

Erik Pekkarinen

**Data-driven measurement of the
background with misidentified tau leptons
in a search for charged Higgs bosons**

School of Science

Thesis submitted for examination for the degree of Master of
Science in Technology.

Espoo 10.3.2015

Thesis supervisor:

Prof. Filip Tuomisto

Thesis advisor:

Dr. Sami Lehti

Author: Erik Pekkarinen

Title: Data-driven measurement of the background with misidentified tau leptons in a search for charged Higgs bosons

Date: 10.3.2015

Language: English

Number of pages: 9+81

Department of Applied Physics

Professorship: Applied Physics

Code: Tfy-56

Supervisor: Prof. Filip Tuomisto

Advisor: Dr. Sami Lehti

In 2012, a neutral Higgs boson with a mass of 125 GeV was discovered by the CMS and ATLAS experiments at the CERN LHC. Models with an extended Higgs sector involving charged Higgs bosons are not excluded by the discovery. A search procedure for the charged Higgs bosons (H^\pm) focusing on the $H^\pm \rightarrow \tau^\pm \nu_\tau$ decay channel in the fully hadronic final state is presented. The covered mass ranges are 80-160 GeV, when the charged Higgs boson is lighter than the top quark and 180-600 GeV, when the charged Higgs boson is heavier than the top quark. The presented search is based on the data recorded by the CMS experiment in 2012 at the center-of-mass energy of 8 TeV. As charged Higgs bosons are not observed in the studied data, the goal of the analysis is the setting of 95 % model-independent upper limits for their production and decay.

To observe a potential signal, it has to be extracted from the backgrounds. Of the three main backgrounds, the QCD multijet events along with the $t\bar{t}$ and electroweak events with genuine τ leptons are measured from data. The background consisting of $t\bar{t}$ and electroweak events with misidentified τ leptons has before this work been estimated from simulation. A data-driven measurement method for this background is studied. Since the QCD multijet background also lacks genuine τ leptons, the composition of these backgrounds is referred to as the *background with misidentified τ leptons* in this work and a method measuring this background inclusively is developed. The results given by the measurement method are in good agreement with simulated estimates and the uncertainties in the backgrounds yields are reduced by the developed method. The background purity obtained with the method is also found to increase when compared to that of the QCD multijet background measurement method. This is important especially for the LHC Run II, which requires using higher selection thresholds.

Keywords: Charged Higgs boson, Background measurement, CERN, CMS

Tekijä: Erik Pekkarinen		
Työn nimi: Väärintunnistettuja tau leptoneita sisältävän taustan datalähtöinen mittaus varattujen Higgsin bosonien etsinnässä		
Päivämäärä: 10.3.2015	Kieli: Englanti	Sivumäärä: 9+81
Teknillisen fysiikan laitos		
Professuuri: Teknillinen fysiikka		Koodi: Tfy-56
Työn valvoja: Prof. Filip Tuomisto		
Työn ohjaaja: FT Sami Lehti		
<p>Vuonna 2012 CERNin LHC:ssä toimivat CMS- ja ATLAS-kokeet löysivät neutraalin Higgsin bosonin, jonka massa on 125 GeV. Löytö ei sulje pois laajennetun Higgsin sektorin sisältäviä malleja, jotka ennustavat varattujen Higgsin bosonien olemassaolon. Tässä työssä esitellään menetelmä varattujen Higgsin bosonien (H^\pm) etsintään, keskittyen $H^\pm \rightarrow \tau^\pm \nu_\tau$ hajoamiskanavaan sekä täysin hadroniseen lopputilaan. Tutkitut massavälit ovat 80-160 GeV, kun varattu Higgsin bosoni on t-kvarkkia kevyempi ja 180-600 GeV, kun varattu Higgsin bosoni on t-kvarkkia raskaampi. Esiteltävä etsintä perustuu CMS-kokeen vuonna 2012 keräämään dataan, jonka massakeskipiste-energia on 8 TeV. Havaintoja varatusta Higgsin bosonista ei näy tutkimuksessa datassa, joten analyysin tavoitteena on asettaa 95 %:n malliriippumattomat ylärajat varatun Higgsin bosonin tuotolle ja hajoamiselle.</p> <p>Mahdollisen signaalin havaitsemiseksi se on erotettava taustoista. Kolmesta päätaustasta QCD moniryöppy-törmäystapahtumat sekä $t\bar{t}$ ja sähköheikot törmäystapahtumat, jotka sisältävät aitoja τ leptoneita, mitataan datalähtöisesti. Väärintunnistettuja τ leptoneita sisältävää $t\bar{t}$ ja sähköheikkoa taustaa puolestaan on aiemmin arvioitu simuloinnin avulla. Tässä työssä tutkitaan datalähtöistä mittausmenetelmää kyseiselle taustalle. Koska myöskään QCD moniryöppy-tausta ei sisällä aitoja τ leptoneita, näiden kahden taustan yhdistelmää kutsutaan tässä työssä <i>väärintunnistettuja τ leptoneita sisältäväksi taustaksi</i> ja kehitetty menetelmä mittaa tämän taustan kokonaisuudessaan. Mittausmenetelmän antamat tulokset ovat hyvässä sopusoinnussa simuloitujen arvioiden kanssa ja epävarmuudet taustojen tuotoissa pienenevät sen myötä. Menetelmä myös kasvattaa taustan puhtautta verrattuna QCD moniryöppy-taustan mittausmenetelmään. Tämä on erityisen tärkeää pitäen silmällä LHC:n uudelleenkäynnistystä suuremmilla energioilla ja sen myötä nostettavia törmäystapahtumien valintaroja.</p>		
Avainsanat: Varattu Higgsin bosoni, Taustamittaus, CERN, CMS		

Preface

This master's thesis is based on the research done at the Helsinki Institute of Physics (HIP) during the latter half of the year 2014 and the beginning of the year 2015. During this time my life took a big turn, as I was able to finally shift my full-time focus from course work to research. Along with practical data analysis skills, this time taught me much about the creative aspect in physics.

I would like to thank my advisor Dr. Sami Lehti for providing me the great opportunity to do research on this particular subject, setting clear and inspiring goals for the work and for always putting the reviewing of the manuscript as his first priority. I thank Dr. Lauri Wendland for the instructive explanations of many matters and for the countless hours of help on technical issues. I would also like to acknowledge Doc. Ritva Kinnunen and Dr. Matti Kortelainen for always having time for questions and support. I would like to thank Santeri Laurila for helping me to get in progress in the beginning of the research work and for reading my thesis. I also offer my special thanks to Prof. Paula Eerola for providing me the possibility to work at the HIP in the first place, and to my supervisor Prof. Filip Tuomisto for the comments regarding this work and for the flexibility in practical matters.

The results of this thesis have been presented at CERN at the Higgs to Taus-working meeting on the 16th of January 2015. I thank Sami and Dr. Kati Lassila-Perrini for the possibility of attending the meeting, along with Dr. Roger Wolf and Dr. Jan Stegmann for the valuable comments in the meeting. I also convey my great appreciation to the whole CMS collaboration for maintaining this outstanding experiment.

I would like to express my deep gratitude to my parents Mika and Päivi for their constant encouragement and support throughout my studies, and also my gratefulness to the rest of the family for their understanding and help at all times: my stepfather Arne, my stepmother Maiju, as well as my siblings Robert, Sofie and Frans. I would also like to thank my friend Pekka for making physics more fun since high school. Jenny, thank you for the kindness, patience, and for making the past years so much richer.

Otaniemi, 10.3.2015

Erik Pekkarinen

Contents

Abstract	ii
Abstract (in Finnish)	iii
Preface	iv
Contents	v
Symbols, abbreviations and conventions	vii
1 Introduction	1
2 Theory of charged Higgs bosons	4
2.1 Standard Model and its problems	4
2.2 Higgs mechanism	5
2.3 Two-Higgs doublet model	9
2.3.1 Minimal supersymmetric standard model	11
2.4 Investigated H^\pm search channel	12
2.4.1 H^\pm production	12
2.4.2 Hadronic τ lepton decays	13
3 Large Hadron Collider and CMS experiment	16
3.1 Large Hadron Collider	16
3.2 CMS detector	17
3.2.1 CMS by layers	18
3.2.2 Trigger system	20
4 Experimental concepts regarding hadron collisions at the LHC	22
4.1 Coordinate system	22
4.2 Rapidity and angular variables	22
4.3 Track parameters	23
4.4 Momentum and energy	24
4.5 Mass variables	25
4.6 Luminosity and cross section	26
4.7 Pile-up	26
5 Event reconstruction and simulation	27
5.1 Event reconstruction	27
5.1.1 Tracking	28
5.1.2 Primary vertices	28
5.1.3 Muons	29
5.1.4 Electrons	30
5.1.5 Photons	30
5.1.6 Jets	31

5.1.7	B tagging	31
5.1.8	Identification of hadronic τ decays	32
5.1.9	Missing transverse energy	33
5.2	Event simulation	35
6	Event selection	36
6.1	Online selection	36
6.2	Offline selection	37
6.2.1	Data quality criteria	37
6.2.2	Main selection cuts	37
6.2.3	Tau polarisation cut	40
6.2.4	Angular cuts	42
6.3	Optimisation of cuts	44
6.4	Fitting of the transverse mass tail	44
7	Background measurements	46
7.1	Data-driven measurement of the QCD multijet background	46
7.1.1	Strategy of the measurement	46
7.1.2	Normalisation	47
7.1.3	QCD purity	49
7.1.4	Systematic uncertainties	50
7.2	Data-driven measurement of the EWK+ $t\bar{t}$ background with genuine τ leptons	51
7.3	Simulation based measurement of the EWK+ $t\bar{t}$ background with misidentified τ leptons	52
8	Data-driven measurement of EWK+$t\bar{t}$ background with misidentified τ leptons	54
8.1	Breakdown of background origin	54
8.2	Principal idea behind the method	54
8.3	Normalisation	55
8.4	Systematic uncertainties	57
8.5	Fake rate probabilities	59
8.6	Fake τ purity	59
8.7	Transverse mass	63
8.8	Closure test	64
9	Results	65
9.1	Event yields	65
9.2	Expected limits	65
10	Conclusions	72
	References	74

Symbols, abbreviations and conventions

Symbols

A	neutral CP-odd (pseudoscalar) Higgs boson
$\mathcal{B}(X)$	branching ratio for process X
b	beauty quark
c	speed of light in vacuum
ΔR	angular distance
Z/γ^*	Drell-Yan process
e	electron
e	elementary charge
E_T^{miss}	missing transverse energy (magnitude)
\vec{E}_T^{miss}	missing transverse energy (vector)
ϵ	efficiency
η	pseudorapidity
γ	photon
h	lightest neutral CP-even Higgs boson <i>or</i> hadron
H	neutral CP-even Higgs boson
H^\pm	charged Higgs boson
I	isospin
\mathcal{L}	Lagrangian density <i>or</i> instantaneous luminosity
L	integrated luminosity
M	invariant mass
m_T	transverse mass
μ	muon
$N_{\text{d.o.f.}}$	number of degrees of freedom
ν	neutrino
ν_τ	tau neutrino
p	proton
p_T	transverse momentum (magnitude)
\vec{p}_T	transverse momentum (vector)
ϕ	azimuthal angle
R_τ	tau lepton polarisation variable
ρ	distance from the z axis
\sqrt{s}	center-of-mass collision energy
$SU(n)$	special unitary group of degree n
σ_X	cross section for process X
t	top quark
$\tan \beta$	ratio of two Higgs doublet vacuum expectation values
τ	tau lepton
τ_h	hadronically decaying tau lepton
θ	polar angle
θ_W	weak mixing angle
$U(n)$	unitary group of degree n
v	vacuum expectation value of Higgs field
V	potential
W^\pm	charged weak boson
y	rapidity
Z	neutral weak boson

Abbreviations

ATLAS	A Toroidal LHC Apparatus
CERN	European Organization for Nuclear Research
CL	Confidence level
CM	Center of mass
CMS	Compact Muon Solenoid
CP	Charge parity
CSC	Cathode strip chamber
CSV	Combined secondary vertex algorithm
DT	Drift tube
ECAL	Electromagnetic calorimeter
EWK	Electroweak theory
GSF	Gaussian sum filter
HCAL	Hadron calorimeter
HLT	High level trigger
HPS	Hadron-Plus-Strips algorithm
LHC	Large Hadron Collider
MC	Monte Carlo simulation method
MET	Missing transverse energy
MSSM	Minimal Supersymmetric Standard Model
QCD	Quantum chromodynamics
SM	Standard Model
PF	Particle-Flow algorithm
PV	Primary vertex
RPC	Resistive plate chamber
VEV	Vacuum expectation value of Higgs field
2HDM	Two-Higgs-doublet model
4FS	4 flavour scheme
5FS	5 flavour scheme

Conventions

Units Natural units, in which the reduced Planck constant and the velocity of light are set to unity ($\hbar = c = 1$), are used throughout this work. Energies, momenta and masses are given in units of electron volts (eV).

Charge-conjugated processes In this work, it is assumed that no strong violation of CP symmetry exists, which means that all cross sections, branching fractions, and decay widths are assumed to be equal for charge-conjugated processes. Hence, whenever particle processes are expressed, it is implied that the charge-conjugated process is considered as well. Additionally, the charges of particles are not written out explicitly when there is no chance of misunderstanding i.e. the process $H^\pm \rightarrow \tau\nu$ refers to both $H^+ \rightarrow \tau^+\nu_\tau$ and $H^- \rightarrow \tau^-\bar{\nu}_\tau$.

Plots showing simulated data In the plots showing simulated data, the simulated event samples are normalised to the integrated luminosity of 21.79 fb^{-1} . The data gathered in the year 2012 is used in the analysis, and this luminosity corresponds approximately to its amount.

The plots involving the simulated charged Higgs boson signal are shown for $\mathcal{B}(t \rightarrow bH^\pm) = 0.01$ and $\mathcal{B}(H^\pm \rightarrow \tau^\pm\nu_\tau) = 1$. Correspondingly, 0.01 is subtracted from the Standard Model top quark pair decay branching ratio $\mathcal{B}(t\bar{t} \rightarrow bW^\pm bW^\mp)$.

1 Introduction

Particle physics is a field of science which investigates the elementary particles of nature and the interactions between them. Theoretically, particle physics is described by the *Standard Model* (SM) [1–3]. It describes with great accuracy a wide range of different phenomena, and it has successfully predicted the existence of several particles observed later: most recently, the last missing piece of the theory, the Higgs boson, discovered in 2012 [4–6]. The observed Higgs boson is likely to be responsible for the mass generation of other particles through the Higgs mechanism [7–12], but further research is required to confirm this. Despite its success, the theory is lacking in a few senses, since it for example does not include gravity nor give an explanation for dark matter. Some extended theory is needed and experimental evidence for the correct extension is highly called for.

In this work, new physics is searched for in the Higgs sector, of which little is known so far. In addition to the recently discovered neutral Higgs boson predicted by the Standard Model, charged Higgs bosons H^\pm might exist as well, and they are the focus of the presented analysis with the aim of proving or ruling out their existence in the electroweak scale (~ 1 TeV). These charged Higgs bosons are predicted by several extensions of the Standard Model [13–20]. In particular, they are predicted by all theories involving additional Higgs doublet fields. Many of these theories involve two doublets and are based on supersymmetry [13, 21].

There are several challenges in reaching experimental conclusions regarding charged Higgs bosons. Elementary particles are experimentally studied by colliding particles with high energies, since according to the well-known equation $E = mc^2$, higher energies are needed to produce particles with more mass. To achieve this, powerful particle accelerators are built to reach increasingly higher energies. In this work, the data collected at the most powerful accelerator built to this date, the Large Hadron Collider (LHC), is used. As the collision processes have a probabilistic quantum mechanical nature, many collision events must be observed in order to reach statistically significant conclusions. This is the reason for having a very high collision frequency at the LHC. In addition, as charged Higgs bosons are expected to have a very short lifetime, their existence must be inferred from their decay products. Also, the masses of the Higgs bosons are not predicted by theory, and this is why they are searched for over a wide mass range of 80 – 600 GeV. The lower threshold is motivated by the model-independent lower limit of $m_{H^\pm} > 78.6$ GeV, set by the experiments at the LEP collider [22–25]. The upper threshold on the other hand is set by the available data: the aim is to raise it close to 1 TeV in the future as more data with higher collision energies is collected at the LHC, making the analysis sensitive for heavier charged Higgs bosons.

The mass of the top quark $m_t \approx 173.07$ GeV [26] divides the analysis of the charged Higgs boson into two mass regions: in the light charged Higgs boson case $m_{H^\pm} \lesssim m_t$, and the opposite situation where $m_{H^\pm} \gtrsim m_t$ is referred to as the heavy charged Higgs boson case. In the former case, charged Higgs bosons are produced in the decays of the top quarks as $t \rightarrow bH^+$ and $\bar{t} \rightarrow \bar{b}H^-$. The shorthand notation $t \rightarrow bH^\pm$ is used for these processes. In the pp collisions at the LHC, the top quarks on the other hand are mostly produced as top-antitop pairs $t\bar{t}$. In the heavy charged Higgs boson case, where this decay is not kinematically allowed, charged Higgs bosons are produced in association with top and bottom quarks through the processes $pp \rightarrow H^+b\bar{t}$ and $pp \rightarrow H^-\bar{b}t$. For these processes, the shorthand notation $pp \rightarrow H^\pm tb$ is used. Of the decay modes of the charged Higgs bosons, the presented analysis is sensitive to the decay processes $H^+ \rightarrow \tau^+\nu_\tau$ and $H^- \rightarrow \tau^-\bar{\nu}_\tau$, or $H^\pm \rightarrow \tau\nu$ in shorthand notation.

The presented analysis is performed on the proton-proton collision data recorded with the Compact Muon Solenoid (CMS) detector of the LHC during the year 2012. During that time, the total integrated luminosity of the recorded events was 21.79 fb^{-1} , and the energy of the collisions was 4 TeV per beam. The aim of the analysis as a whole is to set limits on the probability of the charged Higgs boson production and decay. In terms of the branching ratio \mathcal{B} , this is expressed as $\mathcal{B}(t \rightarrow bH^\pm) \times \mathcal{B}(H^\pm \rightarrow \tau^\pm\nu_\tau)$ in the light charged Higgs boson case. In the heavy charged Higgs boson case, this probability is expressed using the cross section σ as $\sigma_{pp \rightarrow H^\pm tb} \times \mathcal{B}(H^\pm \rightarrow \tau^\pm\nu_\tau)$. The *expected limit* is set on these probabilities by assuming that the data is described by the background, and it describes the sensitivity of the analysis. The *observed limit* on the other hand is used to quantify whether charged Higgs bosons exist or not, but it is not a focus of this work.

The analysis considers only the so called fully hadronic final state that is characterised by a hadronically decaying τ lepton, missing transverse energy resulting from neutrinos and additional jets of which two originate from bottom quarks. This is the *detector fingerprint* of the charged Higgs boson *signal (candidate) events*. In the analysis, these events are searched for using an *event selection*. In addition to the signal events, some particular *background events* that certainly do not contain charged Higgs bosons can also pass the selection. This can happen because they have a similar final state as the signal events: as an example, the top-antitop pair can decay to two W bosons, and the event then passes the event selection due to its similarity with the signal events. Another possibility is that the background events pass the selection because of mismeasurement and misidentification: for example, the energies of the particles can be mismeasured, which can easily lead to the false conclusion that the event contains undetected neutrinos. The aim is to suppress both of these background types as much as possible, and then estimate their effect to extract the signal. The backgrounds can be estimated using simulation or they can be measured directly from the data. The latter approach is called *data-driven* background measurement. The focus of this work is on a background measurement method, so the main backgrounds of the analysis are next reviewed.

The background that may contain a genuine τ lepton consists of the following event types: SM top-antitop events ($t\bar{t}$), in which both top quarks decay to a W boson; events in which a single top quark is produced (single top); events with a single W boson and additional jets ($W + \text{jets}$); diboson events with two weak bosons (WW, WZ, ZZ); and events with a Drell-Yan process ($Z/\gamma^* + \text{jets}$). These processes are collectively referred to as the *EWK+ $t\bar{t}$ genuine τ background* processes throughout this work, because they contain electroweak processes (except for the $t\bar{t}$ and single top backgrounds). Of these background events, the $t\bar{t}$ and single top backgrounds are largely irreducible, while the rest can be partially reduced using b jet tagging and the information given by the spin of the τ lepton, as will be discussed in more detail later. The EWK+ $t\bar{t}$ background with genuine τ leptons is measured from data in the analysis.

The EWK+ $t\bar{t}$ background events can however also pass the selection without containing a genuine τ lepton due to the misidentification of a hadronic jet, an electron, or a muon as a τ lepton. This background is called the *EWK+ $t\bar{t}$ misidentified τ background* or equivalently the *EWK+ $t\bar{t}$ fake τ background*. In the analysis before this work, this background has been estimated with a simulation. The final background type with a significant contribution is the *QCD multijet background*. It contains multiple jets produced only via the strong interaction, governed by the theory of quantum chromodynamics (QCD). These events rarely contain genuine τ leptons or neutrinos, and only pass the event selection because of misidentification and mismeasurement. Although the misidentification and mismeasurement rates are small, this background has to be taken into account because of its large production cross section in the collisions. The QCD background is measured using a data-driven approach.

The focus of this work is on the EWK+ $t\bar{t}$ fake τ background. This background makes up approximately 10 % of the expected background in the presented analysis [27], and it mainly consists of hadronic jets misidentified as τ leptons. The dominating sources for this background are the $t\bar{t}$ ($\sim 90\%$) and single top events. The *first* goal is to develop a method that reliably measures this background from data. As data is always more reliable than simulation, the advantage of doing this is that the systematic uncertainties related to the simulation could be disposed to a large extent. The *second* goal of this work is reducing uncertainties related to the measurement of the EWK+ $t\bar{t}$ fake τ background. This is then expected to increase the signal sensitivity of the analysis and this is studied by calculating the expected limits. Since the EWK+ $t\bar{t}$ fake τ and QCD events have the common feature of not containing genuine τ leptons, the combined background is referred to as the *fake τ background* in this work. The data-driven method to measure the EWK+ $t\bar{t}$ fake τ background is developed by extending the data-driven QCD measurement method so that it measures the whole fake τ background inclusively. The *third* goal of this work is to improve the QCD measurement method having an eye to the LHC Run II. The achievement of the aforementioned goals is discussed in Sections 8, 9 and 10.

2 Theory of charged Higgs bosons

2.1 Standard Model and its problems

The Standard Model (SM) of particle physics is a theory describing the electromagnetic, weak, and strong interactions, which mediate the dynamics of the known elementary particles. These particles can be grouped in quarks, leptons, gauge bosons and the neutral Higgs boson. All SM particles and some of their characteristics are presented in Table 1. The SM is technically formulated as a quantum field theory, in which the Lagrangian density describing its dynamics and kinematics is constructed so that it is invariant under so called *gauge transformations* [28]. This gives rise to the interactions of the theory, that are described by two separate formulations: the electromagnetic and weak interactions are unified and described by the electroweak (EWK) theory, while the strong interaction is explained using quantum chromodynamics (QCD). The masses of the weak bosons are not allowed to be non-zero by the gauge invariance and for them to have mass, the Higgs mechanism is additionally introduced. This theory ‘spontaneously’ breaks the EWK symmetry and requires an additional field, the *Higgs field*, and the related particle, the spinless *Higgs boson* with a non-zero mass.

The current formulation of the SM was finalised in the mid-1970s upon experimental confirmation of the existence of quarks. The discoveries of the top quark (1995), the τ neutrino (2000), and most recently the Higgs boson (2012) have given further assurance for the validity of the theory, that has been observed to be in an astonishing agreement with a wide range of experiments. Yet, the SM has several known issues. For example, CP violation is a necessary ingredient to explain the dominance of matter over antimatter in the present Universe but the CP violating effects arising from the SM are too small. Also, in the SM neutrinos are massless, which is in contradiction with observations of so called *neutrino oscillations*: this phenomenon requires the neutrinos to have at least a very small mass. The theory is also lacking because it gives explanations neither to gravity, dark matter nor dark energy. Additionally, the SM gives no explanation to the question of why the fermions are grouped to *three* families.

There is also a severe problem in the fundamental nature of the theory itself, rising from the radiative corrections to particle masses that are typical to quantum field theories: from the theory one would expect the radiative corrections on the Higgs mass to be huge, comparable with the Planck mass (1.2×10^{19} GeV). However, experimentally it is proven that these corrections are of the order of the W boson mass. This separation of scales is called the *hierarchy problem* and could only be explained by a fine-tuning cancellation between the quadratic radiative corrections and the bare mass.

Table 1: Particle contents of the Standard Model [29].

Particle	Symbol	Spin	Charge (e)	Mass (GeV)
First generation fermions				
Electron neutrino	ν_e	1/2	0	$< 2.2 \times 10^{-9}$
Electron	e	1/2	-1	5.11×10^{-4}
Up quark	u	1/2	2/3	2.4×10^{-3}
Down quark	d	1/2	-1/3	4.8×10^{-3}
Second generation fermions				
Muon neutrino	ν_μ	1/2	0	$< 1.7 \times 10^{-4}$
Muon	μ	1/2	-1	0.1057
Charm quark	c	1/2	2/3	1.27
Strange quark	s	1/2	-1/3	0.104
Third generation fermions				
Tau neutrino	ν_τ	1/2	0	$< 1.55 \times 10^{-2}$
Tau	τ	1/2	-1	1.777
Top quark	t	1/2	2/3	171.2
Bottom quark	b	1/2	-1/3	4.2
Gauge bosons				
Photon	γ	1	0	0
W bosons	W	1	± 1	80.4
Z boson	Z	1	0	91.2
Gluons	g	1	0	0
Higgs boson				
Higgs boson	H	0	0	125.7

The cancellation could more naturally be understood with requirements of higher scale symmetries, which lead to expansions of the Higgs sector. The *two-Higgs doublet model* (2HDM) [30–34] is one of the simplest extensions of the SM, containing two Higgs doublets instead of just one. For example, it solves the matter-antimatter asymmetry by allowing CP violations explicitly in the scalar sector. The 2HDM theory also includes possible explanations for dark matter and the fermion mass spectrum [35,36]. This theory is discussed in more detail in Section 2.3.

2.2 Higgs mechanism

The SM electroweak theory is governed by the gauge symmetry group $SU(2)_L \times U(1)_Y$. This group is generated by the Pauli matrices of weak isospin τ_a and the hypercharge Y . The weak isospin gauge fields of the $SU(2)_L$ couple only to left-handed fermions, which is denoted with the subscript L . In the SM, the electroweak symmetry is ‘spontaneously’ broken by the Higgs mechanism [7–12], which provides arguably the most elegant way to generate mass to fermions and gauge bosons.

In the Higgs mechanism, the masses are generated when a new potential, the Higgs potential, with a linked scalar Higgs field $\phi = \phi(x)$, is added to the Lagrangian of the theory. The field is contained in the free propagation term

$$(D_\mu \phi)^\dagger (D^\mu \phi), \quad (1)$$

where D_μ is the covariant derivative defined as

$$\partial_\mu + \frac{ig}{2} \tau_a W_\mu^a + \frac{ig'}{2} Y B_\mu. \quad (2)$$

Here W_μ and B_μ are the gauge fields and the coupling constants are denoted with g and g' . The potential is included in the term

$$V_{\text{Higgs}} = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (3)$$

where μ is a complex and λ a real constant parameter. These parameters are chosen so that $\mu^2 < 0$ and $\lambda > 0$, which makes the potential bounded from below. The qualitative shape of the resulting potential is shown in Fig. 1. The potential can be seen to depend only on the product $\phi^\dagger \phi$, which means that minima of the potential can be found as

$$\frac{\partial V}{\partial \phi^\dagger \phi} = \mu^2 + 2\lambda \phi^\dagger \phi = 0. \quad (4)$$

This leads to

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}, \quad (5)$$

where v is defined as the non-zero vacuum expectation value (VEV) of the potential.

The Lagrangian for the Higgs field can be obtained from Eqs. (1)-(3) as

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (6)$$

The field associated with the potential is taken to be a complex scalar doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (7)$$

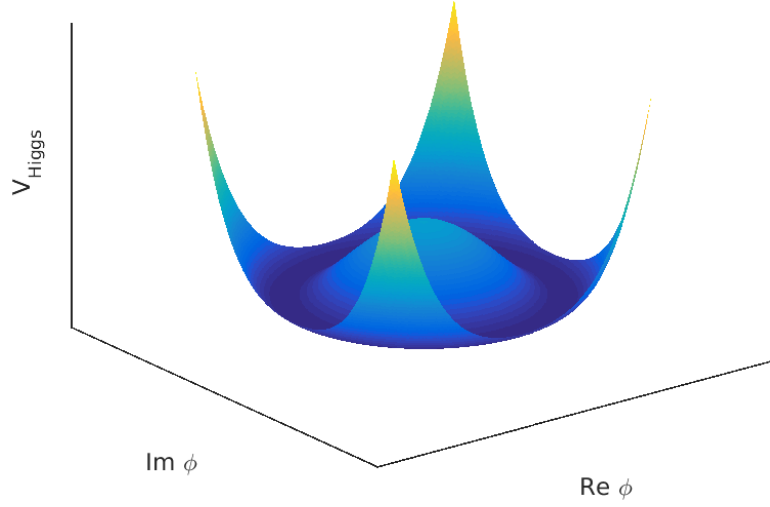


Figure 1: Shape of the Higgs potential.

where ϕ_a are real scalar fields. Substituting this to Eq. (5) gives

$$\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = -\frac{\mu^2}{\lambda}. \quad (8)$$

The potential can now be seen to have an infinite number of minima on a circle of radius v in the ϕ_a plane. These minima of energy correspond by definition the vacuum, which now is not at the origin. This means that the Higgs field has a non-zero vacuum expectation value v . Once any set of ϕ_a satisfying Eq. (8) is chosen, the symmetry of the equation and that of the $SU(2)_L$ and $U(1)_Y$ is spontaneously broken.

Usually ϕ_a are chosen so that $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$, which makes

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (9)$$

the vacuum expectation value of the field doublet. The particle spectrum of the theory can be found by expanding ϕ around the vacuum value v as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (10)$$

which gives

$$\phi^\dagger \phi = \frac{1}{2} (v + H)^2. \quad (11)$$

Inserting the ϕ of the broken symmetry into Eq. (6) yields

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2} (\partial_\mu H) (\partial^\mu H) + \mu^2 H^2 - \frac{g^2 v^2}{8} \left((W_1)_\mu (W_1)^\mu + (W_2)_\mu (W_2)^\mu \right) \quad (12)$$

$$- \frac{v^2}{8} \left(g (W_3)_\mu - g' B_\mu \right) \left(g (W_3)^\mu - g' B^\mu \right) + \mathcal{L}_\chi, \quad (13)$$

where \mathcal{L}_χ contains all the uninteresting interaction and cross-terms. A term in the Lagrangian is said to have the form of a mass term if it contains squares of the fields, like $A_\mu A^\mu$. While e.g. the H^2 term in the above Lagrangian is of this form, the weak isospin and hypercharge gauge fields W_3 and B mix in the last term. In order for them to be mass eigenstates that represent physical particles, neutral gauge fields A_μ and Z_μ are constructed as orthogonal combinations of the original fields through a rotation around the so called weak mixing angle θ_W

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ (W_3)_\mu \end{pmatrix}. \quad (14)$$

The weak mixing angle also links together the coupling constants g and g' with the elementary charge e according to

$$g \sin \theta_W = g' \cos \theta_W \equiv e. \quad (15)$$

The physical charged boson fields are in turn obtained with

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left[(W_1)_\mu \mp i (W_2)_\mu \right]. \quad (16)$$

With the help of these expressions, the final Lagrangian of the Higgs field becomes

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2} (\partial_\mu H) (\partial^\mu H) + \mu^2 H^2 - \frac{g^2 v^2}{4} (W^+)_\mu (W^-)^\mu \quad (17)$$

$$- \frac{1}{2} \left(\frac{g v}{2 \cos \theta_W} \right)^2 Z_\mu Z^\mu + 0 \times A_\mu A^\mu + \mathcal{L}_\chi. \quad (18)$$

It can now be seen that because of the symmetry breaking, the W^\pm bosons have acquired the mass $M_{W^\pm} = gv/2$, while the mass of the Z boson is $M_Z = gv/(2\cos\theta_W)$. The photon in turn remains massless, as it should. Another noteworthy feature of the mechanism is that while the Higgs field introduces 4 new degrees of freedom as a complex doublet, 3 of these are absorbed by the gauge boson fields as the weak bosons W^\pm and Z acquire mass. These fields are the so called *Goldstone modes* G^0 and G^\pm . After this, only one additional degree of freedom is left, which corresponds to a new scalar particle called the Higgs boson.

In the process of acquiring the masses for the vector fields, the mass term for the scalar Higgs field appearing as the second term in Eq. (18) is introduced: this is the mass of the Higgs boson, and it remains as a free parameter in the theory to be determined experimentally. The fermion masses are also generated when the electroweak symmetry is broken and Yukawa interactions between the fermion and Higgs fields are introduced. These masses however remain as free parameters too. The masses of the weak bosons have been measured, which sets the vacuum expectation value of the Higgs potential to be $v \approx 246$ GeV. As discussed earlier, very recently also the Higgs boson has been experimentally observed and its mass has been measured.

2.3 Two-Higgs doublet model

Even though a Higgs boson has recently been discovered, only little is known about the Higgs sector. Investigating models that go beyond the minimal SM Higgs sector is thus necessary. One of the simplest extensions to the SM Higgs sector is obtained by adding another doublet to the theory, which leads to a two-Higgs doublet model (2HDM) [31–34]. The general vacuum structure of the 2HDM is very rich and the most general scalar potential contains 14 free parameters [30]. It can have CP conserving or CP violating minima in addition to charge violations. Usually it is assumed that CP is conserved in the Higgs sector and that CP is not spontaneously broken.

The most general scalar potential under these assumptions is given by [37]

$$\begin{aligned}
 V_{\text{Higgs}} = & m_{11}^2 \phi_1^\dagger \phi_1 + m_{22}^2 \phi_2^\dagger \phi_2 - \left[m_{12}^2 \phi_1^\dagger \phi_2 + \text{h.c.} \right] \\
 & + \frac{\lambda_1}{2} (\phi_1^\dagger \phi_1)^2 + \frac{\lambda_2}{2} (\phi_2^\dagger \phi_2)^2 + \lambda_3 \phi_1^\dagger \phi_1 \phi_2^\dagger \phi_2 + \lambda_4 \phi_1^\dagger \phi_2 \phi_2^\dagger \phi_1 \\
 & + \left\{ \frac{\lambda_5}{2} (\phi_1^\dagger \phi_2)^2 + \left[\lambda_6 (\phi_1^\dagger \phi_2) + \lambda_7 (\phi_1^\dagger \phi_2) \right] (\phi_1^\dagger \phi_2) + \text{h.c.} \right\},
 \end{aligned} \tag{19}$$

where h.c. denotes the hermitian conjugate of the preceeding terms and the Higgs self-coupling parameters λ_i as well as the squared-mass parameters $m_{i,j}^2$,

$i = 1, j = 1, 2$ are assumed to be real. The doublet fields ϕ_1 and ϕ_2 have hypercharge +1.

Under the so called discrete Z_2 -symmetry [38] the interchange of the sign of the other doublet field ($\phi_1 \rightarrow -\phi_1, \phi_2 \rightarrow \phi_2$ or the opposite) leaves the Lagrangian invariant. This symmetry also forbids transitions of the type $\phi_1 \rightarrow \phi_2$. Imposing this symmetry on the potential requires that λ_6 and λ_7 have to be suppressed, because the terms containing them lead to hard Z_2 -violation. An exact Z_2 -symmetry forbids CP-violations, while the soft breaking of the symmetry allows them. The parameter m_{12}^2 governs the soft violation of the Z_2 -symmetry and in a 'realistic' theory this symmetry is violated, that is, $m_{12}^2 \neq 0$.

The minimisation of the potential gives the vacuum expectation values

$$\langle \phi_i \rangle = \frac{v_i}{\sqrt{2}}, \quad i = 1, 2 \quad (20)$$

with the requirement $v^2 \equiv v_1^2 + v_2^2 = (246 \text{ GeV})^2$ [30]. The two complex scalar SU(2) doublets lead to eight fields. The three Goldstone modes G^0 and G^\pm are again absorbed to give mass to the Z^0 and W^\pm gauge bosons. The remaining five fields are physical scalar Higgs fields. Of these, two are charged scalars and they are denoted with H^\pm . Two of the fields correspond to neutral CP-even particles, and these are denoted with h and H . They are the so called *scalars*. The remaining field corresponds to a neutral CP-odd particle A , that is also called the *pseudoscalar*. The heavier of the two neutral scalars is denoted with H .

Diagonalisation of the mass matrix leads to mixing between the CP-even states with mixing angle α ; the charged scalars and the pseudoscalar mix with the angle β . With the help of these angles the original doublets can be rewritten as [37]

$$\phi_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}(G^+ \cos \beta - H^- \sin \beta) \\ v \cos \beta - h \sin \alpha + H \cos \alpha + i(G^0 \cos \beta - A \sin \beta) \end{pmatrix} \quad (21)$$

$$\phi_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}(G^+ \cos \beta - H^- \sin \beta) \\ v \sin \beta + h \cos \alpha + H \sin \alpha + i(G^0 \sin \beta + A \cos \beta) \end{pmatrix} \quad (22)$$

In general, the interactions of the Higgs fields with other fields are determined by the angles α and β . The parameter $\sin(\beta - \alpha)$ in particular controls the approach to the so called decoupling limit, in which the properties of the lightest CP-even Higgs scalar h approach those of the SM Higgs boson [39,40]. If one sets $\sin(\beta - \alpha) = 1$, then the tree-level couplings of h coincide exactly with the SM Higgs boson tree-level couplings. With the discovery of the SM-like Higgs boson, it can be assumed that $\sin(\beta - \alpha)$ is relatively close to unity [40].

An important parameter in the 2HDM is $\tan \beta \equiv v_2/v_1$. As it is the ratio of the vacuum expectation values, it can be considered to determine a particular basis for

the analysis. Using Eqs. (21) and (22), the Higgs masses can be calculated for any basis by requiring that the minimum of the potential corresponds to the vacuum expectation values given in Eq. (20). By introducing four linear combinations of λ_i

$$\lambda \equiv \lambda_1 \cos^4 \beta + \lambda_2 \sin^4 \beta + \frac{1}{2} \lambda_{345} \sin^2 (2\beta) \quad (23)$$

$$\hat{\lambda} \equiv \frac{1}{2} \sin (2\beta) [\lambda_1 \cos^2 \beta - \lambda_2 \sin^2 \beta - \lambda_{345} \cos (2\beta)] \quad (24)$$

$$\lambda_A \equiv \cos (2\beta) (\lambda_1 \cos^2 \beta - \lambda_2 \sin^2 \beta) + \lambda_{345} \sin^2 (2\beta) - \lambda_5 \quad (25)$$

$$\lambda_F \equiv \lambda_5 - \lambda_4, \quad (26)$$

where $\lambda_{345} \equiv \lambda_3 + \lambda_4 + \lambda_5$, these masses can be written as [40]

$$m_h^2 = v^2 \left[\lambda - \frac{\hat{\lambda} \cos (\beta - \alpha)}{\sin (\beta - \alpha)} \right] \quad (27)$$

$$m_H^2 = v^2 \left[\lambda + \frac{\hat{\lambda} \sin (\beta - \alpha)}{\cos (\beta - \alpha)} \right] \quad (28)$$

$$m_A^2 = v^2 \left[\lambda_A + \hat{\lambda} \left(\frac{\sin (\beta - \alpha)}{\cos (\beta - \alpha)} - \frac{\cos (\beta - \alpha)}{\sin (\beta - \alpha)} \right) \right] \quad (29)$$

$$m_{H^\pm}^2 = m_A^2 + \frac{v^2}{2} \lambda_F \quad (30)$$

These expressions are exact. They are tree-level results, but in the general 2HDM this is not a problem as all the masses are independent parameters [40]. From the above expressions, one can directly note that H is more massive than h , while H^+ and H^- have the same mass.

2.3.1 Minimal supersymmetric standard model

Supersymmetry (SUSY) [13,21] is a theoretical framework postulating a symmetry between fermions and bosons. In supersymmetric theories, each SM particle has a superpartner and the two differ by 1/2 spin. If the supersymmetry would be exact, the superpartners would have exactly the same masses and quantum numbers as the corresponding particles. As no evidence of supersymmetry has been found, the symmetry must be broken. Supersymmetry solves the hierarchy problem and can include gravity. The former is achieved by the fact that the symmetry prevents large radiative corrections to the Higgs boson mass as the quadratic loop contributions of the SM particles are exactly cancelled by their superpartners.

The Minimal Supersymmetric Standard Model (MSSM) [41, 42] is the minimal low-energy globally supersymmetric extension of the SM. The MSSM has minimal particle content by containing three generations of fermions (without right-handed neutrinos) and their superpartners. Two Higgs superfields are included to break the electroweak symmetry. The Higgs sector of the MSSM is a special case of the more general 2HDM: in the MSSM the two doublets have hypercharge +1 and -1, which is necessary to generate masses both to the ordinary particles and their superpartners. The reason for this is the following: in the SM the masses of the fermions of a given weak isospin are generated using a scalar field that also generates the masses of the weak bosons W^\pm and Z , while its conjugate field with opposite hypercharge generates the masses of the fermions with opposite isospin. In supersymmetry however, the superpotential involves only superfields and not their conjugate fields. This is why two doublets with opposite hypercharges are needed to generate the masses of all particles.

2.4 Investigated H^\pm search channel

2.4.1 H^\pm production

The production mechanisms for a charged Higgs boson at the LHC depend on its mass with respect to the difference between the top and bottom quark masses $m_t - m_b$: in the ‘light’ charged Higgs case $m_{H^\pm} \lesssim m_t - m_b$; the opposite situation, where $m_{H^\pm} \gtrsim m_t - m_b$, is referred to as the ‘heavy’ charged Higgs case. In practice, because $m_b/m_t \approx 2\%$, the critical mass value against which the charged Higgs boson mass is compared to can be simplified to m_t .

In the Standard Model, top quarks decay almost exclusively to a W boson and a b quark. In the 2HDM or MSSM however, the top quark can also decay to a charged Higgs boson and a b quark if the charged Higgs boson is ‘light’. In pp collisions at the LHC, this production method via top quark decay, $t \rightarrow bH^\pm$, is the most important one if it is kinematically allowed. The top quarks are mostly produced as top-antitop pairs $t\bar{t}$ in pp collisions. The dominant production method is gluon fusion, while a smaller contribution arises from quark-antiquark annihilation [43]. Because the studied events involve a pair of top quarks, there are signal events with two H^\pm (HH) or with one H^\pm and one W (HW). These processes are illustrated in Fig. 2. The rate of the events with two charged Higgs bosons is however negligible, since from previous results it is known that $\mathcal{B}(t \rightarrow bH^\pm)$ is at most in the order of a few percent, and the HH events are in turn expected to be a few percent of this rate [27]. Single top quarks are also produced, but the event selection described in Section 6.2 makes the analysis sensitive only to the decays of single top quarks that are produced in association with a W boson.

In the heavy charged Higgs boson case $m_{H^\pm} \gtrsim m_t$ and the $t \rightarrow bH^\pm$ production method is not kinematically allowed. In this case the main production mechanism

becomes charged Higgs boson production *in association with top quark and bottom quarks*, $pp \rightarrow t(b)H^\pm + X$. Other production methods, such as quark-antiquark annihilation $q\bar{q}' \rightarrow H^\pm$, have suppressed rates. When the initial state contains no b quarks the production corresponds to the so called Higgs-strahlung ($gg \rightarrow t\bar{b}H^\pm$) and direct H^\pm production ($qq \rightarrow t\bar{b}H^\pm$) methods. The cross section of these processes can be calculated in the 4 flavour scheme (4FS) [44]. The opposite process where the initial state can contain b quarks is in the lowest order presented with the process $gb \rightarrow tH^\pm$ and the associated cross section can be calculated in the 5 flavour scheme (5FS) [44]. The next-to-leading order processes include the tree-level processes $gg \rightarrow t\bar{b}H^\pm$ and $qq \rightarrow t\bar{b}H^\pm$. Both the 4FS and 5FS processes are illustrated in Fig. 3. The schemes yield identical result when calculated to all orders of the perturbation theory, but because of the way of ordering, differences follow at finite orders. To take this into account in the practical next-to-leading order calculations, the results are combined with a procedure called *Santander matching* [44].

It is worth noting that the cross sections for the production methods in the heavy charged Higgs boson case are significantly smaller compared to that of the top quark decay process, dominant in the light charged Higgs boson case. This was the reason for starting the search for the charged Higgs bosons via the latter process. However, with the increasing energy and luminosity of the LHC, the analysis is becoming sensitive to the heavy charged Higgs bosons as well. The light charged Higgs boson has recently been excluded in large parts of parameter spaces in many models [27].

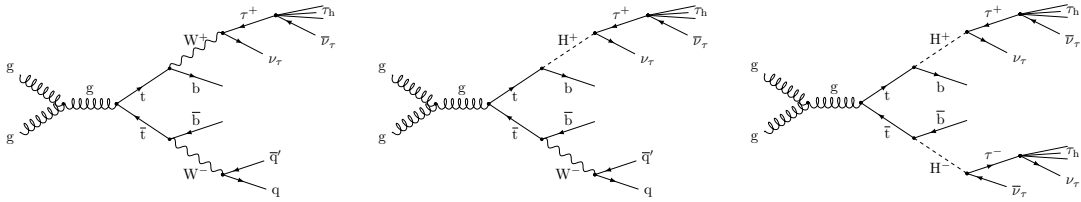


Figure 2: In the SM the top quarks decay to W bosons and b quarks (left). If a charged Higgs boson lighter than the top quark exists, either one (middle) or both top quarks (right) can decay into a charged Higgs boson.

2.4.2 Hadronic τ lepton decays

When a charged Higgs boson is produced, the decay follows promptly. Of the possible decays, the presented analysis focuses on the $H^\pm \rightarrow \tau\nu$ decay channel: light charged Higgs bosons decay mainly via this channel when $\tan\beta > 3$ [45]. This follows from the fact that the coupling strength of the Higgs bosons is proportional to the fermion mass and the $\tau\nu$ is the most massive final state that is kinematically allowed. In the heavy charged Higgs boson case the most favoured final state is $t\bar{b}$, while the $\tau\nu$ channel is subdominant. The $\tau\nu$ -channel is however the most sensitive

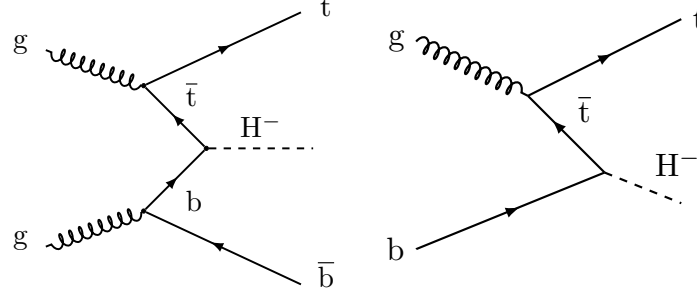


Figure 3: Charged Higgs boson produced directly from the pp collision in 4FS (*left*) and 5FS (*right*).

in this mass region as well, mostly because of the large multijet background that is difficult to suppress in the $t\bar{b}$ channel [27]. The final state of the $\tau\nu$ channel in the heavy charged Higgs boson case is similar to that of the light charged Higgs boson case.

The analysis presented in this work is constructed to be sensitive to the so called fully hadronic final state, in which there are no charged leptons. For this to happen, the τ leptons must decay to hadrons while the W bosons have to decay to pairs of quarks and anti-quarks. The probability of this to happen is about two thirds for both τ leptons and W bosons. The fully hadronic final state still contains τ neutrinos that are decay products of the charged Higgs bosons; these are the neutrinos produced in association with the τ lepton and in the hadronic decays of the τ leptons. The fully hadronic final state implies that the transverse mass of the $\tau\nu$ -system can be reconstructed by assuming that the transverse momentum of this neutrino pair is approximately equal to the E_T^{miss} of the event. The possibility of reconstructing the transverse mass is the primary reason for the high sensitivity of the analysis: it is currently the most sensitive analysis on the charged Higgs boson with $m_{H^\pm} < 500$ GeV [27].

Hadronically decaying τ leptons are often referred to as τ jets, since they in many ways resemble ‘normal’ hadronic jets. The most important τ lepton decay modes are summarised in Table 2. As can be seen from it, τ jets decay through a chain of vector or axial mesons to charged and neutral mesons or alternatively, they can decay directly to one charged meson. As a result of the boost effect, these mesons easily reach the outer layers of the CMS detector before decaying. They can thus be considered as stable particles even though their lifetimes are only in the order of $\sim 10^{-8}$. Because most τ jets contain either one or three charged particles, they are classified accordingly as 1-prong or 3-prong jets. These charged particles are often charged pions and charged kaons. However, as discussed in Section 6.2.2, the 3-prong decays are rejected in the analysis.

Table 2: Most important τ lepton decay modes and their branching ratios [29]. The uncertainty of the branching ratios is 0.1 percent units or smaller. The symbol h denotes here a charged pion or a charged kaon.

τ decay channel	\mathcal{B} (%)	$\Sigma\mathcal{B}$ (%)
leptonic modes		35.9
$\tau^- \rightarrow e + \bar{\nu}_e \nu_\tau$	17.9	
$\tau^- \rightarrow \mu + \bar{\nu}_\mu \nu_\tau$	17.4	
hadronic modes		
1-prong (excl. K^0)		48.1
$\tau^- \rightarrow h^- \nu_\tau$	11.6	
$\tau^- \rightarrow \rho^- \nu_\tau \rightarrow h^- \pi^0 \nu_\tau$	26.0	
$\tau^- \rightarrow a_1^- \nu_\tau \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	9.3	
$\tau^- \rightarrow h^- \nu_\tau + \geq 3\pi^0$	1.3	
3-prong (excl. K^0)		14.6
$\tau^- \rightarrow a_1^- \nu_\tau \rightarrow 2h^- h^+ \nu_\tau$	9.7	
$\tau^- \rightarrow 2h^- h^+ \nu_\tau + \geq 1\pi^0$	5.2	
5-prong (excl. K^0)		0.1
$\tau^- \rightarrow 3h^- 2h^+ \nu_\tau \geq 0\pi^0$	0.1	
hadronic modes (incl. K^0)		2.0
$\tau^- \rightarrow K_S^0 + X$	0.9	
$\tau^- \rightarrow K_L^0 + X$	1.1	

3 Large Hadron Collider and CMS experiment

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator that is located at the Franco-Swiss border, west of the city of Geneva. The accelerator is managed by CERN (European Organization for Nuclear Research) and it was built between the years 1998-2008. It went live for the first time on the 10th of September, 2008. Its main purpose is colliding protons or heavy ions in two beams traversing in opposite directions in a ring-shaped 27 km long pipe located underground. The two beams are collided at four *interaction points*, in which detectors are located to observe the collisions.

At the LHC, superconducting radio frequency cavities are used to accelerate passing particles. Before being directed to the actual 27 km long storage ring, the energies of the protons are raised using a series of *pre-accelerators*. Most of the space in the collider tunnels is occupied by different kinds of magnets. Superconducting dipole magnets producing a homogeneous magnetic field are needed to bend the trajectories of the particles to follow the shape of the accelerator. Quadrupole magnets focus the beams, while additional multipole magnets control and shape the beams. The collider tunnel contains two adjacent parallel beamlines, in which the beams travel, grouped into larger *bunches*. In the proton collisions of 2012, there were 1404 bunches per beam, when the *bunch spacing* was set to 50 ns. Each bunch contains approximately 10^{11} particles.

At the four interaction points, the bunches are set to cross so that some of the protons collide. Seven detectors have been constructed at these interaction points. Of these, ATLAS and CMS are largest and they are general purpose detectors, designed to observe various different phenomena. ALICE focuses on the study of quark–gluon plasma using heavy-ion collisions and LHCb investigates the apparent imbalance of matter and antimatter in the Universe. These four large experiments are located so that there is one at each of the four interaction points. The rest of the detectors are much smaller and are designed for very specialised research. They are situated alongside with the larger experiments. An overview of the experiments is shown in Fig. 4.

The LHC operated at 3.5 TeV per beam in 2010 and 2011 and at 4 TeV in 2012. In February 2013 it went into shutdown for upgrades to increase beam energy to 6.5 TeV per beam, with reopening planned for early 2015. This next collision run is referred to as the Run II. During Run II, the aim is to reach the design collision energy of 7 TeV per beam with 25 ns bunch intervals. However, the energies reached by 2013 already make the LHC the most powerful particle collider ever built. In this work the data gathered at the CMS in the year 2012 with the energy of 4 TeV per beam is used.

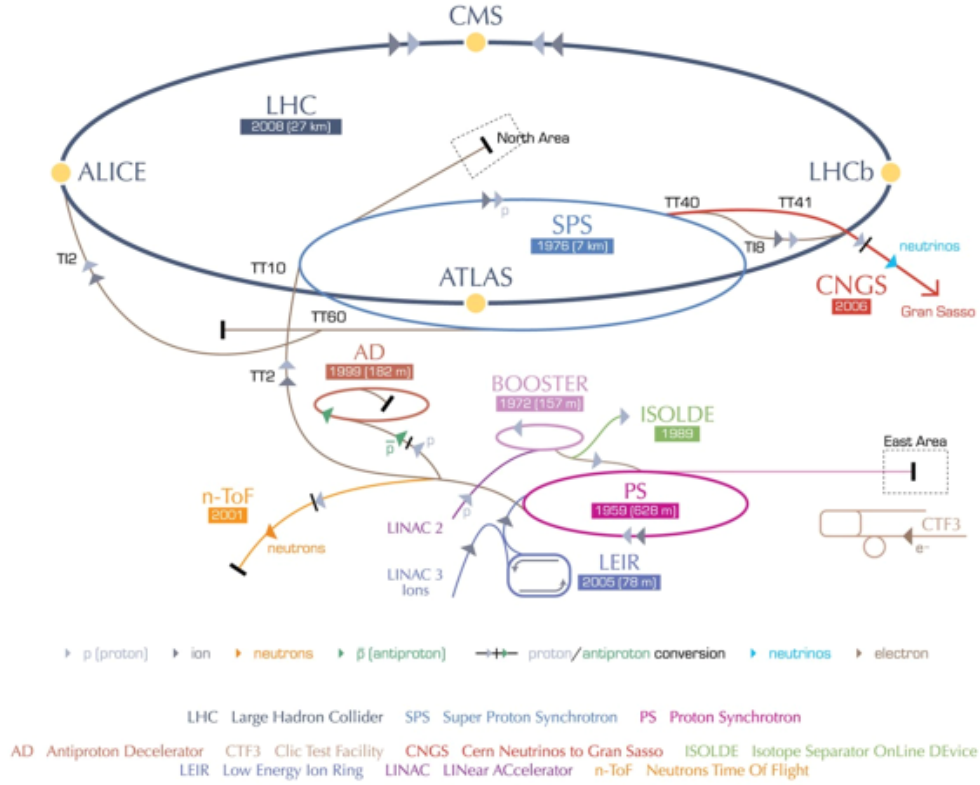


Figure 4: Overview of different accelerators and experiments at CERN [46].

3.2 CMS detector

The Compact Muon Solenoid (CMS) detector [47–49] is located beneath the village of Cessy in France, approximately 100 meters underground. CMS has the shape of a cylinder lying on its side and it is 21.6 metres long, 15 metres in diameter, and weighs about 12,500 tonnes. At the geometric center of the cylinder is the interaction point, into the beams are focused. The luminous region produced by the collisions is referred to as the *beam spot*.

CMS is built around a huge solenoid magnet, which gives the detector its name. The magnet is a cylindrical coil of superconducting cable that generates a homogeneous magnetic field of 4 Teslas to bend the trajectories of charged particles. Embedded inside the magnet is a compact tracking system, an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). Outside the solenoid magnet are the large muon detectors, which are inside the iron return yoke. A schematic overview of the detector is shown in Fig. 5, while the cross section is illustrated in Fig. 6.

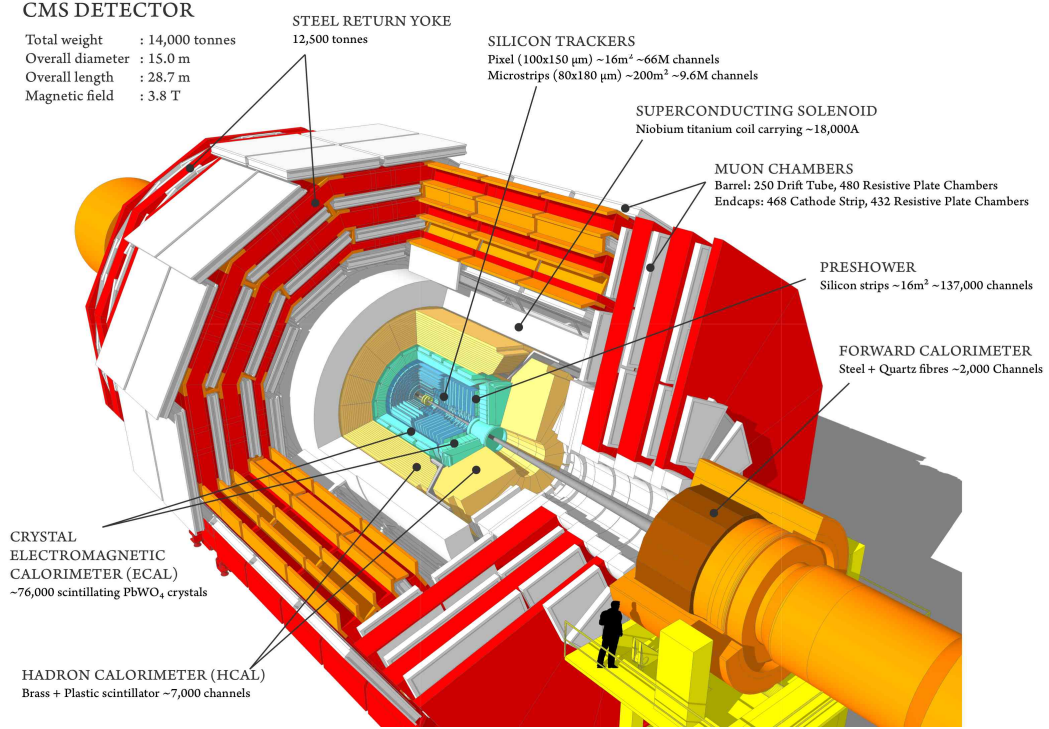


Figure 5: Sectional view of the CMS detector [50].

3.2.1 CMS by layers

The subsystems of the CMS detector are structured in a layer-like manner. They cover almost a full solid angle around the interaction point so that as many as possible of the produced particles are detected. The few particles that are scattered in directions where they cannot be detected (discussed in Ref. [47]) are said to be outside of the *acceptance*. All of the subdetectors consist of *barrel* parts in the side of the cylinder and *endcap* parts in the top and bottom of the cylinder. To detect particles with high pseudorapidities (discussed in Section 4.2), so called forward calorimeters are installed very close to the beam line. They are of particular interest when measuring jets in the forward direction, monitoring the instantaneous luminosity and to improve the \vec{E}_T^{miss} measurement (concepts discussed in Sections 4 and 5.1.6) [51].

The central tracker lies in the heart of the CMS detector and deals with the highest intensity of particles. Its purpose is to measure the tracks of charged particles that bend in the magnetic field. The innermost layers are made of silicon pixels and they are followed by silicon strips in the outer layers. As particles travel through pixels and strips they produce electric signals that are amplified and detected. The resolution of the pixel detector is approximately 10 μm .

The tracker can reconstruct the paths of high-energy muons, electrons and hadrons as well as record tracks coming from the decay of very short-lived particles. This

information is used to reconstruct the exact position of each collision (*primary vertex*). The positions of the decays of the short-lived particles (*secondary vertices*) can also be measured. Measuring the tracks of the particles allows also determining their momenta: the more curved the path in the magnetic field of the solenoid magnet is, the less momentum the particle has. The Lorentz force law allows the precise determination of the momenta of the particle when its radius of curvature is known.

The electromagnetic calorimeter (ECAL) is designed to measure the energies of electrons and photons. To do this, it uses scintillating PbWO_4 crystals which are very dense but optically clear, making them ideal for stopping high energy particles. The material has also fine granularity, it is radiation resistant and emits the light fast. When particles hit the crystals they produce light in proportion to the energies of the particles. Photomultiplier tubes that are attached to the crystals then detect these photon bursts. This kind of system measures accurately the energies of most charged particles and photons. Muons and charged hadrons however pass through the ECAL by depositing only a small fraction of their energy.

The hadron calorimeter (HCAL) measures the energy of hadrons. Additionally it provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos. Because of the large amount of hadronic jets in the LHC, HCAL has an important role in many analyses. The HCAL consists of layers of brass, which is sufficiently dense to cause showering and absorption of particles, interleaved with tiles of plastic scintillators with attached photodiodes. This combination was determined to allow a compact design while also having a maximal amount of absorbing material inside of the magnet coil. The usage of the brass absorber implies that not all of the energy of the particles is deposited in the active medium and detected directly. The correct energy is in turn calculated from the directly observed energies. The resolution of HCAL is 25 times coarser than that of ECAL.

The CMS solenoid magnet is the principal device around which the experiment is built. It is realised as a superconducting solenoid, capable of producing a 3.8 T magnetic field. The high-strength field bends the particle trajectories efficiently and combined with high-precision position measurements made in the tracker and the muon detectors, this allows accurate measurement of the momentum and charge of particles with high energies, even by the tracker. The large magnet also provides most of the structural support of the experiment.

The muon detector system and the return yoke are the only subsystems of the detector located outside of the magnet. The muon detector is one of the most important parts of the CMS detector. Since the mass of muons is large when compared to that of e.g. electrons, muons are not stopped by any of the calorimeters and travel to the edge of the CMS detector, where they are the only particles likely to cause a signal in the muon detector.

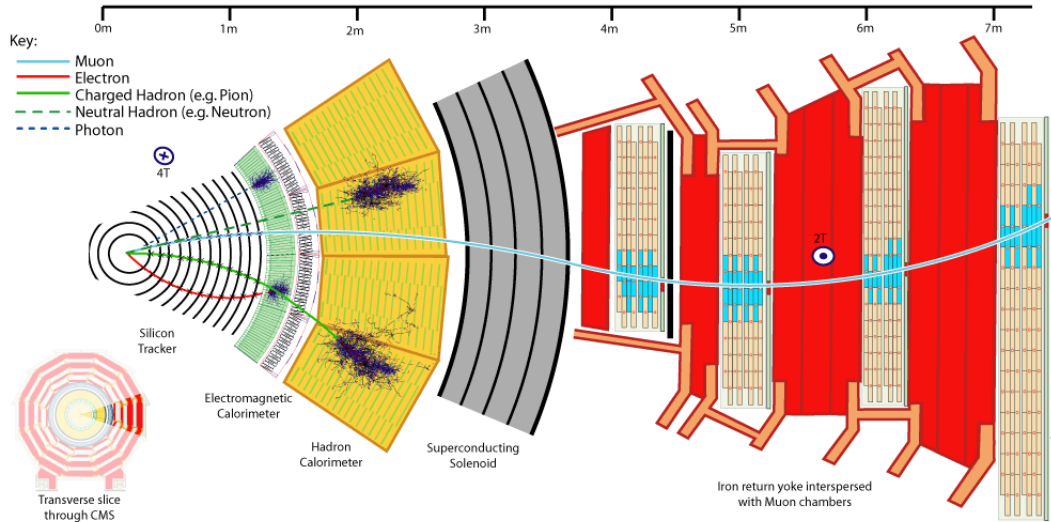


Figure 6: Cross section through CMS with example particle paths [46].

The detector consists of three types of subdetectors, all using gas as the active medium: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The choice of these detector technologies was driven by the very large surface to be covered and by the different radiation environments present. The DTs are used for precise trajectory measurements in the central barrel region. CSCs on the other hand are used in the end caps for the same purpose. The RPCs are installed in both the barrel and the end caps to provide a fast signal and good time resolution when a muon passes through the muon detector, allowing the identification of the bunch crossing that produced the muon.

The different detector types are arranged in layers which are called *muon stations*, and thick iron plates that act as the return yoke are placed between them. The purpose of the yoke is to make the magnetic field more homogeneous: it reduces the spreading of the magnetic field outside the solenoid, which then forces the field lines to be more parallel inside it. A homogeneous magnetic field is crucial because determining the momentum of particles relies on the measurement of their path curvature that depends on the field strength.

3.2.2 Trigger system

The amount of data that can be collected for each collision event is around 1 MB. At the design maximum luminosity of the LHC ($\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$), this means that one second of collisions corresponds to 1 PB (10^{15} B) of storage space. Because of this huge amount of data and the limitations in data storage capacity and rates, selecting events in real time as the collisions occur is necessary. This selection is said to happen *online*, as opposed to the *offline selection* performed later on in the analysis of the collision events. The online selection is also referred to as *triggering*.

and only the events passing the trigger requirements are stored on computer disk for subsequent analysis.

Triggers are sophisticated algorithms used to identify the minuscule fraction of collision events that contain interesting physics. For example, events not containing hard proton collisions are rejected. The trigger system of CMS is capable of reducing the event rate from ~ 40 MHz to ~ 100 Hz. This rate still requires around 100 MB of storage space each second.

The events are triggered hierarchically on several levels so that only the events passing the previous level are considered by the triggers of the next level. The trigger system of CMS can be divided into the following three levels:

- Level-1 (L1): hardware based trigger, short dead time and fixed latency. Selects events with large energy deposits in the calorimeters or hits in the muon chambers, reducing the rate by a factor of ~ 4000 .
- Level-2 (L2): software based trigger. Selects events based on a rudimentary analysis of regions of interest identified at Level-1 and uses calorimeter and muon system data.
- Level-3 (L3): software based trigger. Does a preliminary reconstruction of the entire event using full tracker data. The events that are selected by this trigger are stored for offline analysis.

The Level-2 and Level-3 triggers are collectively referred to as *High-level triggers* (HLT). Their combination and parametrisations are customised according to the physics analysis to be conducted with the data.

4 Experimental concepts regarding hadron collisions at the LHC

A noteworthy feature of the LHC collider is that it does not collide elementary particles, but protons, which consist of two up quarks and one down quark. This complicates the analysis of the collisions. The quarks and gluons making up a hadron are called *partons*, and each of them carries only a fraction of the total momentum of the hadron. These fractions depend on the energy at which the protons are probed. At high energies, not only the three quarks of the proton (*valence quarks*) have momenta, but also gluons and virtual quarks contribute significantly to the total momenta of the proton. Typically only two partons participate in the hard scattering. This is why it is not known which partons take part in the collision and what their momenta were. One can however overcome this challenge by analysing the collisions in terms of certain simplified variables. These quantities and some general concepts related to experimental particle physics are next considered.

4.1 Coordinate system

The CMS experiment uses a right handed coordinate system: the origin is located at the nominal interaction point, while the x axis points towards the geometrical center of the LHC ring, the y axis points upwards perpendicular to the accelerator plane and the z axis points in the counterclockwise direction of the ring. An important concept is the *transverse plane*, which is the x - y plane. Several variables are defined in this plane because of the fact that the total momentum of the initial state in the direction of the beam is unknown (discussion above). On the other hand, in the x - y plane the initial momenta of the beam particles is known to be small, since the beam particles do not move significantly in this plane. Since the principle of momentum conservation can be used in the transverse plane, it is an useful concept in many analyses.

4.2 Rapidity and angular variables

A quantity called *rapidity* is defined for a particle with energy E and momentum \vec{p} as

$$y \equiv \frac{1}{2} \ln \frac{E + |\vec{p}|}{E - |\vec{p}|}. \quad (31)$$

It is used more often in special relativity than in experimental particle physics, since it is related to the Lorentz factor γ via $\gamma = \cosh y$. In high energy collisions, the

particles can be assumed as massless i.e. $E \approx |p|$ and have significant momentum only in the z -direction:

$$y \approx \frac{1}{2} \ln \frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \equiv \eta, \quad (32)$$

which is the definition of *pseudorapidity* η , that is a commonly used quantity in collider physics.

The azimuthal angle ϕ is measured with respect to the x axis in the transverse plane, while the polar angle θ is measured from the z axis. With the latter definition, the pseudorapidity is also determined directly by the polar angle through the expression

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right], \quad (33)$$

but often preferred over it as a measure of the scattering angles of particles. The reason for this is that the particle production is approximately constant as a function of pseudorapidity, i.e. the number of particles ΔN with their direction within the interval $\Delta\eta$ is $\Delta N / \Delta\eta \approx \text{constant}$.

The *angular distance* between two vectors (e.g. particle momenta) can be defined using the above definitions as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (34)$$

Using the angular distance, *cones* can be defined around the tracks of the particles. These cones are then used in the event selection steps to measure the proximity of particles in e.g. matching particles (*matching cone*), requiring particle isolation (*isolation cone*) and identifying decay products (*signal cone*).

4.3 Track parameters

To describe the track of a charged particle travelling in a helix trajectory through the detector, five parameters are needed. These are the *curvature* κ i.e. the radius of helix track, the track azimuthal angle ϕ , the polar angle θ (or pseudorapidity η), the signed *transverse impact parameter* IP_T and the *longitudinal impact parameter* IP_z . The transverse (longitudinal) impact parameter is defined as the transverse (longitudinal) distance of the closest approach of the linearly extrapolated track to the primary vertex. For illustration, the track parameters in the transverse plane are presented in Fig. 7.

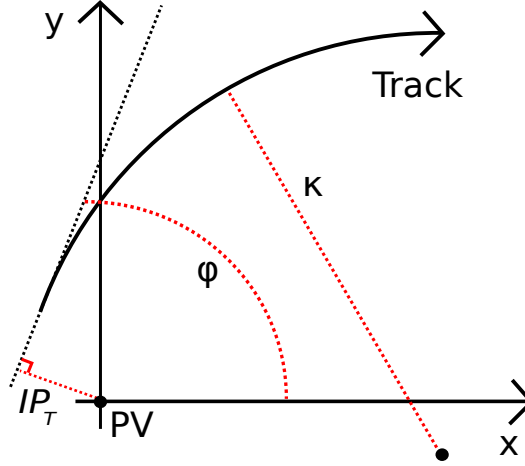


Figure 7: Track parameters in the transverse plane. The primary vertex (PV) is located at the origin.

4.4 Momentum and energy

The collision energies are often given in the center-of-mass frame using the square root of the so called *Mandelstam variable* $\sqrt{s} = |p_{\text{beam } 1} + p_{\text{beam } 2}|$. This refers only to the energy of the proton-proton collisions, not that of the 'true' CM energy of the partons. Since the initial momentum of the colliding particles in the transverse plane is known to be close to zero, the principle of momentum conservation is especially applicable in this plane. This explains the interest in the so called *transverse variables* defined in the transverse plane. Of these, the most noteworthy is the projection of the momentum vector to the transverse plane, the transverse momentum vector, that is denoted by \vec{p}_T . Its absolute value is referred to as transverse momentum p_T . The transverse energy E_T and the corresponding vector quantity \vec{E}_T are also often used concepts in high energy physics. These quantities are defined exactly like the transverse momenta and used interchangeably with them in many contexts, since it is often sufficient to treat the particles as massless, which implies that $E = |\vec{p}|$.

The energy of a particle is measured in the CMS using the electromagnetic and hadronic calorimeters. When the energy deposits in a given region of a calorimeter are summed up over a *calorimeter tower* consisting of many tile layers in depth, the total amount of deposits is a measure of a particle's energy.

Neutrinos interact very weakly with matter, which makes them non-detectable at the LHC. Their energy or momentum can thus not be measured. This fact manifests itself in that the energy in the pp collision does not appear to be conserved based on the observed particles. This *missing energy* can be calculated by taking the vector sum of the momenta of the detected particles and by comparing that to the initial momentum: the difference then reveals the vector sum of the momenta of the undetected particles.

From the discussion regarding the transverse plane it can be deduced that any net momentum in the transverse direction indicates *missing transverse energy* (MET) that can be calculated as

$$\vec{E}_T^{\text{miss}} = - \sum_i^{\text{all detected particles}} \vec{p}_{T,i}. \quad (35)$$

Here it is again assumed that the particles are massless and $E = |\vec{p}|$. It is worth noting that mismeasurements and still unknown particles can also contribute to the missing transverse energy.

4.5 Mass variables

Since the mass of a system can be understood as the difference between its energy and momentum, according to $m^2 = E^2 - p^2$, mass can be considered as an internal feature of the system. As unstable particles appear as peaks in the mass spectrum of their decay products, new particles can then be discovered by finding new peaks that are not explained by the background. As with the previously presented quantities, the *transverse mass* is of specific interest: it is calculated using only transverse momenta and energies. This allows the construction of a physically meaningful mass variable with the help of E_T^{miss} , even though undetected neutrinos are present in the event. The mass variable of most interest in this work is the transverse mass of the system consisting of a τ lepton and its neutrino. The transverse mass of this system is given by

$$m_T(\tau, \nu_\tau) = \sqrt{2E_T^{\tau \text{ jet}} E_T^{\text{miss}} (1 - \cos \Delta\phi(\tau \text{ jet}, \nu_\tau))}. \quad (36)$$

This expression neglects the mass of the τ lepton, which is a justified approximation since it is less than 2 GeV and the experimental resolution and uncertainty are thus larger in comparison. The transverse mass given by Eq. (36) is a key variable in the analysis, since the extraction of the limits on the production and decay of the charged Higgs bosons is based on the shape analysis performed on it.

4.6 Luminosity and cross section

To express the rate of collisions at a particular interaction point of the LHC, a quantity called *instantaneous luminosity* \mathcal{L} is used. Integrating the instantaneous luminosity over the time during which collision events are recorded gives the *integrated luminosity*

$$L = \int dt \mathcal{L}(t). \quad (37)$$

The probability of a given process X to occur is expressed as the *cross section* σ_X . Multiplying the integrated luminosity and the cross section then naturally gives the expected number of detected events involving process X as

$$N_X = \sigma_X L. \quad (38)$$

Because of this simple relation, the integrated luminosity is a useful quantity to characterise the performance of a particle accelerator: a higher integrated luminosity means there is more data available for analysis. In this work the integrated luminosity is expressed in units of fb^{-1} . The total integrated luminosity of the events recorded by the CMS in the year 2012 was 21.79 fb^{-1} and the maximal peak luminosity was 7.67 Hz/nb .

4.7 Pile-up

The high instantaneous luminosity achieved by the LHC means that multiple collisions occur per each bunch crossing, which is called *pile-up*. In 2012 with 8 TeV CM energy, the average number of collisions per bunch crossing was 20. More accurately, this sort of pile-up is referred to as *in-time pile-up*. *Out-of-time pile-up* occurs when particle signatures from consecutive bunch crossings are mixed. This phenomenon occurs because of the short 50 ns bunch interval and increases in the LHC Run II with the used 25 ns bunch interval.

Pile-up gives rise to many complications in triggering and data-analysis, since disentangling effects from different collisions is difficult. Additional charged particle tracks and energy deposits that result from pile-up are always present even though most of the proton-proton interactions are weak. Pile-up is also difficult to simulate correctly. In practice, the discrepancy is corrected by experimental weighting factors to make the simulated event distributions match with real data.

5 Event reconstruction and simulation

5.1 Event reconstruction

Event reconstruction refers to the process of converting the raw output of the CMS detector, voltage readouts, to the corresponding particles causing these signals. The reconstruction combines the information from different subdetectors to construct the particle objects. These objects can be for example tracks of charged particles, and they are then used in the actual physics analyses.

The main tool in the reconstruction process is the Particle-Flow (PF) algorithm [52–55], that combines all the information from the subdetectors with the goal of identifying all stable particles belonging to an event and estimating their momenta. The usage of all gathered data in this way allows extracting information that could not be obtained if only the individual subdetectors were to be considered. The algorithm first divides the collision event into *blocks*, which are made up of signals that are likely to be caused by one or few physical particles. The algorithm then searches for signals in the following order, and upon identification, removes them from the signal blocks:

1. Muons: tracker and muon chamber information
2. Electrons: tracker and ECAL information
3. Charged hadrons: tracker and HCAL information
4. Photons and neutral hadrons: the excess in the charged hadron deposits if they are larger than that of one pion and the remaining blocks (photons in ECAL blocks and neutral hadrons in HCAL blocks).

The basic principle behind the PF algorithm is illustrated in Fig. 8. In the following subsections, the identification processes used in the presented analysis are considered in more detail.

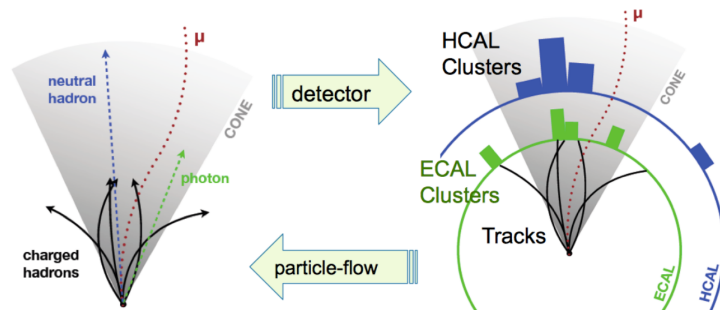


Figure 8: Event reconstruction using the Particle-Flow approach [56].

5.1.1 Tracking

As the charged particles traverse through the tracker and the muon system, they leave a set of so called *hits* behind them. The hits refer to the points of impact on the detector layers. The trajectories are bent by the strong magnetic field of the CMS detector into the shape of a helix. *Track reconstruction* or *track finding* refers to the task of converting these hits into a set of particle tracks. This is done using the pattern recognition algorithm called *combinatorial track finder* (CTF) [57], which is based on a global Kalman filter [58]. An initial estimate for the track, the *track seed*, is obtained by searching for three pixel hits that could form a track (or two hits and a constraint from the position of the nominal interaction point in the transverse plane if a third hit is not found) [59]. The track is then expanded by adding more compatible hits and by simultaneously removing them from the list of unclassified hits, which makes the algorithm computationally lighter. At the end of each iteration round, only the tracks fulfilling the *high purity* criterion are kept [60].

The efficiency of charged-particle track reconstruction has been measured to be 99% for isolated muons [61]. *Track fitting* on the other hand refers to the process of calculating physical quantities, such as momentum, from these tracks. When the performance of the fitting has been studied using J/ψ decays, in which the mass peak positions are extracted from data and simulation, agreements within 1 MeV have been obtained after calibration [62].

5.1.2 Primary vertices

The spatial position where the initial interaction of the collision takes place is called the *primary vertex*. When pile-up occurs, an event can contain multiple primary vertices. They are reconstructed using the tracks of charged particles with the deterministic annealing method [63,64], which has proven to perform well in a high-pile-up environment. The primary vertex with the largest $\sum_i^{\text{tracks}} p_{T,i}^2$ is selected as the signal primary vertex. The vertex is required to fulfil the following requirements:

- $N_{\text{d.o.f.}} > 4$,
- $|\rho| < 2 \text{ cm}$, and
- $|z| < 24 \text{ cm}$.

5.1.3 Muons

In the CMS detector, muons pass through the detector material without depositing large amounts of energy until they reach the muon chambers. The effect in which other types of particles reach the muon chambers (excluding neutrinos that rarely interact with the detector at all) has a small probability and is referred to as a *punch-through*. The search for muon candidates is based on combining the information given by the tracks in the inner tracker and in the muon detector [49,65] and can be done with two methods [66]:

- A *global muon candidate* is found if the tracks in the inner tracker can be matched with the tracks of the muon detector and a *global muon track* can be fitted to the hits.
- A *tracker muon candidate* is found if the tracks in the inner tracker can be extrapolated to at least one short track in the muon detector, referred to as a *muon segment*.

Muons are then identified from these candidates using the *tight muon selection*, which requires that:

- Muon is reconstructed as a global muon, and passes the Particle-Flow identification,
- Number of tracker layers with hits in the global muon track > 5 ,
- At least 1 hit in pixel tracker,
- At least 1 muon chamber hit in the global muon track,
- For the fit of the global muon track $\chi^2/N_{\text{d.o.f.}} < 10$,
- At least 2 muon segments matched to the global muon,
- Transverse impact parameter with respect to the beam spot $IP_T < 0.2 \text{ cm}$ (Section 4.3), and
- Longitudinal impact parameter with respect to the primary vertex $IP_z < 5 \text{ mm}$ (Section 4.3).

5.1.4 Electrons

In the electron identification process, the Gaussian Sum Filter (GSF) method [67,68] is used to select the electron candidates. It relies on the following principle: as electrons move through the tracker, they emit Brehmstrahlung photons. These photons cause hits in the ECAL, that are spread in ϕ due to the bending of the electron trajectory in the magnetic field. The GSF method then searches for superclusters of these hits and propagates their energy-weighted mean positions back to the tracker, which gives track seeds for the electron candidates. The GSF algorithm also takes into account the energy losses of the electrons and the resulting changes in their trajectory caused by the emission of photons. The actual identification is then performed with a simple cut-based method, which relies on quantifying the shape of the photon shower and on track-cluster matching requirements. These cuts and the relevant parameter values are discussed in more detail in Ref. [69].

5.1.5 Photons

As photons are massless and neutral particles they are not bent by the magnetic field and do not interact with the tracker. However, they leave energy deposits in the ECAL. Their detector fingerprint is thus a deposit in the ECAL with no associated particle track in the tracker. Still, an additional phenomenon has to be taken into account, namely the effect of *photon conversion* in which photons interact with the tracker and produce electron-positron pairs [70]. The probability for this to happen is high especially for very energetic photons ($\sim 70\%$) and the signature of the conversion is a pair of oppositely charged tracks with a small opening angle.

The photon identification [71] is based on the construction of ECAL superclusters, which consist of clusters of basic ECAL clusters with some spread in ϕ . This method can be used to determine if a photon was converted or not by introducing a discriminant

$$r_9 \equiv \frac{E_{3 \times 3}}{E_{\text{supercluster}}}, \quad (39)$$

where $E_{3 \times 3}$ is the energy deposited in a 3×3 array of crystals centred around the crystal with the highest energy deposit, referred to as the *seed crystal*. $E_{\text{supercluster}}$ on the other hand is the energy deposited in the supercluster. If $r_9 > 0.94$ (0.95) in the ECAL barrel (endcap), the photon is considered unconverted and the energy of the 5×5 array around the seed crystal is used as the photon energy. Otherwise, the photon is considered to be converted and the energy of the supercluster is used instead.

5.1.6 Jets

Quarks and gluons cannot exist in free form, i.e. without being bound to other quarks and gluons, because of so called *QCD confinement* [28]. A *jet* is a collimated spray of particles that forms when a hadron fragments: before the separating partons are directly detected, the attractive potential between them grows until new quark-antiquark pairs and gluons are produced from the binding energy. These particles form a jet in a process called *hadronisation*. In jet reconstruction, the tracks, energy deposits and particle objects are grouped to a jet and then the momentum of the constructed jet is calculated. In this analysis, jets are reconstructed by clustering all particles reconstructed by the PF-algorithm with the so called *anti- k_T algorithm* [72].

The principle of the anti- k_T jet clustering algorithm is the following: by defining $d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \Delta_{ij}^2 / R^2$ and $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ for two particles i and j , where y denotes the rapidity and R is the jet radius parameter, the algorithm starts with particle i and loops over j particles until it finds the smallest d_{ij} . If this $d_{ij} < p_{T,i}^{-2}$, the particles are merged into the same jet. The looping continues until no particle fulfils the previous condition. At this stage the jet is considered as complete and the algorithm proceeds to cluster the next jet. In the presented analysis, the jet radius parameter R is set to 0.5. The anti- k_T algorithm is both *infrared safe* and *collinear safe*, which are characteristics of a good jet clustering algorithm. The former means that observables remain constant if a parton emits a soft parton, while the latter requires that this holds when a parton splits into two partons travelling in the same direction.

After the jet reconstruction, the jet momenta must still be corrected with multiplicative factors, mainly due to the non-uniform detector response in the (η, p_T) -space, electronics noise, pile-up and the observed discrepancy between simulations and data. In the presented analysis, the jets are corrected with the *L1FastJet*, *L2Relative*, *L3Absolute* and *L2L3Residual* (only for data) jet energy scale corrections [73].

5.1.7 B tagging

Based on the discussion in Section 2.4, identifying jets that originate from b quarks, referred to as *b jets*, is essential in the selection of signal events with the $t\bar{t}$ -like topology. This identification is called *b tagging*, and because of its importance in many analyses, several algorithms have been designed for the task. They are based on the fact that b jets differ drastically from jets with other flavours: b jets have longer lifetimes, higher mass, harder fragmentation and a larger fraction of semileptonic decays.

In the presented analysis, the *combined secondary vertex* (CSV) method [74] is used for b tagging. The output of the algorithm is a discriminator value, which

represents the probability that the investigated jet originates from a b quark. Three different working points are established for the discriminator: *Loose*, *Medium* and *Tight*, which correspond approximately to the probabilities of 10%, 1%, and 0.1%, for jets originating from gluons or other quarks to be misidentified as b jets. Of these, the Tight working point is used in the analysis, because it is most efficient for choosing the $t\bar{t}$ -like topology and it suppresses greatly the $W + \text{jets}$, Z/γ^* and QCD multijet background events.

5.1.8 Identification of hadronic τ decays

Identifying τ leptons is especially important for the presented analysis. As τ leptons decay very quickly to electrons or muons, it is difficult to know if these particles originate from a τ or the initial hard interaction. This is why events containing isolated electrons or muons are rejected in the analysis and only the so called fully hadronic τ decays which result in so called τ jets (discussed in Section 2.4.2) are considered. When a τ lepton decays hadronically, the jet has typically a low track multiplicity and consists of one or three charged hadrons along with possible neutral hadrons. These characteristics of the τ jets can be used to distinguish the genuine τ leptons from the large background of QCD jets.

In this work, τ jets are identified using the *Hadron-Plus-Strips* (HPS) algorithm [75,76]. The HPS algorithm utilises the Particle-Flow algorithm: it starts with a PF jet and searches for visible τ lepton decay products from the jet constituents. The decays considered in the search are $\tau \rightarrow h^- \nu_\tau$, $\tau \rightarrow h^- \pi^0 \nu_\tau$ and $\tau \rightarrow h^- h^+ h^- \nu_\tau$, where h denotes a hadron. The neutral pions decay further to a pair of photons ($\mathcal{B}(\pi^0 \rightarrow \gamma\gamma) \approx 99\%$), which in turn convert to electron-positron pairs in the CMS tracker material via the photon conversion process discussed in Section 5.1.5. This typically broadens the ECAL deposit signatures in the azimuthal direction for the pions as the electron tracks are bent by the magnetic field. This is taken into account by combining the PF photons in *strips* in the azimuthal direction.

To reconstruct the τ jet candidates, the strips are then combined with the charged hadrons so that both are required to lie within a narrow cone of $\Delta R < 2.8/p_T^\tau$. Here, p_T^τ is calculated by summing the transverse momenta of the reconstructed charged hadrons and strips. If multiple decay modes are reconstructed, the mode with the smallest $\sum E_T$ of the jet constituents *not* associated with the τ jet is selected.

The majority of the τ jet candidates reconstructed so far are hadronic jets that happen to have a similar signature as that of a genuine τ decay. To reject these jets, the fact that genuine τ jets are significantly narrower than hadronic jets is used. In practice, this is implemented by requiring isolation of the τ jet candidate. First, an isolation cone with the size $0.3 < \Delta R < 0.5$ is defined around the leading track of the jet and all photons with $E_T^{\text{miss}} > 0.5$ GeV inside the cone are considered. Charged hadrons inside the isolation cone are considered if they fulfil the following requirements:

- Transverse impact parameter with respect to the beam spot $IP_T < 300 \mu\text{m}$ (Section 4.3),
- At least 2 hits in the pixel detector,
- Total of at least 3 hits in the central tracker, and
- $p_T > 1.0 \text{ GeV}$ (*Loose*), $p_T > 0.8 \text{ GeV}$ (*Medium*) or $p_T > 0.5 \text{ GeV}$ (*Tight*).

The choice of the transverse momentum threshold (*working point*) depends on the chosen *isolation scenario*.

The photons and hadrons satisfying the above criteria are then used to construct the discriminator quantity:

$$E_T^{\text{isolation}} = \sum_{IP_z < 2 \text{ mm}} p_T^{\text{charged hadrons}} + \min \left(0, \sum E_T^{\text{photons}} - \frac{1}{2} \times \sum_{IP_z < 2 \text{ mm}} p_T^{\text{charged hadrons}} \right). \quad (40)$$

The second term contains the so called $\Delta\beta$ correction for particles that are likely to come from a pile-up vertex with $IP_z < 2 \text{ mm}$.

Depending on the isolation scenario, the threshold for the isolation requirement is either $E_T^{\text{isolation}} < 2.0 \text{ GeV}$ (*Loose*), $E_T^{\text{isolation}} < 1.0 \text{ GeV}$ (*Medium*) or $E_T^{\text{isolation}} < 0.8 \text{ GeV}$ (*Tight*). In the presented analysis, the Tight isolation scenario is used.

5.1.9 Missing transverse energy

In the presented analysis, the calculation of the missing transverse energy is based on the Particle-Flow approach: the E_T^{miss} is taken to be the negative vector sum of the transverse momentum of all the particles reconstructed by the PF algorithm. This takes into account all detector information and it is referred to as the PF E_T^{miss} .

Missing transverse energy mismeasurement is caused by numerous reasons which include particle momentum mismeasurement, particle misidentification, detector malfunctions, particles impinging on poorly instrumented regions of the detector and cosmic-ray particles [77]. To mention a few other reasons, also the non-linear response and the minimum energy thresholds of the calorimeters, underlying event activity as well as pile-up detector noise cause bias [78].

To bring the measured E_T^{miss} closer to its true value, E_T^{miss} contributions coming from τ jets, unclustered energy, along with isolated high p_T photons, electrons and muons are considered separately and their energy scales are corrected using a

two-step scheme [78]. In the so called *Type-I* corrections, energy scale corrections are applied to all jets with corrected $p_T > 10$ GeV before the PF E_T^{miss} is calculated. To further correct the energies of the remaining soft jets falling below this p_T threshold and the energy deposits not clustered in any jet, the *Type-II* corrections can be applied to the unclustered energy. These corrections have been obtained by measuring $Z \rightarrow ee$ events [78]. In the presented analysis, only Type-I corrections are used and the jet which corresponds to the selected τ jet in the event is ignored in the jet energy correction, because its energy response is assumed to be so close to truth.

5.2 Event simulation

Simulated event samples are mostly used in the analysis for validating the methods. Additionally, in the QCD multijet background measurement, simulated EWK+ $t\bar{t}$ samples are subtracted from data and in the measurement of the EWK+ $t\bar{t}$ background with genuine τ leptons, simulated τ leptons are embedded to the data. Simulation has been used to estimate the EWK+ $t\bar{t}$ background with misidentified τ leptons as well but with the data-driven measurement method presented in this work, the simulation dependency of this background measurement is reduced significantly.

Simulated charged Higgs boson samples are also used in the analysis to predict what a signal could look like. In the light charged Higgs boson case the used mass values are $M_{H^\pm} = 80, 90, 100, 120, 140, 150, 155, 160$ GeV, while in the heavy charged Higgs boson case the used masses are $M_{H^\pm} = 180, 190, 200, 220, 250, 300, 400, 500, 600$ GeV. These mass values are chosen because they are well spread over the studied two mass ranges. Many of the results of this work, including the limits in Section 9.2, are shown for these mass values.

The simulated processes and the employed software are represented in Table 3. TAUOLA [79] is used to simulate all the τ decays, since it includes the τ polarisation effects. This is crucial for the analysis, since polarisation is used in suppressing the background with genuine τ leptons originating from W boson decays. The response of the CMS detector is simulated with the GEANT4 [80] toolkit, using the commonly maintained detector specifications.

Table 3: Overview of the simulated processes and the employed software.

Simulation type	Process	Software
Background events	$t\bar{t}$	MadGraph 5 [81]
	W + jets	MadGraph 5
	$Z/\gamma^* + \text{jets}$	MadGraph 5
	Diboson (WW, WZ, ZZ)	Pythia 6 [82]
	Single top	Powheg [83]
Signal events	HH	Pythia 6
	HW	Pythia 6
Particle processes	τ decays	Tauola
	Fragmentation and hadronisation	Pythia 6
Detector simulation	All	Geant 4

6 Event selection

6.1 Online selection

The online event selection is done using triggers. The trigger system of the CMS is discussed in Section 3.2.2. Because of the topology of the charged Higgs boson production, the triggers can require any combination of the following: a single τ lepton, multiple jets or missing transverse energy. For the low luminosity conditions of the collisions recorded in the year 2012, a single τ lepton and E_T^{miss} trigger is selected, since it allows a low τ p_T trigger threshold while still having an sufficiently high event rate suppression.

The requirements for an event to pass the different trigger levels are:

- Level-1: $E_T^{\text{miss}} > 36 \text{ GeV}$ or $E_T^{\text{miss}} > 40 \text{ GeV}$ (however, some events passing the Level-1 $E_T^{\text{miss}} > 36 \text{ GeV}$ trigger are prescaled [49], so only the events with Level-1 $E_T^{\text{miss}} > 40 \text{ GeV}$ are selected offline).
- High-level trigger E_T^{miss} part:
 - Level-2 $E_T^{\text{miss}} > 70 \text{ GeV}$.
- High-level trigger τ part:
 - Level-2 jet with regional reconstruction in the Level-1 jet and τ object directions, with towers taken in a cone of 0.8, $E_T > 25 \text{ GeV}$, $|\eta| < 3$,
 - Level-3: reconstructed τ object with the fixed cone PF-algorithm, τ $p_T > 35 \text{ GeV}$, $|\eta| < 2.5$,
 - Level-3: leading track finding, leading track $p_T > 20 \text{ GeV}$,
 - Level-3: number of charged hadrons in the signal cone < 3 , and
 - Level-3: loose charged hadron isolation with charged hadron $p_T > 6 \text{ GeV}$, matching cone of 0.2, signal cone of 0.2, and isolation cone of 0.5.

The events passing these trigger requirements are still dominated by the QCD multijet events. This allows the data-driven measurement of the QCD multijet background, described in Section 7.1. The background events are further suppressed in the offline selection steps.

6.2 Offline selection

6.2.1 Data quality criteria

The technical quality of the data has to be ensured before further analysis on it is performed. This is controlled by checking that the following conditions are met:

- Low noise from the calorimeters [84],
- High fraction of hits with the associated reconstructed track (*track purity*). The high purity condition [60] is required for at least 20 % of the tracks for events with more than 10 charged-particle tracks, and
- Requirements on the position of the primary vertex given in Section 5.1.2.

Otherwise the event is rejected.

6.2.2 Main selection cuts

The main event selection steps of the offline analysis can be divided into two parts: the selection and identification of the τ jet and the selection of the correct event topology. An example of a recorded signal candidate event fulfilling these selection requirements is shown in Fig. 9.

Selection and identification of the τ jet The τ objects passing the following criteria are referred to as identified τ jets:

- Offline τ jet candidate matches with $\Delta R < 0.4$ with a HLT τ object,
- τ -jet decay mode is successfully reconstructed with the HPS algorithm,
- τ -jet p_T threshold ($p_T > 41 \text{ GeV}/c$),
- τ -jet η limits ($|\eta| < 2.1$),
- p_T threshold for the leading electrically charged particle (leading track $p_T > 20 \text{ GeV}/c$),
- Rejection of electrons faking τ jets,
- Rejection of muons faking τ jets,
- τ jet is isolated,

- τ jet contains one charged particle track, and
- τ jet passes cut on the τ polarisation with $R_\tau = p^{\text{ldg. charged particle}} / p^{\tau \text{ jet}} > 0.7$ (discussed in Section 6.2.3).

The most discriminative criterion is the isolation of the τ leptons. The isolation is applied on both neutral and charged particles. The 3-prong τ decays are rejected, because they increase the QCD multijet background too much when compared to the gain in signal acceptance.

Selection of the event topology In addition to having an identified τ jet, the events are required to have at least 3 more jets passing the following criteria:

- Jet p_T threshold ($p_T > 30 \text{ GeV}/c$),
- Jet η limits ($|\eta| < 2.4$), and
- One of the jets must originate from a b quark.

The last requirement is fundamental because it is used to select $t\bar{t}$ -like event topologies, which are characterised by the presence of two jets in the final state, originating from b quark hadronisation.

Events with isolated electrons or muons are excluded to select the fully hadronic final state. This also improves the discriminating power of the transverse mass reconstructed from the selected τ jet and the E_T^{miss} . An event is rejected when the identified electrons (muons) have $p_T > 15 \text{ GeV}$ ($p_T > 10 \text{ GeV}$).

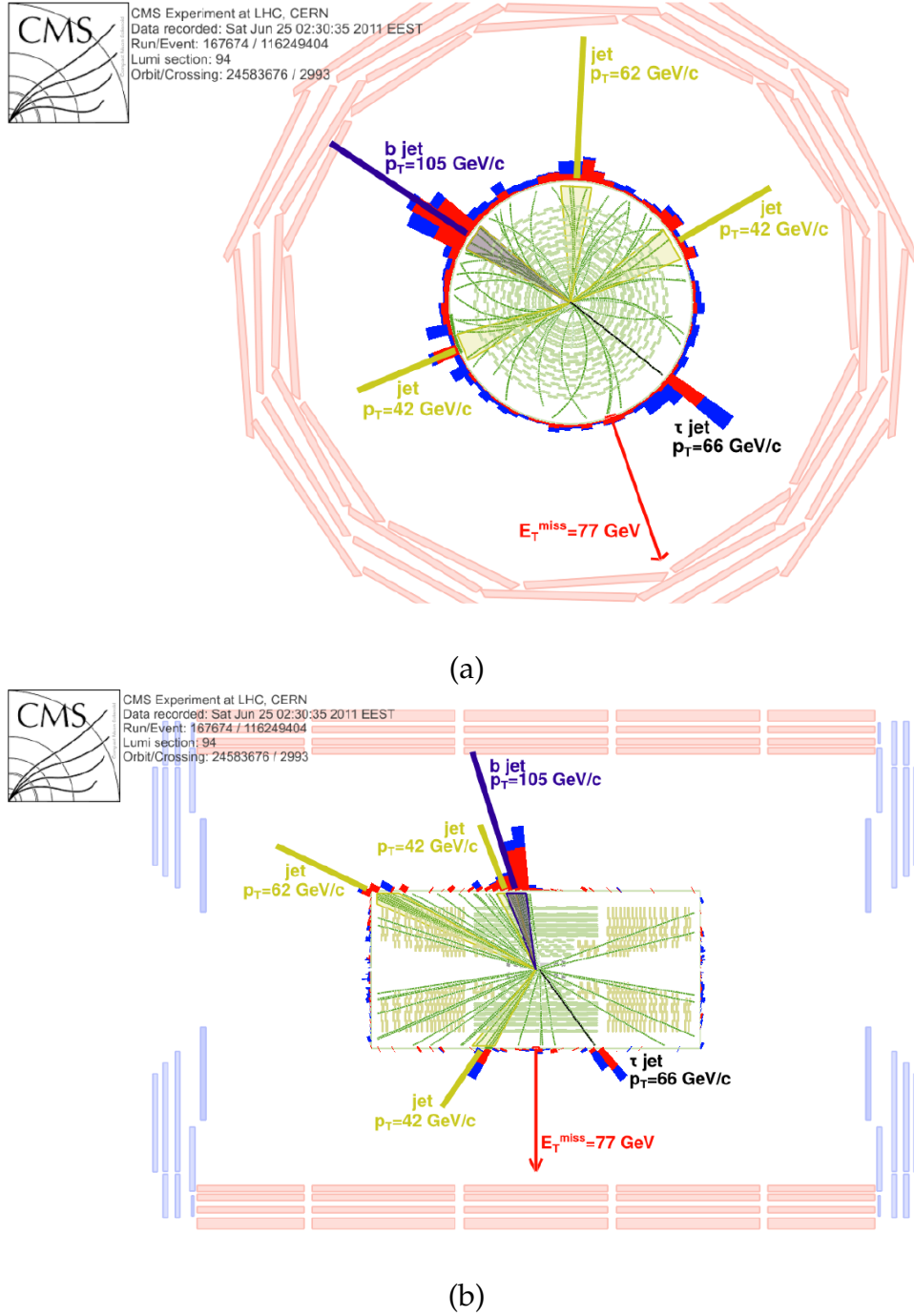


Figure 9: Example of a recorded signal candidate event fulfilling the selection requirements in the r - ϕ (a) and r - z (b) views [85]. Charged particle tracks are shown in green lines and ECAL (HCAL) energy deposits with red (blue) bars. The τ jet is shown as a black line, the light-quark jets in yellow, the b jet in purple and E_T^{miss} is represented by a red arrow. The transverse mass, reconstructed from the τ jet and E_T^{miss} , is $m_T \approx 40$ GeV.

6.2.3 Tau polarisation cut

Some of the irreducible background resulting from a W boson decaying to a τ lepton and a neutrino can be suppressed in the analysis by exploiting the different properties of the W^\pm and H^\pm bosons. Of these differences, spin is the most useful in the analysis and therefore a cut based on it is introduced in order to suppress the irreducible background following from W^\pm bosons [86,87]. The spin is taken into account by relying on the fact that the W^\pm bosons have spin 1 while the H^\pm bosons have spin 0. Because neutrinos are experimentally observed to be always left-handed, the τ leptons produced in the decays of these bosons have opposite helicities. The helicity in turn affects the momentum distributions of the decay products, which can be observed and exploited in the analysis. This effect is referred to as the *helicity correlations*.

As discussed in Section 6.2.2, only the 1-prong τ jets are selected for the analysis. They decay mainly via the following processes:

$$\begin{aligned}\tau^\pm &\rightarrow \pi^\pm \nu_\tau, \\ \tau^\pm &\rightarrow \rho^\pm \nu_\tau \rightarrow \pi^\pm \pi^0 \nu_\tau, \text{ and} \\ \tau^\pm &\rightarrow a_1^\pm \nu_\tau \rightarrow \pi^\pm \pi^0 \pi^0 \nu_\tau.\end{aligned}$$

These modes cover 97.4% of the 1-prong decays when the processes involving neutral kaons K^0 are neglected. The effect of the remaining modes that, among others, involve more than two neutral pions are also negligible. In the helicity correlations one can thus focus on the above decay modes, which are next discussed separately.

The differential decay width of $\tau^\pm \rightarrow \pi^\pm \nu_\tau$ is given by

$$\frac{1}{\Gamma_\tau} \frac{d\Gamma_\pi}{d\cos\theta} = \frac{1}{2} (1 + P_\tau \cos\theta), \quad (41)$$

where P_τ is +1 (-1) when the τ lepton is a decay product of the H^\pm (W^\pm) boson. If the direction of the spin quantisation axis is chosen to be the same as the flying direction of the τ lepton in the laboratory coordinates, θ is the angle between the flight direction of the charged pion and the spin quantisation axis of the τ lepton in its rest frame. The above equation suggests that when the τ lepton is a decay product of a charged Higgs boson, the charged pion will preferably be produced in the direction of the τ lepton. The opposite direction is instead favoured if the τ lepton is coming from a W boson. This implies that the τ jets are typically more energetic when they originate from charged Higgs bosons.

In the processes involving the massive vector mesons ρ^\pm and a_1^\pm , the situation is more complex, as these particles can have both longitudinal (L) and transverse (T) polarisation modes. The decay widths related to these two modes are given by

$$\frac{1}{\Gamma_v} \frac{d\Gamma_{v,L}}{d\cos\theta} = \frac{1}{2} \frac{M_\tau^2}{M_\tau^2 + M_v^2} (1 + P_\tau \cos\theta), \quad (42)$$

$$\frac{1}{\Gamma_v} \frac{d\Gamma_{v,T}}{d\cos\theta} = \frac{1}{2} \frac{M_v^2}{M_\tau^2 + M_v^2} (1 - P_\tau \cos\theta), \quad (43)$$

where θ is now defined as the angle between the flight directions of the vector meson and the τ lepton in the rest frame of the latter and v refers to the considered vector meson (ρ^\pm or a_1^\pm).

The angle θ in the above equations can be expressed as

$$\cos\theta = \frac{2M_\tau^2}{M_\tau^2 + M_{\pi,v}^2} \times \frac{p_{\pi^\pm}}{p_{\tau \text{ jet}}} - 1. \quad (44)$$

From Eqs. (42)-(45) it can be seen that the τ jets are more energetic when they originate from H^\pm bosons as opposed to W^\pm bosons, if the τ lepton decays to a vector meson with longitudinal polarisation. In the case of a vector meson with transverse polarisation, the W^\pm bosons produce more energetic τ jets.

When the vector meson has longitudinal polarisation, uneven momentum distribution between its decay products π^\pm and π^0 is favoured. On the other hand, in the transverse case the momentum tends to be shared evenly between the decay products.

The differential decay width distributions of ρ^\pm and a_1^\pm as a function of the energy fraction X' carried by the charged pion, i.e.

$$X' = E_{\pi^\pm} / E_{\rho^\pm, a_1^\pm}, \quad (45)$$

are shown in Fig. 10. The distributions are clearly peaked near $X' \approx 0$ and $X' \approx 1$ for the longitudinal ρ^\pm and a_1^\pm . For the transverse states, the peaks appear in the middle. In the $\tau^+ \rightarrow \pi^+ \nu_\tau$ decay, the charged pion is the only particle in the jet, and thus the decay would appear as a δ -function at $X' = 1$ on this plot.

The above inference implies that imposing a p_T requirement on the τ jet enhances the longitudinal polarisation states for $H^\pm \rightarrow \tau \nu$ decays, and the transverse states for $W^\pm \rightarrow \tau \nu$ decays. This separates the τ jets from H^\pm and W^\pm boson decays and allows the suppression of the W^\pm boson events with the requirement that

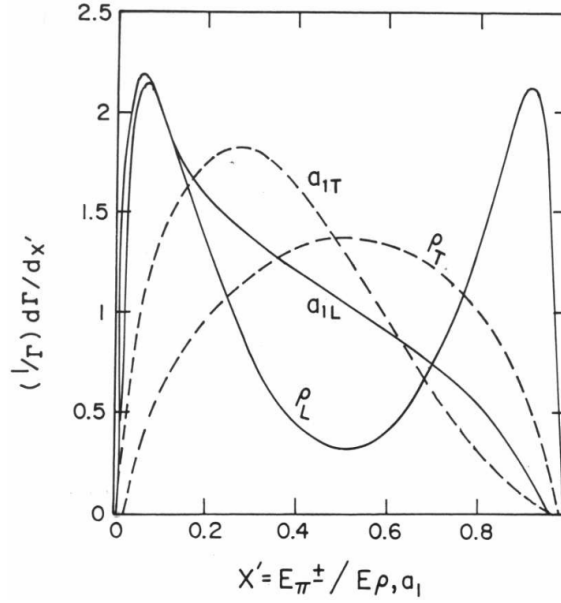


Figure 10: Distributions of the normalised τ lepton decay widths via $\rho_{L,T}^{\pm} \rightarrow \pi^{\pm} \pi^0$ and $a_{1L,T}^{\pm} \rightarrow \pi^{\pm} \pi^0 \pi^0$ as a function of the energy fraction carried by the charged pion [87]. The distributions are calculated in the rest frame of the τ lepton, but are similar in the laboratory frame.

the charged pion carries most of the energy of the visible τ jet. In terms of a discriminator, this requirement can be formulated by defining the quantity

$$R_{\tau} \equiv \frac{p_{\text{leading charged particle}}}{p_{\tau \text{ jet}}}, \quad (46)$$

and by setting a lower threshold (0.7 in the analysis [86]) for it in the event selection. This discriminator makes the assumption that the charged pion produces the *leading track* of the jet, that is, it is assumed to have the highest p_T of the particles associated with the jet.

6.2.4 Angular cuts

To suppress the QCD multijet background, the selected events must have $E_T^{\text{miss}} > 60 \text{ GeV}$ with the direction of the τ jet satisfying $\Delta\phi(\tau \text{ jet}, E_T^{\text{miss}}) < 160^\circ$. Further suppression is obtained with so called *angular selections*: *Cut areas* are defined in the $\Delta\phi(\tau \text{ jet}, E_T^{\text{miss}}) - \Delta\phi(\text{jet}_n, E_T^{\text{miss}})$ planes, where $n = 1, 2, 3$ and jet_n refers to the reconstructed PF jet with the highest ($n = 1$), second highest ($n = 2$) and third highest p_T ($n = 3$). The jet corresponding to the τ jet is included here. The cut areas are illustrated in Fig. 11. The following areas can be identified in the plot and the cuts are subjected to them:

- The topology where a jet is collinear with E_T^{miss} and the τ jet is back-to-back with E_T^{miss} is favoured in QCD multijet events. Only few signal events fall into this region and it is referred to as the *back-to-back corner*.
- The topology where a jet is back-to-back to E_T^{miss} and the τ jet is collinear with E_T^{miss} is most probable in EWK+ $t\bar{t}$ events and signal events. However, the events end up in the small transverse mass values and do not therefore contribute to signal separation from the backgrounds. This region is referred to as the *collinear corner*.

The back-to-back cuts are applied in four different scenarios where the radius of the circular cut is altered and these scenarios along with their names are defined in Table 4.

Table 4: Back-to-back cut scenarios

Scenario name	Back-to-back cut
None	No Cut
Loose	Circular, 40°
Medium	Circular, 60°
Tight	Circular, 80°

In the signal events, a topology where a jet and the τ jet are back-to-back with E_T^{miss} is favoured and it maximises signal separation from EWK+ $t\bar{t}$ events. Events with possibly mismeasured E_T^{miss} in this region are, however, rejected because

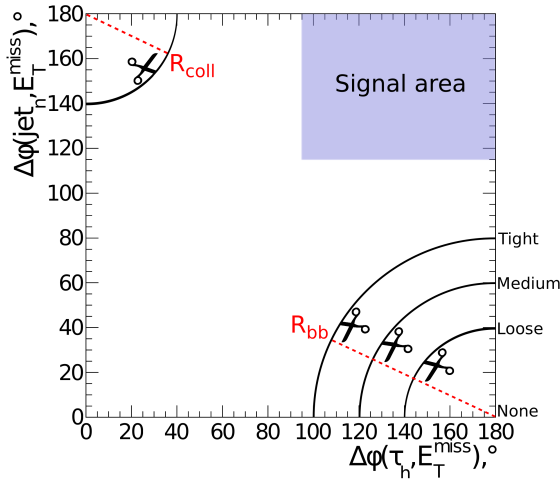


Figure 11: Definitions of the collinear and back-to-back cuts in the $\Delta\phi(\tau \text{ jet}, E_T^{\text{miss}})$ - $\Delta\phi(\text{jet}_n, E_T^{\text{miss}})$ planes: The collinear corner is cut so that the events falling inside the circle quadrant in the upper left corner are rejected. The back-to-back corner is cut so that events falling inside the circle quadrants in the lower right corner (different scenarios) are rejected.

of the requirement of the selected jets to be spatially separated from the τ jet as discussed earlier.

6.3 Optimisation of cuts

The data-driven fake τ measurement method developed in this work is expected to have only a small effect on the sensitivity. Therefore, the optimal cuts yielding the most sensitive limits are not assumed to be altered by the developed method, and the results are presented with the same set of cuts as those optimised and presented in Ref. [27]. Of the cuts used, the following are the most noteworthy:

- The tight requirements on identification of the τ jet, b tagging selection and the number of selected jets are found to be essential for an efficient suppression of the QCD multijet background and these cuts are therefore not further optimised.
- For stability reasons the p_T threshold for hadronic jets is taken to be 30 GeV, although a lower value could improve the sensitivity at some parts of the phase space.
- The values for the E_T^{miss} and τp_T cut thresholds and the option for the angular cut scenario are optimised and the most sensitive results are obtained with $E_T^{\text{miss}} > 60$ GeV, $\tau p_T > 41$ GeV and Loose back-to-back cuts.

6.4 Fitting of the transverse mass tail

There is a small number of events in all the backgrounds in the tail of the m_T distribution. The QCD multijet measurement is especially sensitive to this, since the purity of the method does not allow the usage of finer bins beyond $m_T > 200$ GeV, as can be seen from Section 8.6. When $m_{H^\pm} > 200$ GeV, the majority of the signal events end up in the last transverse mass bin. To enhance the sensitivity of the analysis, the falling parts of the distributions are parametrised so that the original event yield estimates can be replaced with finely binned histograms. The penalty of doing this is that the fit uncertainty involved in the parametrisation needs to be taken into account in the part where the parametrisation is used instead of the original event yield estimate.

By denoting the m_T value with x , the following fit function is used

$$f(x) = c' A^{-B(x-c)}, \quad (47)$$

where A and B are fit parameters, c is the starting point of the fit and c' is the integral of the fitted area (gives additional numerical stability to the fit). This function is used in the range 180 – 700 GeV for all backgrounds and the bin width is set to 20 GeV.

When evaluating the uncertainty of the fit, the eigenvalues and eigenvectors of the fit parameters are used to calculate a base where the parameters are orthogonal to each other. The up and down variations of all the parameters in the orthogonal base are then combined to obtain the uncertainties of the fitted parameters.

7 Background measurements

7.1 Data-driven measurement of the QCD multijet background

7.1.1 Strategy of the measurement

QCD multijet events only rarely contain genuine τ leptons. They can however pass the event selection due to misidentification of a hadronic jet as a τ jet and mismeasurement of the energies of the jets that leads to 'fake' E_T^{miss} . The data-driven measurement of this background is based on the fact that the event sample is dominated by the QCD multijet events before the τ identification selections. The isolation of the τ jet yields one of the largest individual suppressions against the QCD multijet events and is only slightly correlated with the event configuration. This feature is exploited in the measurement method to obtain a so called control sample dominated with QCD multijet events from the events triggered with the τ jet and E_T^{miss} trigger.

The QCD multijet background measurement cut flow with inverted τ jet isolation is illustrated in Fig. 12. The control sample is selected by rejecting the events which pass the τ jet isolation. This τ jet selection, where the event passes all other τ jet selection cuts except the isolation, is called the *Inverted selection*. The rows on the left hand side of Fig. 12 refer to these selections. The rows on the right hand side refer to the nominal selections with τ jet isolation. The so called *Baseline selection* is comprised of these selections. The events where the τ jet does not pass the

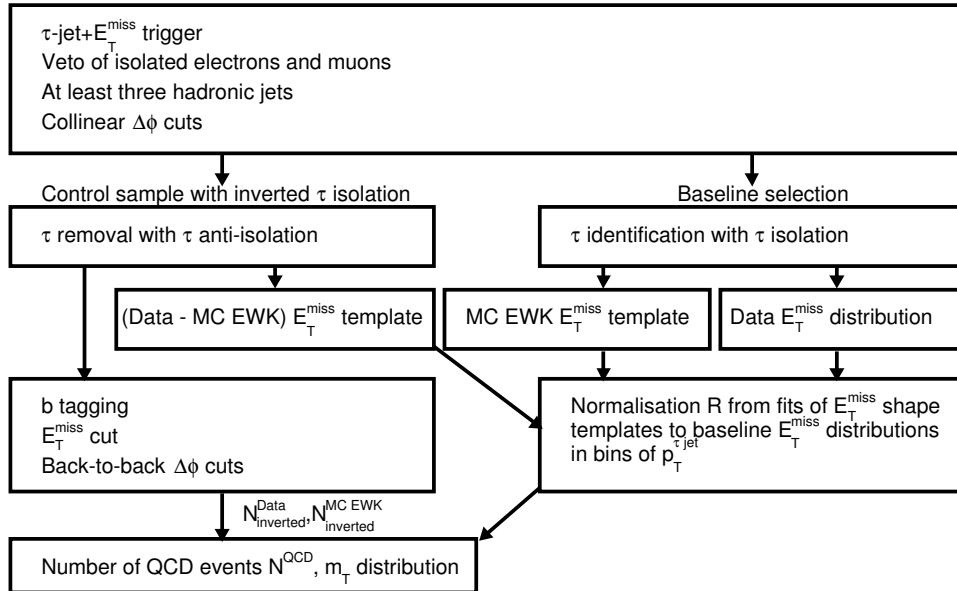


Figure 12: Schematic diagram of the QCD multijet background measurement method with inverted τ jet isolation.

isolation form a control sample orthogonal to the Baseline selection. It is however worth noting that since the offline τ jet candidates are required to match with a HLT τ object, a soft isolation is already applied in the Inverted selection on the offline τ jet candidates although the tighter offline τ isolation is reversed.

All steps in the measurement are performed in τ jet p_T -bins to minimise the correlations between the p_T of the τ jet and the E_T^{miss} variable. The p_T -bin widths are taken to be 10 GeV from 40 GeV up to 80 GeV. The other bins are 80 – 100 GeV, 100 – 120 GeV and > 120 GeV. This choice of binning is made based on the need to have a sufficient number of QCD events in each bin.

The Inverted measurement is based on the so called *fake rate method* [88], and the events are normalised by the *fake rate probability* after all selections to obtain yield estimates in the signal region. The Inverted selection begins by applying all τ jet identification cuts, with the exception that the τ jet isolation cut is inverted (*anti-isolation*). After this, the cuts common to both the Inverted and Baseline selections are applied: all events with isolated electrons and muons are rejected, at least three hadronic jets are required and the collinear cut is performed. The normalisation of the QCD events is determined at this stage, and it is discussed in Section 7.1.2. After this, the events in the Inverted selection are further subjected to the b tagging, E_T^{miss} and back-to-back cuts. The simulated $t\bar{t}$, single top and electroweak backgrounds (W+jets, Z+jets, WW, WZ, ZZ) are processed through the same Inverted selection in the p_T -bins. These events are called simulated EWK events.

The measured number of QCD multijet events expected in the signal region can then be expressed as

$$N^{\text{QCD}} = \sum_i^{\tau\text{-jet candidate } p_T \text{ bins}} \left(N_{\text{inverted}, i}^{\text{Data}} - N_{\text{inverted}, i}^{\text{EWK MC}} \right) \times R_i^{\text{fit}}, \quad (48)$$

where the terms $N_{\text{inverted}, i}^{\text{Data}}$ and $N_{\text{inverted}, i}^{\text{EWK MC}}$ are the number of data and simulated EWK events, respectively, from the Inverted selection after all cuts. The term R_i^{fit} is the fake rate probability that is used as the normalisation factor. It is obtained by fitting QCD and EWK E_T^{miss} templates to the Baseline data E_T^{miss} distributions in each τ -jet p_T -bin, as discussed in more detail in the next section. The index i runs over all τ jet p_T -bins.

7.1.2 Normalisation

The normalisation of the QCD events is determined after the collinear cut with the help of E_T^{miss} templates. These E_T^{miss} templates are obtained by processing the data through the Inverted selection and the simulated EWK events through the Baseline

selection in the τ jet p_T -bins. The normalisation is then found by comparing the E_T^{miss} distributions in the Inverted and Baseline selections in the following manner: the E_T^{miss} templates obtained with fits to data from the Inverted selection and with fits to simulated EWK events from the Baseline selection are used to fit the E_T^{miss} distributions of data from the Baseline selection in each τ jet p_T -bin. This yields the overall normalisation and the fraction of the QCD multijet events for each τ jet p_T -bin as

$$R_i = \frac{N_{\text{Baseline}, i}^{\text{Data}} - N_{\text{Baseline}, i}^{\text{EWK MC}}}{N_{\text{Inverted}, i}^{\text{Data}} - N_{\text{Inverted}, i}^{\text{EWK MC}}}. \quad (49)$$

The data in the Inverted selection is dominated by QCD events and the contribution of the simulated EWK events from the Inverted selection is found to be insignificant regarding the fitting. Thus it is not subtracted from the data. The best description for the Inverted QCD E_T^{miss} distribution is obtained with the sum of Rayleigh, Gaussian and exponential functions as

$$F(x; \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3) = \frac{x}{\sigma_1^2} e^{\frac{-x^2}{2\sigma_1^2}} + \frac{e^{-(x-\mu_2)^2}}{\sigma_2 \sqrt{2\pi}} + \sigma_3 e^{-\mu_3 x}. \quad (50)$$

The EWK+ $t\bar{t}$ E_T^{miss} distribution in the Baseline selection on the other hand is best described by the Gaussian function when $E_T^{\text{miss}} < 160$ GeV

$$F(x; \mu_1, \sigma_1) = \frac{e^{-(x-\mu_1)^2}}{\sigma_1 \sqrt{2\pi}}, \quad (51)$$

and by the exponential function when $E_T^{\text{miss}} > 160$ GeV

$$F(\mu_2, \sigma_2) = \sigma_2 e^{-\mu_2 x}. \quad (52)$$

In the Inverted selection the Gaussian fitting function is replaced by the Landau function when $E_T^{\text{miss}} < 100$ GeV

$$F(x; \mu_1, \sigma_1) = \frac{1}{2\pi\sigma_1 i} \int_{c-i\infty}^{c+i\infty} \exp \left[\frac{s(x-\mu_1)}{\sigma_1} + s \log s \right] ds, \quad (53)$$

where c is any positive real number and the value of F is independent of it. When $E_T^{\text{miss}} > 100$ GeV the exponential function given in Eq. (52) is used. The fits are shown in Fig. 13.

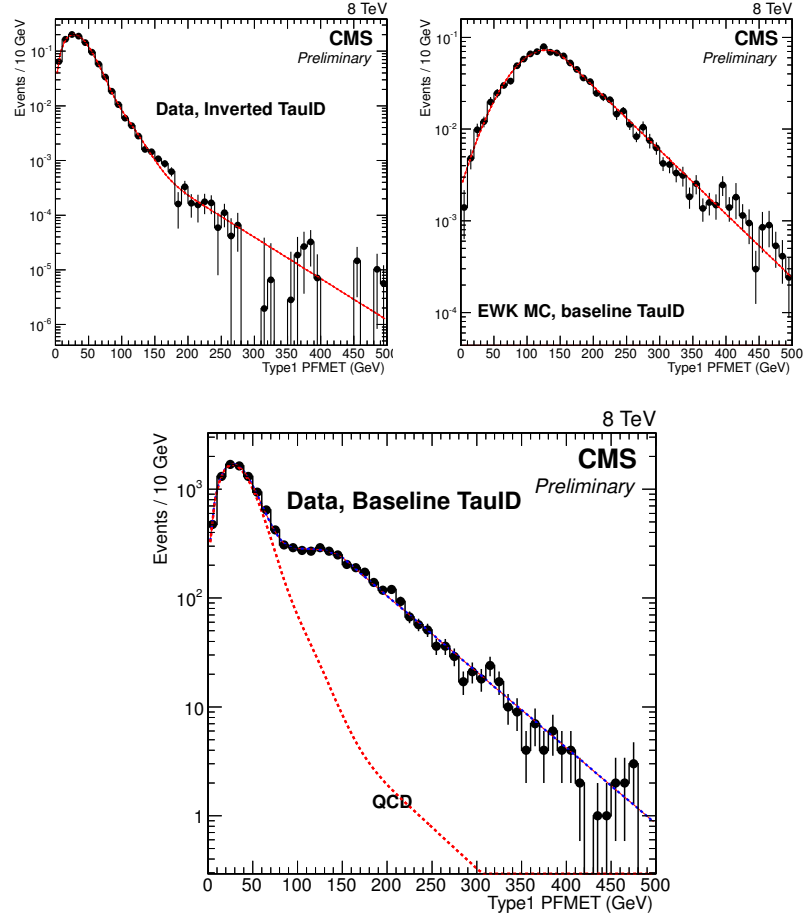


Figure 13: Distribution of the E_T^{miss} template for the QCD multijet data events from the Inverted selection (*top left*), E_T^{miss} template distribution for the simulated EWK events from the Baseline selection (*top right*) and E_T^{miss} distribution for the data from the Baseline selection (*bottom*). The distributions summed over τ jet p_T -bins are shown.

7.1.3 QCD purity

The purity of the selected QCD sample is calculated according to

$$\text{Purity}_{\text{QCD}} = \frac{N^{\text{data}} - N^{\text{MC EWK}}}{N^{\text{data}}}, \quad (54)$$

where N^{data} is the number of data events passing the Inverted selection and $N^{\text{MC EWK}}$ is the number of simulated EWK events passing the Inverted selection. More discussion and plots of the purity are presented in Section 8.6.

7.1.4 Systematic uncertainties

Since the QCD multijet background measurement is closely related to the EWK+ $t\bar{t}$ fake τ background measurement, the systematic uncertainties affecting it are discussed in more detail than those of the other backgrounds. The following uncertainties are taken into account in the measurement:

The simulation uncertainties affect the measurement, since the simulated EWK+ $t\bar{t}$ processes are subtracted from the events passing the Inverted selection. These uncertainties are discussed in more detail in Ref. [27], and of them, the trigger, τ identification, lepton veto, jet energy resolution, b tagging, top p_T -reweighting, cross section, pile-up and luminosity uncertainties, along with the uncertainties of the τ , jet and E_T^{miss} energy scales are taken into account. They are passed through the calculation and scaled down by the purity of the selected sample. Because of this, these uncertainties affect mostly the tail of the transverse mass distribution, where the purity of the selected sample is smaller than that of the bulk part.

The uncertainty from the shape difference of the E_T^{miss} distributions between the Inverted and Baseline selections is taken into account using the following method: first the transverse mass shape for $N^{\text{data}} - N^{\text{simulated EWK}}$ is obtained for both the isolated and anti-isolated τ jet samples. Then the area of these distributions is normalised to unity, and for each bin, the normalisation R_i is calculated. The uncertainty of this normalisation factor R_i is then calculated using error propagation. This is done under the assumption that only statistical uncertainties affect the event counts and that they cover for the shape difference between the isolated and anti-isolated samples. The resulting systematic difference depends on the transverse mass bin and therefore it is taken into account as a shape uncertainty.

The uncertainty from the E_T^{miss} fits is estimated by calculating how much the normalisation changes as the same fit functions are used with different fitting methods. The χ^2 and LogLikelihood methods, and using the integral of the fit function in each bin are used in the estimation. The maximal relative change in the bin-wise normalisations is then taken as the upper limit for the fit uncertainty. This approach gives 3 % as the systematic uncertainty following from the fitting.

The uncertainty resulting from determining the QCD fake rate probability after the collinear cut and not at the end of the selection is additionally taken into account. The reason for determining the normalisation already after the collinear cut is to have enough events to perform the E_T^{miss} -fitting. Of the remaining cuts, only b tagging is found to have a significant effect on the normalisation. To estimate this uncertainty, the fake rate probabilities are calculated before and after the b tagging cut is applied, and the *inclusive* probabilities are then compared to have enough statistics. The relative change in the probability is 4.43 % for the QCD sample and 4.5 % is taken as the systematic uncertainty for determining the normalisation before all cuts are applied.

7.2 Data-driven measurement of the EWK+ $t\bar{t}$ background with genuine τ leptons

The EWK+ $t\bar{t}$ background with genuine τ leptons consists mainly of $t\bar{t}$ and W +jets processes, while the rest arises from Z/γ^* , diboson and single-top-quark events with genuine τ leptons. The data-driven method to measure this background is based on τ *embedding*, in which observed muons are replaced with simulated τ leptons.

The embedding method relies on lepton universality: since the gauge bosons and photons couple to all three lepton flavours with equal strength, there should be identical numbers of electroweak events with τ leptons, muons and electrons, given a certain kinematical configuration. This is exactly true only when all leptons have the same mass, but it is a justified approximation and the errors resulting from the mass differences are small [29]. The method also exploits the fact that the CMS detector is able to reconstruct muons with a good efficiency and high momentum resolution. Because of this, the usage of muons yields better results when compared to electrons, which could also in principle be used in the embedding.

The background with genuine τ leptons can thus be estimated from the collision data with μ +jets events by selecting events that are otherwise identical to the signal events but have a single muon instead of a single τ_h . If charged Higgs bosons exist, the rate at which they decay to muons is very low and contamination from these events is negligible.

The steps of the method are the following: first a control sample of events with $W \rightarrow \mu\nu_\mu$ decays is selected by choosing events that are triggered by a single muon trigger. Events with an isolated muon are chosen offline and the events with more than one muon or additional isolated electrons or muons passing the selection are rejected. At least three hadronic Particle-Flow jets are then required from the event in order to pass the selection. In addition, the selected jets are required to be separated from the selected muon with $\Delta R(\mu, \text{jet}) > 0.5$.

The identified muon is then removed from the event, and a hadronically decaying τ lepton with the same momentum, that is simulated and reconstructed at the Particle-Flow level, is added to the event. This is the actual embedding part, which leads to a *hybrid event*, consisting of both data and simulation. After this, the E_T^{miss} is recalculated and the hybrid events are passed to the event selection described in Section 6.2, after which they are used as the genuine τ background.

The final step in the background measurement is the normalisation of the background shape. This corrects the errors arising from the different trigger and selection efficiencies in the signal and embedding samples. The normalised prediction for the number of EWK+t \bar{t} genuine τ events is then obtained by summing over all events as

$$N^{\text{EWK+t}\bar{t}\tau} = \sum_i^{\text{Selected events}} \mathcal{B}(\tau \rightarrow \text{hadrons}) \times w_{\text{min vis. } p_T, i} \times \left[1 - f_{W \rightarrow \tau \nu_\tau \rightarrow \mu \nu_\mu \nu_\tau}(p_{T,i}^\mu) \right] \\ \times \frac{\epsilon_{\text{trg}}^\tau(p_{T,i}^{\tau \text{ jet}}) \times \epsilon_{\text{L1}}^{E_T^{\text{miss}}}(E_{T,i}^{\text{miss}})}{\epsilon_{\text{trg}}^\mu(\eta_i^\mu) \times \epsilon_{\text{sel}}^\mu(\eta_i^\mu)} \times c_{\text{t}\bar{t}}(m_{T,i}), \quad (55)$$

where the index i refers to the quantities of event i and the following symbols are used:

- $w_{\text{min vis. } p_T}$ is the event weight from the minimum visible p_T requirement,
- $f_{W \rightarrow \tau \nu_\tau \rightarrow \mu \nu_\mu \nu_\tau}(p_T^\mu)$ is the fraction of selected $W \rightarrow \tau \nu_\tau \rightarrow \mu \nu_\mu \nu_\tau$ events,
- $\epsilon_{\text{trg}}^\tau(p_T^{\tau \text{ jet}})$ is the τ trigger efficiency in bins of τ jet p_T ,
- $\epsilon_{\text{L1}}^{E_T^{\text{miss}}}(E_T^{\text{miss}})$ is the L1 E_T^{miss} trigger efficiency in bins of E_T^{miss} ,
- $\epsilon_{\text{trg}}^\mu(\eta^\mu)$ is the muon trigger efficiency in bins of muon η ,
- $\epsilon_{\text{sel}}^\mu(\eta^\mu)$ is the muon offline selection efficiency in bins of muon η , and
- $c_{\text{t}\bar{t}}(m_T)$ is the correction from a difference in the m_T distribution between simulated embedded t \bar{t} and normal t \bar{t} events as a function of m_T .

7.3 Simulation based measurement of the EWK+t \bar{t} background with misidentified τ leptons

The only remaining background contribution to be determined is the EWK+t \bar{t} background without identified genuine τ leptons within the p_T, η acceptance. These events result from an electron, muon, or hadronic jet misidentified as the τ jet or from a genuine τ jet that is misidentified or outside the acceptance. This EWK+t \bar{t} background with misidentified τ leptons has been previously estimated with simulation instead of a data-driven method because its contribution was small in the data with the collision energy of $\sqrt{s} = 7$ TeV. However, its importance is greater with the considered $\sqrt{s} = 8$ TeV data. As discussed in Section 1, this

is the reason for developing the data-driven background measurement method presented in this work. However, the simulated estimates are still needed to a smaller extent to support the data-driven method and to cross-check its validity. This is why the simulated method is presented in the following.

The first step in the simulated method is the application of the full event selection, described in 6.2, on the simulated EWK+ $t\bar{t}$ events. Then the generator particle information is matched to the selected τ jet direction: electrons, muons, and τ leptons with $p_T > 10 \text{ GeV}/c$ are considered in the matching process and if a simulated electron, muon, or unidentified/out-of-acceptance τ lepton is found within $\Delta R < 0.1$ of the selected τ jet, the event is considered to belong to the EWK+ $t\bar{t}$ background with misidentified τ leptons. In case multiple matches are found for the selected τ jet, the τ jet is first checked to match with an electron, then with a muon, and finally with a τ lepton so that the first match is taken as the originator of the selected τ jet. If no match is found, the selected τ jet is assumed to originate from a hadronic jet.

8 Data-driven measurement of EWK+ $t\bar{t}$ background with misidentified τ leptons

8.1 Breakdown of background origin

The EWK+ $t\bar{t}$ background with misidentified τ leptons is defined as in Section 7.3. Simply put, the background consists of EWK+ $t\bar{t}$ events not having an identified genuine τ lepton inside the acceptance. In the following sections, this background is referred to simply as the EWK+ $t\bar{t}$ fake τ background. To gain insight of the background, the simulated estimates described in Section 7.3 can be used. This way, it can be found out that the majority of the EWK+ $t\bar{t}$ fake τ background events have a hadronic jet misidentified as the τ jet in the event. Electrons misidentified as τ jets make up only a few percent of the EWK+ $t\bar{t}$ fake τ background events, while the $\mu \rightarrow \tau$ misidentification rate is negligible. Because of this, jet characteristics have an important effect on the method.

Since the different simulated samples are also used to some extent in the data-driven method, estimating their relative contributions is essential. The dominating source for the simulated background is the $t\bar{t}$ sample ($\sim 90\%$) and then the single top W-channel. The W+jets sample on the other hand is dominant before the b tagging cut, which suppresses it strongly by requiring a $t\bar{t}$ -like event topology.

8.2 Principal idea behind the method

The EWK+ $t\bar{t}$ fake τ background and the QCD multijet background have the common feature that neither contains genuine τ leptons and both pass the selection because of misidentification or mismeasurement, as discussed in Sections 7.1 and 7.3. This gives the motivation to use the QCD multijet background measurement method with inverted τ jet isolation as the basis for the data-driven EWK+ $t\bar{t}$ fake τ background measurement. To achieve this, the Inverted method is extended and the steps of the *fake rate method* [88] are again followed to allow the data-driven measurement of the EWK+ $t\bar{t}$ fake τ background inclusively as well.

Since the EWK+t \bar{t} events, that are subtracted from the data events in the Inverted selection of the QCD multijet measurement method according to Eq. (48), come from simulation, the fake τ events of the EWK+t \bar{t} sample can be identified and filtered out of the subtraction. This simple observation is the starting point of the data-driven measurement of the combined background that contains both the QCD multijet and EWK+t \bar{t} fake τ events according to the equation

$$N^{\text{Fake } \tau} \equiv N^{\text{QCD}} + N^{\text{EWK fake } \tau} \\ = \sum_i^{\tau\text{-jet candidate } p_T \text{ bins}} \left(N_{\text{inverted}, i}^{\text{Data}} - N_{\text{inverted}, i}^{\text{EWK MC genuine } \tau} \right) \times R_i, \quad (56)$$

where the notation is the same as in Eq. (48). In Eq. (56) and throughout the rest of this work, this mixture background is referred to as the *fake τ background* (or equivalently the *misidentified τ background*). As can be seen from Eq. (56), the data-driven measurement of the fake τ background ultimately culminates in finding the correct normalisation factors R_i for it.

In the following sections, the background measurement method where the QCD multijet background is measured from data and the EWK+t \bar{t} fake τ background is estimated with simulation is referred to as the *simulated method* for brevity. The developed measurement method where both of these backgrounds are measured inclusively from data is opposingly referred to as the *data-driven method*.

One of the key assumptions done in developing the method is that the simulation based yield for the EWK+t \bar{t} background is considered reliable, that is, $N_i^{\text{EWK MC } \tau} \approx N_i^{\text{EWK Data } \tau}$. This assumption is also made in the QCD multijet background measurement and in charged Higgs boson searches sensitive to different final states [89] with good results, so the assumption is reasonable. The shape and normalisation of the fake τ background are still measured from data.

8.3 Normalisation

The normalisation process for the mixture background is not as straightforward as in the sheer QCD multijet measurement discussed in Section 7.1.2, since the fake rate probabilities for both QCD and EWK+t \bar{t} fake τ samples are needed and they have to be combined to obtain the final normalisation. The separation of the fake rate probabilities on the other hand is necessary since the different jet compositions of the EWK+t \bar{t} and QCD samples affect their selection efficiencies. This effect is discussed in more detail in Section 8.5.

In the developed method, the EWK+t \bar{t} sample is splitted so that the E_T^{miss} distributions of the EWK+t \bar{t} fake τ sample are fitted separately in the determination of the

EWK+t \bar{t} fake rate probability. As in Section 7.1.2, the E_T^{miss} fits are used to obtain the fake rate probabilities R_i of the QCD and EWK+t \bar{t} fake τ samples according to

$$R_i^{\text{QCD}} \equiv \frac{N_{\text{Baseline},i}^{\text{QCD}}}{N_{\text{Inverted},i}^{\text{QCD}}}, \quad (57)$$

$$R_i^{\text{EWK+t}\bar{t} \text{ fake } \tau} \equiv \frac{N_{\text{Baseline},i}^{\text{EWK+t}\bar{t} \text{ fake } \tau}}{N_{\text{Inverted},i}^{\text{EWK+t}\bar{t} \text{ fake } \tau}}. \quad (58)$$

To obtain the QCD fake rate probability, the procedure is exactly the same as that described in Section 7.1.2: E_T^{miss} templates obtained with fits to data from the Inverted selection and with fits to simulated EWK+t \bar{t} events (genuine and fake τ samples *not* separated) from the Baseline selection are used to fit the E_T^{miss} distributions of data from the Baseline selection in each τ jet p_T -bin. The relevant E_T^{miss} distributions along with the fits for obtaining the QCD fake rate probability were shown in Fig. 13.

A similar scheme to obtain the EWK+t \bar{t} fake rate probability is not possible since QCD and EWK+t \bar{t} fake τ events cannot be separated from data. This is why the fake rate probability is determined from simulation by simply comparing the integrals of the corresponding fits in the Baseline and Inverted selections in each τ jet p_T -bin. The best description of the E_T^{miss} distributions in both the Baseline and Inverted selections is obtained with the fit functions given by Eq. (53) when $E_T^{\text{miss}} < 100$ GeV and Eq. (52) when $E_T^{\text{miss}} > 100$ GeV. The E_T^{miss} distributions along with the fits are shown in Fig. 14.

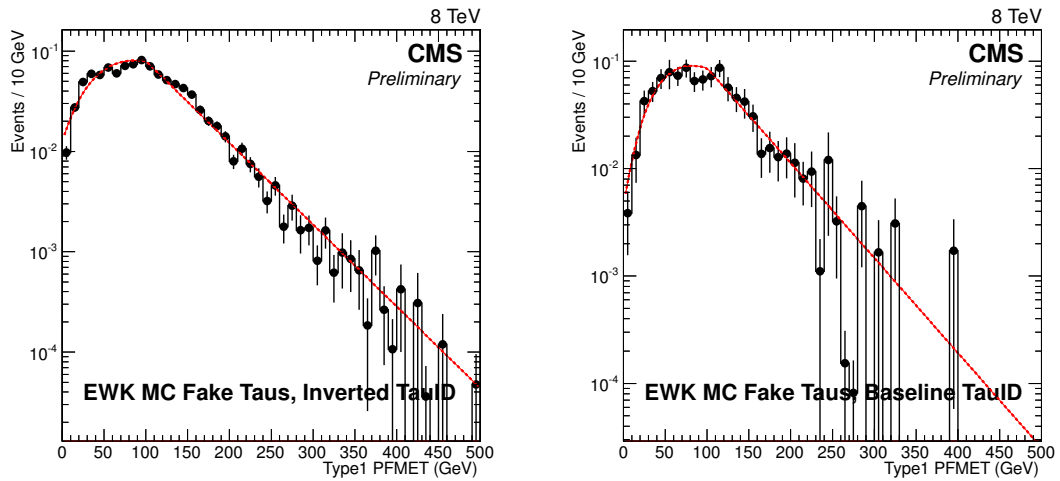


Figure 14: E_T^{miss} distributions for simulated EWK+t \bar{t} fake τ events in the Inverted (*left*) and Baseline (*right*) selections along with the fits. The distributions shown are summed over τ jet p_T -bins.

With the fake rate probabilities given by the E_T^{miss} fits and Eqs. (57) and (58), the probabilities can be combined to yield the final fake rate probability (normalisation) as the weighted sum

$$R_i = w_i R_i^{\text{QCD}} + (1 - w_i) R_i^{\text{EWK}+\text{t}\bar{\text{t}} \text{ fake } \tau}. \quad (59)$$

where w_i is the weighting factor in the i :th τ_h p_T -bin. A natural choice is to calculate it as the relative frequency of Inverted QCD events with respect to all Inverted fake τ events according to

$$w_i = \frac{N_{\text{Inverted},i}^{\text{QCD}}}{N_{\text{Inverted},i}^{\text{Fake } \tau}}. \quad (60)$$

The weights recovered using this scheme in turn give the final normalisation factors R_i . These are then used in the normalisation according to Eq. (56) to finally obtain the number of fake τ events in the signal region. This procedure, defined by Eq. (59) is referred to as the *fake rate probability weighting*.

As discussed in the previous section, the assumption that $N_i^{\text{EWK MC } \tau} \approx N_i^{\text{EWK Data } \tau}$ is made in the method. This is assumed separately for the EWK+t $\bar{\text{t}}$ genuine τ sample in Eq. (56) and for the EWK+t $\bar{\text{t}}$ fake τ sample in Eqs. (58) and (60). It is worth noting that the procedure makes no assumption on what the origin of the misidentified τ leptons is, that is, the method takes equally into account the jets, electrons and muons that are misidentified as τ leptons.

8.4 Systematic uncertainties

Many of the systematic uncertainties affecting the fake τ background measurement can be treated similarly as those related to the QCD background measurement, discussed in Section 7.1.4. The following uncertainties are taken into account:

The simulation uncertainties of the EWK+t $\bar{\text{t}}$ samples are discussed in more detail in Ref. [27]. These uncertainties include the trigger, τ identification, lepton veto, jet energy resolution, b tagging, top p_T -reweighting, cross section, pile-up and luminosity uncertainties, along with the uncertainties of the τ , jet and E_T^{miss} energy scales. To take them into account in the measurement of the fake τ background, the following approach is used: all uncertainties are converted to shape uncertainties, then propagated through Eq. (59) and finally the changes in the transverse mass distribution are recorded.

The uncertainty from the shape difference of the E_T^{miss} distributions between the Inverted and Baseline selections only affects the QCD part of the background

and is therefore scaled down by the fraction of QCD events with respect to all fake τ events in the background measurement.

The uncertainty from the EWK+t \bar{t} fake τ E_T^{miss} -fits is calculated in a similar manner as that related to the QCD E_T^{miss} -fits, discussed in Section 7.1.4. Together with the uncertainties from the QCD E_T^{miss} -fits, the total uncertainty remains under the upper limit of 3 % that was used in the sole QCD measurement.

The uncertainty from determining the EWK+t \bar{t} fake rate probability after the collinear cut and not after all cuts are applied is calculated as in Section 7.1.4 for the fake rate probability of the QCD sample. Especially the b tagging cut could have an effect on the fake rate probabilities of the EWK+t \bar{t} sample: before the cut is applied the sample is W+jets dominated but after the cut it consist mainly of t \bar{t} events. The effect of the b tagging cut was confirmed to be significant, while the other remaining cuts had a negligible effect on the fake rate probability. By comparing the probabilities determined before and after applying the b tagging cut, a relative change of 4.37 % in the inclusive fake rate probability of the EWK+t \bar{t} fake τ sample is observed. A 4.5% uncertainty is then included to take this into account.

The uncertainty from the fake rate weighting is related to the statistics of the weighting procedure and has to be taken into account as a new source of uncertainty in the data-driven fake τ background measurement. Since the rate of having a certain number of QCD events in the set of all fake τ events in the Inverted selection can be considered to be a binomial process, the binomial error

$$\delta w_i = \sqrt{\frac{w_i(1 - w_i)}{N_{\text{Inverted},i}^{\text{Fake } \tau}}}, \quad (61)$$

is used as the error of the weight w_i . The normalisation is then done with the weights $w_i + \delta w_i$, $w_i - \delta w_i$ and the resulting transverse mass distributions are recorded. The differences of these distributions from the nominal distribution are taken as the systematic uncertainties.

The systematic uncertainties affecting the fake τ background measurement are summarised in Table 5. A breakdown of the systematic uncertainties as a function of m_T is shown in Fig. 15. A noteworthy feature in the systematic uncertainties affecting the QCD multijet background is that the subtraction of simulated EWK+t \bar{t} genuine τ events from data causes the simulation related systematic uncertainties of the QCD sample to be anti-correlated with the same uncertainties of other samples. The anti-correlation of the uncertainties between the QCD multijet background and the simulated EWK+t \bar{t} background turns out to be a key component in the reduction of the systematic uncertainties regarding the data-driven fake τ background measurement, as will be discussed in Section 9.1.

The largest sources of uncertainties concerning the QCD multijet background are in order the jet $\rightarrow \tau$ misidentification, top p_T -reweighting, normalisation after collinear cuts, b tagging and QCD shape difference uncertainties, while these are the jet $\rightarrow \tau$ misidentification, top p_T -reweighting, probabilistic b tag, $t\bar{t}$ cross section and jet energy scale uncertainties for the EWK+ $t\bar{t}$ fake τ background. On the other hand, the largest sources of uncertainties concerning the fake τ background are the normalisation after collinear cuts, E_T^{miss} fits, jet $\rightarrow \tau$ misidentification, fake rate weighting and the QCD shape difference uncertainties. It can be seen that the data-driven method reduces all the systematic uncertainties except the uncertainties related to the E_T^{miss} -fitting and performing the normalisation after collinear cuts, which increase. Additionally, the uncertainty of the fake rate probability weighting naturally affects only the data-driven method.

8.5 Fake rate probabilities

In Fig. 16 the fake rate probability weight w (*left*) and resulting total fake rate probability (*right*) are shown as a function of τ_h p_T -bins. The fact that the weight increases with p_T is a direct consequence of the QCD fake τ leptons having more transverse momenta (*harder events*) than the EWK+ $t\bar{t}$ fake τ leptons.

To understand the behaviour of the fake rate probabilities in Fig. 16, it is worth noting that as discussed in Section 7.3, most of the fake τ background consists of misidentified jets. The jet composition is known to have an effect on the τ identification efficiency as described in Ref. [90], and accordingly, the rate for misidentifying a quark jet is higher than it is for a gluon jet. This explains the higher fake rate probabilities for the EWK+ $t\bar{t}$ sample that is quark jet enriched while the probabilities are smaller for the QCD sample that mostly consists of gluon jets. The different quark and gluon jet selection efficiencies are thus taken into account in the separated fake rate probabilities.

8.6 Fake τ purity

The purity of the selected fake τ sample is calculated according to

$$\text{Purity}_{\text{Fake } \tau} = \frac{N^{\text{data}} - N^{\text{MC EWK genuine } \tau}}{N^{\text{data}}}, \quad (62)$$

where N^{data} is the number of data events passing the Inverted selection and $N^{\text{MC EWK genuine } \tau}$ is the number of simulated EWK+ $t\bar{t}$ events with genuine τ leptons passing the Inverted selection.

Table 5: Systematic uncertainties (relative to the background yield, in %) for the backgrounds involving misidentified τ leptons for the analysis with $\sqrt{s} = 8$ TeV data. The systematic uncertainties of the QCD multijet and EWK+ $t\bar{t}$ fake τ backgrounds are shown on the left, along with the combined uncertainty (quadratic sum) of the mixture background. The systematic uncertainties of the fake τ background are shown on the right. The uncertainties, which depend on the final distribution bin, are marked with (S) and for them the maximum contracted value of the negative or positive variation is displayed. Empty cells indicate, that the uncertainty does not apply to the sample. The $t\bar{t}$ cross section, luminosity, lepton veto and probabilistic b tag uncertainties are calculated as shape uncertainties for the QCD (if affected) and fake τ samples and as scalar uncertainties for the EWK+ $t\bar{t}$ fake τ sample.

Source	QCD multijet	EWK+ $t\bar{t}$ fake τ (MC)	Total	Fake τ
τ trigger, data	0.5	1.3	0.6	0.1
τ trigger, sim.	0.2	0.6	0.3	0.1
L1ETM, data	1.2	2.5	1.2	0.2
L1ETM, sim.	< 0.1	0.1	< 0.1	< 0.1
τ ID (S)	0.8	-	0.5	0.5
$e \rightarrow \tau$ mis-ID (S)	0.1	3	1.1	< 0.1
$\mu \rightarrow \tau$ mis-ID (S)	< 0.1	0.8	0.3	< 0.1
jet $\rightarrow \tau$ mis-ID (S)	6.9	16.3	7.5	3
τ ES (S)	1.8	2.3	1.4	0.3
Jet ES	1.4	4.8	2	0.3
Jet energy resolution	0.2	2.3	0.9	0.2
Unclustered E_T^{miss} ES	0.5	1.9	0.8	0.1
b tagging (S)	3.1	4.3	2.5	0.6
Probabilistic b tag (S/no S)	-	6.8	2.5	1.1
Lepton veto (S/no S)	-	1.6	0.6	0.1
Top p_T weighting (S)	6.8	10.2	5.7	0.6
QCD shape (S)	1.3	-	0.8	0.8
E_T^{miss} fits	3	-	1.9	3
Norm. after coll. cuts	4.5	-	2.8	4.5
Fake rate weighting (S)	-	-	-	1.6
pile-up (S/no S)	0.1	2.6	1	< 0.1
$t\bar{t}$ cross section (S/no S)	1.7	5.2	2.2	0.3
luminosity	0.7	2.6	1.1	0.2

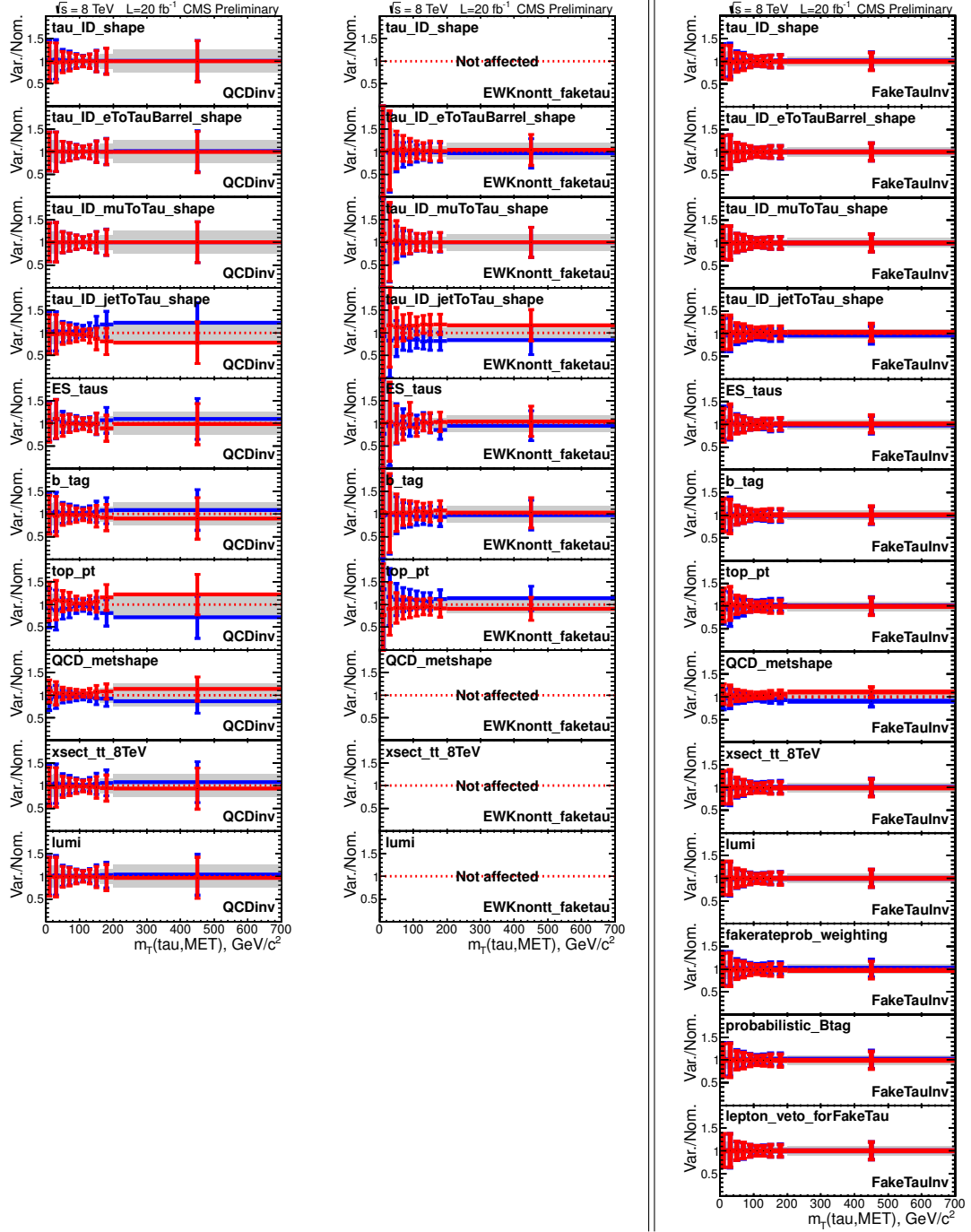


Figure 15: Breakdown of the systematic uncertainties affecting the shape of the m_T distribution of the data-driven QCD multijet background (*left*), simulated EWK+ $t\bar{t}$ fake τ background (*middle*) and data-driven fake τ background (*right*). The up and down variations relative to nominal are denoted by red and blue colors, respectively. The horizontally and vertically adjacent shape uncertainties are treated as fully correlated and fully uncorrelated, respectively. The $t\bar{t}$ cross section, luminosity, lepton veto and probabilistic b tag uncertainties are calculated as shape uncertainties for the QCD (if affected) and fake τ samples and as scalar uncertainties for the EWK+ $t\bar{t}$ fake τ sample.

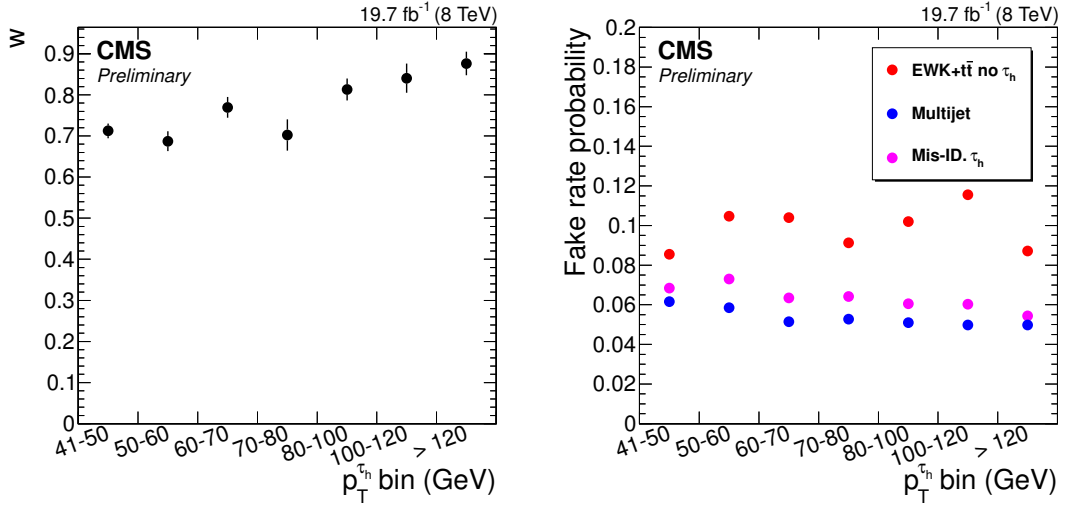


Figure 16: Fake rate probability weight w (left) and resulting total fake rate probability (right) as a function of τ_h p_T -bins.

The QCD purity and fake τ purity calculated according to Eqs. (54) and (62) as functions of transverse mass, for loose, medium and tight back-to-back cuts are shown in Fig. 17. The purity can be seen to increase especially when $m_T > 120$ GeV with the developed method.

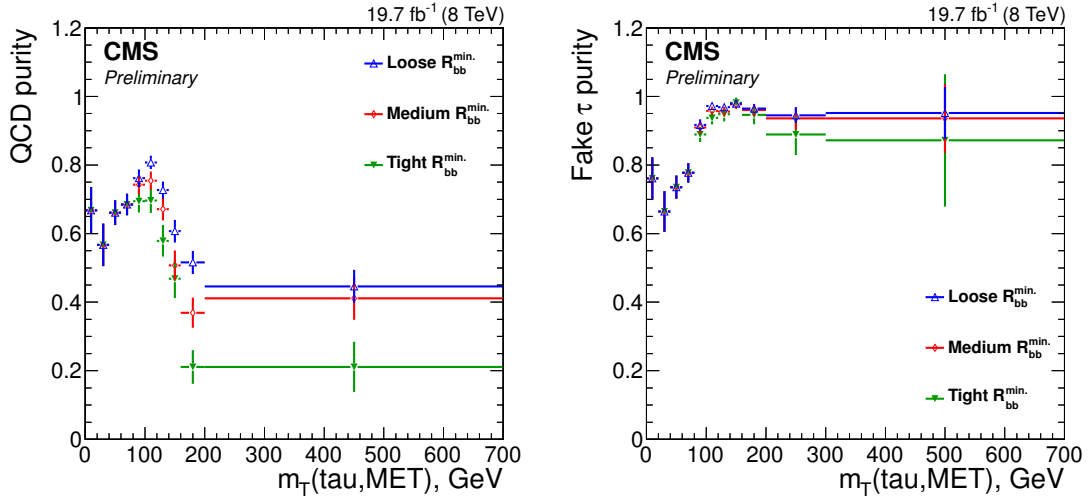


Figure 17: Distributions of QCD purity (left) and fake τ purity (right), defined according to Eqs. (54) and (62) as functions of transverse mass, for loose, medium and tight back-to-back cuts.

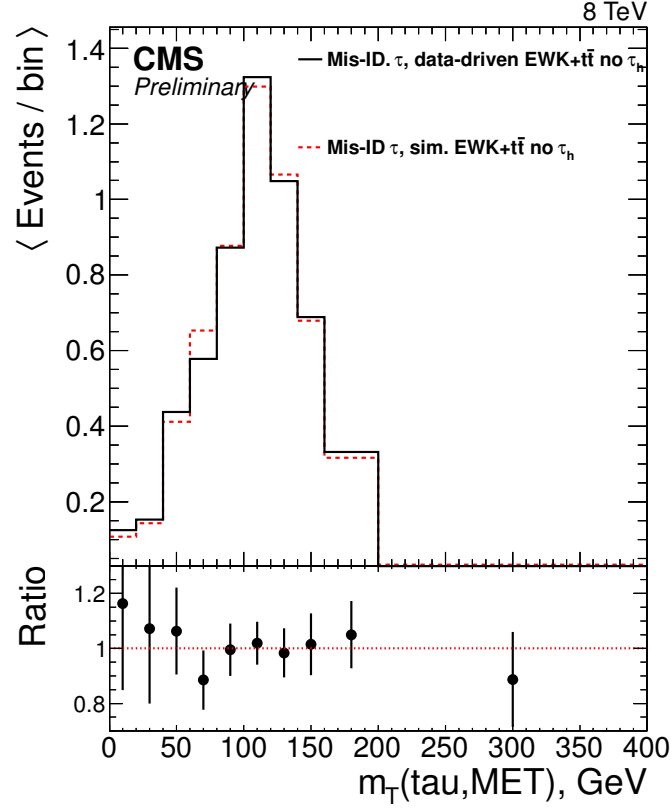


Figure 18: Final transverse mass distributions for the fake τ background obtained when the EWK+t \bar{t} fake τ background is simulated (*sim.*) and measured from data (*data-driven*).

8.7 Transverse mass

The transverse mass distribution of the fake τ background, obtained using the data-driven fake τ measurement is shown in Fig. 18. To allow comparison, the same distribution obtained using the simulated EWK+t \bar{t} fake τ background estimate is also shown. This mixture background is obtained by measuring the QCD background from data and summing it with the simulated EWK+t \bar{t} fake τ background. The distributions can be seen to agree within statistical uncertainties.

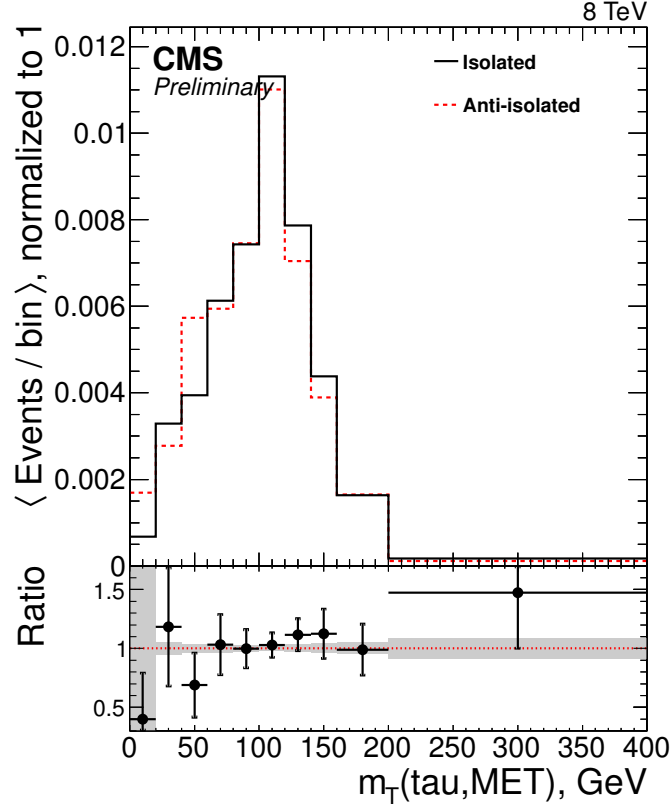


Figure 19: Comparison of the transverse mass distribution of the measured fake τ events from the Inverted selection compared to that of the events from the Baseline selection subtracting the simulated EWK+t \bar{t} events with a genuine τ lepton at the end of the selection and vetoing on tagged b jets.

8.8 Closure test

To give further validation for the developed method, the transverse mass distribution of the measured fake τ events is compared to that of the events from the Baseline selection subtracting the simulated EWK+t \bar{t} events with a genuine τ lepton at the end of the selection. To avoid signal contamination, a veto on tagged b jets is applied. If the data-driven method is valid, the distributions should agree in different areas of the phase space. The comparison is shown in Fig. 19 and it shows that the distributions agree within statistical uncertainties.

9 Results

9.1 Event yields

A summary of the expected fake τ background event yields obtained with the simulated and data-driven methods is shown in Table 6. The event yields of the backgrounds as functions of the transverse mass are shown in Fig. 20. The simulated charged Higgs boson signal for $m_{H^\pm} = 300$ GeV is also shown.

It can be seen from Table 6 that the yields match within 0.1 %. The relative statistical uncertainty is reduced from 3.4 % to 3.0 % with the data-driven method. The systematic uncertainties on the other hand are reduced from $^{+14.2}_{-14.5}$ % to $^{+9.4}_{-9.4}$ %. The statistical uncertainties are thus reduced by 12 % and the systematic uncertainties by 34 %.

The reduction in the statistical uncertainties is a natural consequence of the increased sample sizes with the data-driven method. The systematic uncertainties reduce because of the anti-correlation in the simulation related uncertainties between the QCD multijet and EWK+ $t\bar{t}$ fake τ samples: the uncertainties of the latter sample only affect the final normalisation of the fake τ background, and the anti-correlation cancels their contributions partially as the EWK+ $t\bar{t}$ fake τ background is measured inclusively with the QCD multijet background in the calculation of the fake rate weight.

Table 6: Summary of the expected event yields from the background measurements.

Source	$N_{\text{events}} \pm \text{stat.} \pm \text{syst.}$
QCD multijet background (data-driven)	$78.2 \pm 2.6 \begin{smallmatrix} +9.6 \\ -10.3 \end{smallmatrix}$
EWK+ $t\bar{t}$ fake τ (sim.)	$46.7 \pm 2.2 \begin{smallmatrix} +10.5 \\ -10.2 \end{smallmatrix}$
Total	$124.9 \pm 3.4 \begin{smallmatrix} +14.2 \\ -14.5 \end{smallmatrix}$
Fake τ (data-driven)	$124.8 \pm 3.0 \begin{smallmatrix} +9.4 \\ -9.4 \end{smallmatrix}$

9.2 Expected limits

The calculation of upper limits on $\mathcal{B}(t \rightarrow bH^\pm) \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ in the light charged Higgs boson case and on $\sigma_{pp \rightarrow H^\pm tb} \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ in the heavy charged Higgs boson case is based on the modified frequentist CL_s criterion [91–93]. The ‘LHC approach’ for specifying the test statistic and treatment of systematic uncertainties is used. This approach has been agreed on by the CMS and ATLAS collaborations [94]. A complete description of the statistical method used to set

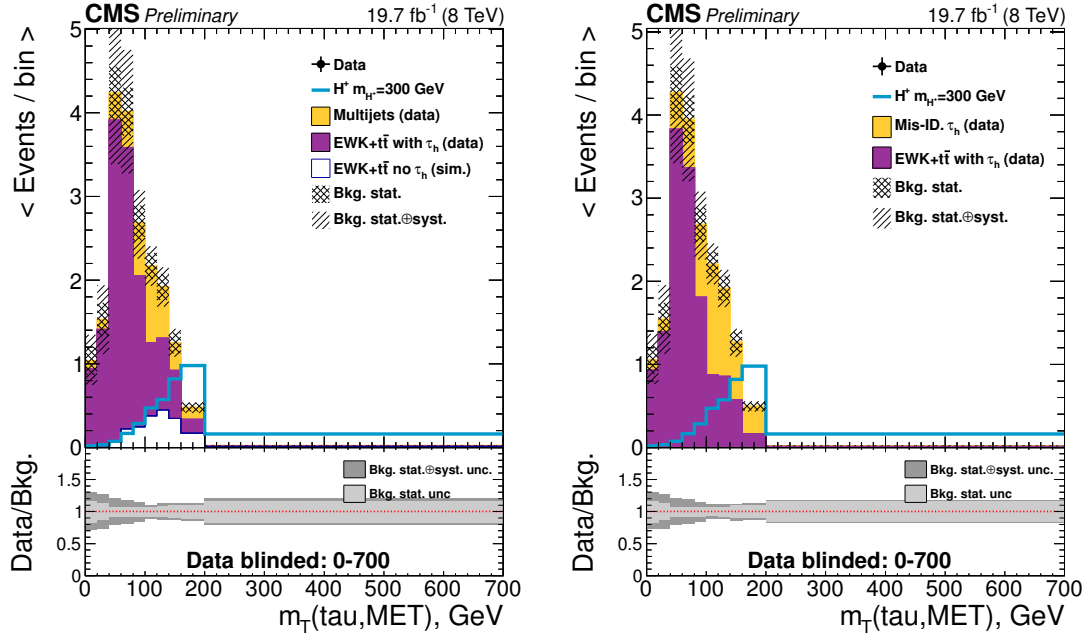


Figure 20: Transverse mass distributions for the measured backgrounds after the full event selection. The expected event yield for $H^\pm \rightarrow \tau^\pm \nu_\tau$ decays is shown as the blue line for $m_{H^\pm} = 300 \text{ GeV}/c^2$, normalised according to $\sigma_{pp \rightarrow H^\pm tb} \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau) = 1 \text{ pb}$. The resulting distributions when the EWK+ $t\bar{t}$ fake τ background is simulated (*left*) and measured from data (*right*) are shown.

these limits is given in Ref. [85]. In practice, the calculation is done with the CMS Higgs combination tool, Combine [95], using the asymptotic LHC-type CL_s approach. The calculation is based on determining a binned maximum likelihood fit on the transverse mass distribution calculated from the selected τ jet and E_T^{miss} and shown in the previous section. The expected limits allow the comparison of the sensitivity reached when the EWK+ $t\bar{t}$ fake τ background is estimated from simulation and when it is measured from data. As sensitivity is the primary quantity of interest, only the expected limits are considered.

The set of cuts described in Section 6.3 have been optimised to give the best signal sensitivity in Ref. [27]. The developed method is assumed to have a small effect on the sensitivity, so these cuts are used in this work as well. In the limit calculation, the HH events are not simulated because of their negligible effect (discussed in Section 2.4.1).

The resulting model-independent expected limits obtained with the data-driven method for the m_{H^\pm} search range 80 – 600 GeV are shown in Fig. 21. The expected median limits and the widths of the uncertainty bands obtained with the simulated and data-driven methods are tabulated in Tables 7 and 8. To allow comparison, the median expected limits obtained with both methods along with their ratio is shown in Fig. 22. The widths of the uncertainty bands of the expected limits are shown in Fig. 23.

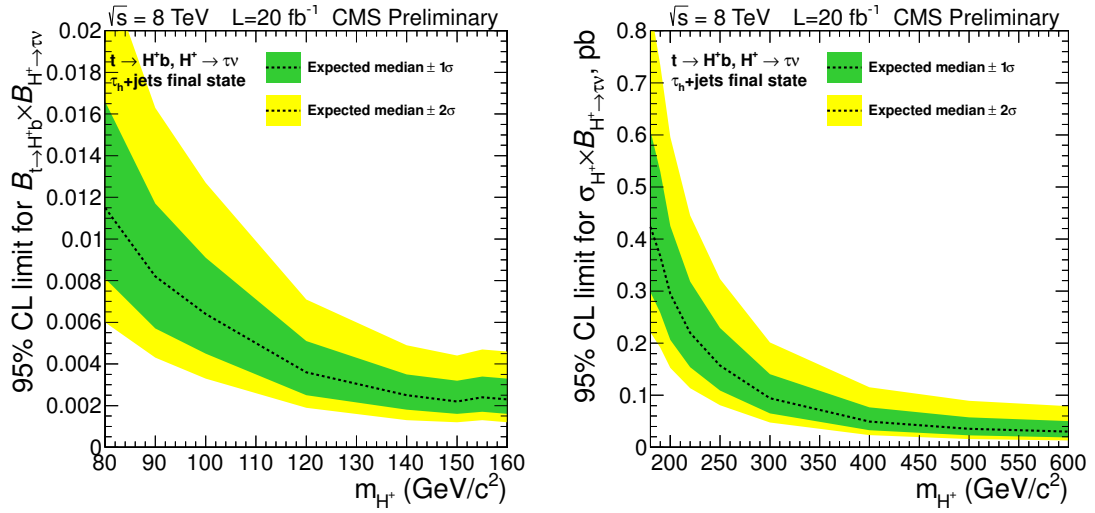


Figure 21: Expected upper limits on the branching fraction $\mathcal{B}(t \rightarrow bH^\pm) \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ for the charged Higgs boson mass hypotheses in the 80–160 GeV/c² region (left) and on $\sigma_H^\pm \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ for the charged Higgs boson mass hypotheses in the 180–600 GeV/c² region (right) obtained when the EWK+ $t\bar{t}$ fake τ background is measured from data.

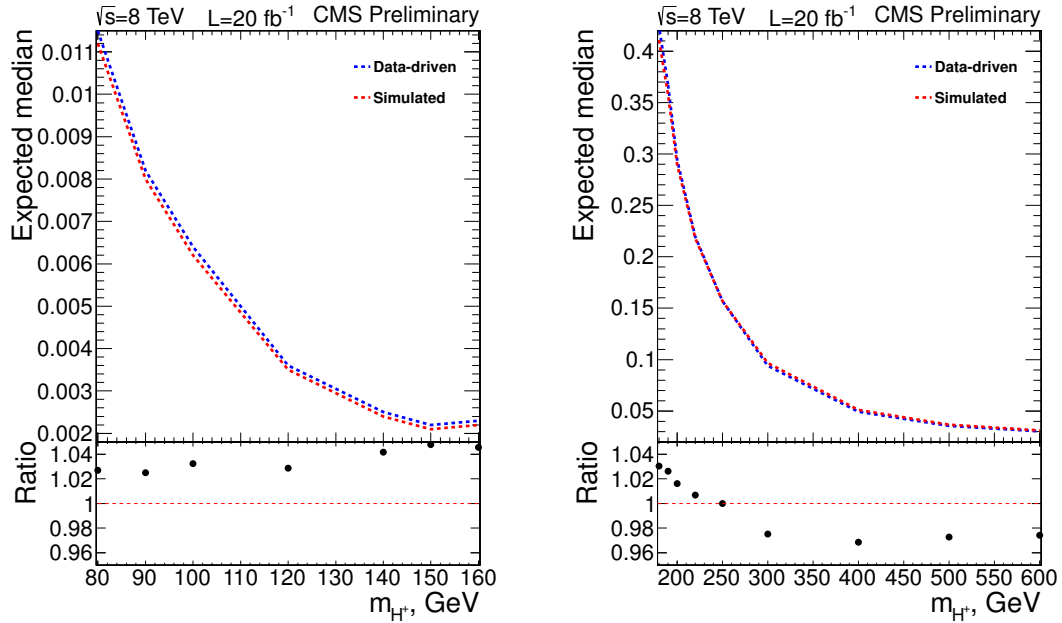


Figure 22: Expected median upper limits obtained with the data-driven and simulated methods along with their ratio.

Table 7: 95% CL median upper limits on $\mathcal{B}(t \rightarrow bH^\pm) \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ for $m_{H^\pm} = 80 - 160$ GeV and the widths of the uncertainty bands.

Simulated EWK+t \bar{t} fake τ					
m_{H^\pm} (GeV)	Expected limit				
	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
80	0.0053	0.0033	0.0112	0.0050	0.0114
90	0.0038	0.0024	0.0080	0.0035	0.0079
100	0.0029	0.0018	0.0062	0.0027	0.0061
120	0.0017	0.0011	0.0035	0.0015	0.0034
140	0.0011	0.0007	0.0024	0.0010	0.0023
150	0.0010	0.0006	0.0021	0.0010	0.0022
155	0.0011	0.0007	0.0023	0.0010	0.0023
160	0.0010	0.0006	0.0022	0.0010	0.0023

Data-driven EWK+t \bar{t} fake τ					
m_{H^\pm} (GeV)	Expected limit				
	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
80	0.0055	0.0034	0.0115	0.0051	0.0116
90	0.0039	0.0025	0.0082	0.0035	0.0081
100	0.0031	0.0019	0.0064	0.0027	0.0063
120	0.0017	0.0011	0.0036	0.0015	0.0035
140	0.0012	0.0007	0.0025	0.0010	0.0024
150	0.0010	0.0006	0.0022	0.0010	0.0022
155	0.0011	0.0007	0.0024	0.0010	0.0023
160	0.0011	0.0007	0.0023	0.0010	0.0023

Table 8: 95% CL median upper limits on $\sigma_{\text{pp} \rightarrow \text{H}^\pm \text{tb}} \times \mathcal{B}(\text{H}^\pm \rightarrow \tau^\pm \nu_\tau)$ for $m_{\text{H}^\pm} = 180 - 600$ GeV and the widths of the uncertainty bands.

Simulated EWK+t $\bar{\text{t}}$ fake τ					
m_{H^\pm} (GeV)	Expected limit				
	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
180	0.197	0.121	0.411	0.182	0.410
190	0.171	0.105	0.358	0.158	0.361
200	0.140	0.087	0.290	0.128	0.296
220	0.105	0.065	0.218	0.098	0.227
250	0.076	0.047	0.157	0.073	0.167
300	0.048	0.030	0.097	0.046	0.108
400	0.027	0.017	0.051	0.028	0.067
500	0.021	0.013	0.037	0.022	0.054
600	0.018	0.012	0.031	0.020	0.049

Data-driven EWK+t $\bar{\text{t}}$ fake τ					
m_{H^\pm} (GeV)	Expected limit				
	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
180	0.202	0.125	0.423	0.184	0.421
190	0.175	0.109	0.367	0.163	0.367
200	0.142	0.088	0.294	0.131	0.302
220	0.106	0.066	0.220	0.098	0.225
250	0.076	0.048	0.157	0.072	0.166
300	0.046	0.029	0.094	0.046	0.107
400	0.025	0.016	0.049	0.028	0.066
500	0.020	0.013	0.036	0.022	0.053
600	0.017	0.011	0.030	0.020	0.049

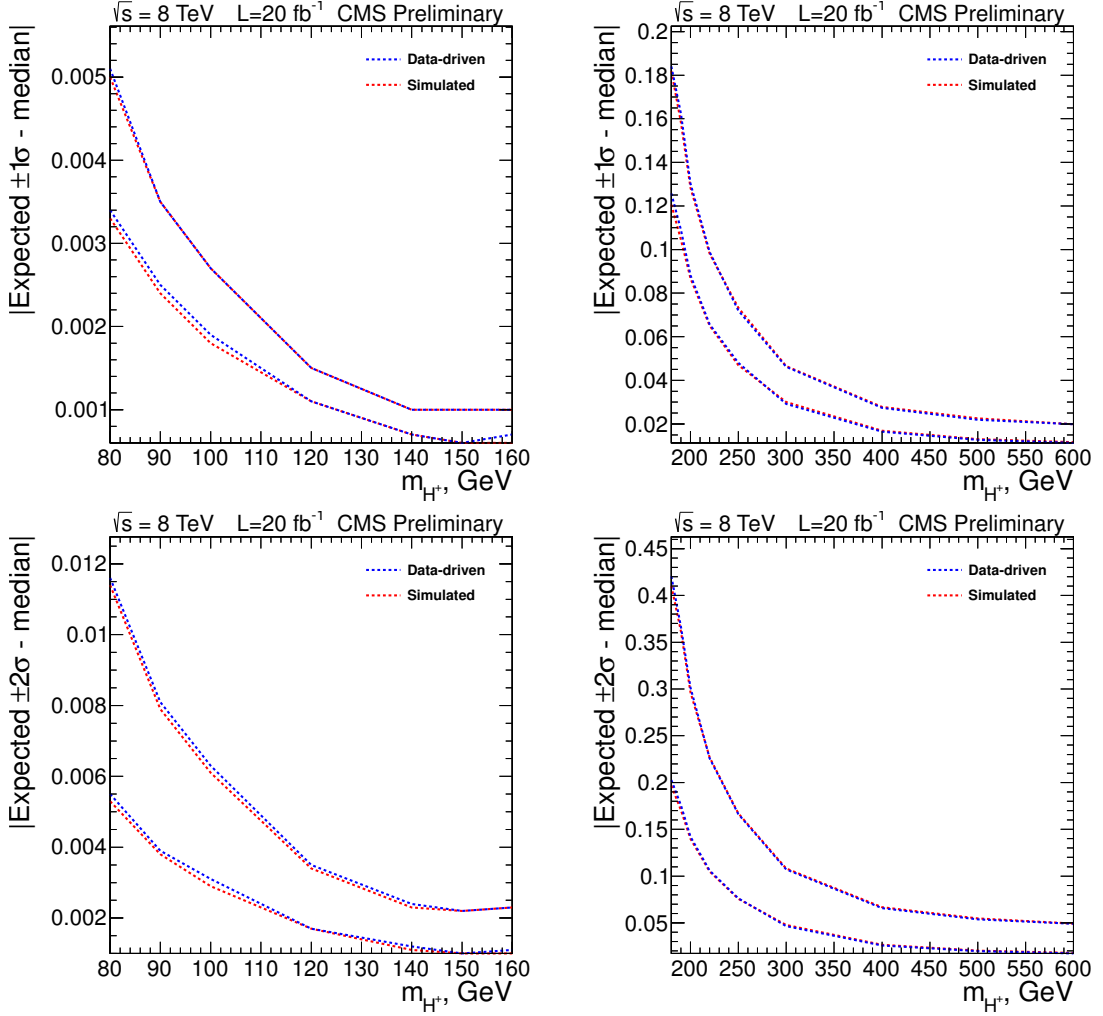


Figure 23: Comparison of the width of the expected one (*top*) and two (*bottom*) standard deviation bands obtained with the data-driven and simulated methods.

Overall, the expected limits can be seen to change little with the developed data-driven method. In the light charged Higgs boson mass range, the expected median limits and their uncertainty bands can be seen to be approximately 3 % higher with the data-driven method. In the heavy charged Higgs boson case, the expected median limit is few percent higher when $m_{H^\pm} < 250$ GeV, but lowered up to ~ 3 % in the opposite case. The uncertainties can be seen to behave similarly: the data-driven method increases the uncertainties when $m_{H^\pm} < 250$ GeV and decreases them when $m_{H^\pm} > 250$ GeV, but overall the changes remain within ~ 3 %.

To gain understanding of the behaviour of the expected limits, the total uncertainties of all the backgrounds obtained with the data-driven and simulated methods are compared as a function of transverse mass in Fig. 24. It can be seen that the uncertainties are generally lower with the data-driven method as the reduced uncertainties in the yields also suggest. The uncertainties can also be seen to

decrease with growing m_T . However, at the mass points $m_T = 110, 150$ GeV and a few other points to a lesser extent, the data-driven method gives larger total uncertainties than the simulated method. This follows from the systematic uncertainties: the 'spike' on the left is caused by the plus-signed systematic uncertainty, while the one on the right follows from the minus-signed uncertainty. The EWK+ $t\bar{t}$ fake τ background has the largest contribution in the transverse mass region $m_T = 100 - 200$ GeV according to Fig. 20, so the changes in this region are anticipated. The variations are relatively small (up to $\sim 6.5\%$) and no single significant source was found for them, so they are assumed to be random fluctuations. Because these fluctuations are located in the primary signal region when $m_{H^\pm} \lesssim 250$ GeV, the expected limits and their uncertainties are higher in this case. When $m_{H^\pm} \gtrsim 250$ GeV, the signal appears at higher transverse mass values, where the data-driven method yields lower uncertainties and the limits are consequently lowered.

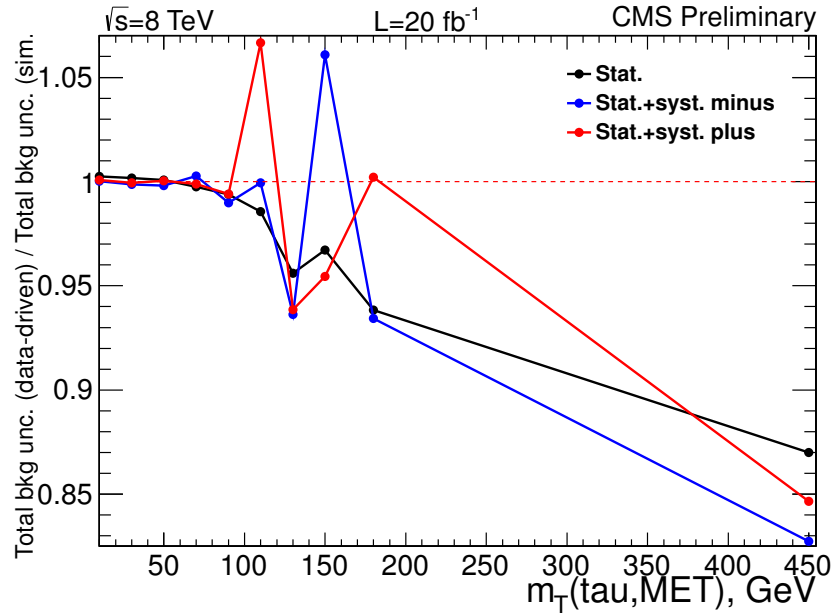


Figure 24: Comparison of the total uncertainty of all the backgrounds obtained with the data-driven and simulated methods as a function of the transverse mass.

10 Conclusions

A method for measuring the EWK+t \bar{t} background with misidentified τ leptons from data has been presented as an alternative for the simulation based estimation method for the background. These background estimation methods were found to be in good agreement, since the transverse mass distributions agree within statistical uncertainties in various different parts of the phase space. This indicates that the developed method is valid.

As the EWK+t \bar{t} fake τ background is measured using the developed data-driven method, all the backgrounds in the analysis are measured from data. Data-driven measurement is generally to be considered to give more reliable estimates than simulation, which was confirmed by the smaller uncertainties in the event yields obtained with the data-driven method: the statistical uncertainties are reduced by 12 % and the systematic uncertainties by 34 %. The reduction of the statistical uncertainty is a natural consequence of the increased sample sizes. The reduction of the systematic uncertainties is a consequence of the anti-correlation of the simulation related uncertainties between the QCD multijet and EWK+t \bar{t} fake τ samples: the latter affects only on the final normalisation of the fake τ background and as these samples are measured inclusively, the contributions of the uncertainties cancel each other partially. The objective of lowering the uncertainties related to the background measurement can thus be considered reached in the light of the event yields.

A preliminary analysis using data from proton-proton collisions at $\sqrt{s} = 8$ TeV recorded by the CMS experiment, corresponding to an integrated luminosity of approximately 19.7 fb^{-1} has been performed to calculate the expected limits on $\mathcal{B}(t \rightarrow bH^\pm) \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ for the charged Higgs boson mass hypotheses in the 80 – 160 GeV region and on $\sigma_H^\pm \times \mathcal{B}(H^\pm \rightarrow \tau^\pm \nu_\tau)$ for the charged Higgs boson mass hypotheses in the 180 – 600 GeV region. In the light charged Higgs boson case the developed data-driven method leads to median expected limits of 1.15 % to 0.23 % on the production branching ratio, while the corresponding limits obtained with the simulated EWK+t \bar{t} fake τ background estimation method range from 1.12% to 0.22%. In the heavy charged Higgs boson case the data-driven limits on the production cross section range from 0.423 pb to 0.030 pb, while the corresponding limits obtained with the simulated EWK+t \bar{t} fake τ background estimation method vary from 0.411 pb to 0.031 pb.

The expected limits obtained with the simulated and data-driven methods are overall very similar, and differ only by few percent. The higher limits and their larger uncertainties obtained with the data-driven method especially in the light charged Higgs boson mass range result from the larger systematic uncertainties in the signal region, when compared to the uncertainties given by the simulated method. These larger uncertainties are assumed to be random fluctuations. In the heavy charged Higgs boson mass range, the data-driven method gives generally

lower uncertainties for the background than the simulated method, which consequently lowers the limits and decreases their uncertainties. A future objective is the closer examination and possible reduction of the larger background uncertainties obtained with the data-driven method in the signal region of the light charged Higgs boson.

An important reason for preferring the data-driven method over the simulation based method is that it increases the background purity obtained with the Inverted method significantly. This is vital for the use of the Inverted method in the LHC Run II, since the increased trigger thresholds used for data taking at higher luminosities are expected to decrease the purity obtained with the method, making it potentially unusable otherwise. The objective of enhancing the performance of the QCD multijet background measurement method can thus also be considered achieved. The future goals accordingly include testing the performance of the method with increased cut thresholds (e.g. $E_T^{\text{miss}} > 80$ GeV) and with the new τ identification algorithm developed for Run II [96].

References

- [1] S. L. Glashow, “Partial-symmetries of weak interactions”, *Nuclear Physics* **22** (1961), no. 4, 579 – 588. [doi:10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [2] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266. [doi:10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [3] A. Salam, “Weak and Electromagnetic Interactions”, in *Elementary particle theory*, N. Svartholm, ed., pp. 367–377. Almquist & Wiksell.
- [4] S. Chatrchyan et al., “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Physics Letters B* **716** (2012), no. 1, 30 – 61. [doi:10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [5] G. Aad et al., “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B* **716** (2012), no. 1, 1 – 29. [doi:10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [6] 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”, [arXiv:1412.8662](https://arxiv.org/abs/1412.8662).
- [7] 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](https://doi.org/10.1103/PhysRevLett.13.321).
- [8] P. Higgs, “Broken symmetries, massless particles and gauge fields”, *Physics Letters* **12** (1964), no. 2, 132 – 133. [doi:10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).
- [9] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. [doi:10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [10] 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](https://doi.org/10.1103/PhysRevLett.13.585).
- [11] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons”, *Phys. Rev.* **145** (May, 1966) 1156–1163. [doi:10.1103/PhysRev.145.1156](https://doi.org/10.1103/PhysRev.145.1156).
- [12] T. W. B. Kibble, “Symmetry Breaking in Non-Abelian Gauge Theories”, *Phys. Rev.* **155** (Mar, 1967) 1554–1561. [doi:10.1103/PhysRev.155.1554](https://doi.org/10.1103/PhysRev.155.1554).
- [13] J. Wess and B. Zumino, “Supergauge transformations in four dimensions”, *Nuclear Physics B* **70** (1974), no. 1, 39 – 50. [doi:10.1016/0550-3213\(74\)90355-1](https://doi.org/10.1016/0550-3213(74)90355-1).
- [14] P. Fayet, “Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino”, *Nuclear Physics B* **90** (1975), no. 0, 104 – 124. [doi:10.1016/0550-3213\(75\)90636-7](https://doi.org/10.1016/0550-3213(75)90636-7).

- [15] P. Fayet, “Supersymmetry and weak, electromagnetic and strong interactions”, *Physics Letters B* **64** (1976), no. 2, 159 – 162.
[doi:10.1016/0370-2693\(76\)90319-1](https://doi.org/10.1016/0370-2693(76)90319-1).
- [16] P. Fayet, “Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions”, *Physics Letters B* **69** (1977), no. 4, 489 – 494. [doi:10.1016/0370-2693\(77\)90852-8](https://doi.org/10.1016/0370-2693(77)90852-8).
- [17] S. Dimopoulos and H. Georgi, “Softly broken supersymmetry and SU(5)”, *Nuclear Physics B* **193** (1981), no. 1, 150 – 162.
[doi:10.1016/0550-3213\(81\)90522-8](https://doi.org/10.1016/0550-3213(81)90522-8).
- [18] K. Inoue, A. Kakuto, H. Komatsu et al., “Low Energy Parameters and Particle Masses in a Supersymmetric Grand Unified Model”, *Progress of Theoretical Physics* **67** (1982), no. 6, 1889–1898. [doi:10.1143/PTP.67.1889](https://doi.org/10.1143/PTP.67.1889).
- [19] K. Inoue, A. Kakuto, H. Komatsu et al., “Aspects of Grand Unified Models with Softly Broken Supersymmetry”, *Progress of Theoretical Physics* **68** (1982), no. 3, 927–946. [doi:10.1143/PTP.68.927](https://doi.org/10.1143/PTP.68.927).
- [20] K. Inoue, A. Kakuto, H. Komatsu et al., “Renormalization of Supersymmetry Breaking Parameters Revisited”, *Progress of Theoretical Physics* **71** (1984), no. 2, 413–416. [doi:10.1143/PTP.71.413](https://doi.org/10.1143/PTP.71.413).
- [21] D. Volkov and V. Akulov, “Is the neutrino a goldstone particle?”, *Physics Letters B* **46** (1973), no. 1, 109 – 110. [doi:10.1016/0370-2693\(73\)90490-5](https://doi.org/10.1016/0370-2693(73)90490-5).
- [22] A. Heister et al., “Search for charged Higgs bosons in e^+e^- collisions at energies up to $\sqrt{s} = 209$ GeV”, *Physics Letters B* **543** (2002), no. 1–2, 1 – 13.
[doi:10.1016/S0370-2693\(02\)02380-8](https://doi.org/10.1016/S0370-2693(02)02380-8).
- [23] P. Achard et al., “Search for charged Higgs bosons at LEP”, *Physics Letters B* **575** (2003), no. 3–4, 208 – 220. [doi:10.1016/j.physletb.2003.09.057](https://doi.org/10.1016/j.physletb.2003.09.057).
- [24] “Search for charged Higgs bosons at LEP in general two Higgs doublet models”, *The European Physical Journal C - Particles and Fields* **34** (2004), no. 4, 399–418. [doi:10.1140/epjc/s2004-01732-6](https://doi.org/10.1140/epjc/s2004-01732-6).
- [25] A. G. et al., “Search for charged Higgs bosons in e^+e^- collisions at $\sqrt{s} = 189 - 209$ GeV”, *The European Physical Journal C* **72** (2012), no. 7.,
[doi:10.1140/epjc/s10052-012-2076-0](https://doi.org/10.1140/epjc/s10052-012-2076-0).
- [26] Particle Data Group Collaboration, “Review of Particle Physics (RPP)”, *Phys.Rev.* **D86** (2012) 010001. [doi:10.1103/PhysRevD.86.010001](https://doi.org/10.1103/PhysRevD.86.010001).
- [27] CMS Collaboration Collaboration, “Search for charged Higgs bosons with the $H^+ \rightarrow \tau\nu$ decay channel in the fully hadronic final state at $\sqrt{s} = 8$ TeV”, Technical Report CMS-PAS-HIG-14-020, CERN, Geneva, (2014).

- [28] M. E. Peskin and D. V. Schroeder, “An Introduction to Quantum Field Theory; 1995 ed.”. Westview, Boulder, CO, 1995.
- [29] K. Olive and P. D. Group, “Review of Particle Physics”, *Chinese Physics C* **38** (2014), no. 9, 090001. doi:[10.1088/1674-1137/38/9/090001](https://doi.org/10.1088/1674-1137/38/9/090001).
- [30] G. Branco, P. Ferreira, L. Lavoura et al., “Theory and phenomenology of two-Higgs-doublet models”, *Phys.Rept.* **516** (2012) 1–102, [arXiv:1106.0034](https://arxiv.org/abs/1106.0034). doi:[10.1016/j.physrep.2012.02.002](https://doi.org/10.1016/j.physrep.2012.02.002).
- [31] N. G. Deshpande and E. Ma, “Pattern of symmetry breaking with two Higgs doublets”, *Phys. Rev. D* **18** (Oct, 1978) 2574–2576. doi:[10.1103/PhysRevD.18.2574](https://doi.org/10.1103/PhysRevD.18.2574).
- [32] J. F. Donoghue and L.-F. Li, “Properties of charged Higgs bosons”, *Phys. Rev. D* **19** (Feb, 1979) 945–955. doi:[10.1103/PhysRevD.19.945](https://doi.org/10.1103/PhysRevD.19.945).
- [33] H. Haber, G. Kane, and T. Sterling, “The fermion mass scale and possible effects of Higgs bosons on experimental observables”, *Nuclear Physics B* **161** (1979), no. 2–3, 493 – 532. doi:[10.1016/0550-3213\(79\)90225-6](https://doi.org/10.1016/0550-3213(79)90225-6).
- [34] J. F. Gunion and H. E. Haber, “Higgs bosons in supersymmetric models (I)”, *Nuclear Physics B* **272** (1986), no. 1, 1 – 76. doi:[10.1016/0550-3213\(86\)90340-8](https://doi.org/10.1016/0550-3213(86)90340-8).
- [35] A. Goudelis, B. Herrmann, and O. Stål, “Dark matter in the inert doublet model after the discovery of a Higgs-like boson at the LHC”, *Journal of High Energy Physics* **2013** (2013), no. 9,. doi:[10.1007/JHEP09\(2013\)106](https://doi.org/10.1007/JHEP09(2013)106).
- [36] H.-J. He, N. Polonsky, and S. Su, “Extra families, Higgs spectrum, and oblique corrections”, *Phys. Rev. D* **64** (Jul, 2001) 053004. doi:[10.1103/PhysRevD.64.053004](https://doi.org/10.1103/PhysRevD.64.053004).
- [37] D. Eriksson, J. Rathsman, and O. Stal, “2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual”, *Comput.Phys.Commun.* **181** (2010) 189–205, [arXiv:0902.0851](https://arxiv.org/abs/0902.0851). doi:[10.1016/j.cpc.2009.09.011](https://doi.org/10.1016/j.cpc.2009.09.011).
- [38] I. F. Ginzburg and M. Krawczyk, “Symmetries of two Higgs doublet model and CP violation”, *Phys.Rev.* **D72** (2005) 115013, [arXiv:hep-ph/0408011](https://arxiv.org/abs/hep-ph/0408011). doi:[10.1103/PhysRevD.72.115013](https://doi.org/10.1103/PhysRevD.72.115013).
- [39] H. E. Haber, “Nonminimal Higgs sectors: The Decoupling limit and its phenomenological implications”, [arXiv:hep-ph/9501320](https://arxiv.org/abs/hep-ph/9501320).
- [40] H. E. Haber, “Toward a set of 2HDM benchmarks”, (2013). Available at <http://scipp.ucsc.edu/~haber/webpage/thdmbench.ps>.
- [41] I. Simonsen, “A Review of minimal supersymmetric electroweak theory”, [arXiv:hep-ph/9506369](https://arxiv.org/abs/hep-ph/9506369).

- [42] A. Djouadi, “The anatomy of electroweak symmetry breaking Tome II: The Higgs bosons in the Minimal Supersymmetric Model”, *Physics Reports* **459** (2008), no. 1–6, 1 – 241. doi:[10.1016/j.physrep.2007.10.005](https://doi.org/10.1016/j.physrep.2007.10.005).
- [43] F.-P. Schilling, “Top Quark Physics at the LHC: A Review of the First Two Years”, *Int.J.Mod.Phys. A* **27** (2012) 1230016, arXiv:[1206.4484](https://arxiv.org/abs/1206.4484). doi:[10.1142/S0217751X12300165](https://doi.org/10.1142/S0217751X12300165).
- [44] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 2. Differential Distributions”, arXiv:[1201.3084](https://arxiv.org/abs/1201.3084). doi:[10.5170/CERN-2012-002](https://doi.org/10.5170/CERN-2012-002).
- [45] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”, arXiv:[1101.0593](https://arxiv.org/abs/1101.0593). doi:[10.5170/CERN-2011-002](https://doi.org/10.5170/CERN-2011-002).
- [46] “CERN Document Server”. <http://cds.cern.ch/>. Accessed: 5.1.2015.
- [47] CMS Collaboration, “The CMS experiment at the CERN LHC”, *Journal of Instrumentation* **3** (2008), no. 08, S08004. doi:[10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [48] CMS Collaboration, “CMS, the Compact Muon Solenoid: Technical proposal”, CERN-LHCC-94-38, CERN-LHCC-P-1 (1994).
- [49] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, Geneva, 2006.
- [50] “CMS detector design”. <http://cms.web.cern.ch/news/cms-detector-design>. Accessed: 9.9.2014.
- [51] G. Bayatian, A. Sirunian, I. Emelyanchik et al., “Design, performance and calibration of the CMS forward calorimeter wedges”, *Eur.Phys.J. C* **53** (2008) 139–166. doi:[10.1140/epjc/s10052-007-0459-4](https://doi.org/10.1140/epjc/s10052-007-0459-4).
- [52] CMS Collaboration, “Particle-flow commissioning with muons and electrons from J/ψ and W events at 7 TeV”, Technical Report CMS-PAS-PFT-10-003, CERN, Geneva, (2010).
- [53] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, (2009).
- [54] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, Technical Report CMS-PAS-PFT-10-001, 2010.

- [55] CMS Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, Technical Report CMS-PAS-PFT-10-002, CERN, Geneva, (2010).
- [56] J. Incandela, “Results from the CMS experiment at the LHC”, (2012). Talk in the 521. WE-Heraeus-Seminar: The First Results from the Large Hadron Collider.
- [57] W. Adam, B. Mangano, T. Speer et al., “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Geneva, (Dec, 2006).
- [58] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987), no. 2–3, 444 – 450. [doi:10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [59] S. Cucciarelli, M. Konecki, D. Kotlinski et al., “Track reconstruction, primary vertex finding and seed generation with the Pixel Detector”, Technical Report CMS-NOTE-2006-026, CERN, Geneva, (Jan, 2006).
- [60] CMS Collaboration, “Tracking and Vertexing Results from First Collisions”, Technical Report CMS-PAS-TRK-10-001, CERN, 2010. Geneva, (2010).
- [61] CMS Collaboration, “Measurement of Tracking Efficiency”, Technical Report CMS-PAS-TRK-10-002, CERN, 2010. Geneva, (2010).
- [62] CMS Collaboration, “Measurement of Momentum Scale and Resolution of the CMS Detector using Low-mass Resonances and Cosmic Ray Muons”, Technical Report CMS-PAS-TRK-10-004, CERN, 2010. Geneva, (2010).
- [63] W. Erdmann, “Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering”, *CMS IN-2011/014*.
- [64] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239. [doi:10.1109/5.726788](https://doi.org/10.1109/5.726788).
- [65] C. Collaboration, “Performance of CMS muon reconstruction in cosmic-ray events”, *Journal of Instrumentation* **5** (2010), no. 03, T03022. [doi:10.1088/1748-0221/5/03/T03022](https://doi.org/10.1088/1748-0221/5/03/T03022).
- [66] T. C. collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *Journal of Instrumentation* **7** (2012), no. 10, P10002. [doi:10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002).
- [67] R. Frühwirth, “Track fitting with non-Gaussian noise”, *Computer Physics Communications* **100** (1997), no. 1–2, 1 – 16. [doi:10.1016/S0010-4655\(96\)00155-5](https://doi.org/10.1016/S0010-4655(96)00155-5).

- [68] W. Adam, R. Frühwirth, A. Strandlie et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.
[doi:10.1088/0954-3899/31/9/N01](https://doi.org/10.1088/0954-3899/31/9/N01).
- [69] CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7$ TeV”, *CMS PAS EGM-10-004* (2010).
- [70] CMS Collaboration Collaboration, “Studies of Tracker Material”,.
- [71] CMS Collaboration Collaboration, “Isolated Photon Reconstruction and Identification at $\sqrt{s} = 7$ TeV”,.
- [72] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063.
[doi:10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [73] CMS Collaboration Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002, [arXiv:1107.4277](https://arxiv.org/abs/1107.4277). [doi:10.1088/1748-0221/6/11/P11002](https://doi.org/10.1088/1748-0221/6/11/P11002).
- [74] C. Weiser, “A Combined Secondary Vertex Based B-Tagging Algorithm in CMS”, Technical Report CMS-NOTE-2006-014, CERN, Geneva, (Jan, 2006).
- [75] CMS, “Performance of τ -lepton reconstruction and identification in CMS”, *Journal of Instrumentation* **7** (January, 2012) 1001, [arXiv:1109.6034](https://arxiv.org/abs/1109.6034).
[doi:10.1088/1748-0221/7/01/P01001](https://doi.org/10.1088/1748-0221/7/01/P01001).
- [76] CMS Collaboration, “Study of tau reconstruction algorithms using pp collisions data collected at $\sqrt{s} = 7$ TeV”, *CMS Physics Analysis Summary CMS-PAS-PFT-10-004* (2010).
- [77] T. Sakuma, “Missing E_T (MET) Reconstruction with the CMS Detector”, (2012). Available at <https://indico.ads.ttu.edu/getFile.py/access?contribId=76&sessionId=6&resId=0&materialId=slides&confId=3>.
- [78] T. C. collaboration, “Missing transverse energy performance of the CMS detector”, *Journal of Instrumentation* **6** (2011), no. 09, P09001.
[doi:10.1088/1748-0221/6/09/P09001](https://doi.org/10.1088/1748-0221/6/09/P09001).
- [79] S. Jadach, Z. Was, R. Decker et al., “The tau decay library TAUOLA: Version 2.4”, *Comput. Phys. Commun.* **76** (1993) 361–380.
[doi:10.1016/0010-4655\(93\)90061-G](https://doi.org/10.1016/0010-4655(93)90061-G).
- [80] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl.Instrum.Meth.* **A506** (2003) 250–303.
[doi:10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [81] F. Maltoni and T. Stelzer, “MadEvent: Automatic event generation with MadGraph”, *JHEP* **02** (2003) 027, [arXiv:hep-ph/0208156](https://arxiv.org/abs/hep-ph/0208156).

- [82] T. Sjostrand, L. Lonnblad, S. Mrenna et al., “Pythia 6.3 Physics and Manual”, *LU TP* **03-38** (2003) [arXiv:hep-ph/0308153](#).
- [83] S. Alioli, P. Nason, C. Oleari et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *Journal of High Energy Physics* **2010** (2010), no. 6,. [doi:10.1007/JHEP06\(2010\)043](#).
- [84] C. Collaboration, “Identification and filtering of uncharacteristic noise in the CMS hadron calorimeter”, *Journal of Instrumentation* **5** (2010), no. 03, T03014. [doi:10.1088/1748-0221/5/03/T03014](#).
- [85] M. Kortelainen, “Search for a Light Charged Higgs Boson in the CMS Experiment in pp Collisions at $\sqrt{s} = 7$ TeV”, *HIP-2012-06* (2012). Doctoral dissertation.
- [86] D. P. Roy, “The hadronic tau decay signature of a heavy charged Higgs boson at LHC”, *Phys. Lett.* **B459** (1999) 607–614, [arXiv:hep-ph/9905542](#). [doi:10.1016/S0370-2693\(99\)00724-8](#).
- [87] D. Roy, “Looking for the charged Higgs boson”, *Mod.Phys.Lett.* **A19** (2004) 1813–1828, [arXiv:hep-ph/0406102](#). [doi:10.1142/S0217732304015105](#).
- [88] J. Allison, “The Road to Discovery”, ch. 9. Springer Theses, 2014.
- [89] CMS Collaboration Collaboration, “Search for a heavy charged Higgs boson in proton-proton collisions at $\sqrt{s}=8$ TeV with the CMS detector”, Technical Report CMS-PAS-HIG-13-026, CERN, Geneva, (2014).
- [90] CMS Collaboration, “Performance of τ -lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, [arXiv:1109.6034](#). [doi:10.1088/1748-0221/7/01/P01001](#).
- [91] A. L. Read, “Presentation of search results: the CL_s technique”, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693. [doi:10.1088/0954-3899/28/10/313](#).
- [92] A. L. Read, “Modified frequentist analysis of search results (The $CL(s)$ method)”,.
- [93] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl.Instrum.Meth.* **A434** (1999) 435–443, [arXiv:hep-ex/9902006](#). [doi:10.1016/S0168-9002\(99\)00498-2](#).
- [94] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, (Aug, 2011).

- [95] “Combine”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit>. Accessed: 8.10.2014.
- [96] “Software Guide for Tau Reconstruction: Tau ID 2014 (preparation for Run II)”. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePFTauID#Tau_ID_2014_preparation_for_Run. Accessed: 26.1.2015.