jet to the distribution observed in data. As this method is not applicable for the $\tau \to \pi^\pm\nu_\tau$ decay mode, a fit in the invariant $\mu\tau_\mathrm{h}$ mass is used. The uncertainty on the energy scale is estimated to 3%. It is applied as a shape uncertainty to the signal and $\mathrm{Z} \to \tau\tau$ templates. The event selection uses p_T cuts on the τ_h candidate in both the trigger and offline. This is affected by variations in the energy scale and reflected by a yield uncertainty.

Jets are the fundamental constituents of the event categorization. As described in Section 1.3.4, corrections to the jet energy scale, JEC, are applied on simulated events. Uncertainties on the JECs give rise to rate uncertainties for all signal and background processes that are based on MC predictions. The corresponding nuisance parameters are correlated between all categories. Their values are obtained by propagating variations on the JECs to the final analysis and amount to 1−16% depending on the process and category.

Reconstruction of $E_\mathrm{T}^\mathrm{miss}$ is based on jets and other objects in the event. The usage of $E_\mathrm{T}^\mathrm{miss}$ in the event selection via the m_T variable gives rise to a rate uncertainty for MC-based signals and backgrounds. Similar to the JEC uncertainties, a fully correlated set of nuisance parameters accounts for the $E_\mathrm{T}^\mathrm{miss}$ uncertainties. Yield uncertainties between 1 and 5% are implemented.

Backgrounds from Z bosons are normalized to a yield that was obtained in a $Z \to \mu\mu$ selection. A systematic uncertainty of 3% is applied to this method. The $Z \to \tau\tau$ yield in the individual categories is obtained by extrapolating the inclusive yield with the efficiency of the category selection. This extrapolation is subject to effects in the event modeling using the embedding method that have been studied by comparison with DY events generated by MC. Additional statistical uncertainties caused by the small number of events in the high-sensitivity categories also impact the extrapolation factors. Accordingly, the rate uncertainty caused by the DY extrapolation is set to 5% for the 0 and 1-jet categories, while its value is 10% (13%) for the loose (tight) VBF category. The background originating from $Z \to \mu\mu$ events relies on the knowledge of the misidentification rate, i.e. how often muons are reconstructed as τ_h. This is taken into account by a yield uncertainty of 30%, which reflects the limited amount of events in the tag-and-probe methods deployed to measure the misidentification rate. An additional 26−70% accounts for the statistical uncertainty on the estimate of the $Z \to \mu\mu$ background in the VBF categories. The energy scale of muons that are misidentified as τ_h can differ in data and simulation. As $Z \to \mu\mu$ events dominate the signal in the 0-jet categories and peak in the same $m_{\tau\tau}$ region, it is important to control this energy scale. Accordingly, a shape uncertainty that reflects a 2% shift in the energy scale is deployed.

Normalization of the W + jets background is performed via a high-m_T sideband individually for each category. The extrapolation to the signal region relies on the m_T shape in the Monte Carlo sample and on statistical uncertainties on the number of events in the sideband. Correlated nuisance parameters account for these effects in the different categories. The uncertainty values range from 11 to 51%, with higher uncertainties in categories with a small number of events.

The background introduced by $t\bar{t}$ events is suppressed mainly by a b-jet veto. The corresponding templates are affected by the systematic uncertainty on the b-tagging algorithm. Its impact is quantified to lie between 2% for the low-significance and 6% for the high-significance categories. The expected yield of events with top-quarks is not controlled by sidebands and connected to a yield uncertainty between 10% and 35%. Similarly, diboson samples are subject to yield uncertainties of 15% in the 0- and 1-jet categories, and up to 50% in the VBF categories.

Background from QCD multijet events is hard to model. A yield uncertainty of 6−70% is applied to the QCD template, which is mainly driven by the amount of events in the sideband regions. Additional shape uncertainties account for possible differences in the $m_{\tau\tau}$ shape obtained from the sideband region and events in the signal region. This affects mainly events with small values of $m_{\tau\tau}$. Shape uncertainties are applied to the QCD templates in the 1-jet medium, 1-jet high, and loose VBF categories.

To account for statistical fluctuations in individual bins of the background templates, so-called bin-by-bin uncertainties are applied. These are implemented as shape uncertainties that modify the content of a single bin, while keeping the normalization of the template identical. Bin-by-bin uncertainties are applied to all bins with a relative statistical uncertainty of more than 10%. The MC samples used in this work to estimate W + jets and Z + jets backgrounds include any number of additional jets from the initial state. The number of generated jets corresponds to the SM expectation. Accordingly, the event yields in these samples are small for categories with additional jets in the final state. This leads to larger bin-by-bin uncertainties on these samples compared to the reference analysis presented in Ref. [102], which exploits dedicated $V + n$ jets samples, with $n = 1 \ldots 4$.

3.4.3. Exclusion limits

The $\mathrm{H} \rightarrow \tau\tau$ analysis is set up as a search for a predictable signal on top of known backgrounds. To judge whether the signal exists, the same statistical methods are deployed by both ATLAS and CMS [106]. Exclusion limits are computed by means of the $\mathrm{CL_s}$ method [107, 108], which compares the observed data with both a background-only model and a model that contains background plus the signal.

The prediction of the combined background is denoted as b. The signal model that is to be tested gives rise to an expected yield, which is labeled s. In this work, s corresponds to the predicted distribution of a SM Higgs boson with a specific mass. The signal strength modifier μ is used to scale the yield of the predicted signal distribution, $\mu \cdot s$. Accordingly, a value of $\mu = 0$ denotes that the signal is vanishing, while $\mu = 1$ corresponds to the SM prediction. While μ is a free scalar, s and b are distributions that depend on a set of uncertainties θ. These uncertainties are considered as nuisance parameters. Their nominal values $\tilde{\theta}$ and associated probability density functions (pdfs) $p(\tilde{\theta}|\theta)$ are obtained externally. The likelihood to observe the given data d can then be given as

$$\mathcal{L}(d|\mu,\theta) = \prod_i \frac{(\mu s_i + b_i)^{d_i}}{d_i!} \mathrm{e}^{-(\mu s_i + b_i)} \cdot p(\tilde{\theta}|\theta) \, . \tag{3.7}$$

Here, the probability to find the data d_i in a given bin i is modeled using a Poisson distribution.

Upper exclusion limits on the signal strength μ are set using a test statistic q_μ. It is constructed as profile likelihood ratio and takes the form

$$q_\mu = -2 \ln \frac{\mathcal{L}(d|\mu,\hat{\theta}_\mu)}{\mathcal{L}(d|\hat{\mu},\hat{\theta})} \quad \text{with } 0 \leq \hat{\mu} \leq \mu \, . \tag{3.8}$$

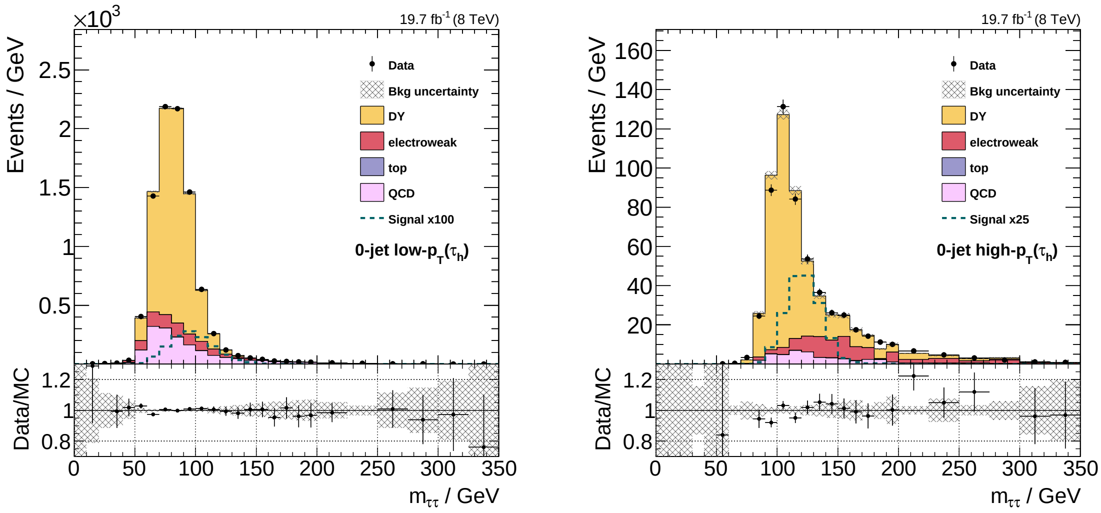
In this equation μ is the tested hypothesis, while $\hat{\theta}_\mu$ denotes the set of values of the nuisance parameters that maximize the likelihood for the given value of μ. The denominator describes the global maximum of the likelihood, with $\hat{\mu}$ and $\hat{\theta}$ being the corresponding set of parameters. Negative signal strength parameters that would correspond to a negative number of events are unphysical and excluded. To obtain a one-sided confidence interval that can be used to obtain upper limits, only signal strength parameters above the best fit $\hat{\mu}$ are tested. With $q_\mu \geq 0$, small values of the test statistic indicate good agreement between data and the signal hypothesis. If q_μ is large, the data is not well compatible with the chosen value of μ.

To obtain the pdf for q_μ, $p(q_\mu)$, a set of toy MC samples can be used. In these, the values of μ and θ are fixed, and q_μ is obtained for each toy MC by selecting a possible nominal value $\tilde{\theta}$ according to the known $p(\tilde{\theta}|\theta)$. This procedure is computationally expensive for analyses with many bins and nuisance parameters. For large data samples, the result of $p(q_\mu)$ obtained from toy MC is asymptotically described by an analytic formula [109]. The limits presented in this work make use of this asymptotic description.

To set limits, the cumulative distribution for $p(q_\mu)$ has to be determined,

$$\mathrm{CL_{s+b}} = \int_{q_\mu^{\mathrm{obs}}}^{\infty} p(q_\mu|\mu,\hat{\theta}_\mu)\mathrm{d}q_\mu \, . \tag{3.9}$$

The index s+b denotes that the underlying model contains both signal and background. Usage of $\mathrm{CL_{s+b}}$ can give counter-intuitive results in regions with a small signal on top of a large



(a) 0-jet low-$p_T(\tau_h)$ (b) 0-jet high-$p_T(\tau_h)$

Figure 3.9: Postfit distributions of $m_{\tau\tau}$ in the 0-jet categories of the $\mu\tau_h$ channel.

background. Here, an under-fluctuation of background can lead to strict limits on the signal contribution. To correct for this, CL_{s+b} is normalized to a background-only probability CL_b. The definition of CL_b follows Equation (3.9), but with $\mu \equiv 0$, resulting in

$$CL_b = \int_{q_\mu^{obs}}^{\infty} p(q_\mu|0, \hat{\theta}_0) dq_\mu \ . \tag{3.10}$$

The figure of merit used for limit setting is thus defined as

$$CL_s = \frac{CL_{s+b}}{CL_b} \ . \tag{3.11}$$

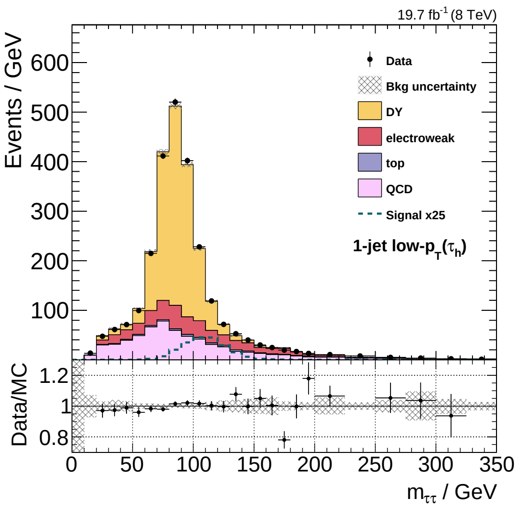
A tested value of μ is excluded at a confidence level $\alpha = 1 - CL_s$. A typical value of α is 95%. In this context, a tested value μ is excluded at 95% CL, if the corresponding value of CL_s is smaller than 5%.

Expected limits are used to judge on the sensitivity of an analysis to exclude a possible signal. They are constructed by inserting a background-only hypothesis as pseudo-data into the limit calculation procedure. Accordingly, they give the limit that an experiment would achieve if the observed data contained no signal.

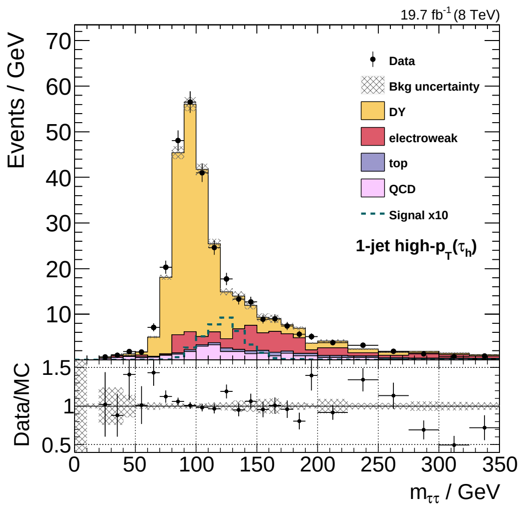
3.5. Results in the $\mu\tau_h$ channel

Postfit distributions of the $m_{\tau\tau}$ variable in the seven event categories of the $\mu\tau_h$ channel are presented in Figures 3.9, 3.10, and 3.11. The term *postfit* is used to emphasize that the event yields and shapes of the templates correspond to the result of the global fit, which is executed simultaneously in all $\mu\tau_h$ categories. The shaded uncertainty bands in the histograms include the values and correlations of all nuisance parameters. The distribution of a possible Higgs boson signal with a mass of 125 GeV is displayed as dashed line.

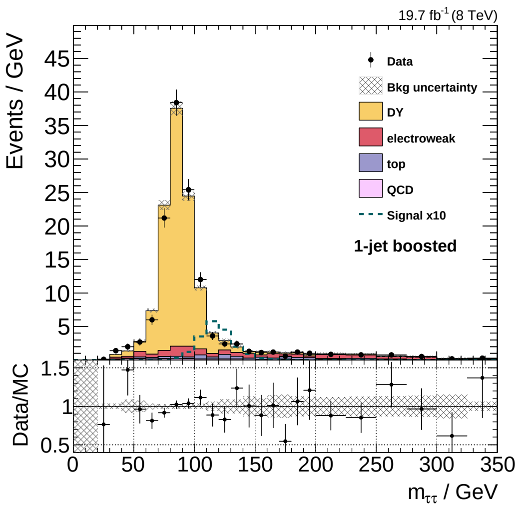
A possible signal in the 0-jet categories presented in Figure 3.9 is concealed by the large backgrounds. The 0-jet low-p_T category presented in Figure 3.9(a) is dominated by $Z \to \tau\tau$



(a) 1-jet low-$p_T(\tau_h)$

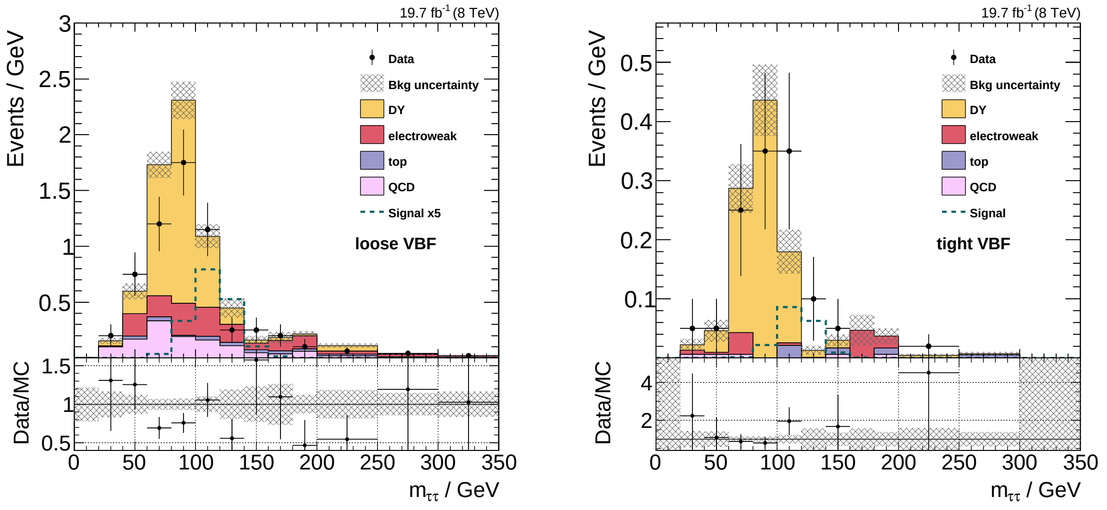


(b) 1-jet high-$p_T(\tau_h)$



(c) 1-jet boosted

Figure 3.10: Postfit distributions of $m_{\tau\tau}$ in the 1-jet categories of the $\mu\tau_h$ channel.



(a) loose VBF (b) tight VBF

Figure 3.11: Postfit distributions of $m_{\tau\tau}$ in the VBF categories of the $\mu\tau_h$ channel.

events. After the fit, the data in the DY peak region is well modeled. Misidentified muons from $Z \to \mu\mu$ decays contribute to the right flank of the DY distribution.

In the 1-jet categories (Figure 3.10), the signal contribution is marginal in the low-p_T category, which is mainly intended to constrain the background nuisances. A sizable signal yield is obtained especially in the 1-jet boosted category, which in turn is affected by sizable statistical uncertainties.

The VBF categories shown in Figure 3.11 contain the smallest event yields. Given the large statistical fluctuations, their contribution to the constraint of the nuisance parameters can be neglected. With the best signal-to-background ratios among all categories, they offer a high sensitivity to the signal.

A comparison of the sensitivity of the individual categories is presented in Figure 3.12 by means of their expected 95% CL limit. Assuming $m_H = 125\,\text{GeV}$, the best limit is obtained in the tight VBF category, which alone excludes Higgs boson production with cross sections above $1.7 \cdot \mu$. Boosted ggF Higgs production targeted in the 1-jet boosted category allows to exclude values of $\mu > 1.9$. The loose VBF, 1-jet high-p_T, and 0-jet high-p_T categories obtain similar sensitivities with upper limits around $\mu = 3.0$. The 0-jet and 1-jet low-p_T categories play no role in the combination. The expected 95% CL upper limit of the full $\mu\tau_h$ channel is obtained as $0.89^{+0.41}_{-0.28}$.

The observed limit as function of m_H is shown in Figure 3.13. For a Higgs boson with $m_H = 125\,\text{GeV}$, it is obtained as 1.4. The excess compared with the expected limit hints towards a Higgs boson signal. The most probable value of the signal strength modifier as determined by the fit is

$$\mu_{\text{obs}} = 0.60^{+0.43}_{-0.36}\,. \tag{3.12}$$

The value is obtained for a Higgs boson mass of 125 GeV. In the $\mu\tau_h$ channel alone, as implemented in this work, the observed signal strength is compatible with a SM Higgs boson within 1σ. The null hypothesis is compatible with the observed signal strength within 2σ.

The results presented in this work differ from the values published by CMS in Ref. [102]. The

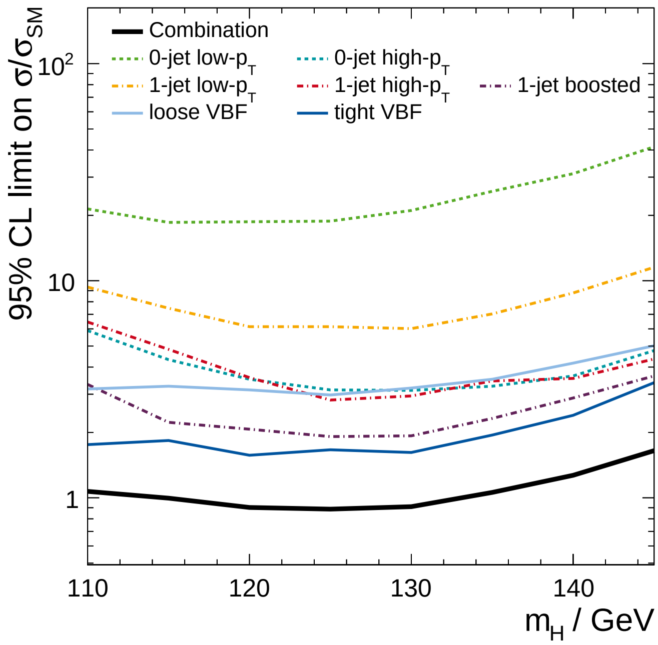


Figure 3.12: Expected 95% CL limits on $\mu = \frac{\sigma}{\sigma_{\mathrm{SM}}}$ obtained in the individual event categories of the $\mu\tau_{\mathrm{h}}$ channel. A smaller expected limit corresponds to a higher sensitivity. The limit is shown as a function of the Higgs boson mass hypothesis. The combination of all categories is shown as thick curve.

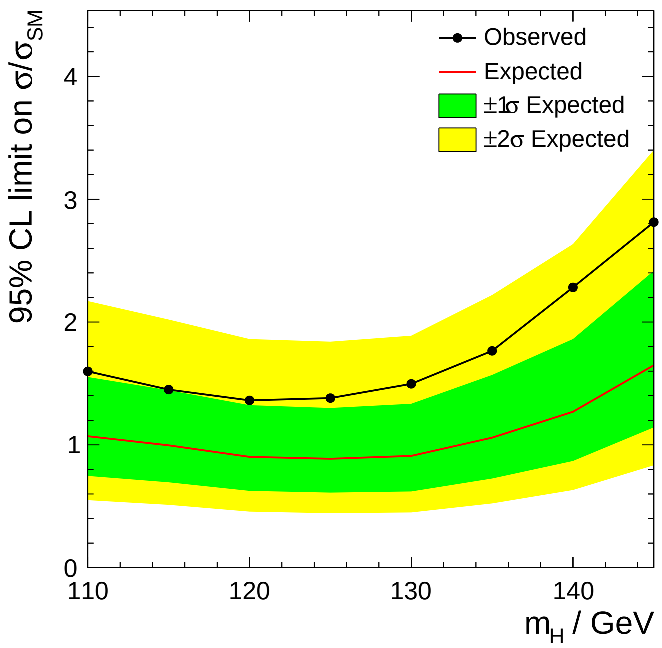


Figure 3.13: Expected and observed 95% CL limit on $\mu = \frac{\sigma}{\sigma_{\mathrm{SM}}}$ for different Higgs boson mass hypotheses in the $\mu\tau_{\mathrm{h}}$ channel. The green and yellow bands represent the 1σ and 2σ uncertainties on the expected limits. The expected limit is obtained with the assumption of no Higgs signal in the data.

best-fit value for μ given in the reference for the $\mu\tau_h$ channel is 1.01 ± 0.41. The corresponding expected limit for $m_H = 125\,\text{GeV}$ is $0.85^{+0.38}_{-0.6}$. These deviations originate from a different implementation, including changes in the scale factors applied to the embedding sample and usage of MC samples that are not enriched with additional jets. The results obtained in this work, which are compatible with the published version, are used in Chapter 4 to study modifications to the analysis.

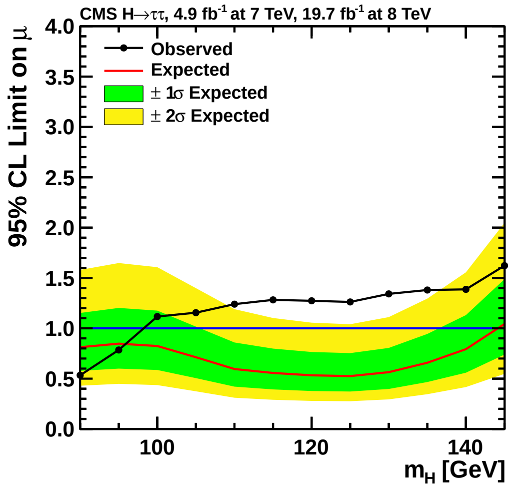
3.6. Evidence for fermionic Higgs couplings

The results obtained by the CMS collaboration using the LHC Run-I data provide evidence for the decay $H \to \tau\tau$ [102]. It combines the results obtained in seven different channels that aim at Higgs bosons produced via ggF, VBF, and VH. A selection of important results from this analysis is assembled in Figure 3.14. If there had been no Higgs boson, the analysis would have excluded a possible signal for all studied masses, as is shown by the expected limit in Figure 3.14(a). The observed limit contradicts a background-only hypothesis and is compatible with the SM prediction. To quantify the incompatibility with pure background, a local p-value is calculated. Its expected and observed distribution is presented in Figure 3.14(b). The expected p-value is a measure of the sensitivity of the search for a SM Higgs boson with a given mass. The maximal expected sensitivity of 3.7 standard deviations is obtained for $m_H = 125\,\text{GeV}$. The p-value observed in data is minimal for $m_H = 120\,\text{GeV}$ and corresponds to a significance of 3.3 standard deviations, providing evidence for the decay $H \to \tau\tau$.

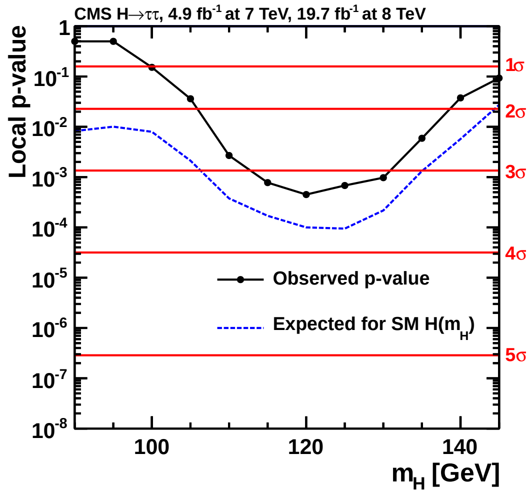
A complex statistical investigation of the data is performed to extract the signal. Accordingly, giving a visual representation of the data and the contained signal is challenging. The mass distribution presented in Figure 3.14(c) combines the results of four individual search channels that use $m_{\tau\tau}$ as final discriminant. Bins from the underlying histograms in each of the channels are weighted with the signal over signal-plus-background ratio ($S/(S+B)$) of the respective category, a measure for their contribution to the final result. The inlay compares the distribution of a SM Higgs boson with data that has been cleaned from the background component. A clear excess in data, compatible with the SM expectation, is visible.

The best fit value of the signal strength parameter μ obtained from a combination of all channels is 0.78 ± 0.27. Figure 3.14(d) shows the values obtained from the individual channels. All channels are compatible with each other as well as with the SM value $\mu = 1$. The uncertainties on the individual μ values are a measure for the sensitivity of the channel and thus for its contribution to the combined result. The $\mu\tau_h$ channel offers the best sensitivity. It is a sensible choice to study the impact of alterations on the search strategy, as is done in Chapter 4. Second best sensitivity is provided by both the $e\tau_h$ and $\tau_h\tau_h$ channels. The latter is based on the trigger developments described in Chapter 2 of this work, which allowed to collect fully-hadronic event signatures with reasonably low p_T thresholds.

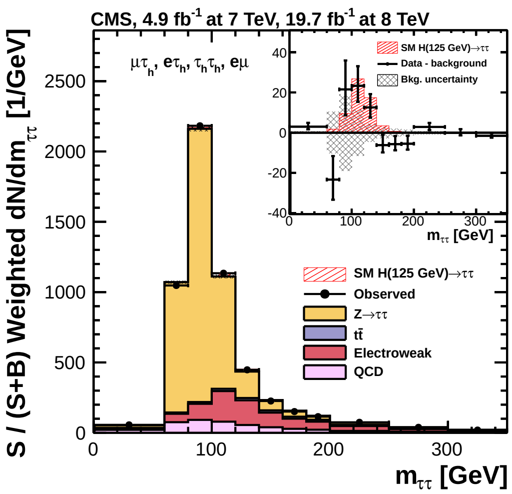
Couplings of the Higgs boson to fermions are a fundamental part of the standard model. Decays into $\tau\tau$ and bb final states allow to study the couplings to down-type fermions, which are not well constrained by observations in $\gamma\gamma$, ZZ, and WW signatures. A combination of the searches for $H \to \tau\tau$ and $H \to bb$ was published by CMS in Ref. [43]. The significance of Higgs decays into down-type fermions corresponds to 3.8 standard deviations, where 4.4 are expected. The ratio of the two decay modes corresponds to the predictions of the Yukawa couplings.



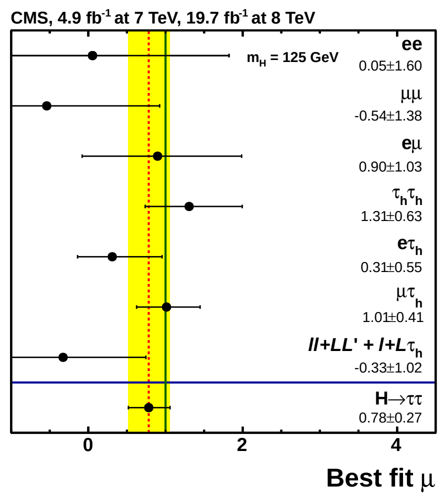
(a) Expected and observed 95% CL limit on μ. Without signal contribution, the analysis is expected to exclude H $\to \tau\tau$ processes over the full analyzed mass range. The excess in the observed limit indicates a signal.



(b) Expected and observed local p-values as a function of the Higgs-boson mass hypothesis. At $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, the observed (expected) p-value corresponds to a significance of 3.2 (3.7) standard deviations.



(c) Distribution of the Higgs boson mass estimator $m_{\tau\tau}$. The histogram is constructed by combining the mass distributions of four channels, weighted by the signal over signal-plus-background ratio in each category. The inlay shows a clear excess over the background that is compatible with the signal expectation.



(d) Best fit value for the signal strength modifier μ as obtained in the individual search channels. All values are compatible with a SM Higgs boson signal. The combined best fit value of μ is 0.78 ± 0.27.

Figure 3.14: Results obtained by the CMS collaboration in the search for H $\to \tau\tau$ decays using the Run-I data set [102]. Figures (a) and (b) show the combination of all considered channels. Figure (c) includes only those channels that deploy $m_{\tau\tau}$ as final discriminating variable.

Chapter 4

Higgs boson decays into 3-prong taus

The analysis of $\mathrm{H} \to \tau\tau$ decays at CMS, as described in Chapter 3, relies on the reconstruction of the τ decay products. The visible component of leptonic τ decays, an electron or muon, is identified and reconstructed with high performance. Hadronic τ decays give rise to various signatures in the detector: Depending on the underlying hadronic resonance, one, three, or more charged hadrons emerge, which may be accompanied by neutral hadrons. A description of the tau decay phenomenology was given in Section 1.1.4. The broad spectrum of hadronic tau decay modes is one of two reasons that τ_h are less potent in the $\mathrm{H} \to \tau\tau$ event selection than leptonic decays. The other reason is their similarity to other hadronic signatures that frequently appear in QCD interactions.

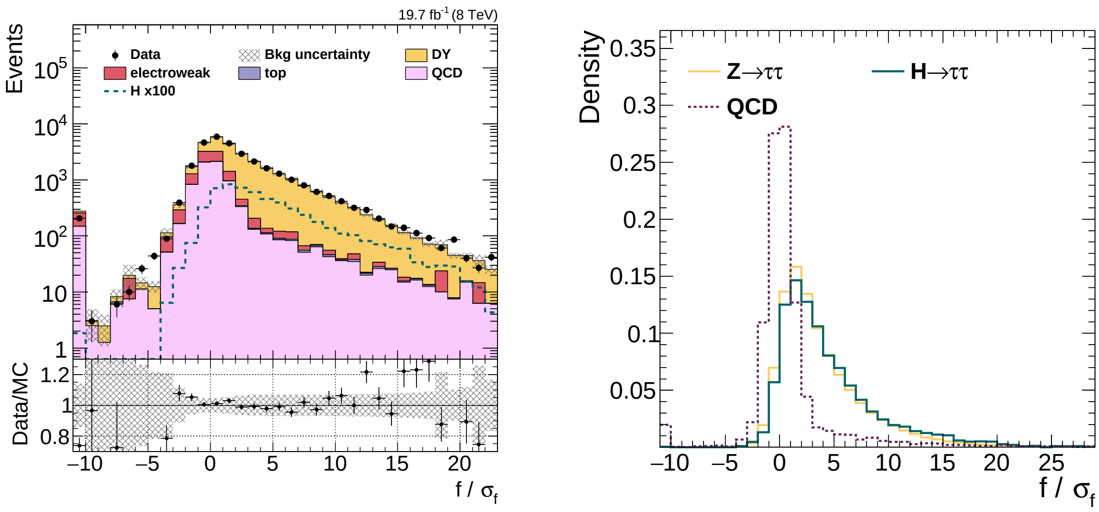
As described in Section 1.3.6, CMS addresses these challenges with the HPS algorithm. It is capable to efficiently identify and reconstruct hadronic tau lepton decays, while suppressing background from electrons, muons, and parton jets. Based on the reconstruction of individual, low-level objects by the PF algorithm, the HPS algorithm establishes the underlying hadronic decay mode of the τ candidate. This allows to apply reconstruction and identification techniques that exhibit the structure of a given τ decay. By doing so, the lower identification efficiency due to various signatures is mitigated by the access to additional information that remains hidden for leptonic tau decays. A detailed study of the individual τ_h modes is a prerequisite for this.

This work concentrates on the $\tau \to \pi^\pm\pi^\pm\pi^\pm\nu_\tau$ decay. The three charged pions are reconstructed as tracks and allow the reconstruction of the tau decay vertex. With this additional information, background from quark and gluon jets can be suppressed. Additionally, the full τ momentum can be reconstructed, including the neutrino component. Studies are performed to measure the impact of these methods on the $\mathrm{H} \to \tau\tau$ analysis.

4.1. Tau decay vertex

The three charged particles in the final state of $\tau \to \pi^\pm\pi^\pm\pi^\pm\nu_\tau$ decays are well measured by the CMS tracking system. They originate from a common secondary vertex (SV) that is identified with the position of the τ decay. A displacement of the SV from the pp collision vertex is expected, corresponding to the flight length of the τ lepton (see Section 1.1.4). For τ leptons from Z (H) decays, where the resonance was in rest in the laboratory frame, the mean distance is 2.2 mm (3.1 mm). Given the CMS vertex resolution of $\mathcal{O}(10\,\mu\mathrm{m})$ (see Section 1.3.1), a separation of the vertices is feasible.

In contrast, τ_h candidates that do not originate from a genuine τ decay are not expected to show this particular feature. If a quark or gluon jet is misidentified as τ_h, the signal tracks



(a) Event distribution. (b) Shape comparison.

Figure 4.1: Distribution of the signed flight length significance f/σ_f for events with 3-prong taus passing the $\mu\tau_\mathrm{h}$ baseline selection. Events without reconstructed SV are assigned to the underflow bin.

originate from the primary vertex (PV). An exception to this are b jets, where heavy B mesons have a sizable lifetime. At the LHC, these processes are rare in comparison to the total cross section of QCD dijet events (compare Figure 1.6). Should some of the tracks originate from PU or other hadronic activity, it is unlikely that they can be fitted to one common SV.

The SV reconstruction is not part of the default HPS algorithm. Here, this information is added for HPS taus that are identified as a 3-prong decay mode. The SV position is obtained by fitting the three signal tracks as identified by the HPS algorithm. The fit implements the adaptive vertex fitter algorithm [110].

During the CMS event reconstruction procedure, vertex fitting is performed before τ_h are identified. Accordingly, the PV that is connected to a τ candidate includes the three signal tracks. This introduces a bias on the position of the vertex. To remove this bias, the PV is refitted after removal of the three signal tracks from the track collection [52, 111]. The fitting technique used for the refitted PV is based on the Kalman filter formalism [70], similar to the standard PV reconstruction at CMS (see Section 1.3.1).

4.1.1. Flight length significance

The flight length of a τ candidate is given by the distance between the reconstructed production and decay vertices, $|\vec{SV} - \vec{PV}|$. Information can be added by considering the direction of flight: For a genuine τ the SV has to lie along the direction of the τ momentum. This consistency is tested by projection of the flight direction $(\vec{SV} - \vec{PV})$ on the momentum of the charged decay products $\vec{p}(\tau_\mathrm{h})$. Here, $\vec{p}(\tau_\mathrm{h})$ is used as an approximation for $\vec{p}(\tau)$. Accordingly, the signed flight length is defined as

$$f = \mathrm{sgn}\left[\vec{p}(\tau_\mathrm{h}) \cdot (\vec{SV} - \vec{PV})\right] \cdot |\vec{SV} - \vec{PV}|\,. \tag{4.1}$$

To incorporate the uncertainties in the vertex reconstruction, the signed flight length significance f/σ_f is used as discriminating variable. The uncertainty σ_f is obtained from the co-

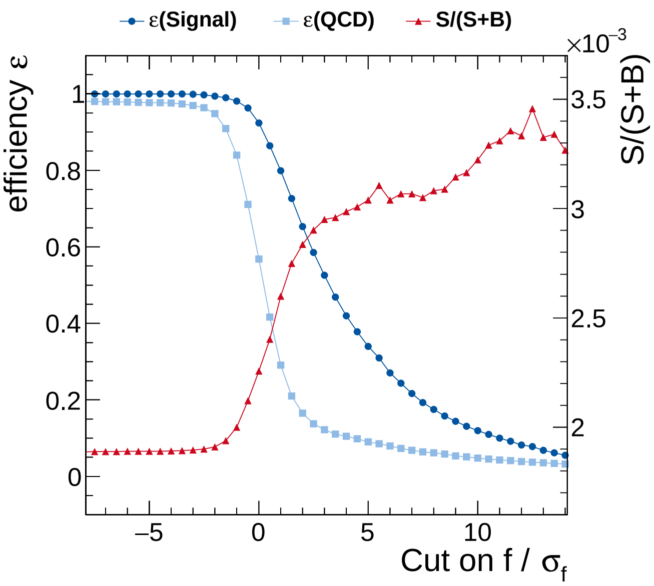
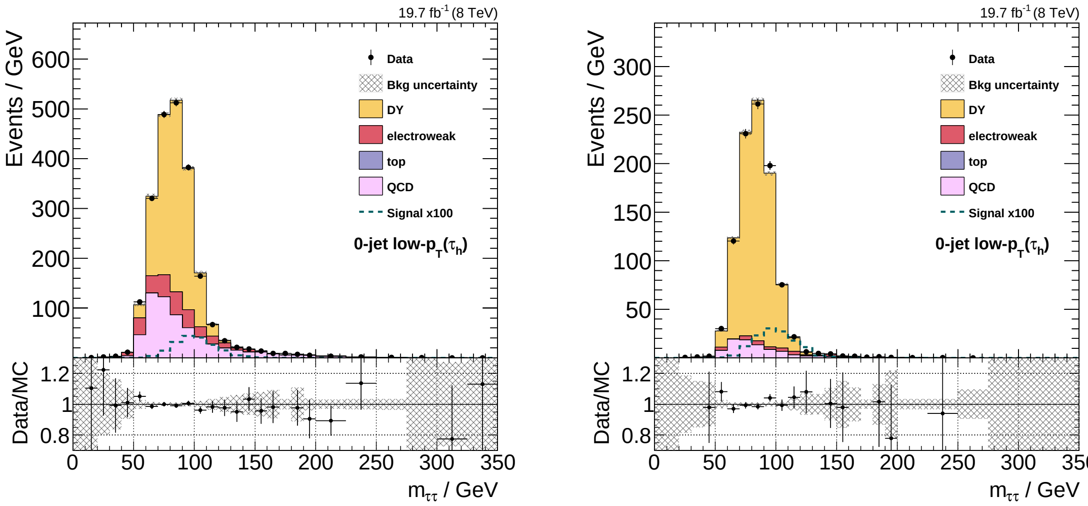


Figure 4.2: Impact of a cut on f/σ_f on the event selection. Shown is the selection efficiency ε for a Higgs boson signal (circles) and for background from QCD multijet events (squares). The reduction of backgrounds from misidentified jets increases the $S/(S+B)$ ratio (triangles) significantly.

variance matrices of the two vertices. The distribution of f/σ_f is shown in Figure 4.1. Only events are included here that pass the $\mu\tau_{\rm h}$ baseline selection presented in Section 3.2 and where the $\tau_{\rm h}$ candidate is identified as a 3-prong decay. Events where no SV could be built from the three signal tracks are included in the underflow bin. Figure 4.1(a) shows that the observed data is well modeled. The processes with genuine τ leptons, namely $\mathrm{H} \to \tau\tau$ and $\mathrm{Z} \to \tau\tau$, exhibit the expected exponential decay distribution. A comparison of the shapes in Figure 4.1(b) substantiates the slightly longer decay length of τ leptons from Higgs boson decays compared to Z decays. Although the difference is too small to apply a meaningful cut, this information could be included in an MVA to separate Higgs from Z bosons.

Backgrounds from misidentified quark and gluon jets show a much sharper cut-off. Most events originating from QCD and W + jets processes are found within a symmetric Gaussian around 0. Accordingly, this background category can be efficiently suppressed by a cut on f/σ_f. The impact of such a cut on the event selection is studied in Figure 4.2. Shown are the efficiencies ε to pass this additional selection for a SM Higgs boson signal sample (circles) and the QCD background component (squares) as a function of the cut value. Background from QCD multijet events is suppressed significantly: Even if a loose cut value of 0.0 is chosen, it is reduced by more than 40%, while the signal is diminished by less than 8%. Also shown is the ratio of signal over the sum of signal and background, $S/(S+B)$, i.e. the share of signal in all selected events. This variable has been used in Ref. [102] as a measure for the sensitivity of the selection. The background B contains all considered processes (see Section 3.2.3) including the irreducible DY background. The small absolute values are mainly due to this particular component, which dominates the selection.

For $f/\sigma_f \gtrsim 2.5$ a long-lived component becomes visible in the QCD background. It causes



(a) No cut on f/σ_f.

(b) $f/\sigma_f > 2.0$.

Figure 4.3: Postfit distribution of $m_{\tau\tau}$ in the 0-jet low-$p_\mathrm{T}(\tau_\mathrm{h})$ category for events with 3-prong taus (a) without and (b) with additional cut on f/σ_f.

a change of slope in the QCD distributions shown in Figure 4.1 as well as in the $S/(S+B)$ curve in Figure 4.2. Cutting tighter than this value leads only to small improvements in $S/(S+B)$, while the signal reduction is unchanged. The signal efficiency is crucial especially for the H $\rightarrow \tau\tau$ categories with small event yields, which offer the highest sensitivities. Accordingly, a cut is applied as

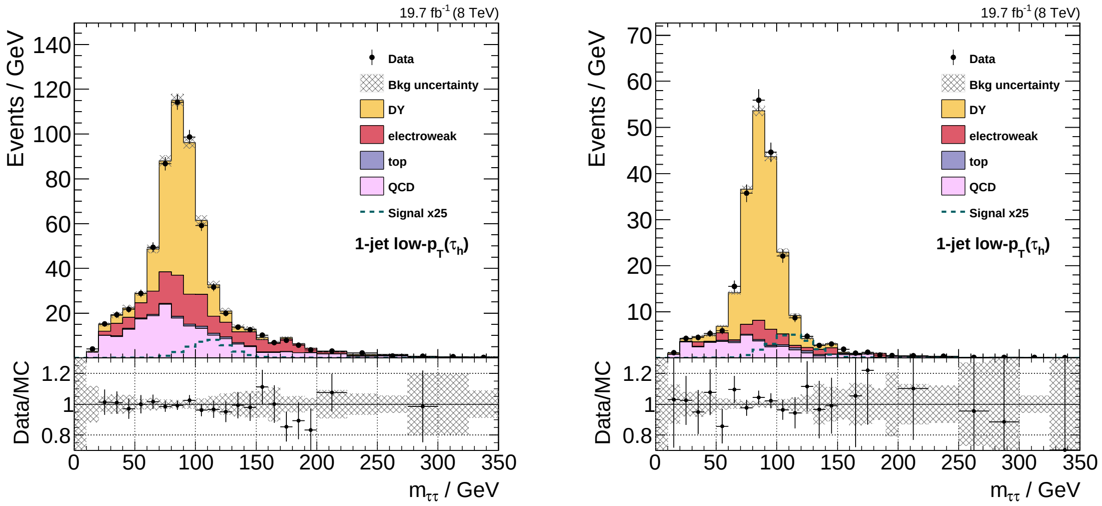
$$f/\sigma_f > 2.0\,. \tag{4.2}$$

This diminishes QCD by 84%, while signal is reduced by 35%. It results in an increase of the $S/(S+B)$ ratio by 52%.

To study the impact of the flight length significance cut on the $\mu\tau_\mathrm{h}$ channel of the H $\rightarrow \tau\tau$ search, the analysis described in Chapter 3 is repeated with additional selection criteria. One-prong decays do not offer access to the SV and are therefore neglected for this investigation. In a first implementation, the baseline selection is adapted to only accept τ_h candidates which are identified as 3-prong τ decays by the HPS algorithm. The obtained results describe the contribution of 3-prong τ decays to the $\mu\tau_\mathrm{h}$ channel and are used as reference. A second implementation adds the flight length requirement given in Equation (4.2) to the 3-prong τ baseline selection.

For both implementations the used background methods, systematic uncertainties, and the signal extraction technique remain unchanged. This ensures that comparisons to the results described in Chapter 3 are valid. The used methods rely on statistically meaningful inputs: The event yields in each category and the corresponding sidebands have to be sufficiently large to obtain a stable result. For this reason, only the 0-jet low-$p_\mathrm{T}(\tau_\mathrm{h})$, 1-jet low-$p_\mathrm{T}(\tau_\mathrm{h})$, and 1-jet high-$p_\mathrm{T}(\tau_\mathrm{h})$ categories are considered here. They contain among the highest signal yields, but contribute only marginally to the total sensitivity (compare Figure 3.12). A large number of signal events is hidden in these categories. Uncovering them could significantly improve the total sensitivity.

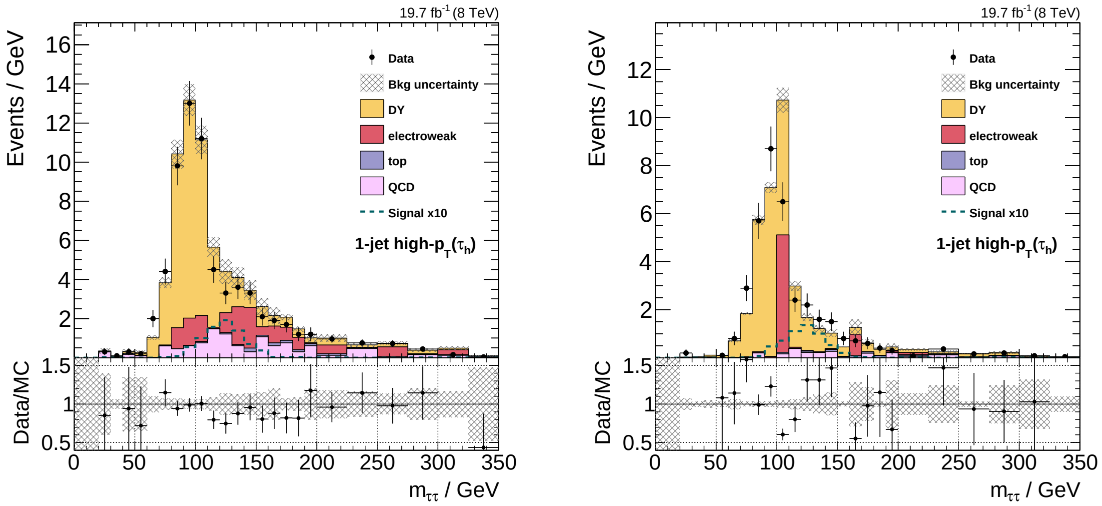
Figures 4.3, 4.4, and 4.5 show the postfit $m_{\tau\tau}$ distributions obtained in the 0-jet low-$p_\mathrm{T}(\tau_\mathrm{h})$, 1-jet low-$p_\mathrm{T}(\tau_\mathrm{h})$, and 1-jet high-$p_\mathrm{T}(\tau_\mathrm{h})$ categories, respectively. In each figure, panel (a) gives



(a) No cut on f/σ_f.

(b) $f/\sigma_f > 2.0$.

Figure 4.4: Postfit distribution of $m_{\tau\tau}$ in the 1-jet low-$p_T(\tau_h)$ category for events with 3-prong taus (a) without and (b) with additional cut on f/σ_f.



(a) No cut on f/σ_f.

(b) $f/\sigma_f > 2.0$.

Figure 4.5: Postfit distribution of $m_{\tau\tau}$ in the 1-jet high-$p_T(\tau_h)$ category for events with 3-prong taus (a) without and (b) with additional cut on f/σ_f.

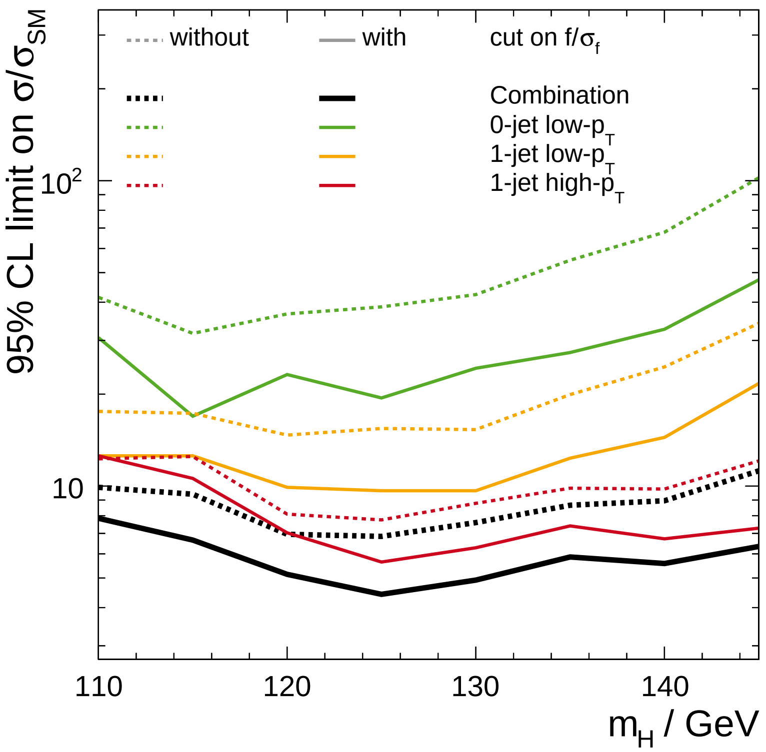


Figure 4.6: Exclusion limits obtained from events with 3-prong τ decays without (dashed lines) and with (solid lines) a cut on $f/\sigma_f > 2.0$. Only the three categories with the highest event yields are shown and used in the combination.

the result without flight length cut, while panel (b) has the cut applied. In each considered category, background from QCD and $\mathrm{W}+$jets is significantly reduced by the cut on f/σ_f. To give an example, the event yield of the QCD ($\mathrm{W}+$jets) background in the 0-jet low-$p_\mathrm{T}(\tau_\mathrm{h})$ category is diminished by 85% (89%), while 65% of the signal remains (compare Figure 4.3). The irreducible DY background is highly dominating after the cut (Figure 4.3(b)). Accordingly, the nuisance parameters related to this background component can be well constrained, as other systematic uncertainties scarcely contribute. As the DY uncertainties have a big impact on the high sensitivity categories, the impact of the background reduction is twofold: direct via the smaller QCD and $\mathrm{W}+$jets background yield, and indirect via better constraints on other nuisance parameters.

The postfit $m_{\tau\tau}$ distribution in the 1-jet high-$p_\mathrm{T}(\tau_\mathrm{h})$ category including the cut on f/σ_f (Figure 4.5(b)) indicates the effects of small event yields. The dominant component of the electroweak background, $\mathrm{W}+$jets events, has entries in only four bins. As described in Section 3.2.3, the $\mathrm{W}+$jets background yield is estimated via extrapolation from a m_T sideband, while its shape is taken from MC. The used MC sample does not contain enough events to model the shape accurately after the additional rejection introduced by the flight length cut. As the yield is fixed and can only be altered within its uncertainty by the global fit, the individual entries are scaled to match the estimated yield. This results in the observed spikes. Accordingly, results obtained from this category can give a merely qualitative indication on the impact of the flight length cut. The categories with even smaller yields that have been neglected in this section would be affected even more.

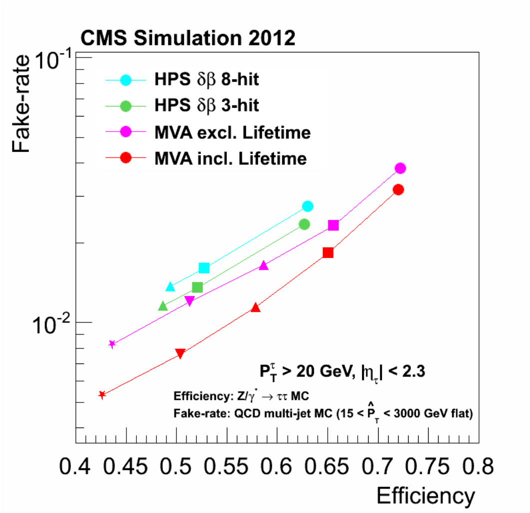
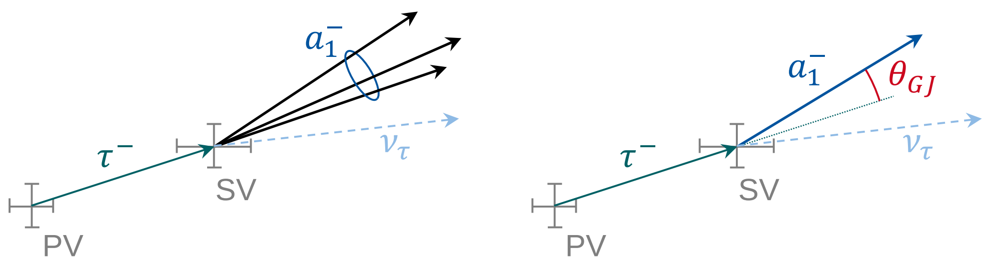


Figure 4.7: Comparison of two cut-based (blue and green) and two MVA-based (magenta and red) τ-ID discriminators [112]. The different markers denote the working points defined by the CMS collaboration. The MVA that includes the lifetime information (red) shows a significant reduction in fake-rate for similar signal efficiencies.

The expected 95% CL exclusion limits on SM Higgs boson production obtained from events with 3-prong τ decays are presented in Figure 4.6. For each category, limits are presented for a selection without (dashed lines) and with (solid lines) the cut on f/σ_f applied. All considered categories show a significant improvement. For $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, the combined expected limit of the three categories improves from 6.8 to 4.4.

As was shown in this section, splitting 3-prong τ decays into separate categories improves the limit in categories with sufficient event yields. Simultaneously, it leads to critically small yields in the high-sensitivity categories. It is therefore reasonable to exploit the τ lifetime information at an earlier stage of the analysis chain. The version of the HPS algorithm that was used for the Run-I $\mathrm{H} \to \tau\tau$ search [102] did not include τ lifetime information. For Run-II, the lifetime was incorporated into an MVA based τ_{h} identification variable [90]. Several variables that contain lifetime information are used in this MVA. For both 1-prong and 3-prong decays, the transverse impact parameter d_0 of the p_{T}-leading track within the τ_{h} candidate is used, as well as its significance d_0/σ_{d_0}. If a SV is reconstructed, the flight length f and its significance f/σ_f is used. The impact of these variables on the τ_{h} identification is shown in Figure 4.7. The signal efficiency obtained from a $\mathrm{Z} \to \tau\tau$ sample is given on the x-axis. The y-axis shows the fake rate obtained from QCD multi-jet MC events. Here, fake rate denotes the likelihood of a parton jet to be identified as τ_{h} candidate. For each considered variable multiple working points are provided, which are depicted using different symbols. The cut-based, $\delta\beta$-corrected isolation variables are shown in blue and green. The 3-hit $\delta\beta$-corrected isolation (green) is used in the Run-I analysis described in Ref. [102] and Chapter 3 of this work. Incorporation of the isolation variables into an MVA (magenta) allows for a small improvement. The addition of lifetime information (red) reduces the background from QCD events by about 40% for a similar signal efficiency. This MVA-based τ identification is foreseen to be used in $\mathrm{H} \to \tau\tau$ analyses of Run-II data.



(a) Sketch of the particles in the τ decay. (b) Definition of θ_{GJ}.

Figure 4.8: Sketch of the objects appearing in a 3-prong τ decay. (a) While the a_1 momentum is measured via the three tracks, the τ flight direction is obtained from the PV and SV positions. Panel (b) shows the definition of the Gottfried-Jackson angle θ_{GJ} as angle between the flight directions of the τ lepton and the a_1 meson in the laboratory frame.

4.2. Kinematic reconstruction of 3-prong tau decays

By default, CMS treats the hadronic tau decay products τ_{h} as a physical object. The tau lepton, τ, is not accessible for the object reconstruction, as the neutrino escapes the detector unseen. For the $\mathrm{H} \to \tau\tau$ search, the neutrino is modeled on a later stage for the determination of $m_{\tau\tau}$. This includes information from other event properties. For 3-prong tau decays, a reconstruction of the tau momentum is possible without access to other event properties. The method was first described in Ref. [113]. The additional information that allows to do so is the direction of flight of the tau lepton. It is reconstructed as the difference of the SV and PV positions,

$$\vec{e}_\tau = \frac{\vec{SV} - \vec{PV}}{|\vec{SV} - \vec{PV}|} \,. \tag{4.3}$$

Here, $\vec{SV}$ is the vector pointing to the tau decay vertex, while $\vec{PV}$ corresponds to the vector pointing to the location of the tau production.

Figure 4.8 shows sketches of the decay. At this stage, several components are measured. The three tracks originate from the decay of the intermediate a_1 meson; the meson's momentum is therefore known with high accuracy. The refitted vertices allow access to $\vec{e}_\tau$, i.e. the direction of flight of the τ itself. The *Gottfried-Jackson angle* θ_{GJ} is defined as the angle between the τ and the a_1 directions of flight as measured in the laboratory frame (Figure 4.8(b)). Accordingly, its value is also known. Finally, the masses of the τ lepton and the neutrino are taken from literature.

Using these inputs, the momentum of the τ lepton can be computed. Starting point is four-momentum conservation,

$$m_\nu^2 = 0 = (P_\tau - P_{\mathrm{a}_1})^2 \,, \tag{4.4}$$

where P_τ and P_{a_1} denote the four-vectors of the corresponding particles. The τ momentum is then given by

$$\vec{p}_\tau = |\vec{p}_\tau| \cdot \vec{e}_\tau \tag{4.5}$$

with

$$|\vec{p}_\tau| = \frac{(m_{\mathrm{a}_1}^2 + m_\tau^2)|\vec{p}_{\mathrm{a}_1}| \cos\theta_{\mathrm{GJ}} \pm \sqrt{(m_{\mathrm{a}_1}^2 + \vec{p}_{\mathrm{a}_1}{}^2)\left[(m_{\mathrm{a}_1}^2 - m_\tau^2)^2 - 4m_\tau^2 \vec{p}_{\mathrm{a}_1}{}^2 \sin^2\theta_{\mathrm{GJ}}\right]}}{2(m_{\mathrm{a}_1}^2 + \vec{p}_{\mathrm{a}_1}^2 \sin^2\theta_{\mathrm{GJ}})} \,. \tag{4.6}$$

The τ momentum is reconstructed with a two-fold ambiguity. It corresponds to two different configurations in the τ rest frame that are indistinguishable after the boost into the laboratory frame. Figure 4.9 shows sketches of two τ decays that lead to the same value θ_{GJ}. In Figure 4.9(a) $|\vec{p}_\tau|$ is large and the a_1 is emitted against the τ direction of flight. If $|\vec{p}_\tau|$ is smaller but the Gottfried-Jackson angle in the τ rest frame, θ^*_{GJ}, is also small, the same angle θ_{GJ} is measured. The curves presented in Figure 4.9(c) show the two possible τ momenta $|\vec{p}_\tau|$ as a function of the measured θ_{GJ}, as described by Equation (4.6). The upper branch is denoted $+$ solution, the lower branch $-$ solution, corresponding to the sign of the square root in the formula. To decide which of the solutions is correct, external information from other parts of the event has to be used.

The case in which the square root equals zero is called 0 solution. It corresponds to a τ decay with $\theta^*_{\mathrm{GJ}} = \pi/2$. The value of $|\vec{p}_\tau|$ for this configuration is unambiguous. The measured Gottfried-Jackson angle is maximal, $\theta^{\max}_{\mathrm{GJ}}$. From Equation (4.6) this maximum allowed value can be calculated as

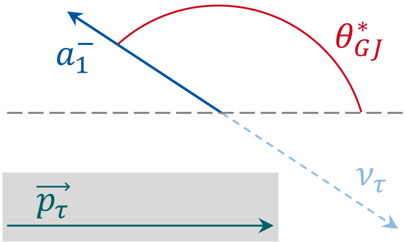
$$\theta^{\max}_{\mathrm{GJ}} = \arcsin \frac{m_\tau^2 - m_{\mathrm{a}_1}^2}{2m_\tau|\vec{p}_{\mathrm{a}_1}|} \,. \tag{4.7}$$

Tau leptons occurring in LHC collisions are highly boosted. Accordingly, the Gottfried-Jackson angle is small. Typical values are in the order of 10 mrad. While the tracking information used to determine the a_1 momentum allows for this accuracy, the determination of θ_{GJ} is sensible to the vertex reconstruction. A mismeasurement of the vertex position can lead to a measured angle $\theta^{\mathrm{obs}}_{\mathrm{GJ}} > \theta^{\max}_{\mathrm{GJ}}$. To recover such unphysical signatures, the τ direction of flight is rotated towards the a_1 candidate until $\theta^{\mathrm{obs}}_{\mathrm{GJ}} = \theta^{\max}_{\mathrm{GJ}}$. A sketch of this procedure is shown in Figure 4.10. The reconstructed τ momentum is then given by the 0 solution. The rotation introduces a bias on the reconstructed kinematic quantities, which is small compared to the resolution.

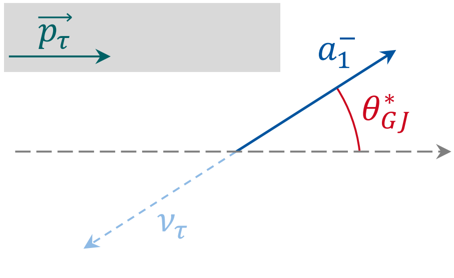
Reconstructed kinematic quantities of the $\tau_{3\pi}$ candidates with ambiguity are shown in Figure 4.11. While a cut $f/\sigma_f > 2.0$ has been applied, these distributions have not been subject to the global fit described in Section 3.4.1. Accordingly, the agreement between data and background predictions is worse than in the postfit $m_{\tau\tau}$ distributions presented in Section 4.1. The energy of $\tau_{3\pi}$ candidates reconstructed using the $-$ solution (Figure 4.11(a)) is significantly smaller than for the $+$ solution (Figure 4.11(b)): mean values of 63 GeV and 100 GeV are observed, respectively, in $\mathrm{Z} \to \tau\tau$ events. The difference in the two ambiguous solutions lies only in the absolute value of the momentum, while the direction is fixed by the vertices. The distributions of the pseudorapidity η and the azimuthal angle φ for $\tau_{3\pi}$ with ambiguity are presented in Figure 4.11(c) and Figure 4.11(d), respectively.

Similar distributions for $\tau_{3\pi}$ reconstructed with $\theta^{\mathrm{obs}}_{\mathrm{GJ}} \geq \theta^{\max}_{\mathrm{GJ}}$ are presented in Figure 4.12. By construction, the reconstructed energy lies between the $-$ and $+$ solutions. For $\mathrm{Z} \to \tau\tau$ events, the mean value of the energy distribution is measured as 73 GeV. The symmetry in φ is not affected by the rotation applied to $\tau_{3\pi}$ in the 0 solution (see Figure 4.12(c)). The pseudorapidity distribution for rotated $\tau_{3\pi}$ candidates shown in Figure 4.12(b) is more centralized compared to the ambiguous solutions (Figure 4.11(c)).

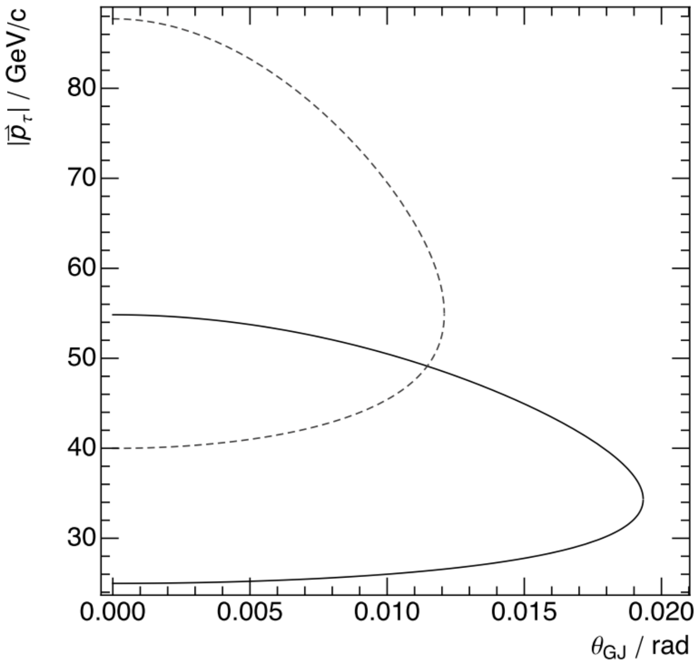
For the reconstruction of the invariant mass $m_{\tau\tau}$ of a ditau system, the $\tau_{3\pi}$ energy resolution is crucial. This will be discussed in more detail in Section 4.3. Figure 4.13(a) shows the difference of reconstructed and true energy $E(\tau_{3\pi}) - E(\tau_{\mathrm{gen}})$ for $\tau_{3\pi}$ in $\mathrm{Z} \to \tau\tau$ events. Only events that pass the baseline selection and the flight length cut are considered. To understand the impact of the underlying τ momentum distribution, Figure 4.13(b) presents the relative



(a) The positive sign in Eq. (4.6) (+ solution) corresponds to a larger τ momentum.



(b) The negative sign in Eq. (4.6) (− solution) corresponds to a smaller τ momentum.



(c) Distribution of the τ momentum as a function of the Gottfried-Jackson angle in the laboratory frame. The solid (dashed) curve corresponds to a_1 parameters $m_{a_1} = 1.2\,\text{GeV}$ and $p_{a_1} = 25\,\text{GeV}$ ($p_{a_1} = 40\,\text{GeV}$) [52, 111].

Figure 4.9: The ambiguity in the calculation of the τ momentum corresponds to two different configurations in the τ rest frame that cannot be distinguished in the laboratory frame: (a) a large angle θ^*_{GJ} combined with a large τ momentum p_τ and (b) a small angle θ^*_{GJ} combined with a small value of p_τ. The sketches are given in the τ rest frame except for the τ momentum. Panel (c) shows the reconstructed value $|\vec{p}_\tau|$ as a function of the measured angle θ_{GJ} for two sets of a_1 parameters. For each curve, the upper part corresponds to the configuration shown in (a), the lower part corresponds to (b).

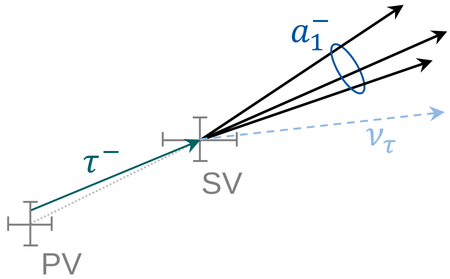


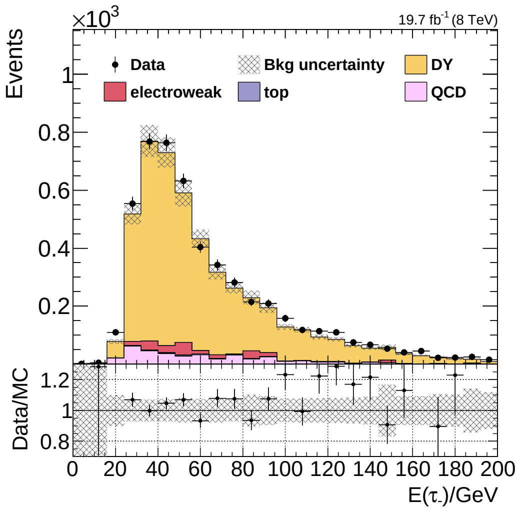
Figure 4.10: If the observed value of the Gottfried-Jackson angle is bigger than kinematically allowed, $\theta_{\mathrm{GJ}}^{\mathrm{obs}} > \theta_{\mathrm{GJ}}^{\mathrm{max}}$, the direction of flight is rotated until $\theta_{\mathrm{GJ}}^{\mathrm{obs}} = \theta_{\mathrm{GJ}}^{\mathrm{max}}$.

resolution $[E(\tau_{3\pi}) - E(\tau_{\mathrm{gen}})]/E(\tau_{\mathrm{gen}})$. Objects reconstructed using the 0 solution (orange) overestimate the true τ energy by 9.0%, while the distribution is symmetric around this mean value. The width (RMS value) is 19.4% or 15.3 GeV. If the $\tau_{3\pi}$ is reconstructed with $\theta_{\mathrm{GJ}}^{\mathrm{obs}} < \theta_{\mathrm{GJ}}^{\mathrm{max}}$, one solution has to be chosen. The dark blue (light blue) histogram shows the energy resolution if always the $-$ solution ($+$ solution) is selected. A clear distinction between the two solutions is visible. Usage of information from the MC generator to select the correct solution (green dashed line) results in a peak around the nominal value with an RMS of 15.7 GeV. The τ phase space is restricted by the τ_{h} object selection: τ leptons that decay with a $+$ solution are more likely to pass the event selection, and thus the distribution is asymmetric with larger tails towards higher energies. This effect is weak compared to the impact of a wrongly assigned sign for the $+$ and $-$ distributions, which gives rise to large tails towards high (low) energies for assigned $+$ solutions ($-$ solutions). The corresponding RMS values are 17.5% for the $-$ and 30.0% for the $+$ solution, respectively. The difference in shape between the two solutions originates from the asymmetry of the ambiguous term in Equation (4.6), which is visualized in Figure 4.9(c). A comparison of the $+$ and $-$ curves with the curve for the correct choice shows that solving the ambiguity is an important ingredient to use the $\tau_{3\pi}$ reconstruction in a physics analysis. For this, other information of the event has to be utilized. One approach is discussed in Section 4.3.1.

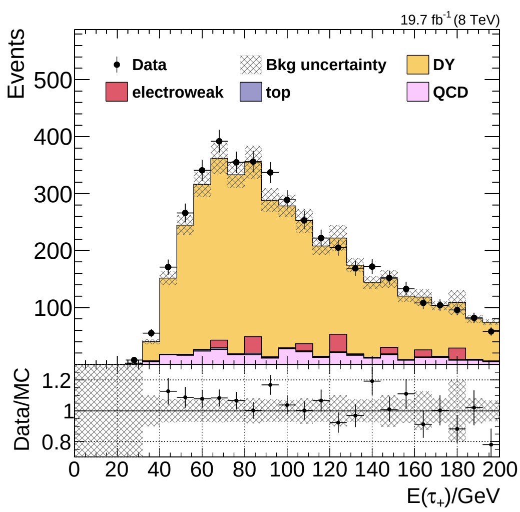
The kinematic reconstruction of $\tau_{3\pi}$ decays pursues two objectives: obtain access to the full τ momentum and suppress background from non-genuine τ_{h} candidates. The feasibility of both is shown in Ref. [113], where information from the 3-prong τ decay is used to measure the τ polarization in Z decays. To scrutinize the understanding of the Higgs boson, the method could be applied to study properties of the Higgs boson like its *CP* state [114]. For this, a large number of $\mathrm{H} \to \tau\tau$ events is needed, which will be available only after several more years of LHC operation. On a shorter timescale, the $\tau_{3\pi}$ reconstruction can be used to suppress QCD background. As a first step, the lifetime information obtained from the SV and the refitted PV described in Section 4.1 was already introduced into the default HPS τ_{h} identification. An additional criterion for this purpose, the rotation significance, is discussed below. Finally, knowledge of the τ momentum can improve the $m_{\tau\tau}$ reconstruction. This is discussed in Section 4.3.

4.2.1. Rotation significance

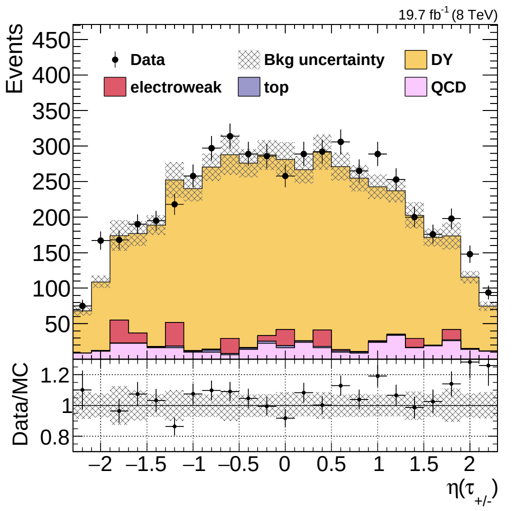
For genuine $\tau_{3\pi}$ decays with $\theta_{\mathrm{GJ}}^{\mathrm{obs}} > \theta_{\mathrm{GJ}}^{\mathrm{max}}$, the disagreement of the measurement with the physical bound originates from mismeasurements of the vertices and the a_1 momentum. Accordingly, the necessary rotation must be compatible with the uncertainties connected to these measurements. If a $\tau_{3\pi}$ candidate does not originate from a τ lepton decay, this is not



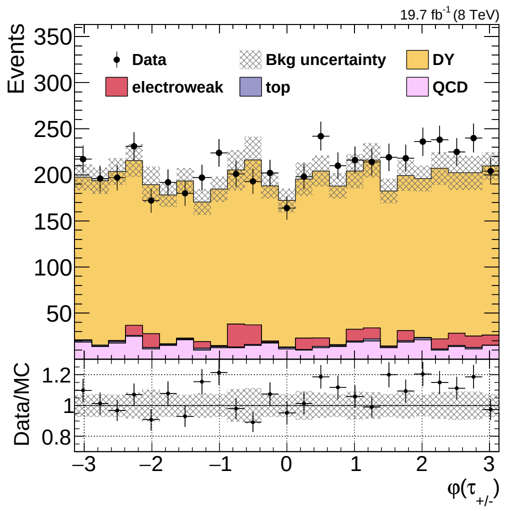
(a) Energy distribution of τ leptons reconstructed using the $-$ solution.



(b) Energy distribution of τ leptons reconstructed using the $+$ solution.

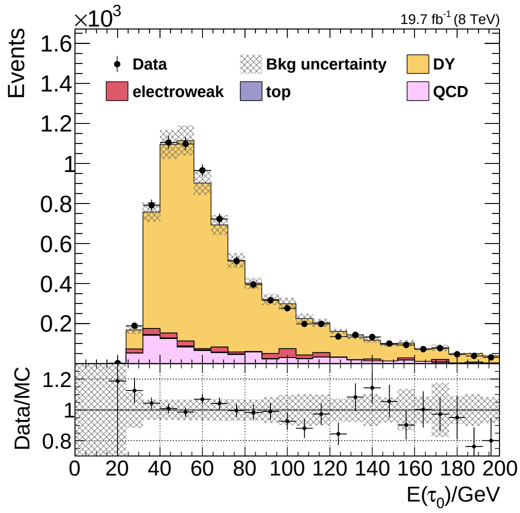


(c) Pseudorapidity distribution of τ leptons reconstructed in the $-$ or $+$ solution.

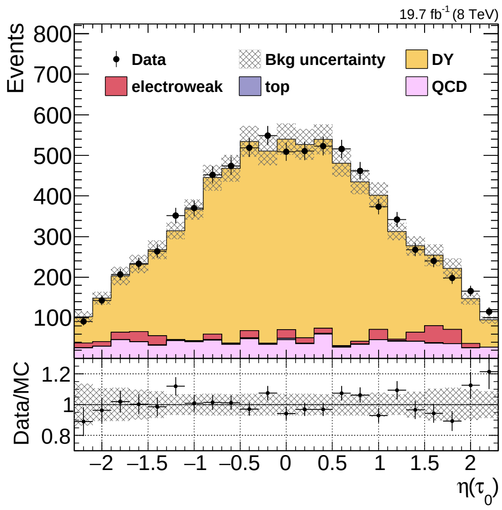


(d) Azimuthal distribution of τ leptons reconstructed in the $-$ or $+$ solution.

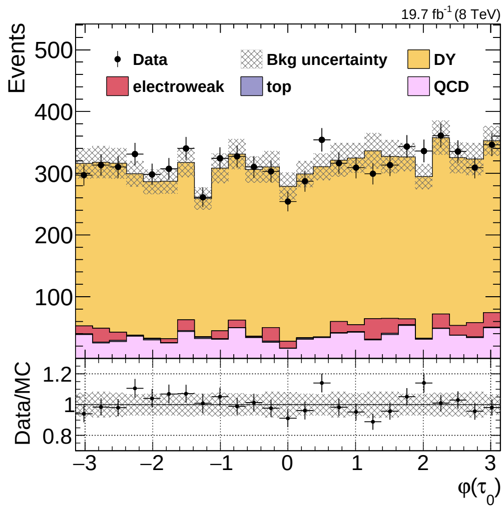
Figure 4.11: Properties of τ leptons as obtained by the kinematic reconstruction of 3-prong taus. The upper panels show the reconstructed energies when choosing the $-$ or $+$ solution. The angle of the reconstructed τ does not depend on the ambiguity; the angular distributions presented in the lower panels are thus valid for both solutions.



(a) Energy distribution of τ leptons reconstructed using the 0 solution.

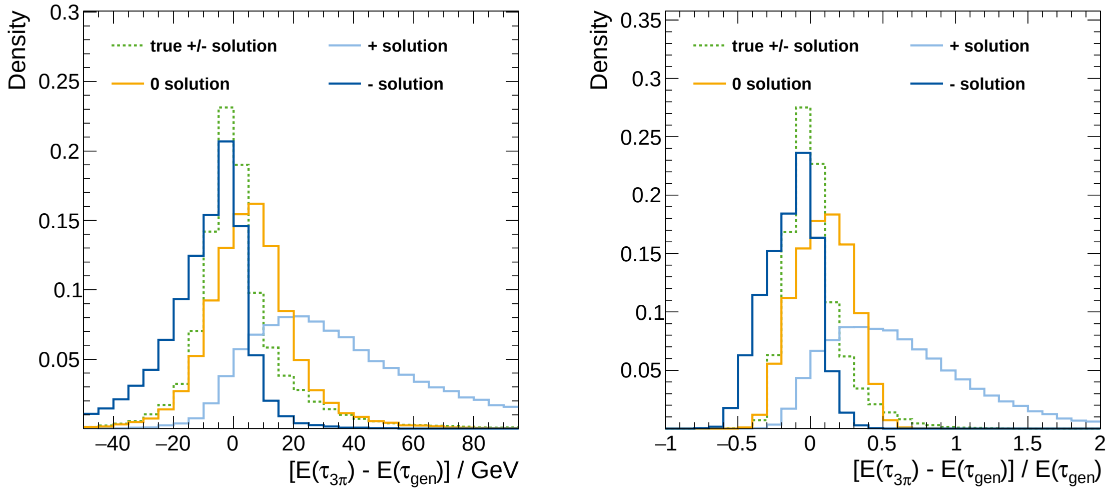


(b) Pseudorapidity distribution of τ leptons reconstructed in the 0 solution.



(c) Azimuthal distribution of τ leptons reconstructed in the 0 solution.

Figure 4.12: Properties of τ leptons as obtained by the kinematic reconstruction of 3-prong taus for cases with $\theta_{GJ}^{obs} \geq \theta_{GJ}^{max}$.



(a) Absolute resolution. (b) Relative resolution.

Figure 4.13: Energy resolution of $\tau_{3\pi}$ candidates reconstructed with 0 ambiguity (orange) or with $+$ and $-$ solutions (light and dark blue). If the candidate is reconstructed with an ambiguity, the correct solution can be determined using MC information (green dashed line).

necessarily true. The rotation variable

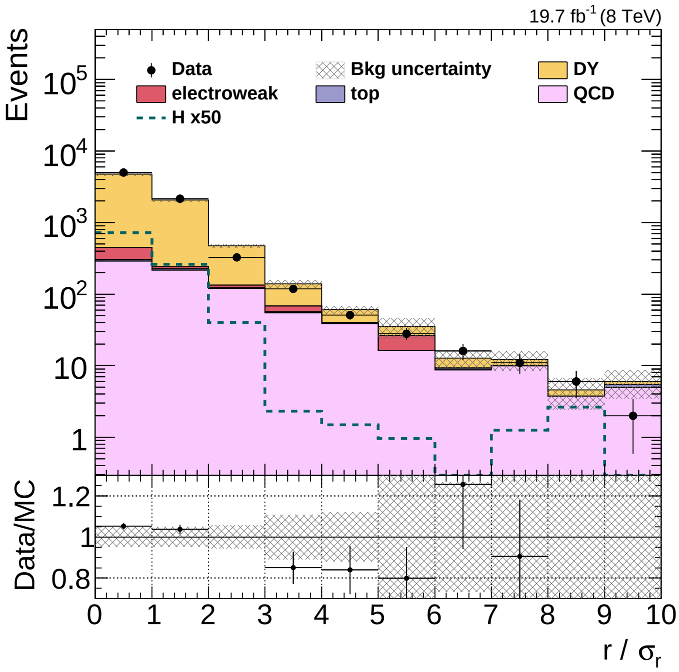
$$r = \theta_{\mathrm{GJ}}^{\mathrm{obs}} - \theta_{\mathrm{GJ}}^{\mathrm{max}} \tag{4.8}$$

is introduced to quantify the disagreement between the measured value and the physical bound. While $\theta_{\mathrm{GJ}}^{\mathrm{obs}} = \theta_{\mathrm{GJ}}^{\mathrm{obs}}(\vec{PV}, \vec{SV})$ depends on the reconstructed vertices, the maximum Gottfried-Jackson angle $\theta_{\mathrm{GJ}}^{\mathrm{max}} = \theta_{\mathrm{GJ}}^{\mathrm{max}}(\vec{p}_{\mathrm{a}_1}, m_{\mathrm{a}_1})$ is calculated from the measured a_1 parameters according to Equation (4.7). To incorporate the uncertainties on these measurements, the rotation significance r/σ_r is analyzed.

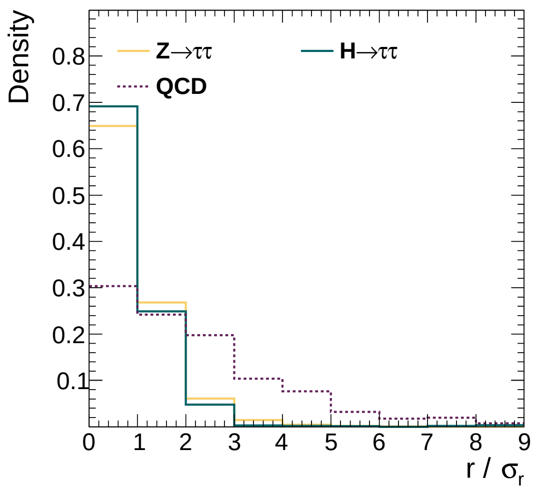
Figure 4.14(a) shows the observed distribution[1] for events passing the $\mu\tau_{\mathrm{h}}$ baseline selection with a 3-prong τ_{h} candidate that meets $f/\sigma_f > 2.0$. A rotation of the τ direction is only applicable for candidates reconstructed in the 0 solution. Accordingly, only events with $\theta_{\mathrm{GJ}}^{\mathrm{obs}} \geq \theta_{\mathrm{GJ}}^{\mathrm{max}}$ are included in the figures. Out of all $\mathrm{Z} \to \tau\tau$ events passing the selection, approximately 55% contain $\tau_{3\pi}$ candidates reconstructed with the 0 solution, while 45% are found within the physical bounds on θ_{GJ}. In contrast, 68% of QCD events are reconstructed with the 0 solution. Accordingly, the sample of events with 0 solution shown in Figure 4.14(a) is enriched with QCD events.

It is feasible to reduce this background by a cut on r/σ_r. Figures 4.14(b) and (c) show comparisons of signal and background shapes for events selected without and with a cut $f/\sigma_f > 2.0$, respectively. Between 65 and 70% of processes with genuine τ decays are found within $r/\sigma_r < 1$. This is expected for Gaussian uncertainties and considering that most of $\tau_{3\pi}$ decays appear with Gottfried-Jackson angles close to the allowed maximum in the laboratory frame [52, 111]. The distribution of genuine τ decays remain unchanged when a cut on the flight length is applied. For QCD background, less than 40% of events are found within one

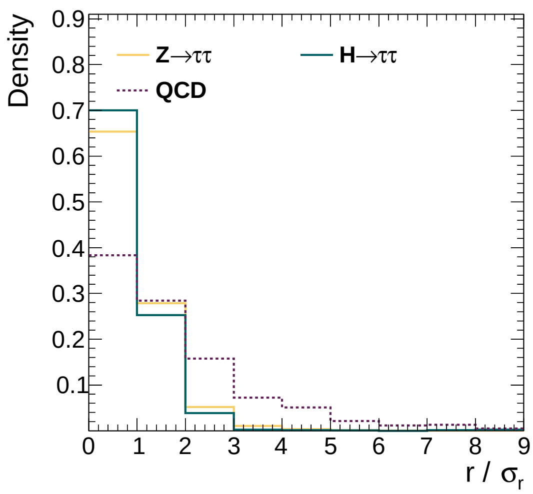
[1]At large values of r/σ_r the signal sample distribution is affected by statistical fluctuations that lead to empty bins.



(a) Event distribution.



(b) Shape comparison before cut on f/σ_f.



(c) Shape comparison after cut on f/σ_f.

Figure 4.14: Rotation significance r/σ_r for 3-prong τ decays reconstructed with 0 solution. For panel (a), a cut $f/\sigma_f > 2.0$ has been applied. The lower panels give a comparison of the shapes for different processes (b) without and (c) with cut on f/σ_f.

standard deviation of the central value. A small correlation between r/σ_r and f/σ_f is observed for QCD background: The amount of events with $r/\sigma_r < 1.0$ rises from 30% to 38% if the flight length cut is applied. Inclusion of the rotation significance into the MVA-based HPS τ identification is advisable.

4.3. Reconstruction of $m(\tau_\mu \tau_{3\pi})$

Reconstruction of the invariant $\tau\tau$ mass with good resolution is the key ingredient to separate Higgs boson from Z boson events. At CMS, this task is performed by SVFIT with good performance. It was described in Section 3.3. The kinematic reconstruction of $\tau_{3\pi}$ decays offers access to additional information that is not used by SVFIT. This is, firstly, the τ direction of flight obtained from the refitted primary and secondary vertices. Secondly, the τ decay model in SVFIT does not include the matrix elements for hadronic decays. As given in Equation (3.4), hadronic tau decays are modeled to decay uniformly within the allowed phase space. If this additional information is implemented, the $m_{\tau\tau}$ resolution will improve for all $\tau\tau$ decay channels that include hadronic τ decays.

4.3.1. Implementation of the reconstructed $\tau_{3\pi}$ into the mass calculation

The SVFIT software accepts four different types of τ decay products as input. The first three are τ_h, e, and μ as described in Section 3.3. These objects are used in the $\mathrm{H} \to \tau\tau$ analysis. A fourth type is intended to model prompt leptons, i.e. objects that do not originate from a τ decay. It can be used in analyses with only one τ in the final state, like lepton flavor violating Higgs decays.

The fourth type is also an entry point for the kinematically reconstructed $\tau_{3\pi}$. Since the reconstructed $\tau_{3\pi}$ includes the neutrino, it does not have to be modeled by SVFIT. For $\mu\tau_{3\pi}$ events, the SVFIT likelihood is given by

$$\mathcal{L} = \mathcal{L}_{\tau_\mu}(\vec{p}_\mu, \vec{a}_{\tau_\mu}) \cdot \mathcal{L}_\nu(\vec{p}_{\nu\nu}, \vec{a}_{\tau_\mu}) \,. \tag{4.9}$$

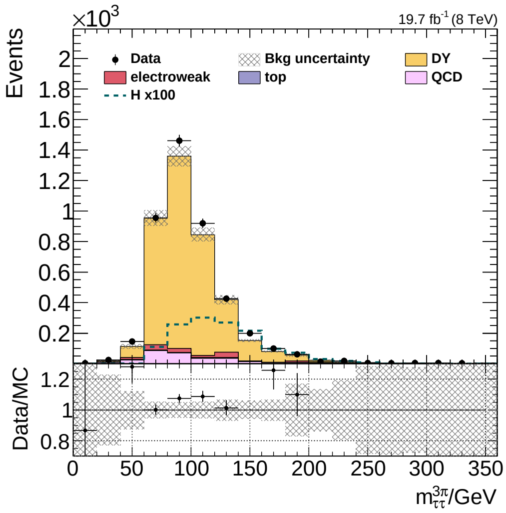
Compared to the general SVFIT likelihood given in Equation (3.2) it is simplified. The term corresponding to the hadronic τ decay, $\mathcal{L}_{\tau_{3\pi}}$, is neglected as all its parameters $\vec{a}_{\tau_{3\pi}}$ are fixed by the kinematic $\tau_{3\pi}$ reconstruction. Accordingly, the only free parameters $\vec{a}_{\tau_\mu}$ that remain belong to the τ_μ decay. The measured $\vec{E}_\mathrm{T}^\mathrm{miss}$ is replaced with the momentum of the neutrino system in the τ_μ decay,

$$\vec{p}_{\nu\nu} = \vec{E}_\mathrm{T}^\mathrm{miss} - \vec{p}_\nu(\tau_{3\pi}) \,. \tag{4.10}$$

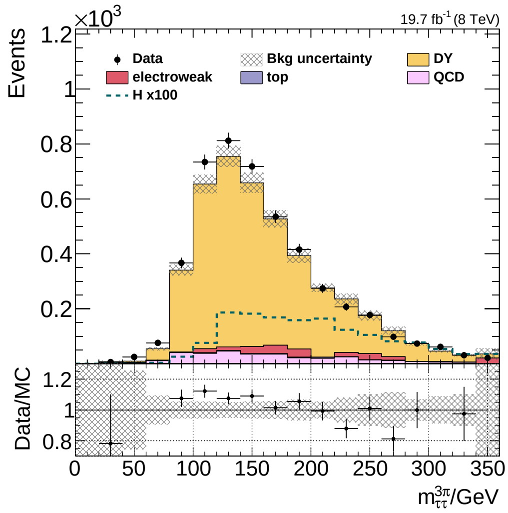
Here, $\vec{p}_\nu(\tau_{3\pi})$ denotes the momentum of the neutrino occuring in the hadronic τ decay. It is obtained from the kinematic $\tau_{3\pi}$ reconstruction as $\vec{p}_\nu(\tau_{3\pi}) = \vec{p}_\tau - \vec{p}_{\tau_\mathrm{h}}$. In total, the likelihood in Equation (4.9) is described by three free parameters from the τ_μ decay and two constraints from the measured $\vec{p}_{\nu\nu}$.

The reconstructed ditau mass using this likelihood is denoted $m_{\tau\tau}^{3\pi}$. Its distribution is depicted in Figure 4.15. A global fit to constrain the nuisance parameters was not applied. As the correct solution for the ambiguity is not known at this stage, the panels (a) and (b) show distributions that contain the same events, reconstructed using the $-$ or $+$ solution, respectively. The corresponding distribution for events with $\theta_\mathrm{GJ}^\mathrm{obs} \geq \theta_\mathrm{GJ}^\mathrm{max}$ is shown in panel (c).

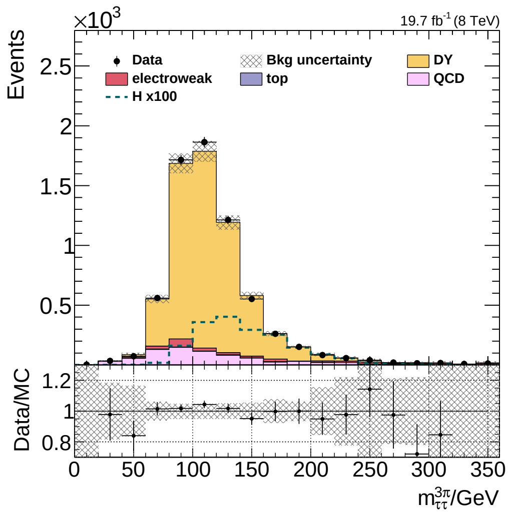
The critical property of the ditau mass variable is its resolution. A high mass resolution is the only handle that allows to disentangle Higgs boson from Z boson events. Compared



(a) − solution.

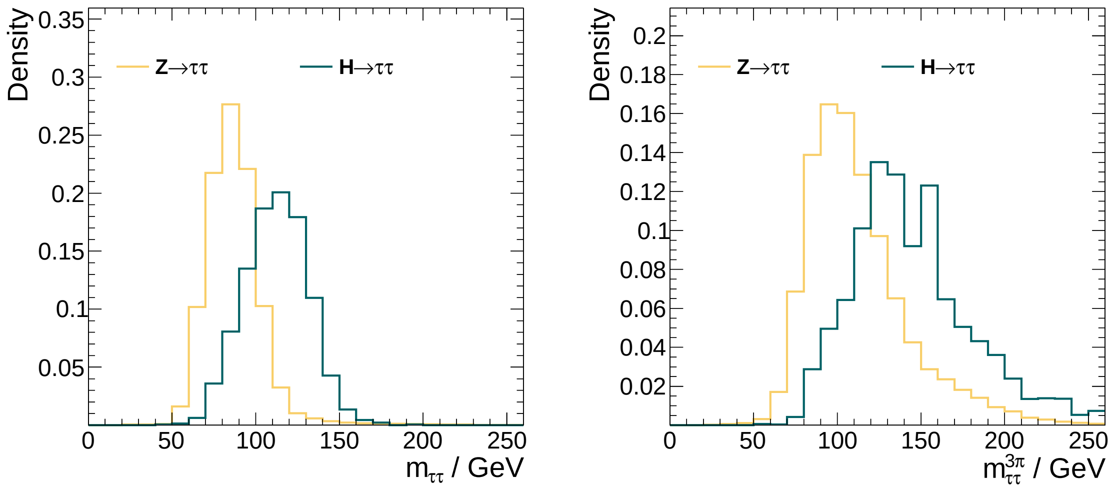


(b) + solution.



(c) 0 solution.

Figure 4.15: Distributions of the reconstructed ditau mass $m_{\tau\tau}^{3\pi}$ for events reconstructed (a) with the − solution, (b) with the + solution, and (c) with the 0 solution. A global fit to constrain the nuisance parameters is not applied.



(a) Default SVFIT mass $m_{\tau\tau}$.

(b) Modified SVFIT mass $m_{\tau\tau}^{3\pi}$.

Figure 4.16: Shape comparison of the invariant ditau mass (a) obtained using the default SVFIT algorithm and (b) using the modified, 3-prong based SVFIT approach. Shown are $Z \to \tau\tau$ and $H \to \tau\tau$ events that pass the $\mu\tau_h$ baseline selection with a 3-prong τ that fulfills $f/\sigma_f > 2.0$. Panel (b) shows only events where $\theta_{GJ}^{obs} < \theta_{GJ}^{max}$; the variable $m_{\tau\tau}^{3\pi}$ was computed using the correct ambiguity as determined via MC information.

to the $m_{\tau\tau}$ distributions presented in Figures 4.3-4.5, the modified SVFIT variable $m_{\tau\tau}^{3\pi}$ has a significantly worse resolution. The difference can neither be explained by the categorization that was applied in Section 4.1.1, nor by the global fit, which mainly affects the normalization. A comparison of the shapes for the default SVFIT mass $m_{\tau\tau}$ and the modified approach $m_{\tau\tau}^{3\pi}$ is presented in Figure 4.16. The shapes presented in Figure 4.16(b) use the correct solution of the ambiguity as obtained from generator information. Nevertheless, the mass resolution of 24% (27%) is significantly larger than the corresponding value of 17% (20%) for the default SVFIT mass, measured in $H \to \tau\tau$ ($Z \to \tau\tau$) events.

The mass reconstruction gets worse although more information is added. This effect originates from the energy resolutions of the $\tau_{3\pi}$ objects presented in Figure 4.13. The SVFIT likelihood given in Equation (3.2) does not consider inaccuracies in the momentum measurement of the input objects. In the default SVFIT approach, the uncertainty on $\vec{E}_T^{miss}$ clearly dominates the momentum uncertainties of the e, μ, or τ_h candidates, which are measured with high precision. Accordingly, a likelihood term for these objects can safely be neglected. The τ momentum, as obtained by the $\tau_{3\pi}$ reconstruction, does not only depend on well measured tracks, but also on vertices. The momentum resolution is in the same order of magnitude than the uncertainty on $\vec{E}_T^{miss}$, and neglecting it is no longer feasible. As a consequence, the observed resolution in the $\tau_{3\pi}$ energy directly translates to $m_{\tau\tau}^{3\pi}$.

The distribution shown in Figure 4.16(b) uses generator information to select the correct solution for the $\tau_{3\pi}$ momentum. When using data, this information is not available, and the ambiguity must be chosen from event properties. For this purpose, the modified SVFIT likelihood given in Equation (4.9) can be exploited, when the algorithm is executed twice. The maximum likelihood value $\mathcal{L}_{max}$ when using the $\tau_{3\pi}$ with the $-$ solution as input is compared to the corresponding value for the $+$ solution. The correct solution should be better compatible with the SVFIT physics model and thus give a higher likelihood. Table 4.1

Table 4.1: Migration matrix between the true and reconstructed ambiguity solution for genuine 3-prong τ decays via the a_1 resonance. Decays with $\theta_{GJ}^{obs} \geq \theta_{GJ}^{max}$ are sorted into the 0 solution. All other decays are reconstructed as $-$ or $+$ solution, depending on the larger $\mathcal{L}_{max}$.

True solution	$-$	$+$
Reconstructed solution		
0	64%	65%
$-$	28%	27%
$+$	8%	8%

shows which percentage of events is reconstructed in each of the three possible solutions. Only events where the $\tau_{3\pi}$ originates from a genuine 3-prong τ decay via the a_1 resonance are considered. No significant preference for the correct solution is observed.

4.3.2. Prospects to improve the $m(\tau_\mu \tau_{3\pi})$ reconstruction

A combination of the $\tau_{3\pi}$ reconstruction and the SVFIT algorithm must be based on a proper treatment of all sizable uncertainties. The ad-hoc method described in Section 4.3.1 neglects the uncertainty on the $\tau_{3\pi}$ momentum, which results in a poor resolution for the invariant ditau mass.

A straightforward solution to this problem appears to exist, as the measured $\tau_{3\pi}$ momentum can be treated similarly to the $\vec{E}_T^{miss}$ measurement. The likelihood given in Equation (4.9) is then extended to

$$\mathcal{L} = \mathcal{L}_{\tau_\mu}(\vec{p}_\mu, \vec{a}_{\tau_\mu}) \cdot \mathcal{L}_{\tau_{3\pi}}(\vec{p}_{\tau_{3\pi}}, \vec{a}_{\tau_{3\pi}}) \cdot \mathcal{L}_\nu(\vec{p}_{\nu\nu}, \vec{a}_{\tau_\mu}, \vec{a}_{\tau_{3\pi}}) \,. \tag{4.11}$$

Here, $\vec{p}_\mu$ denotes the measured momentum of the muon, i.e. the τ decay product, and $\mathcal{L}_{\tau_\mu}$ describes the kinematic properties of a leptonic τ decay. For the $\tau_{3\pi}$ the kinematic properties are already reconstructed. Accordingly, $\vec{p}_{\tau_{3\pi}}$ describes the momentum of the τ, and $\mathcal{L}_{\tau_{3\pi}}$ describes the compatibility of the reconstructed momentum $\vec{p}_{\tau_{3\pi}}$ with the calculated momentum $\vec{y}_{\tau_{3\pi}} = \vec{y}_{\tau_{3\pi}}(\vec{a}_{\tau_{3\pi}})$. As the covariance matrix $\mathbf{M}$ of the reconstructed $\tau_{3\pi}$ is known, the likelihood term is given as

$$\mathcal{L}_{\tau_{3\pi}}(p_{\tau_{3\pi}}) = \frac{1}{2\pi\sqrt{|\mathbf{M}|}} \exp\left[-\frac{1}{2}(\vec{p}_{\tau_{3\pi}} - \vec{y}_{\tau_{3\pi}})^\intercal \, \mathbf{M}^{-1} \, (\vec{p}_{\tau_{3\pi}} - \vec{y}_{\tau_{3\pi}})\right] \,. \tag{4.12}$$

There are complications connected to the likelihood given in Equation (4.11). First, the technical implementation of SVFIT is based on the premise that leptons are perfectly measured. To adopt it for $\tau_{3\pi}$, substantial changes have to be applied to the framework. While this is in principle feasible, the likelihood in Equation (4.11) can only be applied for 3-prong τ decays. The majority of τ decays have 1-prong final states, which would still be reconstructed using the default likelihood given in Equation (3.2). Usage of two fundamentally different models for 1-prong and 3-prong decays is not desirable, as the corresponding uncertainties would be different and therefore harder to constrain. Additionally, it was shown in Section 4.1 that a separation of 1-prong and 3-prong final states leads to small event yields in the high sensitivity categories of $H \to \tau\tau$ analyses. It is thus crucial to handle 1-prong and 3-prong decays identically as far as possible.

An alternative to the approach presented in Equation (4.11) is to extent the original SVFIT likelihood given in Equation (3.2) by a term $\mathcal{L}_{\mathrm{SV}}(\vec{SV} - \vec{PV}, \vec{a}_{\tau_{\mathrm{h}}})$. It describes the compatibility of the measured SV with the parameters $\vec{a}_{\tau_{\mathrm{h}}}$ of the hadronic tau decay. In this case the likelihood for 1-prong and 3-prong decays would be identical except for $\mathcal{L}_{\mathrm{SV}}$. The additional information used in the $\tau_{3\pi}$ reconstruction, the matrix element for 3-prong tau decays, could be implemented in the term $\mathcal{L}_{\tau_{\mathrm{h}}}$ (see Equation (3.4)).

An additional complication is caused by the construction of the likelihood from individual components. It affects all considered likelihoods: the default form given in Equation (3.2), the modified version in Equation (4.11), and the default likelihood extended by $\mathcal{L}_{\mathrm{SV}}$. The full likelihood is not based on first principles. The corresponding probability does not correspond to a physically meaningful variable. Accordingly, the individual likelihood terms are not properly normalized with respect to each other. This can introduce arbitrary, intrinsic dependencies that give preference to lower or higher masses. For SVFIT, such a term is observed which is proportional to $\log(m_{\tau\tau})$. It can be corrected for with a corresponding penalty term. For analyses which study a small mass range, like searches for a SM Higgs boson, the logarithmic dependency can be ignored, as it is weak compared to each of the three likelihood terms. This statement does not hold for a likelihood term that implements information from a secondary vertex, which itself is only a weak constraint.

A similar observation was made for the global event fit (GEF) described in Ref. [113, Chapter 8]. It uses a likelihood method to reconstruct the full kinematics of $\mathrm{Z} \rightarrow \tau\tau$ events in the $\mu\tau_{3\pi}$ final state and is based on the same $\tau_{3\pi}$ reconstruction exploited in this work. Nonetheless, it cannot be applied directly to $\mathrm{H} \rightarrow \tau\tau$ analyses, as it utilizes the Z mass as a constraint. To disentangle H from Z events, the GEF can be executed twice, using the H mass or the Z mass. The underlying process is then determined by the solution with the larger likelihood. Alternatively, if the mass of the resonance is unknown, a scan over a mass range can be executed, and the mass value with the highest likelihood is selected. The GEF in its current implementation always prefers the solution with the highest mass. It is reasonable to assume that this effect is similar to the observed $\log(m_{\tau\tau})$ term in SVFIT.

For $\mathrm{H} \rightarrow \tau\tau$ analyses, usage of a different mass reconstruction for 3-prong taus should be avoided. The fit that constraints nuisance parameters relies on correlations between channels and categories. Different mass reconstruction techniques introduce additional uncertainties, which cannot be constraint in low-sensitivity categories with high event yields. Accordingly, it is advisable to use SVFIT for all τ decay modes. The secondary vertex offers additional information that can be used independently for the identification of τ_{h} and in the SVFIT likelihood. For the latter, an improved formalism has to be derived that is based on first principles, namely the probability of a physics process to be produced in the LHC and, subsequently, observed by CMS.

Chapter 5

Summary and conclusion

This thesis is devoted to the analysis of Higgs bosons that decay into two τ leptons. The decay into fermions allows to probe the Yukawa couplings of the Higgs boson. A focus has been put on hadronic τ decay modes. Trigger paths have been developed and maintained that collected data during the year 2012 and were used in the analysis that reported evidence for $H \to \tau\tau$ decays [102]. The 2012 data set, consisting of $19.7\,\text{fb}^{-1}$ collected at a center-of-mass energy of 8 TeV at CMS, has been studied in the $\mu\tau_h$ channel of the $H \to \tau\tau$ analysis.

Compared to 2011, the instantaneous luminosity increased significantly in the year 2012. As fully-hadronic $\tau\tau$ decay channels are subject to large backgrounds from QCD dijet events, it was unclear if dedicated, inclusive trigger paths for $\tau_h\tau_h$ events could be deployed. The first part of this thesis has described τ_h identification and reconstruction in the Level-1 and high level trigger systems. Both systems have been investigated for improvements. Eventually, an inclusive $\tau_h\tau_h$ trigger path was included in the data parking campaign. It accepted events with two reconstructed τ_h candidates with $p_T > 35\,\text{GeV}$ and $|\eta| < 2.1$. Its development, proposal, and maintenance have been executed within the scope of this work. While the data collected by this path was not accessible during 2012, it has been used for the CMS analysis that reported evidence for $H \to \tau\tau$ decays [102] in 2014.

A set of performance parameters of various trigger paths that include τ_h has been determined. During data collection, the trigger rate must stay within bounds. The deployed L1 algorithm used as seed for $\tau_h\tau_h$ paths collected data with a rate of 6.6 kHz at an instantaneous luminosity of $\mathcal{L} = 5 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$. At HLT, the $\tau\tau$ parking trigger reached a maximum rate of 52.7 Hz in early 2012. Improvements in the HLT τ_h identification have allowed to reduce the rate. For the largest part of the 2012 data-taking period the trigger rate of the parking trigger, normalized to $\mathcal{L} = 5 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$, has been constant at about 32 Hz.

Angular and momentum resolutions have been measured for trigger objects at different trigger levels. The L1 trigger overestimates $p_T(\tau_h)$ by 23 GeV on average, so that the thresholds of 44 GeV on L1 tau jets correspond to physical objects with $p_T \gtrsim 25\,\text{GeV}$. The momentum resolution of trigger objects built exclusively from calorimeter information, i.e. L1 and L2 objects, is in the order of 10 GeV. Level-3 τ_h objects are reconstructed using the PF approach, which leads to significantly better resolutions. For the data collected in 2012, the central part of their resolution curve can be modeled as a sum of two Gaussians with widths of 0.2 GeV and 2.0 GeV.

The efficiency of τ_h objects in trigger paths has been measured independently for the individual trigger levels. At L1, an inefficiency caused by the absolute tau isolation criterion has been mitigated by accepting L1 central jets with a higher threshold. For this reason the L1

$\tau_h\tau_h$ algorithm reaches a plateau of 100% efficiency for $p_T \gtrsim 60\,\text{GeV}$. The energy resolution of trigger objects translates into the slope of the efficiency curve. The final trigger efficiency combines the efficiency curves of the individual trigger levels. It has been used to correct for trigger effects in the first search for $H \to \tau_h\tau_h$ performed by CMS [100]. The efficiency reaches a plateau above 80% for $p_T(\tau_h) > 50\,\text{GeV}$. A systematic uncertainty of 4.5% has been assigned to the trigger efficiency measurement for each τ_h leg.

The $\mu\tau_h$ channel offers the best sensitivity for $H \to \tau\tau$ analyses at CMS. The event selection and reconstruction for this channel has been implemented independently for this work. A result compatible with the CMS publication [102] is obtained. Subsequently, it has been used to study improvements for 3-prong τ decays, which offer access to additional information that can be useful in the context of Higgs boson analyses. A reconstruction of the secondary vertex allows to exploit the lifetime of the τ lepton. Application of a cut on the signed flight length significance of $f/\sigma_f > 2.0$ reduces background from QCD multijet events by 84%, while 65% of the signal remains. Consequently, the combined expected 95% CL limit of three investigated categories improves from 6.8 to 4.4 for a Higgs boson mass hypothesis of 125 GeV. For $H \to \tau\tau$ analyses using the LHC Run-II data set, the lifetime information was included in an MVA τ_h identification variable.

A second quality criterion for 3-prong tau decays has been discussed: the rotation significance r/σ_r. It is not yet used in the official τ_h ID and is only defined for τ_h candidates where the τ direction of flight, taken as the direction from the primary to the secondary vertex, is not physically compatible with the reconstructed a_1 meson. This occurs for 55% of genuine $\tau_{3\pi}$ decays. Out of these decays, 70% are reconstructed with a disagreement of less than one standard deviation. In contrast, 68% of QCD multijet events are reconstructed with incompatible vertices and a_1 direction. The disagreement is greater than 1σ for more than 60% of these events. Accordingly, the rotation significance is proposed as useful additional component of the MVA τ_h ID.

It has been shown that 3-prong τ decays can be kinematically reconstructed from the tracks and the primary vertex and secondary vertex positions. An ad-hoc implementation of such objects into the ditau mass reconstruction algorithm SVFIT leads to significantly larger $m_{\tau\tau}$ resolutions. The cause of this effect was traced to the likelihood, which is inconsistent with leptons that are reconstructed with sizable uncertainties. As information from the secondary vertex is not yet included in SVFIT, suggestions have been made how the likelihood could be adapted. Although this needs a thorough redesign of the SVFIT framework, it is worth the effort: 3-prong τ decays offer good additional handles to suppress QCD background and to reconstruct the kinematics of the event.

Appendix A

Level-1 Trigger Hardware

This section presents sketches depicting the data procession in the Level-1 trigger system. The different hardware components and their interactions are shown. Given the focus on τ_h triggers, a description of the muon trigger system is omitted.

Figure A.1 is taken from Ref. [68] and gives an overview of the full L1 trigger. The following figures offer a more detailed representation of different trigger subsystems and have been created as part of this thesis. Data from the detector front-end electronics is combined to form trigger primitives (Figure A.2), which are then further processed in several layers. The RCT (Figure A.3) investigates the information on a small angular scale. It is followed by the GCT (Figure A.4), which uses detector symmetries as depicted in Figure A.5 to analyze several regions simultaneously. In a final step, the GT (Figure A.6) gathers all available information, calculates the trigger algorithms, and organizes the data readout in case of a positive trigger decision. The figures show the hardware as it was deployed during LHC Run-I. An upgrade of the L1 hardware was installed for Run-II [115], but is not considered here.

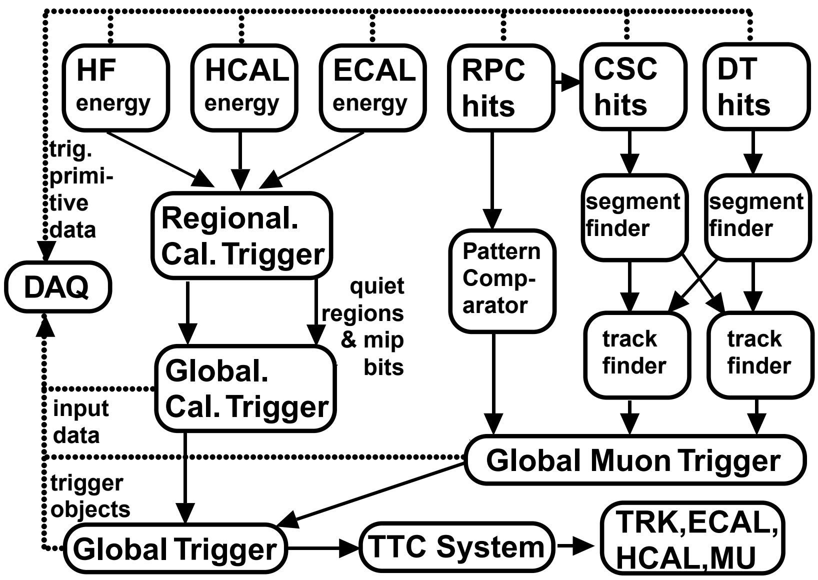


Figure A.1: Overview of the L1 trigger system [68, modified].

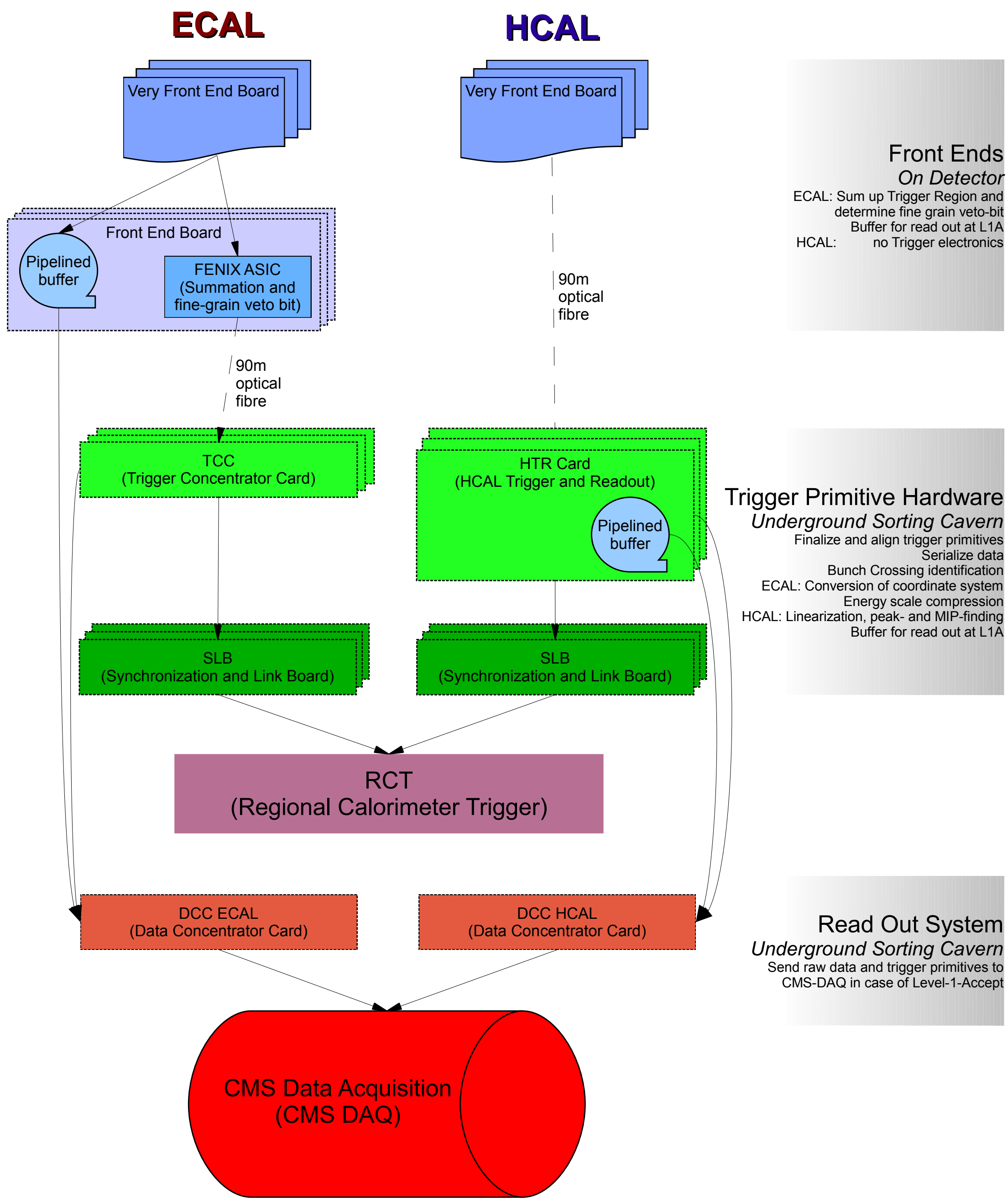


Figure A.2: Electronic components involved in the generation of L1 calorimeter trigger primitives.

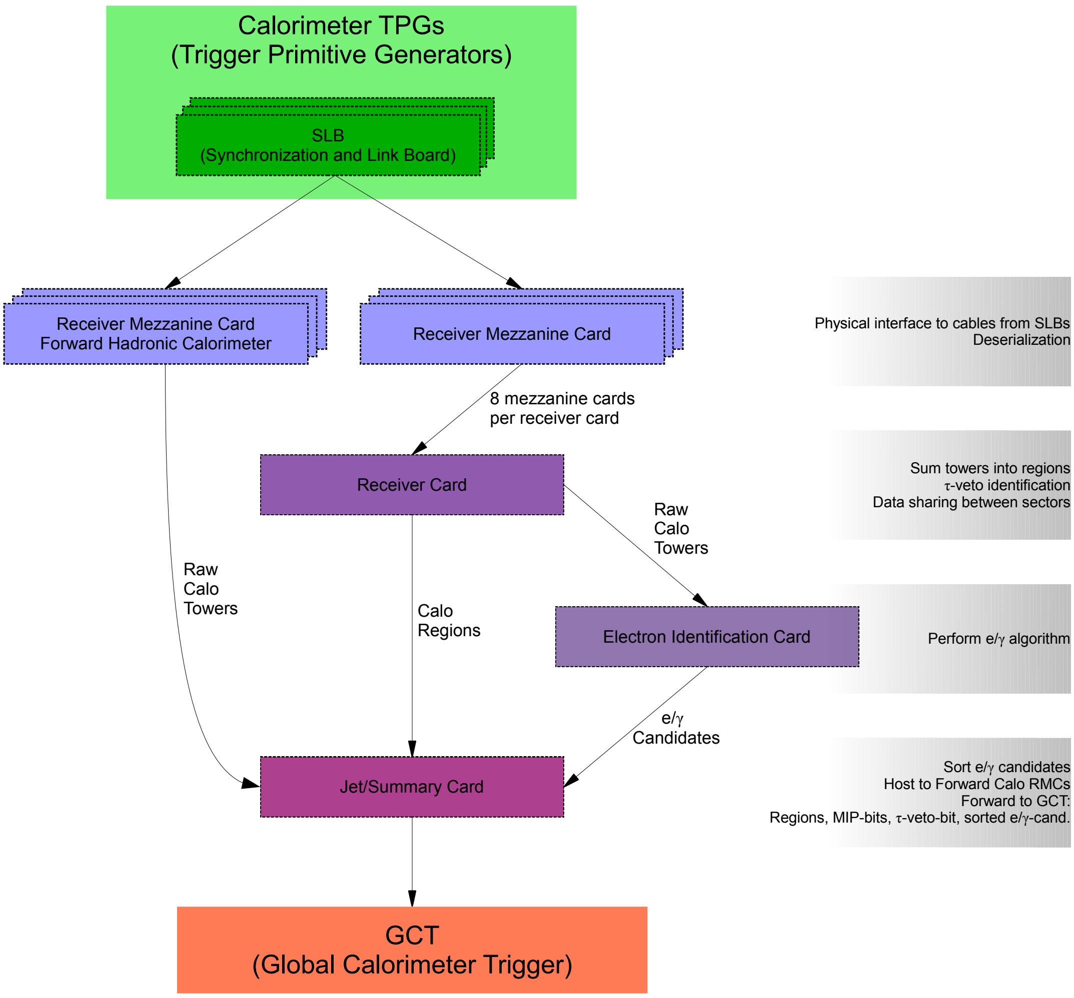


Figure A.3: Electronic components of the L1 regional calorimeter trigger (RCT).

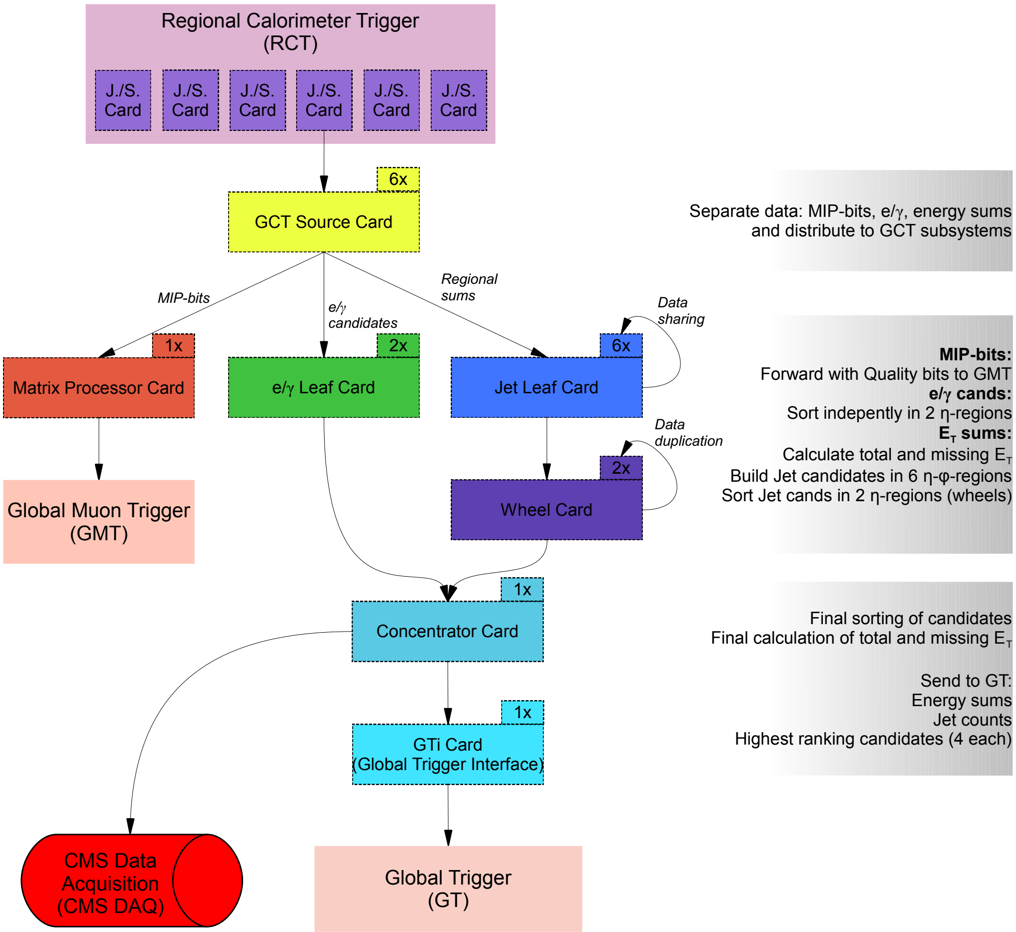


Figure A.4: Electronic components of the L1 global calorimeter trigger (GCT).

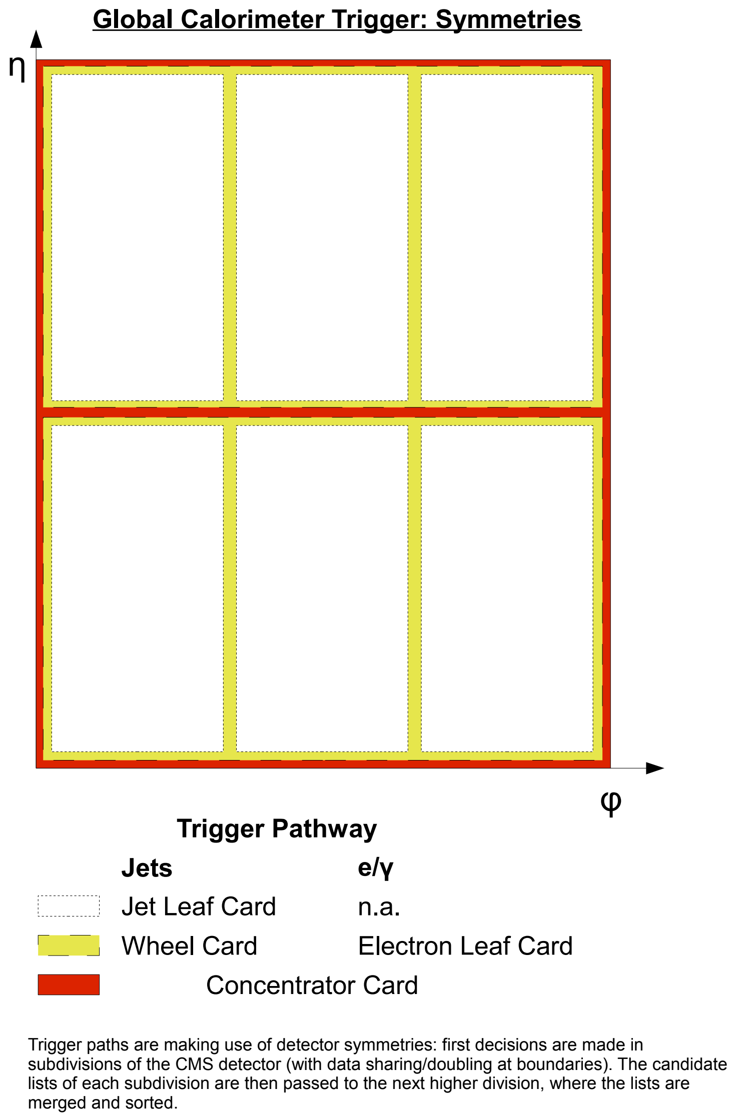


Figure A.5: Detector symmetries are exploited by the GCT to process data simultaneously. The information is then combined for further analysis.

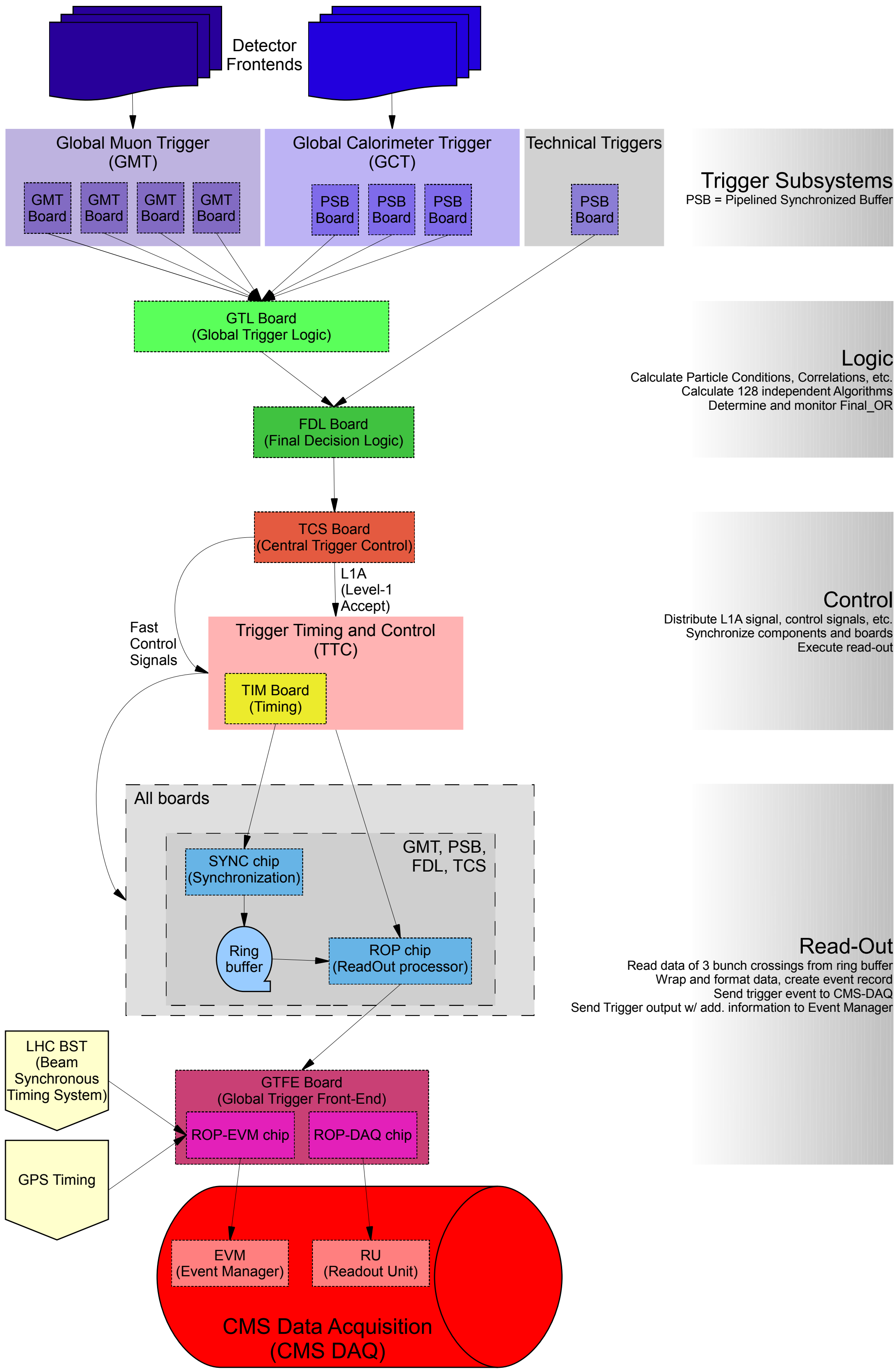


Figure A.6: Electronic components of the L1 global trigger (GT) and the process leading to the trigger decision and possible data readout.

Bibliography

[1] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[3] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[4] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$=7 and 8 TeV", *J. High Energy Phys.* **2013** (jun, 2013) 81, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.

[5] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[6] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (sep, 1964) 132–133, doi:10.1016/0031-9163(64)91136-9.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[8] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

[9] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories", *Phys. Rev.* **155** (1967) 1554, doi:10.1103/PhysRev.155.1554.

[10] J. J. Thomson, "Cathode Rays", *Philos. Mag. Ser. 5* **44** (oct, 1897) 293–316, doi:10.1080/14786449708621070.

[11] A. Compton, "A Quantum Theory of the Scattering of X-rays by Light Elements", *Phys. Rev.* **21** (may, 1923) 483–502, doi:10.1103/PhysRev.21.483.

[12] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2.

[13] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (nov, 1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[14] A. Salam, "Weak and electromagnetic interactions", in *Elementary particle physics: relativistic groups and analyticity*, N. Svartholm, ed., p. 367. Almqvist & Wiksell, Stockholm, 1968.

[15] M. Gell-Mann, "A schematic model of baryons and mesons", *Phys. Lett.* **8** (feb, 1964) 214–215, doi:10.1016/S0031-9163(64)92001-3.

[16] G. Zweig, "An SU_3 model for strong interaction symmetry and its breaking", CERN Theory Report CERN-TH-401, CERN, Geneva, jan, 1964.

[17] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the color octet gluon picture", *Phys. Lett. B* **47** (nov, 1973) 365–368, doi:10.1016/0370-2693(73)90625-4.

[18] D. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (jun, 1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[19] H. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (jun, 1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[20] Y. Fukuda et al., "Measurements of the Solar Neutrino Flux from Super-Kamiokande's First 300 Days", *Phys. Rev. Lett.* **81** (aug, 1998) 1158–1162, doi:10.1103/PhysRevLett.81.1158.

[21] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", *Astron. & Astrophys.* **571** (oct, 2014) A16, doi:10.1051/0004-6361/201321591.

[22] J. Berger, "Search for the Higgs Boson Produced via Vector-Boson Fusion in the Decay Channel $H \to \tau\tau$". PhD thesis, Karlsruhe Institute of Technology, 2014.

[23] R. D. Ball et al., "Impact of heavy quark masses on parton distributions and LHC phenomenology", *Nucl. Phys. B* **849** (aug, 2011) 296–363, doi:10.1016/j.nuclphysb.2011.03.021.

[24] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields", *Nucl. Phys. B* **44** (jul, 1972) 189–213, doi:10.1016/0550-3213(72)90279-9.

[25] Y. Nambu, "Quasi-Particles and Gauge Invariance in the Theory of Superconductivity", *Phys. Rev.* **117** (feb, 1960) 648–663, doi:10.1103/PhysRev.117.648.

[26] J. Goldstone, "Field theories with 'Superconductor' solutions", *Nuovo Cim.* **19** (jan, 1961) 154–164, doi:10.1007/BF02812722.

[27] K. A. Olive and Particle Data Group, "Review of Particle Physics", *Chinese Phys. C* **38** (aug, 2014) 090001, doi:10.1088/1674-1137/38/9/090001.

[28] J. Ellis, M. K. Gaillard, and D. Nanopoulos, "A phenomenological profile of the Higgs boson", *Nucl. Phys. B* **106** (jan, 1976) 292–340, doi:10.1016/0550-3213(76)90382-5.

[29] A. Djouadi, "The anatomy of electroweak symmetry breaking", *Phys. Rep.* **457** (feb, 2008) 1–216, doi:10.1016/j.physrep.2007.10.004, arXiv:0503172.

[30] K. Nakamura and Particle Data Group, "Review of Particle Physics", *J. Phys. G Nucl. Part. Phys.* **37** (jul, 2010) 075021, doi:10.1088/0954-3899/37/7A/075021.

[31] LHC Higgs Cross Section Working Group, "LHC Higgs Cross Section WG Picture Gallery". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGCrossSectionsFigures. Accessed 11 Dec. 2015. The website provides

additional pictures for [32–34].

[32] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 1. Inclusive observables", CERN Report CERN-2011-002, Geneva, 2011. doi:10.5170/CERN-2011-002, arXiv:1101.0593.

[33] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", CERN Report CERN-2012-002, Geneva, 2012. doi:10.5170/CERN-2012-002, arXiv:1201.3084.

[34] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 3. Higgs Properties", CERN Report CERN-2013-004, Geneva, 2013. doi:10.5170/CERN-2013-004, arXiv:1307.1347.

[35] W. J. Sterling, "Parton Luminosity and Cross Section Plots". http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html. Accessed 11 Dec. 2015.

[36] LEP Working Group for Higgs Boson Searches et al., "Search for the Standard Model Higgs boson at LEP", *Phys. Lett. B* **565** (jul, 2003) 61–75, doi:10.1016/S0370-2693(03)00614-2.

[37] B. Tuchming and CDF and D0 Collaborations, "Tevatron Higgs results", *EPJ Web Conf.* **60** (nov, 2013) 02003, doi:10.1051/epjconf/20136002003, arXiv:1307.4873.

[38] CDF Collaboration, "Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (apr, 1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626.

[39] D0 Collaboration, "Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s}$= 1.8 TeV", *Phys. Rev. Lett.* **74** (mar, 1995) 2422–2426, doi:10.1103/PhysRevLett.74.2422.

[40] M. Baak et al., "Updated status of the global electroweak fit and constraints on new physics", *Eur. Phys. J. C* **72** (may, 2012) 2003, doi:10.1140/epjc/s10052-012-2003-4.

[41] ATLAS Collaboration and CMS Collaboration, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-15-002, CERN, Geneva, 2015.

[42] ATLAS Collaboration, "Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector", *J. High Energy Phys.* **2015** (apr, 2015) 117, doi:10.1007/JHEP04(2015)117, arXiv:1501.04943 [hep-ex].

[43] CMS Collaboration, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions", *Nat. Phys.* **10** (jun, 2014) 557–560, doi:10.1038/nphys3005, arXiv:1401.6527 [hep-ex].

[44] ATLAS Collaboration and CMS Collaboration, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* **114** (2015), no. 19, 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589 [hep-ex].

[45] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. journal. C, Part. fields* **75** (jan, 2015) 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.

[46] ATLAS Collaboration, "Evidence for the spin-0 nature of the Higgs boson using ATLAS data", *Phys. Lett. B* **726** (oct, 2013) 120–144, doi:10.1016/j.physletb.2013.08.026, arXiv:1307.1432.

[47] 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** (2013) 81803, doi:10.1103/PhysRevLett.110.081803, arXiv:1212.6639.

[48] L. D. Landau, "On the angular momentum of a system of two photons", *Dokl. Akad. Nauk Ser. Fiz.* **60** (1948), no. 2, 207–209, doi:10.1016/B978-0-08-010586-4.50070-5.

[49] C. N. Yang, "Selection Rules for the Dematerialization of a Particle into Two Photons", *Phys. Rev.* **77** (jan, 1950) 242–245, doi:10.1103/PhysRev.77.242.

[50] 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.

[51] L. Evans and P. Bryant, "LHC Machine", *J. Instrum.* **3** (aug, 2008) S08001–S08001, doi:10.1088/1748-0221/3/08/S08001.

[52] L. Perchalla, "Kinematic Tau Reconstruction and Search For The Higgs Boson in Hadronic Tau Pair Decays with the CMS Experiment". PhD thesis, RWTH Aachen, 2011.

[53] C. Lefevre, "LHC: the guide". CERN Brochure, 2009.

[54] S. Myers, "The Large Hadron Collider 2008-2013", *Int. J. Mod. Phys. A* **28** (oct, 2013) 1330035, doi:10.1142/S0217751X13300354.

[55] CMS Collaboration, "CMS Luminosity - Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed 06 Jan. 2016.

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

[57] T. Sakuma, "SketchUpCMS Gallery". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMSGallery. Accessed 11 Jan. 2016.

[58] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV", *J. Instrum.* **8** (sep, 2013) P09009–P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016.

[59] P. Adzic et al., "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *J. Instrum.* **2** (apr, 2007) P04004, doi:10.1088/1748-0221/2/04/P04004.

[60] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *J. Instrum.* **8** (nov, 2013) P11002–P11002,

doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.

[61] CMS Collaboration, "Measurement of CMS Luminosity", CMS Physics Analysis Summary CMS-PAS-EWK-10-004, CERN, Geneva, 2010.

[62] D. Froidevaux and P. Sphicas, "General-Purpose Detectors for the Large Hadron Collider", *Annu. Rev. Nucl. Part. Sci.* **56** (nov, 2006) 375–440, doi:10.1146/annurev.nucl.54.070103.181209.

[63] CMS HCAL/ECAL Collaborations, "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", *Eur. Phys. J. C* **60** (mar, 2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5.

[64] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *J. Instrum.* **7** (oct, 2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[65] A. Bell et al., "Fast beam conditions monitor BCM1F for the CMS experiment", *Nucl. Instrum. Methods Phys. Res., Sect. A* **614** (mar, 2010) 433–438, doi:10.1016/j.nima.2009.12.056, arXiv:0911.2480.

[66] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[67] J. Brooke, "Performance of the CMS Level-1 Trigger", *PoS* **ICHEP2012** (2013) 508, arXiv:1302.2469.

[68] CMS Collaboration, "The CMS Trigger System", (2016). To be published.

[69] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *J. Instrum.* **9** (may, 2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[70] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Methods Phys. Res., Sect. A* **262** (dec, 1987) 444–450, doi:10.1016/0168-9002(87)90887-4.

[71] P. Billoir, "Progressive track recognition with a Kalman-like fitting procedure", *Comput. Phys. Commun.* **57** (dec, 1989) 390–394, doi:10.1016/0010-4655(89)90249-X.

[72] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system", *Nucl. Instrum. Methods Phys. Res., Sect. A* **395** (aug, 1997) 169–184, doi:10.1016/S0168-9002(97)00705-5.

[73] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the Kalman filtering method", *Nucl. Instrum. Methods Phys. Res., Sect. A* **294** (sep, 1990) 219–228, doi:10.1016/0168-9002(90)91835-Y.

[74] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at LHC", *J. Phys. G Nucl. Part. Phys.* **31** (jun, 2003) N9–N20, doi:10.1088/0954-3899/31/9/N01.

[75] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proc. IEEE* **86** (1998), no. 11,

2210–2239, doi:10.1109/5.726788.

[76] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, Geneva, 2009.

[77] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, Geneva, 2010.

[78] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CERN, Geneva, 2010.

[79] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-003, CERN, Geneva, 2010.

[80] F. Beaudette, "The CMS Particle Flow Algorithm", in *Proceedings, Int. Conf. Calorim. High Energy Front. (CHEF 2013)*, pp. 295–304. 2013. arXiv:1401.8155.

[81] D. Barney, "CMS slice image with transverse/longitudinal/3-D views". CMS Document Database CMS-doc-5697-v2, 2012.

[82] G. P. Salam, "Towards jetography", *Eur. Phys. J. C* **67** (may, 2010) 637–686, doi:10.1140/epjc/s10052-010-1314-6, arXiv:0906.1833.

[83] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *J. High Energy Phys.* **2008** (apr, 2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[84] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (mar, 2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[85] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[86] CMS Collaboration, "8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS", CMS Perfomance Note CMS DP-2013/033, CERN, Geneva, 2013.

[87] CMS Collaboration, "Pileup Jet Identification", CMS Physics Analysis Summary CMS-PAS-JME-13-005, CERN, Geneva, 2013.

[88] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (apr, 2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[89] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *J. Instrum.* **10** (feb, 2015) P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[90] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS", *J. Instrum.* **11** (jan, 2016) P01019–P01019, doi:10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.

[91] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 physics and manual”, *JHEP* **05** (2006) 26, doi:10.1088/1126-6708/2006/05/026.

[92] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 40, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[93] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 70, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[94] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 43, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[95] S. Alioli et al., “Jet pair production in POWHEG”, *JHEP* **04** (2011) 81, doi:10.1007/JHEP04(2011)081, arXiv:1012.3380.

[96] J. Alwall et al., “MadGraph 5: going beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[97] CMS Collaboration, “Measurement of the underlying event activity at the LHC with $\sqrt{s}$= 7 TeV and comparison with $\sqrt{s}$ = 0.9 TeV”, *JHEP* **09** (2011) 109, doi:10.1007/JHEP09(2011)109, arXiv:1107.0330.

[98] S. Jadach, J. H. Kühn, and Z. Was, “TAUOLA - a library of Monte Carlo programs to simulate decays of polarized tau leptons”, *Comput. Phys. Commun.* **64** (1991) 275, doi:10.1016/0010-4655(91)90038-M.

[99] S. Agostinelli and Others, “GEANT4 – a simulation toolkit”, *Nucl. Instrum. Methods Phys. Res., Sect. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[100] CMS Collaboration, “Search for the standard model Higgs boson decaying to tau pairs”, CMS Physics Analysis Summary CMS-PAS-HIG-12-043, CERN, Geneva, 2012.

[101] CMS Collaboration, “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, Technical Report CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, CERN, Geneva, jun, 2015.

[102] CMS Collaboration, “Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons”, *J. High Energy Phys.* **2014** (may, 2014) 104, doi:10.1007/JHEP05(2014)104, arXiv:1401.5041.

[103] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, “Higgs boson production at the LHC: transverse momentum resummation effects in the $H \to \gamma\gamma$, $H \to WW \to l\nu l\nu$ and $H \to ZZ \to 4l$ decay modes”, *JHEP* **06** (2012) 132, doi:10.1007/JHEP06(2012)132, arXiv:1203.6321.

[104] A. Burgmeier and G. Quast, “Position Resolution and Upgrade of the CMS Pixel Detector and Search for the Higgs Boson in the $\tau^+\tau^-$ Final State”. PhD thesis, Karlsruhe U., EKP, 2014.

[105] B. K. Bullock, K. Hagiwara, and A. D. Martin, “Tau polarization and its correlations as a probe of new physics”, *Nucl. Phys. B* **395** (1993) 499, doi:10.1016/0550-3213(93)90045-Q.

[106] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report ATL-PHYS-PUB 2011-11, CMS NOTE 2011/005, CERN, 2011.

[107] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Methods Phys. Res., Sect. A* **434** (sep, 1999) 435–443, doi:10.1016/S0168-9002(99)00498-2.

[108] A. L. Read, "Presentation of search results: the CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[109] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (feb, 2011) 1–19, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727.

[110] R. Frühwirth, P. Vanlaer, and W. Waltenberger, "Adaptive Vertex Fitting", CMS Note CMS Note 2007/008, mar, 2007.

[111] P. Sauerland, "Kinematic reconstruction of tau leptons and test for lepton universality in charged weak interactions with the CMS experiment". PhD thesis, RWTH Aachen, 2011.

[112] CMS Collaboration, "Tau ID Performance Plots", CMS Performance Note CMS-DP-2014-015, CERN, Geneva, 2014.

[113] V. Cherepanov, "Measurement of the polarization of τ-leptons produced in Z^0 decays at CMS and determination of the effective weak mixing angle $\sin^2\theta_{eff}$". Phd thesis, RWTH Aachen University, 2016.

[114] S. Berge, W. Bernreuther, and J. Ziethe, "Determining the *CP* Parity of Higgs Bosons via Their τ Decay Channels at the Large Hadron Collider", *Phys. Rev. Lett.* **100** (apr, 2008) 171605, doi:10.1103/PhysRevLett.100.171605.

[115] B. Kreis et al., "Run 2 upgrades to the CMS Level-1 calorimeter trigger", *J. Instrum.* **11** (jan, 2016) C01051–C01051, doi:10.1088/1748-0221/11/01/C01051.

Acronyms

2HDM two-Higgs-doublet model 15
ALICE A Large Ion Collider Experiment 18, 19
AP associated production 10, 65
ATLAS A Toroidal LHC Apparatus 1, 2, 13–15, 18, 19, 80
BCM beam conditions monitor 29
BDT boosted decision tree 36, 40
BSM beyond the standard model 15, 19
CDF Collider Detector at Fermilab 13
CERN European Organization for Nuclear Research vii, 17, 18, 26, 30
CL confidence limit viii, 13, 81, 83, 84, 86, 93, 108
CMS Compact Muon Solenoid vii, viii, 1, 2, 12–16, 18–35, 37, 38, 43, 50, 51, 55, 60–65, 70, 80, 83, 85–88, 93, 94, 102, 106–108
CPU central processing unit 30, 47, 50, 51, 62
CSC cathode strip chamber 27, 28
CSV combined secondary vertex 36, 66
D0 D0 experiment 13
DA deterministic annealing 32
DAQ data acquisition 45
DQM data quality monitoring 44
DT drift tube 27, 28
DY Drell-Yan 9, 50, 65–67, 69–72, 75–77, 79, 83, 89, 92
EBH Englert-Brout-Higgs 2, 5, 13–15
ECAL electromagnetic calorimeter vii, 12, 21, 25, 26, 32, 33, 40, 45, 46
FSR final state radiation 40
GCT global calorimeter trigger viii, 45, 109, 112, 113
GEF global event fit 106
ggF gluon fusion 8–10, 50, 65, 67, 69, 70, 75, 78, 83, 85
GSF Gaussian-sum filter 31
GT global trigger viii, 45, 109, 114
HB hadron barrel 26, 27
HCAL hadronic calorimeter 21, 25–27, 33, 40, 45, 46
HE hadron endcap 26, 27
HEP high energy physics 1, 17
HF hadron forward 25–27, 29, 45
HLT high level trigger 30, 43, 44, 47, 48, 50–53, 56–58, 62, 66, 107
HO hadron outer 26, 27
HPS hadrons plus strips vii, 38–40, 47, 50, 52, 54, 58, 66, 87, 88, 90, 93, 97, 102
ID identification viii, 36, 67, 78, 93, 108
ISR initial state radiation 9, 40, 69
JEC jet energy correction 35, 36, 57, 78
L1 Level-1 trigger vii, viii, 29, 30, 43–48, 50, 53–58, 60, 107, 109–112, 114
LEP Large Electron-Positron Collider 13, 17
LHC Large Hadron Collider vii, 1, 2, 4, 8, 10–15, 17–22, 24, 26, 29, 30, 33, 35, 37, 39–41, 44, 48, 52, 62, 65, 67, 70, 85, 88, 95, 97, 106, 108, 109, 123
LHCb Large Hadron Collider beauty 18, 19
Linac 2 Linear accelerator 2 18
MC Monte Carlo simulation 31, 35, 36, 38–41, 44, 70–72, 78–80, 85, 92, 93, 97, 100, 104
MET missing transverse energy 36
MVA multivariate analysis 36, 37, 40, 52, 66, 75, 89, 93, 102, 108
NNLO next-to-next-to-leading order 70
OOT-PU out-of-time pileup 21
PDF parton distribution function vii, 4, 5, 78
pdf probability density function 80
PF particle flow 31–33, 35–38, 47, 50, 53, 58, 70, 87, 107
PLT pixel luminosity telescope 29
PS Proton Synchrotron 18

PSB Proton Synchrotron Booster 18
PU pileup vii, 19, 20, 23, 32, 35–37, 39–41, 50, 51, 55, 58, 62, 67, 71, 88
PV primary vertex 23, 32, 34, 36, 66, 88, 94, 97, 102, 108
QCD quantum chromodynamics 2, 4, 12, 16, 25, 28, 35, 50, 52, 54, 55, 57, 71, 72, 79, 87–90, 92, 93, 97, 100, 102, 107, 108
QED quantum electrodynamics 2
RCT regional calorimeter trigger viii, 45, 109, 111
RF radio frequency 18
RMS root mean square 57, 58, 97
RPC resistive plate chamber 27, 28
SLAC Stanford Linear Accelerator Center 16
SM standard model of particle physics vii, ix, 1–5, 7, 8, 10–16, 19, 72, 77, 79, 80, 83, 85, 86, 89, 93, 106
SPS Super Proton Synchrotron 17, 18
SSB spontaneous symmetry breaking 2, 3, 6, 7
SV secondary vertex 23, 24, 36, 87–90, 93, 94, 97, 102, 106, 108
TEC tracker endcap 23, 24
TIB tracker inner barrel 23, 24
TID tracker inner disk 23, 24
TOB tracker outer barrel 23, 24
ttH associated production with top quarks 8–10, 12
VBF vector boson fusion viii, 8–10, 12, 26, 35, 50, 64, 65, 67, 69–72, 75, 78, 79, 83, 85
VH Higgs strahlung 8–10, 12, 85