RWTH AACHEN UNIVERSITY

DOKTORARBEIT

Search for new physics in the final state with a tau lepton and missing transverse momentum at CMS

Autor:
M. Sc. Christoph SCHULER

Erstgutatcher:
Prof. Dr. rer. nat. Thomas HEBBEKER

Zweitgutachter:
Prof. Dr. rer. nat. Oliver POOTH

Disseration zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegt

der

Fakultät für Informatik, Mathematik und Naturwissenschaften
RWTH Aachen University
III. Physikalisches Institut A

30th January 2023

Declaration of Authorship

I, M. Sc. Christoph SCHULER, declare that this thesis titled, 'Search for new physics in the final state with a tau lepton and missing transverse momentum at CMS' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

RWTH AACHEN UNIVERSITY

Abstract

Fakultät für Informatik, Mathematik und Naturwissenschaften
III. Physikalisches Institut A

Doctor of Science

Search for new physics in the final state with a tau lepton and missing transverse momentum at CMS

by M. Sc. Christoph SCHULER

This thesis presents a search for new physics using data taken by the Compact Muon Solenoid (CMS) at the CERN Large Hadron Collider (LHC). Data from proton-proton collisions taken in the years 2016 to 2018, corresponding to an integrated luminosity of 138 fb^{-1}, are analyzed. It is searched for a sign of new physics in events which exhibit kinematics similar to that of Standard Model W boson production, where the W boson decays to a tau lepton and a neutrino. The tau lepton decays quickly, and only its decay products can be measured with the CMS detector. This thesis focuses on the hadronic decay channels of the tau lepton. Neutrinos are not reconstructable with the CMS detector. Therefore, events which contain missing transverse momentum, a quantity which provides an indirect handle for such particles, are selected.

The measured data is compared to the background expected from known Standard Model processes. No significant deviation is observed. The results are interpreted in the terms of various models. Special interest lies on interpretations which predict enhanced couplings of intermediate particles to tau leptons. A lower mass limit of 4.7 TeV is set on the production of a SM W boson like new heavy charged vector bosons. In this context, limits are placed on varying coupling strengths of such a particle w.r.t. the SM boson, and on a non-universal gauge interaction model, which predicts enhanced branching fractions of such a new boson towards decays of tau leptons. Production of charged quantum black holes is excluded up to masses of 6.6 TeV. Additionally, an exclusion limit on different coupling types of an effective field theory model and a model-independent limit are provided. An additional outlook for future prospects of this analysis during the high luminosity LHC era is given. This study was performed for a proton-proton data set corresponding to an integrated luminosity of 3000 fb^{-1} at a center of mass energy of 14 TeV. A simplified analysis approach is used to estimate the sensitivity which can be achieved by an analysis using this data set.

RWTH AACHEN UNIVERSITY

Abstract

Fakultät für Informatik, Mathematik und Naturwissenschaften
III. Physikalisches Institut A

Doctor of Science

Search for new physics in the final state with a tau lepton and missing transverse momentum at CMS

by M. Sc. Christoph SCHULER

In dieser Dissertation wird eine Suche nach neuer Physik vorgestellt, basierend auf Daten, welche mit den Compact Muon Solenoid (CMS) am Large Hadron Collider (LHC) am CERN aufgenommen wurden. Es werden Daten von Proton-Proton Kollisionen, welche in den Jahren 2016 bis 2018 aufgenommen wurden, analysiert. Die Menge der Daten entspricht einer integrierten Luminosität von 138 fb^{-1}. Die Suche nach neuer Physik fokussiert sich auf solche Ereignisse, welche kinematisch ähnlich sind zu solchen, welche man von W Boson Produktion aus den Standard Modell der Teilchenphysik erwarten würde, wenn das W Boson in ein Tau Lepton und ein Neutrino zerfällt. Das Tau Lepton zerfällt schnell und nur dessen Zerfallsprodukte können mit dem CMS Detektor gemessen werden. Diese Dissertation betrachtet nur solche Events, in welchen das Tau Lepton hadronisch zerfällt. Mithilfe des CMS Detektors sind Neutrinos nicht rekonstruierbar. Eine Möglichkeit, diese Objekte dennoch indirekt nachweisen zu können, bietet fehlender transversaler Impuls, welcher in den Ereignissen die hier analysiert werden, vorkommen muss.

Die gemessenen experimentellen Daten werden verglichen mit der Hintergrunderwartung von Prozessen die bereits aus dem Standard Modell bekannt sind. Keine signifikante Abweichung von dieser Erwartung wurde gemessen. Dieses Ergebnis wird nun interpretiert im Rahmen verschiedener Modelle, welche neue Physik vorhergesagt haben. Besonders berücksichtigt werden in diesem Rahmen Modelle, welche verstärkte Kopplungen von intermediären Teilchen zu Tau Leptonen haben. Die Produktion eines dem SM W-Boson ähnlichen neuen, schweren und geladenen Vektor Bosons wird bis zu einer Masse dieses Teilchens von 4.7 TeV ausgeschlossen. Ebenfalls untersucht wird der Einfluss der Kopplungsstärke eines solchen neuen Teilchens, relativ zu der Kopplungsstärke des SM W-Bosons. Betrachtet wird auch ein Modell, in welchem die Kopplungen zu Leptonen nicht universell sind, sondern Zerfälle zu Tau Leptonen präferiert sind. Ausschlussgrenzen werden für beide dieser Modelle bereitgestellt. Die Produktion von geladenen mikroskopischen schwarzen Löchern wird bis zu einer Grenzwertmasse von 6.6 TeV ausgeschlossen. Zusätzlich werden noch Ausschlussgrenzen auf die Kopplungsstärken in einer effektiven Feldtheorie ausgerechnet, sowie eine modell-unabhängige Ausschlussgrenze bestimmt. Zusätzlich enthält diese Dissertation einen Ausblick auf Ergebnisse, welche eine ähnliche Analyse mit dem erreichen kann, was in der Zeit des high luminosity LHC gemessen werden soll. Diese Studie wurde entwickelt unter der Annahme, dass ein Datensatz von Proton-Proton Kollisionen entsprechend 3000 fb^{-1} bei einer Schwerpunktsenergie von 14 TeV aufgenommen werden kann. Eine vereinfachte Studie enthält die Sensitivität, die mit einem solchen Datenvolumen erreicht werden kann.

Acknowledgements

I wanted to express my gratitude to all the people who helped me during my time when preparing this thesis. First and foremost I want to thank Prof. Hebbeker for welcoming me into the III A. Institute and allowing me to work on this thesis in the first place. It cannot be overstated how good and free the atmosphere for scientific research is at the institute. I also want to thank Prof. Pooth for agreeing to be the second reviewer of this thesis. I also want to say a thank you to all the members of the institute, and explicitly thank specific persons. First to mention is Kerstin Hoepfner, who helped me from start to finish with discussions and input for my thesis and took quite an effort to help me in the lengthy and arduous publication procedure in the CMS collaboration. Thank you to Sebastian Wiedenbeck, with whom I developed all the code which runs the NanoAOD analysis from scratch. In team work and discussions we helped each other and constantly improved upon our individual results. I also want to thank Swagata Mukherjee, who kick-started me into procedures and methods used within the analysis, and was open for questions from my side at any time. Of course, I want to thank all the members of the CMS experiment for providing all the resources necessary to analyze their high quality data. This thesis would not have been possible in this way without the decades of work made by the members of this group. Special thanks to the members of the EXO group and the TauPOG, whose members helped me in progressing the analysis. Of course, the highly computing intensive work of a full CMS data analysis necessitates a highly functional network and computing environment. For their continuous work in maintaining and updating all the necessary tools, special thanks to the members of the Aachen computing system, explicitly Thomas Kress and Andreas Nowack. At last, many thanks to my family who endured me when my nerves ran out and supported me through all the highs and lows of the publication and the writing of my thesis.

Contents

Chapter 1

Introduction

1.1 Preface

This thesis presents a search for new physics in the particle physics field of research. This field of research historically evolved from the search for the fundamental constituents of the universe. It was found that macroscopic objects consist of microscopic particles. Particle physicists try to understand these tiny objects and their interactions, and many experiments are and have been conducted which offered insight into the fundamental laws of nature. In time, scientists have developed a highly performant theory which describes these laws. However, there are still unanswered questions, and extension and modifications to the existing theory have been proposed to provide answers. In this context, it is interesting to search for new, possibly unthought of interactions and particles. This thesis presents such a search for new physics by analyzing the particles produced in proton-proton collisions.

The basic principle to search for such objects is to analyze the data and compare it to the expectation of what we already learned in previous experiments. By finding differences between expectations and experimental data, possibly new insights can be gained. Many steps are necessary to ensure that the data is reliable and that the estimation of the expected background is accurate. These are described in the context of this thesis before analyzing the results with regard to the existence of new physics.

There are, in general, two different approaches to search for new physics. Dedicated searches, as presented in this thesis, look quite precisely into specific sub-parts of experimental data where an occurrence of differences to our expectation is well motivated. A more generic approach, searching for differences across hundreds of different possible combinations of particles, is also interesting and pursued [1, 2]. This second method excels at broad coverage but is by design not as sensitive as dedicated searches.

In this thesis, a dedicated search is presented, analyzing hundreds of millions of single proton-proton collision data events, measured with the Compact Muon Solenoid (CMS) detector located at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN)[1]. The analyzed data set was measured in the years 2016, 2017 and 2018 - commonly referred to the Run-2 era of the CMS. A schema of the amount of data taken by the detector is shown in Fig. 1.1.

1.1.1 The final state

When two protons of sufficient energy collide, different interactions producing various particles take place in a very short time. The first categorization of the results of such a collision is to list detectable particles (e.g. the so called τ *lepton*) on the one side and non-detectable (*invisible*) objects on the other side. In the context of particle physics detectors, an

[1] The acronym CERN originates from the founding of the organization, which was initiated by the *Conseil européen pour la recherche nucléaire*.

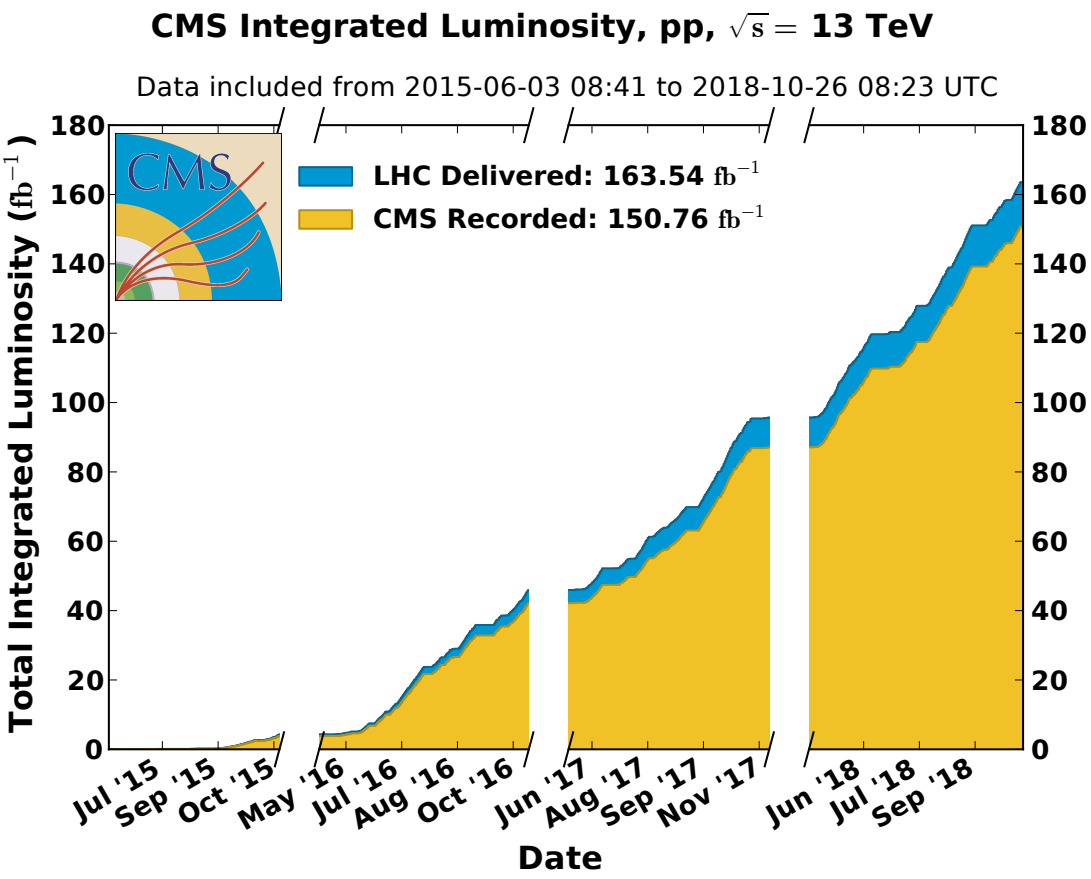

FIGURE 1.1: Integrated luminosity taken by the CMS detector in the Run-2 era. The data taking period shown here includes data from 2015, which will not be part of the analysis presented here. Image taken from [3].

invisible object is something which cannot be measured by the detector used. A known invisible particle for the CMS detector is the neutrino - it can be measured by large, dedicated detectors (e.g. [4, 5]), but its interaction probability is so small, that large fluxes of neutrinos are necessary. The CMS detector is capable of measuring strong and electromagnetic interacting particles - the occurrence of undetected particles is inferred via *missing transverse momentum*. This thesis looks in the part of data, where a τ lepton and missing transverse momentum are reconstructed. The most compelling reason to search in explicitly this final state is that recent results from other experiments have found interesting evidence towards new physics in decay ratios of particles which decay (among others) to τ leptons [6, 7, 8, 9, 10]. This might originate from new physics at a high energy scale, which would lead to an enhanced number of events in the final state with a τ lepton and missing transverse momentum. Various different models which predict such event rates are explored in the scope of this thesis.

1.1.2 Objects at collider experiments

Compared to 'light' charged leptons (electrons and muons), prompt τ leptons are not measured directly at the CMS detector. The tau lepton has mass of $\approx$ 1.78 GeV, a short lifetime (mean lifetime $t = 290.3 \pm 0.5 \times 10^{-15}$) and therefore a small typical decay length of $ct = 87.03$ ¯m s [11]. Therefore, a tau lepton decays very quickly and only its decay products can be directly measured. A distinction is made between the possible decays the τ can undergo. In $\approx$ 35% of cases, it decays to a light lepton and two neutrinos [11]. In all other cases, it decays hadronically - which includes either one or three charged hadrons being emitted. Decay modes with 5 or more charged hadrons are very rare ($<$ 0.01%) and are neglected here. Due to its short lifetime, a leptonically decaying τ can hardly be distinguished from events with just a light lepton and missing transverse momentum. This is due to the fact, that the additional neutrinos from the τ decay all just contribute to one combined sum of missing transverse momentum. This thesis looks into the large decay branch of hadronically decaying τ leptons (τ_h).

Neutrinos only interact via the weak force, which makes them very difficult to detect. It is very unlikely to measure a promptly produced neutrino at the CMS. However, one can use the fact that colliding protons have no (significant) momentum radially to the beam pipe. Conservation of momentum therefore requires that after the collision, this still is the case. By analyzing all outgoing particles and summing their momentum, the transverse momentum of all non-detected particles can therefore be inferred. This is referred to as missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$). This is schematically shown in Fig. 1.2.

For this analysis, this missing transverse momentum originates from two neutrinos: one from the decay of the initial interaction and one more from the decay of the tau lepton. The decay products of the tau lepton, including the neutrino, are highly collimated around the axis of the decaying tau lepton. This means that neutrinos from the tau decay travel roughly back-to-back with the neutrino of the original interaction.

Structure of this thesis

Following this introductory chapter, the theoretical basis necessary to understand the proceedings of this thesis will be explained. Chapters 3 and 4 will focus on the experimental setup and the methods used to extract information on kinematic variables of reconstructed objects. A precise overview on the analyzed data set and the simulation of the expected background in this search is given in chapter 5. This is followed by a small overview on the different simulated signal models in order to find good variables to discern them from the

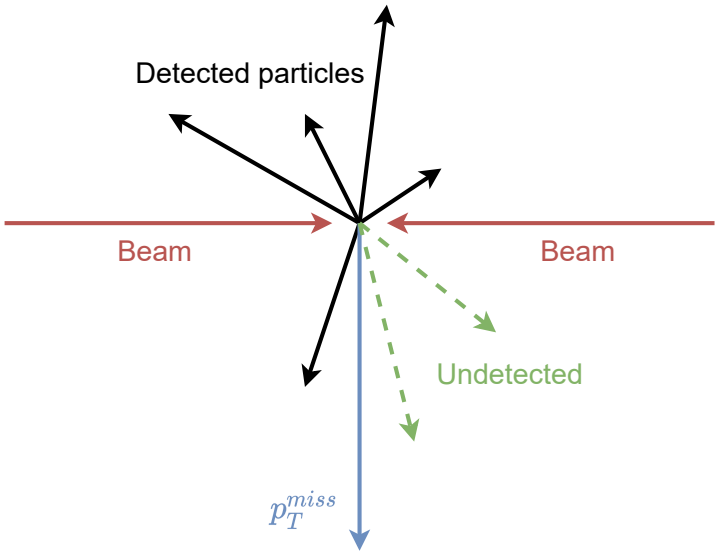

FIGURE 1.2: Schematic sketch of the indirect reconstruction of (the sum of all) invisible particles. Dashed green lines represent undetected objects while bold black lines show detected particles. $p_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as the negative sum of the transverse momenta of all detected particles. Compared to the 'true' vector sum of the missing object, only its projection on the transverse plane can be estimated.

expected background. Chapter 6 contains an in-depth discussion on the analysis strategy, including all considered systematic uncertainties. The results shown at the end of this chapter will then be statistically interpreted with respect to the different signal types in chapter 7. This chapter also contains a small outlook on what can be expected in future LHC searches in a similar analysis. A summary of all results can be found in the final chapter 8.

This thesis covers the analysis of three data taking years. The per-year data sets have been analyzed individually, before being combined to the presented Run-2 result. Many procedures are identical between the years - in this case, exemplary distributions from a selected year are shown. In case the procedure differs between years, it is shown explicitly. The results of this thesis are being made public by the CMS Collaboration [12]. Figures which are not included in the CMS publication are labeled as '*Private Work*'. The CMS publication also includes the analysis of an additional *leptoquark* signal model [13, 14, 15]. While the analysis of that signal model is based on the results of this search, most of the work made there was not performed by me, so it is not included in this thesis.

Chapter 2

The Standard Model and Beyond

Concepts and expressions commonly used in particle physics are referred to frequently in this thesis. This chapter is intended to give a short overview over the most relevant aspects of the current understanding of the fundamental particles and forces of nature categorized under to the so called 'Standard Model' (SM) of particle physics. This model has been developed starting in the 1960s and its predictions have been confirmed experimentally numerous times (see below). At this point in time, the model has such a strong explanatory power that most new theories in particle physics are designed as additions or enhancements to the SM. Some of these enhancing theories, which go beyond the standard model (BSM), are covered in the second part of this chapter. The focus is put on BSM models which are of interest for the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state, but there are many more models which all have their own appealing concepts and explanations.

2.1 Introduction to the Standard Model

The Standard Model of particle physics describes the properties and interactions of the most fundamental particles of nature. These non-composite particles are the fermions (half-integer spin particles) and bosons (integer spin particles). The described forces between these particles are the electromagnetic (EM) force, the weak force and the strong force. There are 12 different fermions, split into six quarks and six leptons. The tau lepton is the one with the highest mass amongst the leptons with a mass of $\approx$ 1776 MeV/c^2. Each particle has a corresponding anti-particle with the same mass but inverse charge(s), e.g. for the tau lepton, which has an electrical charge of -1, an anti-tau lepton with an electrical charge of +1 exists[1]. The leptons can be further split into charged leptons (electrons e, muons μ, and taus τ) and uncharged leptons, called neutrinos (ν_e, ν_μ, ν_τ). Different particle types interact with different forces: the strong force only interacts with quarks (and gluons), the EM force interacts with quarks and charged leptons and the weak force interacts with all these particles.

Each force in the SM is transmitted through mediator particles, notably the photon γ is the mediator for the electromagnetic force, gluons g are the mediators for the strong force and the W and Z boson carry the weak force. From these mediators, only the W and Z bosons have a mass, due to their interaction with the Higgs field.

The Standard Model is widely successful. Many of its predictions have been confirmed experimentally, and in some cases agree with the data to an almost ludicrous precision. Some of the largest achievements include the prediction of the Higgs boson, which was predicted over 50 years ago [16, 17] and finally observed by the LHC experiments in 2012 [18, 19]. Another impressive achievement is the precise theoretical prediction of the magnetic dipole moment of the electron, which agrees with experimental data to a precision higher

[1] Since particles and anti-particles are otherwise identical under charge conjugation w.r.t. decay branching fractions, masses and lifetimes, they are no longer distinguished for the purposes of this thesis unless specifically noted.

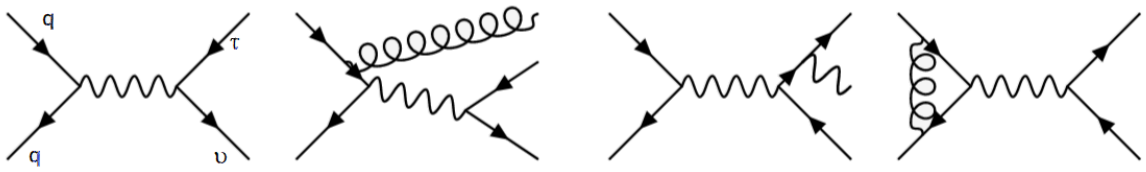

FIGURE 2.1: Examplatory display of possible next-to leading order contributing s-channel Feynman diagrams compared to the leading order Feynman diagram for the $qq \rightarrow W \rightarrow \tau\nu$ process.

than on in a billion [11]. This precision is in and of itself a statement about both the accuracy from a theoretical point, and about the experimental accuracy achievable in some of these experiments conducted to test the SM[2]. Many other aspects of the SM have been tested and confirmed, making it the probably most tested theory in human history.

2.1.1 Mathematical Description

Fundamentally, the SM is a quantum field theory, whose fields fulfil the laws of special relativity. All particles and interactions originate to corresponding fields, which pervade space-time. The SM is a gauge theory consisting of three gauge groups, the $U(1)_Y$, $SU(2)_L$ and a $SU(3)_C$. The corresponding charges are called the weak hypercharge Y, the weak isospin L and the color charge C. The $U(1)_Y \times SU(2)_L$ group structure gives rise to the photon, Z and W bosons.

Calculations of cross sections are performed using perturbation theory. This is possible, since higher orders are suppressed by k^2, where k is the coupling strength of the specific interaction. For electroweak interactions, this value is much smaller than 1, making the *leading order* (LO) calculation a good approximation. Higher order calculations (e.g. *next-to*-LO (NLO) and *next-to-next-to*-LO (NNLO) calculations) have many more contributing processes, since the number of contributing Feynman diagrams increases exponentially. An example for this is given in Fig. 2.1. Therefore, exact computations become very complex rapidly and therefore time and resource consuming.

The two aspects of weak interactions and electromagnetic interactions are combined to a unified electroweak theory. In this approach, there are four vector bosons: three vector bosons W_i which couple with coupling strength g to the weak isospin and one additional boson B which couples with coupling strength $g'/2$ to the weak hypercharge. The two coupling strengths are not independent on each other, but have to fulfil $g\sin(\theta_w) = g'\cos(\theta_w)$, with θ_w being the weak mixing angle. In this approach, the physical bosons $\gamma(A), W^{\pm}, Z$ can be expressed as combinations:

$$A = B\cos(\theta_w) + W_3\sin(\theta_w) \tag{2.1}$$

$$Z = -B\sin(\theta_w) + W_3\cos(\theta_w) \tag{2.2}$$

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.3}$$

This emergence of physical fields from an underlying common electroweak force requires additional explanation. It is not clear why in this description the photon is massless while the W and Z bosons are so heavy. This can be explained by a spontaneously broken symmetry of the underlying gauge group $\mathcal{SU}(2)$. This is achieved by introducing a scalar

[2]It has to be mentioned though, that a similar measurement has been performed for muons and a difference of several standard deviations w.r.t. the theoretical prediction has been reported [20, 21].

field Φ, which gives rise to a new fundamental particle - the Higgs boson. The interaction with this elementary spin-0 particle generates the masses of the heavy gauge bosons and all fermions. Since photons and gluons do not interact with the Higgs boson, they remain massless.

While quantum chromodynamics is a large and interesting topic, for the purpose of this thesis only basic concepts are necessary to understand the proceedings. The mediators of this force, gluons, carry the charge they interact with themselves. This leads to the fact that the strong coupling α_s has a strong dependence on the energy scale - it becomes very small for high energies but $\mathcal{O}(1)$ at low energies. This means that the perturbative approach to calculate strong interactions becomes non-feasible at this energy scale. One consequence of this is *confinement*: quarks and gluons cannot be observed freely, only color charge neutral particles are detected. Another effect is the formation of *jets* at collider experiments: High energy outgoing quarks/gluons quickly create a large shower of hadrons, which travel roughly into the same direction. Due to the LHC being a hadron collider, many of such events are created in one collision, and cross sections for such processes are high.

Another important aspect to mention is the fundamental nature of collisions discussed in this thesis. Protons are composite objects and the high beam energy makes it possible to resolve their substructure. For the purposes of this thesis, *hard* interactions are of interest, which are those events, where large momentum transfer occurs. For such events $ab \to n$ the interaction probability can be expressed as [22]:

$$\sum_{a,b} \int_0^1 dx_a dx_b \int f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\hat{\sigma}_{ab\to n}(\mu_F, \mu_R) \tag{2.4}$$

In this equation, f_a, f_b are parton distribution functions (PDFs), which denote the likelihood of finding a certain parton (quark or gluon) with a relative momentum x inside a proton. μ_F and μ_R are the factorization and renormalization scales, respectively. These are both 'artificial' scales introduced to avoid divergences of the integral, which appear either due to loop integrations (in case of μ_R) or due to infrared corrections (for μ_F). Both these parameters have no physical meaning, but are absorbed by other physical quantities. The coupling constant α_S can be defined dependent on the renormalization scale, while the factorization scale is embedded in the PDFs. If one were to sum the full perturbation series, both scales would vanish. Since this is not feasible, the choice of these scales have an unresolved impact on the result of the calculation. In order to address this problem, both these scales introduce uncertainties, which have to be accounted for (see Sec. 6.4).

The PDFs itself are unfortunately not derivable on a theoretical basis alone. The functions depend on the scale Q^2 and the Bjorken-x variables, and while Q^2 can be derived theoretically, the dependence on x has to measured. This has been done in the past by various experiments, with todays knowledge originating from deep inelastic scattering data from HERA [23] and those of previous LHC measurements [24]. A significant amount of work goes into the determination of these variables and estimations of PDFs at the LHC is a topic of constant improvement and review [25].

2.2 Beyond the Standard Model

While the Standard Model has been hugely successful in predicting the outcome of many experiments, there are some observed phenomena which have no explanation in the SM. One of these phenomena is the measurement of neutrino flavor oscillations [26]. It was measured, that the contribution of non-electron neutrinos originating from the sun is much larger than expected, indicating that the original electron neutrinos changed their lepton

flavor. This observation implies that neutrinos have a finite mass, which is not modeled in the SM.

Another significant deviation from the SM prediction was observed in the measurement of the branching fractions of B meson decays, by the BaBar [27, 28], LHCb [29, 30, 31, 32, 33, 34, 35] and Belle collaborations [8, 36, 37]. In various decay channels it was observed that the rate of decays seems significantly enhanced towards decays with tau leptons compared to decays to light leptons. This is unexpected from the SM perspective, which implies lepton universality, i.e. differences with regard to different lepton flavors should only originate to the different masses of the leptons. While these measurements are quite significant, they are not yet strong enough to rule out statistical fluctuations.

A fascinating topic is the existence of dark matter - different observations in astrophysics indicate the existence of large quantities of non-luminous matter in the universe. The first measurement on this topic comes from 1933. From analyzing the redshifts from the *Coma* cluster it was estimated that the mass of that cluster is 400 times higher than what could be seen [38]. More recently, dark matter could explain measurements of unexpected rotational speeds of galaxies [39]. All this indirect evidence suggests that there might be some neutral particle with significant mass left over from the Big Bang.

Additionally, there is a more theoretical issue on the structure of the SM called the *hierarchy problem*. This problem is related to the vastly different scales of the gravitational force and the electroweak force. The measured Higgs mass value of ≈ 125 GeV is not even close to its expected scale near the Planck mass $M_{Pl} \approx 10^{19}$ GeV. This can be explained with effects originating from higher order loop diagrams, but this requires a significant amount of *fine-tuning* of parameters to work. While one might argue, that this is purely a cosmetic problem, it could be an indication of a missing piece of the puzzle and warrants further investigation.

This thesis is interested in physics beyond the Standard Model which would predominantly occur in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state. Thus, various models which predict the simultaneous creation of a τ lepton and an invisible particle (e.g. a neutrino) are looked into. Of special interest are models which predict enhanced branching fractions towards τ leptons compared to electrons or muons, which could explain the aforementioned observations from LHCb and Belle. Typically, new physics models predict the existence of some kind of new particle, which can then be searched for. In the following subsections, several models are introduced, which can address one or more of the abovementioned open questions. The different model types discussed here are the Sequential Standard Model (SSM), non-universal gauge interaction models (NUGIMs), a quantum black hole (QBH) interpretation and an effective field theory description (EFT).

2.2.1 Heavy charged vector boson W'

The most straightforward attempt at a new particle which could be found in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state is a W boson as already known to the SM, but with a different mass. Such a particle is often referred to as a W'. The corresponding Feynman diagram for this process is shown in Fig. 2.2. In its most simplistic approach, this new particle has all the same coupling values and parameters as the SM W boson except for its mass. Therefore, it can (amongst others) decay leptonically to $\tau + \nu_\tau$. The benchmark model assuming the existence of such a W' boson (and a logically connected Z' boson) is called the Sequential Standard Model [40]. Due to the assumed high mass of the new W' boson, additional decay channels become possible, like the decay to a top and a bottom quark or to W+Z bosons. The branching fraction of the latter decay channel highly depends on the exact model parameters. For this thesis, the WZ decay is assumed to be suppressed w.r.t. the leptonic decay modes. In this case, the differential cross section can be calculated as described in Refs. [41, 42]:

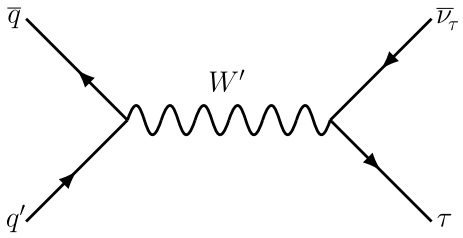

FIGURE 2.2: Leading order Feynman diagram showing the production of a W' boson and its decay to a tau lepton and its neutrino.

$$\frac{d\sigma}{d\Omega} \cdot s \propto \left(\frac{g_{SM}^2}{s - m_W^2}\right)^2 + \left(\frac{g_{W'}^2}{s - m_{W'}^2}\right)^2 \pm 2\left(\frac{g_{SM}^2}{s - m_W^2} \cdot \frac{g_{W'}^2}{s - m_{W'}^2}\right) \tag{2.5}$$

neglecting the decay width. The last term in this equation represents the interference between the SM W and the SSM W' boson. The sign of the interference term depends on the coupling type. If the coupling to the initial and final state particles is similar, the term will be added, else subtracted. Previous analyses have shown that the effect of a possible interference cannot be resolved in the $\tau_h + p_{\mathrm{T}}^{\mathrm{miss}}$ channel [43, 44]. Therefore, this effect reduced to a cross section modification and will be neglected further on.

The partial decay width for each of the SSM W' bosons decay channels can be calculated as

$$\Gamma_{W' \to f f'} = m_{W'} \cdot \frac{g_{W'}^2 C_{f f'}}{96\pi} \cdot \mathcal{F}\left(\frac{m_f}{m_{W'}}, \frac{m_{f'}}{m_{W'}}\right) \tag{2.6}$$

where $C_{f f'}$ is the color factor (1 for leptons and 3 for quarks) and $\mathcal{F}$ is

$$\mathcal{F}(x_1, x_2) = (2 - x_1^2 - x_2^2 - (x_1^2 - x_2^2)^2)\sqrt{(1 - (x_1 + x_2)^2)(1 - (x_1 - x_2)^2)} \tag{2.7}$$

For high masses of the W' boson, the masses of the fermions are relatively small, resulting in $\mathcal{F}(0, 0) = 2$. Summing over all decay fermionic decay channels then yields the total decay width of

$$\Gamma_{W'} = m_{W'} \cdot \frac{g_{W'}^2}{96\pi} \cdot 24 = m_{W'} \cdot \frac{g_{W'}^2}{4\pi} \tag{2.8}$$

This can be, for $g_{W'} = g_W$ be rewritten as

$$\Gamma_{W'} = \frac{4}{3} \frac{m_{W'}}{m_W} \cdot \Gamma_W \tag{2.9}$$

The factor of 4/3 originates from the fact, that the decay to a top-bottom pair is not possible for the SM W boson, as its mass is not high enough.

This model can be easily extended to different values of the coupling $g_{W'}$ of the hypothesized W' boson to leptons. Extending the analysis to a two-dimensional parameter scan in $m_{W'}$ and $g_{W'}/g_W$ allows for direct probe of both these parameters. It has to be stressed that this model assumes identical couplings to all lepton families. This does not favor the tau channel, since this object is typically much harder to reconstruct efficiently compared to electrons or muons at CMS. In terms of sensitivity for this model, the tau channel is complementary but likely not competitive in its results. However, this is a benchmark model of theoretical interest, making the specific analysis of the tau channel valuable.

2.2.2 Non-universal gauge interaction

This model-type originates to a common assumption of the existence of a Higgs doublet. There are many variations on these Two Higgs Doublet Models (2HDM), but we will focus on those, where the second Higgs doublet only couples to third generation leptons [45]. This type of interaction would then be able to explain the vastly different mass values between fermion families. Details on the specific phenomenology of the W' can be found in [46].

In this model, the electroweak gauge group is assumed to be $\mathcal{SU}(2)_l \times \mathcal{SU}(2)_h \times \mathcal{U}_Y$, where l and h stand for the light (first and second) and heavy fermion (third) families, respectively. It is assumed that $\mathcal{SU}(2)_l \times \mathcal{SU}(2)_h$ symmetry is spontaneously broken at an energy scale u to the diagonal $\mathcal{SU}(2)_L$ group. The mixing angle θ_E of the extended gauge group is defined such that $g_l \sin\theta_E = g_h \cos\theta_E = g$, where g is the coupling of the surviving $\mathcal{SU}(2)_L$ and g_L, g_R are those of the original $\mathcal{SU}(2)_l$, $\mathcal{SU}(2)_h$ groups, respectively.

Two Higgs doublets Φ_l and Φ_h emerge, which assign mass to the light and heavy fermions. From this point on, the description is similar to that of the SM electroweak description, except for the existence of the Higgs doublets with different vacuum expectation values v_1 and v_2. Assuming $v_2/v_1 := \tan(\beta) \gg 1$, a different coupling behavior for the light and heavy fermion families is expected.

The gauge bosons in this model are produced similarly to those in the SM. The masses of these particles can be expressed as a function of $\epsilon^2 = v^2/u^2$, where v is the SM vacuum expectation value.

$$m_W^2 = \frac{1}{2}g^2v^2 - \frac{1}{2}g^2v^2(\sin^2(\beta) - \sin^2(\theta_E)) \cdot \epsilon^2 + \mathcal{O}(\epsilon^4) \tag{2.10}$$

$$m_{W'}^2 = \frac{1}{2}g^2u^2\frac{1 + (\sin^2(\beta) - 2\sin^2(\beta)\sin^2(\theta_E) + \sin^4(\theta_E)) \cdot \epsilon^2}{\sin^2(\theta_E)\cos^2(\theta_E)} + \mathcal{O}(\epsilon^4) \tag{2.11}$$

As can be seen, in first order the SM W boson mass is replicated.

In comparison to the SSM W', the decay width and branching fractions of the heavy charged vector boson introduced in the NUGIM depend on the mixing angle θ_E and by that can differ between light leptons and the tau lepton. An additional decay channel for the produced W' is the W+H decay. In the high $\tan(\beta)$ region, the contribution is relatively small $BR(W' \to WH) \propto \frac{1}{4}BR(W' \to \tau\nu)$ and will be neglected in the following, since it has no significant impact on the phenomenology of the W'. With these conditions, the total width of the W' will be

$$\Gamma_{W'} = \Gamma_{W'}^{SSM} \cdot \frac{4.25\cot^2(\theta_E) + 8\tan^2(\theta_E)}{12.25} \tag{2.12}$$

and the branching fractions to light generation (BR_l) and third/heavy generation leptons (BR_h) are

$$BR_l = BR_{SSM} \cdot \tan^2(\theta_E) \cdot \Gamma_{W'}^{SSM}/\Gamma_{W'} \tag{2.13}$$

$$BR_h = BR_{SSM} \cdot \cot^2(\theta_E) \cdot \Gamma_{W'}^{SSM}/\Gamma_{W'} \tag{2.14}$$

For the W', the branching fractions as a function of $\cot(\theta_E)$ are shown in Fig. 2.3. Also shown is the decay width of this boson for different mass hypotheses dependent on $\cot(\theta_E)$.

For high values of $\cot\theta_E$, the branching fraction favors decays to tau leptons by approximately a factor of four compared to light leptons, making the $\tau_h + p_T^{\text{miss}}$ final state interesting

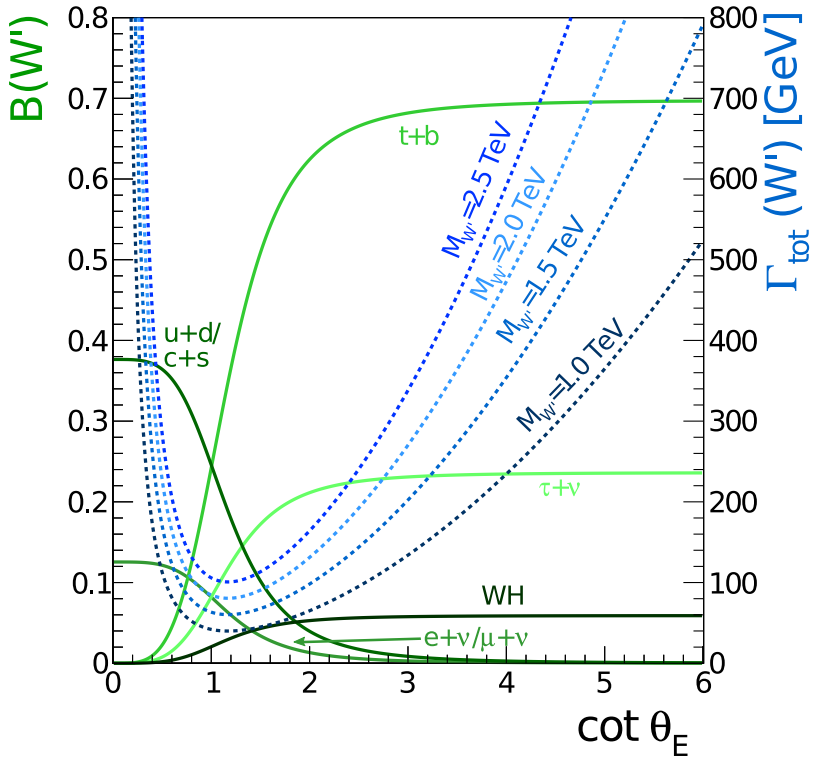

FIGURE 2.3: Enhanced coupling to third generation shown as a function of $\cot\theta_E$. As can be seen, for high values of $\cot\theta_E$, a heavy charged vector boson would exclusively decay to either top+bottom (around 70%), a tau and a tau neutrino (24%), or a W and a Higgs boson (6%). In this model, the light lepton decay possibilities are suppressed for high values of $\cot\theta_E$, where this analysis can provide sensitivity.

for this model. Additionally, the total decay width of the W' increases significantly for high values of $\cot(\theta_E)$, resulting in very broad signal shapes.

2.2.3 Quantum Black-hole

Another attempt to solve the hierarchy problem is to bring the fundamental gravitational scale close to the electroweak scale. This can be achieved by new theories that invoke extra spatial dimensions, which can offer a TeV region fundamental Planck scale, i.e. the energy at which gravitational interactions become strong. Models with $n > 1$ additional dimensions correspond to the Arkani–Hamed–Dimopoulos–Dvali (ADD) model [47]. The additional spatial dimensions could be, in principle, of any size. However, as long as the SM fields are confined to our observable dimensions, these additional dimensions would not be in conflict with previous experiments. If the fundamental Planck scale is in the TeV region, it becomes possible to produce microscopic black holes at the LHC [48, 49]. However, the production of thermal black holes, which subsequently decay into many energetic particles, is disfavored, if the production energy is not far beyond this energy scale. Such a production is therefore unlikely to be achieved at the LHC, since experimental results have set exclusion limits on this scale [50, 51, 52]. However, the effects of strong coupling gravity might still be apparent in the production of non-thermal microscopic quantum black holes (QBHs) [53].

It is to be noted that there are inevitable limitations in any model of quantum black holes since this is precisely the regime in which gravity becomes strongly coupled and the theory is no longer perturbative. Thus, the model and its predictions should be viewed as a dimensional analysis in which it is assumed that some extrapolations from the classical domain will carry over to quantum black holes. In spite of not having a complete theory of quantum gravity, it is still possible to gain insight into the signatures of quantum black holes at the LHC based on some fundamental principles and a few assumptions.

Quantum black holes are expected to form if two partons from a proton-proton collision at the LHC satisfy the hoop conjecture [54]. They predominantly decay into pairs of particles, and it is assumed, that local gauge symmetries of color and electric charge are conserved during production and decay of such a QBH. The production of a QBH occurs only beyond a certain threshold mass m_{th}, which is the main parameter of this model. The cross section for QBH production depends on m_{th} in n additional spatial dimensions. This thesis specifically analyzes the $n = 4$ case, which is a common reference point amongst searches for this model.

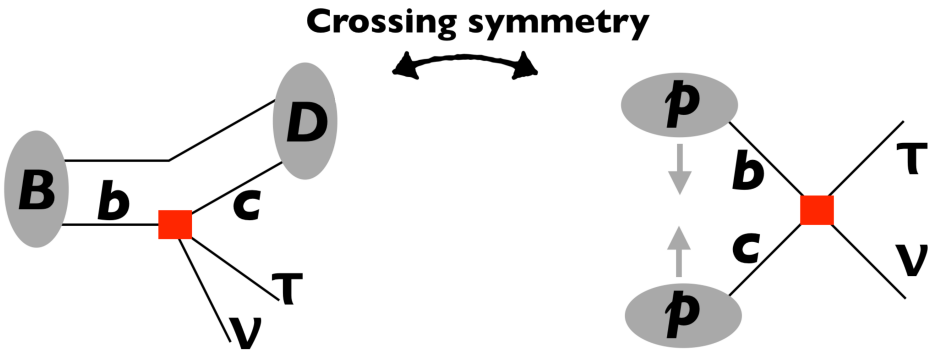

FIGURE 2.4: Crossing symmetry between $b \to c\tau\nu$ transitions and inclusive production of $\tau + p_{\mathrm{T}}^{\mathrm{miss}}$ at the LHC. The red dot indicates, that no particular assumption on the specifics of the interaction is made. Image taken from [55].

2.2.4 Generic field theory approach

Various experiments have measured the ratios of branching fractions of semi-tauonic B-meson decays $R_{D*} = \Gamma(B \to D^{(*)}\tau\nu)/\Gamma(B \to D^{(*)}\ell\nu)$ (with $\ell = e$ or μ). This ratio has been reported to be enhanced with respect to the Standard Model prediction by approximately thirty percent, with a global significance of 4σ [6, 7, 8, 9, 10]. In this model, we look at the possibility, that this enhancement is due to new physics which could also be possibly visible in the high momentum physics searches. All models introducing a new heavy mediator which could explain these enhanced ratios would have a specific phenomenology. They have to connect the $b \to c\tau^-\bar{\nu}$ decay and the $b\bar{c} \to \tau^-\bar{\nu}$ scattering process via crossing symmetry as shown in Fig. 2.4.

In order to include this phenomenon in a generalized way, an effective field theory (EFT) has been proposed [55]. This type of models uses minimal assumptions on the actual interaction process which occurs, but focuses on the incoming and outgoing particles and their interaction parameters. It introduces a low-energy EFT of new physics with terms $\propto \epsilon_\Gamma U_\Gamma$, where U_Γ are the currents and ϵ_Γ are the so-called Wilson coefficients (WCs), which scale as $\epsilon_\Gamma \propto v^2/\Lambda^2$. Here, $v \approx 246$ GeV is the electroweak symmetry breaking scale and Λ is the characteristic scale of NP. At low energies, the operators introduced this way lead to the enhanced semi-tauonic B decays. One can use the measured branching fraction ratios R_{D*} to fit the values of the WCs to its preferred solutions. At high energies, it leads to an additional event yield, which is sensitive to the Lorentz structure of the interaction. The possible structures analyzed here are vector-like coupling ($\epsilon_{\mathrm{L}}^{\mathrm{cb}}$), tensor-like coupling ($\epsilon_{\mathrm{T}}^{\mathrm{cb}}$) and scalar-tensor like coupling ($\epsilon_{\mathrm{S_L}}^{\mathrm{cb}}$). The cross section for this effective field theory model scales with the square of the coupling for each coupling type.

Chapter 3

Experimental Setup and Data Sets

This thesis analyzes experimental data collected with the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) at CERN. In this chapter, these devices are described. This includes a short description about the data processing used in this context.

3.1 The Large Hadron Collider

The LHC [56] is located at the CERN near Geneva, Switzerland. It is (at the time of writing this thesis) the highest-energy particle collider built by humans. It was commissioned in order to explore particle physics at and beyond the electroweak scale. It was designed to probe the mechanism of electroweak symmetry breaking and possibly discover new (unexpected) particles. In order to achieve this goal, the LHC needs to provide a high collision energy and large instantaneous luminosity. The LHC resembles a circular synchrotron with a circumference of approx. 27 km, and it is built into the tunnel, which contained the large electron-positron collider (LEP) accelerator previously.

Two types of collisions are created at the LHC: proton-proton collisions and heavy ion collisions. This thesis focuses on the former, for which the LHC was designed to collide the beams at a center-of-mass energy of 14 TeV and an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Strong dipole magnets are necessary in order to keep the high energy protons on the intended circular orbit. The LHC uses superconducting NbTi dipole magnets operated at a temperature of 2 K, which achieve a high central magnetic field of up to 8.3 T. Additionally, higher order magnets (quadrupole, sextupole etc.) are used to adjust for higher order beam corrections. Clouds of protons, so-called *bunches*, travel around the ring in two separate beam pipes with independent dipole magnetic fields. These bunches are brought to collision at specific interaction points, where designated experiments are located in order to analyze the products of these collisions. One implication of colliding clouds of protons is *pileup* - multiple proton-proton interactions which happen in such small time frames, that the detector of choice is not able to disambiguate them in time.

Its biggest success (so far) is the discovery of the Higgs boson and the analysis of this new fundamental new particle [18, 19]. It was found in the *Run-I* data taking period of the LHC in the years 2011 and 2012. Due to technical reasons, the design performance of the LHC was not met yet, and the machine was operated at $\sqrt{s} = 7$ and 8 TeV. It delivered a data set of approximately 30 fb^{-1} in these years.

This analysis is based on data recorded with the CMS detector, which corresponds to an integrated luminosity of 138 fb^{-1}, collected in the data taking years from 2016 to 2018.

3.2 The CMS Detector

The CMS detector [57] is a multipurpose detector designed to measure charge, momentum and energy of the various types of particles originating from proton-proton collisions. It has

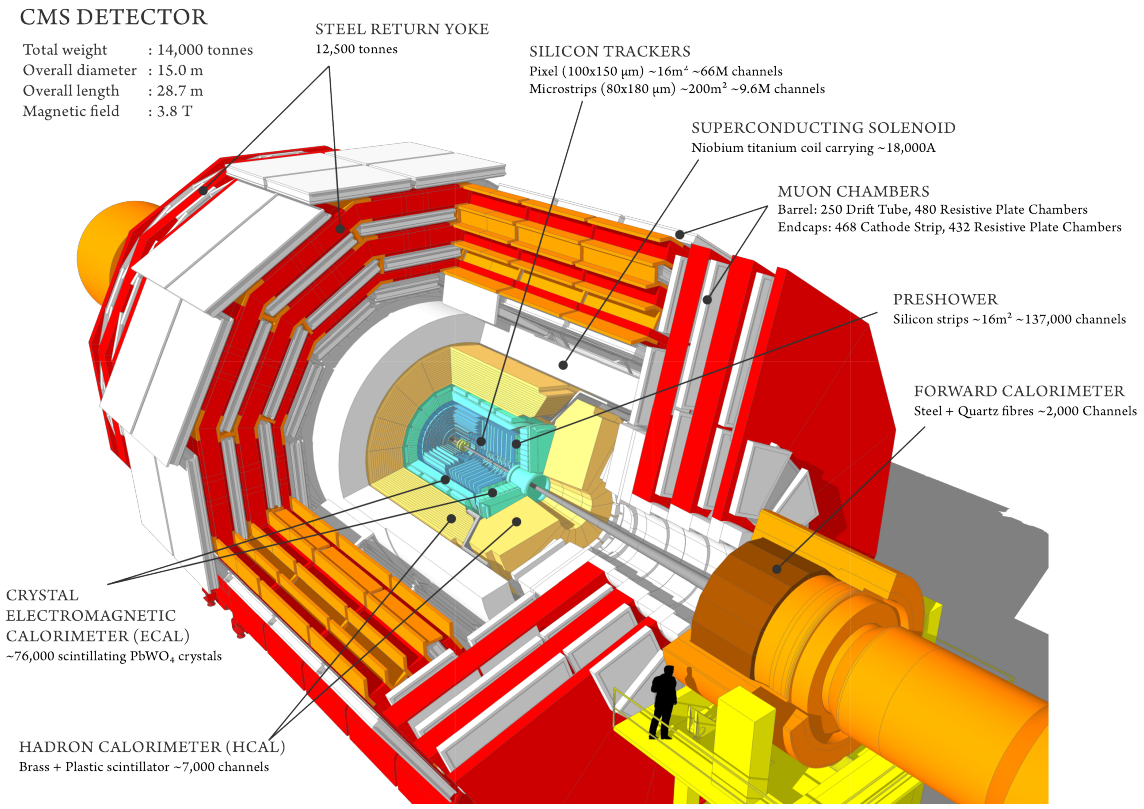

FIGURE 3.1: Schematic view of the CMS detector. Image taken from [58].

a roughly cylindrical shape and is divided into five cylindrical, central *barrel wheels* and two outer *endcap disks*. These can be moved if access to specific modules is necessary. The CMS detector consists of multiple sub-detectors, which are located in concentric layers around the central collision point. These sub-detectors are used to measure the tracks of charged particles close to the collision point, determine the energy of charged particles, photons and neutral hadrons, and to track the outward going muons. The key component of the CMS detector is the central solenoid magnet, which provides a magnetic field of 3.8 T. This strong magnetic field is used to determine the charge of a particle as well as a charged particle's momentum. A sectional view of the detector is given in Fig. 3.1.

Due to the shape of the detector, a cylindrical coordinate system is used, with the z-axis pointing along the direction of the beam pipe, the radial coordinate r perpendicular to it and the azimuthal angle Φ in the plane transverse to the beam axis.

In order to achieve optimal performance, all information from the different sub-detectors is combined in order to reconstruct the properties of all outgoing particles. This section describes all sub-detectors in more detail, as all components of the detector are necessary to understand the reconstruction of physics objects. The sub-detectors are constantly tested and possibly replaced or upgraded - this chapter describes the CMS detector in its 2016 configuration and points out major upgrades of the sub-detectors.

3.2.1 Magnet System

In order to measure precisely the momentum of high-energy charged particles, large bending power is necessary. The CMS magnet system provides a magnetic field of up to 3.8 T and simultaneously provides most of the detector's structural support. In order to provide the structural integrity and to withstand the forces due to the large magnetic field, the system is

built very robust. The CMS detector contains a superconducting solenoid with four layers of winding made from a stabilized reinforced NbTi conductor. The cable is mechanically reinforced with an aluminum alloy. During operation, the solenoid is cooled down to 5 K using liquid helium. While at that temperature, the coil has an inner diameter of 6.3 m and with its nominal current of 19.14 kA contains a stored energy of 2.6 GJ. The geometry of the solenoid leads to a magnetic field which only bends particles travelling in the transverse plane of the beam pipe. The magnetic field within the solenoid, which contains the tracking detectors and calorimeters, is homogeneous. Outwards of the solenoid, the magnetic flux is returned via a massive 10000 t heavy iron yoke divided into the aforementioned endcap and barrel disks. The yoke has three layers of iron, which provide structure to mount the muon detectors.

3.2.2 Tracking Detectors

The innermost layers outside the beam pipe are the tracking detectors. They have to measure the trajectories of all charged particles emerging from collisions quickly and precisely[1]. In order to achieve this, a high granularity detector is necessary, which needs to have a small response time as not to lose tracks. The hit density near the beam pipe at design luminosity is estimated to 1 MHz/ mm^2, which becomes less the further away from the beam pipe one gets. Therefore, CMS uses two types of tracking detectors: the *pixel* detector in the innermost part followed by the outer *strip* detector. The layout of the tracking system of CMS is given in Fig. 3.2. The pixel detector has three layers in the barrel and two disks in the endcap. It covers a pseudorapidity range of up to ± 2.5 and is essential for reconstruction of secondary vertices from b quark and tau lepton decays. The system provides precise tracking points in the $r - \Phi$ plane and the z-axis with 66 million individual pixels. Each pixel has a small pixel size of $150 \times 100\,\mu$m in the $r\phi \times z$ direction, allowing for a spatial resolution of $15 - 20\,\mu$m.

The silicon strip tracker has 10 barrel detection layers divided into the two regions of the *Tracker Inner Barrel* (TIB) and *Tracker Outer Barrel* (TOB). The strip width increases with distance from the beam pipe from 80 to 180 μm. In the high pseudorapidity region of the barrel, the *Tracker Inner Disks* (TID) with 3 layers of silicon strip detectors are placed. The *Tracker Endcaps* (TEC) are also silicon strip detectors and complete the tracking system, with a total size of 564 cm in length and 113.5 cm outer radius (22.5 cm inner radius) surrounding the collision point. It allows for at least $\approx$ 9 hits, where at least $\approx$ 4 hits provide two-dimensional measurements in the region of $|\eta| < 2.4$. The inner tracking detectors achieve a single point resolution in r (ϕ) of 23 μm (35 μm) in the barrel and 100 μm (141 μm) in the endcaps. The outer layers achieve a single point resolution of 53μm for r and 35 μm for ϕ in the barrel.

CMS Pixel Upgrade

The different sub-detectors of the CMS are constantly tested, optimized and are subject to change. Specifically the tracking detectors are disposed to severe radiation damage due to their proximity to the beam pipe. In the year-end technical stop of 2016, the pixel detector of the CMS was upgraded in order to achieve better performance at a very high instantaneous luminosity and to provide a better coverage towards high pseudorapidities [59]. Additionally, the new pixel tracker is placed closer to the interaction point, which is possible due to a smaller radius of the beam pipe in CMS (23 mm instead of the original 30 mm) which was already installed during the long shutdown 1 in 2013/14. A sketch comparing the original setup with the upgraded one is given in Fig. 3.3. The hit position resolution of e.g. barrel pixel detector layer 3 was measured to 9.5 μm in the *r*-ϕ direction and 22.2 μm in the z

[1] At LHC design luminosity, these are estimated to 1000 particles every 25 ns.

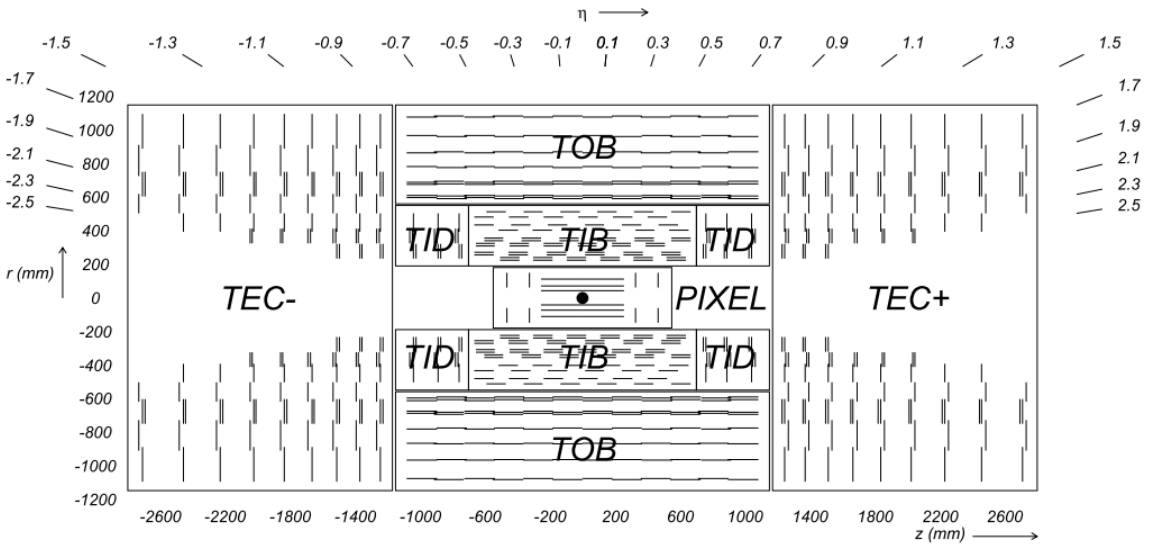

FIGURE 3.2: Schematic view of the CMS tracking system in its 2016 configuration, where each line represents a single detector module. The beam pipe is at $r = 0$ and the collision point is located at the center. The sub-detector systems of the PIXEL Image taken from [57].

direction. The upgraded system allows four-hit coverage up to a pseudorapidity value of $|\eta| = 2.5$.

3.2.3 Electromagnetic Calorimeter

The next layer in the CMS detector is the electromagnetic calorimeter (ECAL) [57, 60]. Its purpose is to measure the energy of electrons and photons by stopping them within the detector. This is achieved by the use of lead tungstate crystals ($PbW0_4$) arranged homogeneously. Lead-tungstate has a short radiation length of 0.89 cm and a small transversal expansion factor for electromagnetic showers (Molière radius) of 2.2 cm. The visually transparent material scintillates when electrons or photons pass through it, producing blue-green light. The system is optically clear, fast and resistant to radiation damage. It is split into a barrel part (EB, $|\eta| < 1.479$) and an endcap part (EE) and extends to pseudorapidities of $|\eta| < 3.0$. The scintillation light of the crystals are measured with photodetectors, which are fast, radiation tolerant and have to function within the large magnetic field of the CMS. They also have to amplify in order to compensate for the small light yield of the crystals. In the barrel region, avalanche photodiodes are used while in the endcaps vacuum phototriodes are mounted.

Additionally, in the endcap region, a *preshower detector* is mounted. Its purpose is to identify neutral pions in the endcaps and to improve the granularity in this region. It consists of two layers of lead absorbers followed by silicon strip sensors. The different strip layers are placed orthogonally to each other, allowing for two-dimensional position measurement. The preshower detector is relatively thin, with a total thickness of 20 cm. The material thickness of the detector corresponds to three radiation lengths.

The energy resolution of the ECAL was measured in test beams for incident electrons of 120 GeV to be 0.5%.

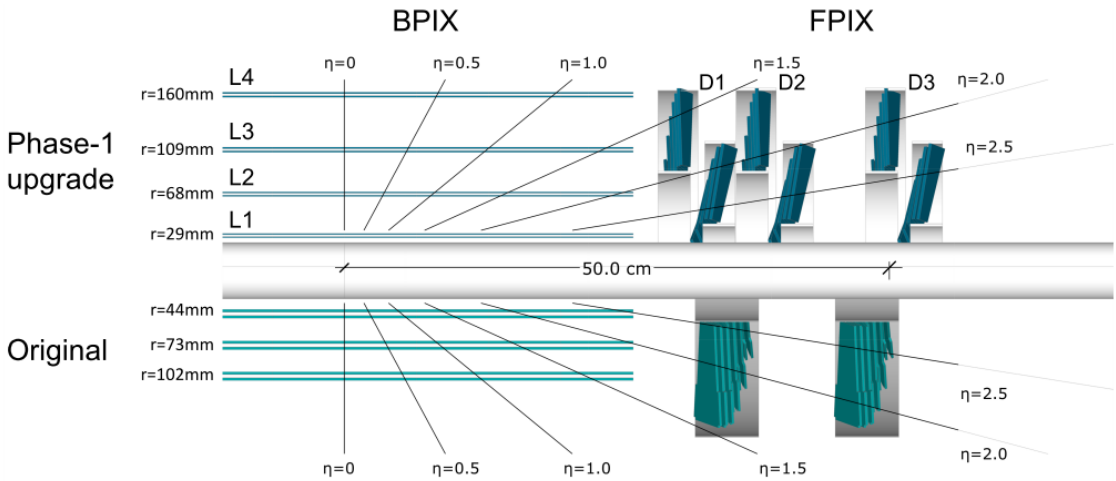

FIGURE 3.3: Comparative layout sketch of the CMS upgraded pixel detector with the original detector layout. The additional layers in the *barrel pixel* (BPIX) detector and the *forward pixel* (FPIX) disks lead to improved performance and better redundancy in track measurements. Image taken from [59].

3.2.4 Hadronic Calorimeter

The hadronic calorimeter (HCAL) measures the energy of hadron jets by stopping them within the sub-detector. It is split into multiple sub-detectors, defined by their location: the calorimeter in the barrel (HB), the calorimeter in the endcap (HE), the outer calorimeter (HO) and the calorimeter in the forward region (HF). A schematic view of the different sub-detectors is given in Fig. 3.4.

The resolution of hadronic particles of the CMS calorimeter system depends on both the ECAL and HCAL subsystems, as most particles start showering in the ECAL. The hadronic energy resolution of HB and EB combined was measured with test beams to [61]

$$\left(\frac{\sigma}{E}\right) = \frac{84.7\%}{\sqrt{E}} + 7.4\% \tag{3.1}$$

where E is in measured in GeV. When combining information from the HCAL with those from the tracker and ECAL, jet energy resolutions of 15% at 10 GeV, 8% at 100 GeV and 4% at 1 TeV have been achieved [62].

Hadron Barrel Detector

The HB covers a pseudorapidity range of $|\eta| < 1.3$ and is divided into the half barrels HB+ and HB-. Each one consists of 18 identical wedges arranged around the beam pipe (in Φ). Each wedge contains two steel absorber plates at the front and the back for structural support which surround fourteen brass absorber plates. The HBs effective absorber thickness increases with higher pseudorapidity values, ranging from 5.82 interaction lengths up to 10.6 interaction lengths. The ECAL in front of the HB provides approximately 1.1 interaction lengths. Energy deposits are measured using plastic scintillators, divided into 16 regions in η.

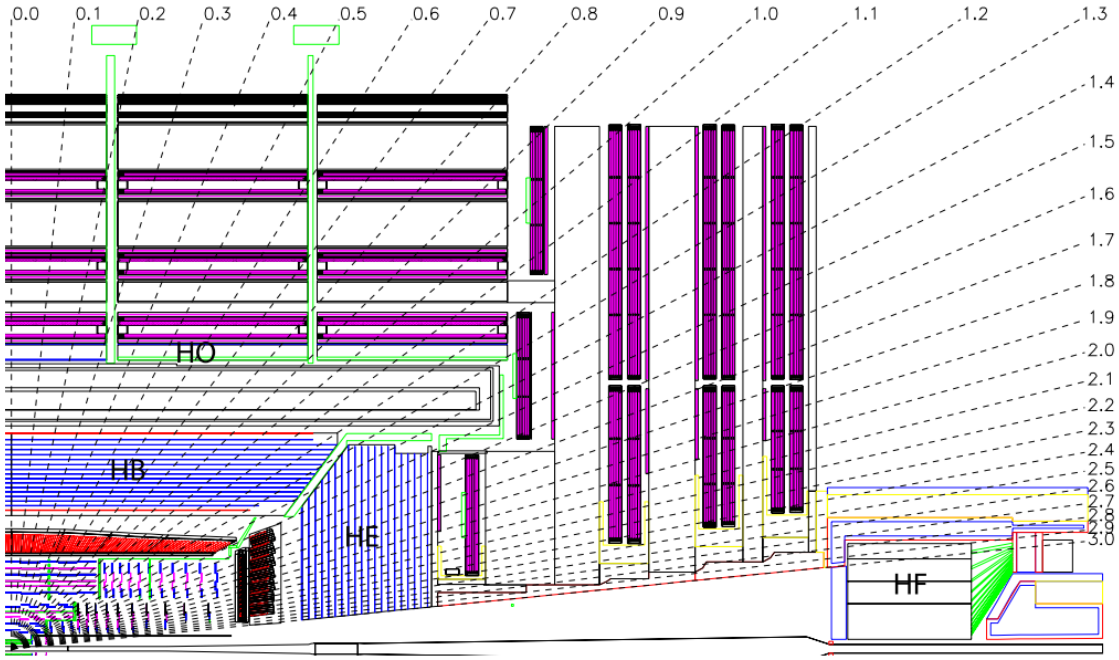

FIGURE 3.4: Schematic longitudinal view of the CMS detector with the highlighted HCAL detector system with the different sub-detectors (HB, HE, HO and HF). The beam pipe goes along the x-axis of the image and dashed lines give values for η. Image taken from [57].

Hadron Endcap Detector

The hadron endcap sub-detector system covers a pseudorapidity range of $1.3 < |\eta| < 3$. It is mounted on the endcap of the iron yoke, close to the endcap muon system. One can divide the hadron endcap region by the positive and negative parts in pseudorapidity (HEP and HEM). Similar to the HB, brass is used as an absorber material. Scintillation light is collected by wavelength shifting fibers. Outside the scintillator material, these fibres are spliced to clear fibres. In the original HCAL design, these were read out using hybrid photodiodes. As part of the *Phase-I* upgrade of CMS, these were planned to be replaced by silicon photomultipliers [63]. The detector was upgraded in this way for the endcap section of the HCAL in late 2016 / early 2017. The barrel section was upgraded after 2018.

Hadron Outer Detector

As the stopping power of the calorimeters is lowest in the most central region due to the geometry of the setup, an additional hadron calorimeter is placed outside the solenoid, which is used as an additional absorber. It is placed just before the first muon stations and extends up to $|\eta| < 1.3|$. Otherwise, its inner setup is similar to that of the HB, using brass as an absorber and plastic scintillator for measurements. Its main purpose is to ensure full detection of high energy jets or late starting parton showers in the central region.

Hadron Forward Detector

In order to cover very high pseudorapidity regions for hadron detection, the CMS detector has a hadron calorimeter in the forward direction. In this region, a very high particle flux is expected to hit the detector, much higher than for the more central region. Therefore, the

leading design guideline of the HF is its survival in these harsh conditions. As active material, quartz fibers were chosen due to its high radiation hardness. This sub-detector extends the range of the HCAL system to the pseudorapidity range $3 < |\eta| < 5.2$. It extracts a signal when charged shower particles with sufficient energy generate Cherenkov light. As there is no dedicated electromagnetic calorimeter in this forward region, the HF is designed to differentiate between hadronic and electromagnetic showers. It achieves this by using quartz fibres of different length: Half of the fibres run over the full depth of the steel absorber structure, while the other half starts at a depth of 22 cm from the front of the detector. Showers from electrons and photons are expected to deposit most of their energy within this first region while hadronic showers produce nearly equal signals in both sections.

3.2.5 Muon System

One of the defining features of the CMS detector is its muon system. Its purpose is to trigger and identify muons as well as to determine their kinematic properties. Similar to all other components, the muon system is split into a barrel and an endcap region. In the barrel region, multiple layers of drift chambers are used, which are arranged such that all outgoing muons (ideally) have multiple hits in these chambers to ensure precise muon timing and track reconstruction. In the endcap region, six layers of cathode strip chambers are used, which have a fast response time and are resistant to radiation damage. The cathode strips in each chamber are arranged radially from the beam pipe for precise $r - \Phi$ measurements, while the anode wires provide measurements in the η variable. The muon system covers a range of $|\eta| < 2.4$ and has no gaps. Due to the large amount of material in the calorimeters, non-muon objects rarely reach the muon chambers and are negligible for the systems design (*punch-through*).

In the barrel region, the muon system consists of muon drift tubes (DTs) - these tubes contain a streched wire within a gas mixture. When a muon (or any charged particle which reaches the drift tubes) passes through a drift tube, it ionizes the gas atoms or molecules. The resulting electrons are then collected on the wire and allow for muon detection. Arranging multiple overlapping layers of DTs allow for a precise reconstruction of muon positions. In the endcap disks of the CMS, where the magnetic field is inhomogeneous and particle flux is high, cathode strip chambers (CSC) are used. These are large chambers which consist of multiple positively charged wires perpendicularly arranged with negatively charged wires, contained within a gas mixture. This grid of wires allows for precise track measurement even in these harsher conditions. In addition to these two detector systems, resistive plate chambers (RPCs) are used in CMS in order to trigger muon events with a precise timing. These sub-detectors consist of two oppositely charged plates and collecting metallic strips, have a very small response time (≈ 1 ns) and allow for immediate online triggering.

While the muon system has many more intricacies to it, this analysis is only mildly interested in muons. The interested reader can find much more information here [57, 64].

3.3 Data Acquisition

The CMS detector has many components with multiple subparts and channels, which combined produce a large amount of data ($\mathcal{O}(1$ MB)), even for a single event. Combined with the high frequency of 40 MHz of collisions the LHC provides, the amount of data collected with the detector is much too high for any readout or storage system. At the same time, most events produced during proton-proton collisions are low momentum QCD scattering events, which are of low interest in terms of studying the Higgs mechanism or for searches

of new heavy particles. In order to reduce the large data volume and quickly scan for interesting physics events, a two-tiered triggering system is used in CMS [65, 66].

Level-1 Trigger

The Level-1 trigger (L1) is a fast, hardware based triggering system, designed to reduce the event rate from 40 MHz to $<$ 100 kHz. To achieve this, it uses information from both calorimeters and the muon system as input, combines them and selects events with interesting signatures. It does not use information based on the tracker, as this would take too much time to run. A fast algorithm composes a list of trigger object inputs for muons, e/γ, central and forward jets and tau jets and checks whether they fulfil certain conditions like a minimum E_T or p_T. All events which have passed the L1 trigger system are sent to a computer farm for the second level of trigger processing - the high-level trigger (HLT).

High-Level Trigger

The HLT triggering system uses a quick reconstruction algorithm to identify all physics object candidates in each event which passed the L1 trigger. It applies identification methods and requirements to select events which are of interest for further processing. As the event rate at this point is still quite high at 100 kHz, a large processor farm is installed with an average processing time of 90 ms per event. The processing is structured based on an HLT *path* - individual processing steps run in a predefined order which reconstruct and identify certain physics objects. I.e. information from muon systems and calorimeters is processed first, before accessing the more CPU-intensive tracker information used for isolation or global event reconstruction techniques. Based on the output of this triggering system, events are grouped into several non-exclusive streams of data, which are stored locally and eventually transferred to the primary CMS computing center (Tier-0) for initial processing. The limiting factor on the performance of the HLT is the size of the events and the processing speed of the downstream systems. The event rate during the full HLT process is reduced to $\approx$ 1 kHz.

Data Storage

The CMS data storage system defines multiple storage layers for data processing. The first level is the Tier-0 storage, where events from the HLT processor farm are sent to for storage and initial ("prompt") reconstruction. These events contain all detector information for each event that passed the HLT step. Based on the HLT trigger paths, the events are then sorted into different *data sets* - e.g. all events which passed any tau related HLT path are sorted into the Tau data stream. The resulting data sets, presorted by their event content, are then duplicated and stored on the second Tier-1 storage level for redundancy of the raw information of the event content. These storages can be used to reconstruct the raw data again when all calibrations are available. Additionally, Tier-1 data centers provide storage for simulated background data. Tier-2 data centers are distributed world-wide between CMS related working groups. They provide computing power to their local analysis needs and to official CMS production tasks like background sample production (which are then sent to and stored in Tier-1 centers).

Chapter 4

Object reconstruction

The two objects of interest in this thesis are the tau lepton and a neutrino. The former decays very close to the interaction point due to the short lifetime of the tau lepton [11] — thus only its decay products are detected. The latter is not detected by the CMS detector and results in missing transverse momentum. This chapter focuses on the methods used to reconstruct jets, muons and electrons at CMS and how possible τ_h and $p_{\mathrm{T}}^{\mathrm{miss}}$ candidates are determined.

4.1 The Particle Flow algorithm

The segmentation of the CMS detector allows for a highly complex and very efficient reconstruction algorithm: the Particle Flow (PF) algorithm [62]. This algorithm uses information from all sub-detectors, analyzes and combines them to identify a comprehensive list of final state particle candidates and global event variables on an event-to-event basis. Particles from pileup interactions are identified and mitigated. The algorithms fundamental ingredients are the ability to reconstruct tracks from the tracking detectors and extrapolating them to the calorimeters, where it identifies clusters of energy deposits. Topologically connected elements are identified and first elements of electron and muon identifications are applied in order to reconstruct the particle candidates. The PF algorithm classifies candidates for photons, electrons and muons as well as neutral and charged hadrons.

Jets are reconstructed from PF candidates, using the *anti*-$\mathrm{k_T}$ algorithm [67]. This has become the preferred reconstruction method for jets, as it is both infrared and collinear safe[1]. The default cone size for these jet candidates at CMS is $\Delta R = 0.4$ (*AK4-jets*). These jets are used as input seeds for tau reconstruction.

4.2 Reconstruction of hadronically decaying taus

The tau lepton is the heaviest lepton known in the Standard Model and has a lifetime of only a few hundred nanoseconds. Thus, it decays before it reaches the detector. In $\approx 65\%$ of tau decays, one or three charged hadrons (*1-prong* and *3-prong*) are emitted together with a tau neutrino and possibly neutral hadrons. An overview of the different tau decay modes is given in Tab. 4.1. The neutrino will contribute to missing transverse momentum, while the hadrons combine to a single jet which is measured in the HCAL. Proper reconstruction of hadronically decaying tau leptons is a complex task. The main background which needs to be discriminated against are gluon and quark jets due to their similar signature. The most significant difference between these two is the *hadron multiplicity*: tau jets are expected to consist of one to three charged hadrons while quark and gluon jets have a high expected

[1] *Infrared safety*: Properties of a reconstructed jet should not depend on the presence of some low energy parton.
Collinear safety: Jet properties should not change if we replace a constituent by multiple collinear constituents which carry the same momentum (in sum).

τ Decay Mode	Intermediate resonance	Branching Fraction [%]
$\tau^- \to h^-\nu_\tau$		11.5
$\tau^- \to h^-\pi^0\nu_\tau$	$\rho(770)$	25.9
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^-h^+h^-\nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^-h^+h^-\pi^0\nu_\tau$	$\rho(1450)$	4.8
$\tau^- \to e^-\bar{\nu}_e\nu_\tau$		17.8
$\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$		17.4

TABLE 4.1: Branching fractions of different tau decays (decay modes) - only the largest fractions are given. $h^\pm$ denotes charged hadrons (pions or kaons) which the tau can decay to. The table is symmetric for τ^+ with respect to charge. Values for branching fractions are rounded and taken from [11].

number of constituents. As there is a high abundance of quark and gluon jets coming from proton-proton collisions, identification of hadronically decaying tau leptons needs to be very efficient.

Due to the short lifetime of the tau lepton, it is usually not possible to discern prompt charged light leptons from those which originate from a tau decay. The only difference is the occurrence of missing transverse momentum due to the neutrinos in the event.

4.2.1 Hadron-plus-Strips algorithm

The hadron-plus-strips algorithm [68, 69, 70, 71] is designed to reconstruct hadronic tau decay candidates and assigns them their visible four-momentum (excluding the invisible neutrino). It accounts for tau decays with and without the emission of neutral hadrons by estimating all combinations of charged hadronic tracks with (and without) nearby neutral hadronic signatures. The algorithm uses anti-k_T jets with a distance parameter of 0.4 as input seeds. The minimum requirement in terms of detector acceptance for these seeds is a transverse momentum $p_T > 14$ GeV and a pseudorapidity $|\eta| < 2.5$. Tau leptons outside this parameter range are not reconstructed at CMS due to the geometrical limits of the detector. Using this seed as starting input, the HPS algorithm considers all tracks and energy deposits within a cone of $\Delta R = 0.5$ for processing and determine the best tau candidate for each seeding jet.

The name of the algorithm is quite indicative of its functionality. It uses jet substructure information for charged hadrons within the jet and combines it with so called 'strips', as shown in Fig. 4.1. These are expected detector signatures coming from π^0-decays. These particles quickly decay themselves into either two photons ($\approx$ 98.8%) or to an electron, a positron and a photon ($\approx$ 1.1%). In either case, these are objects detected in the ECAL. Due to the large magnetic field in CMS, electrons and positrons created in the electromagnetic shower are bent into different directions and produce slightly elongated lines in Φ. These lines are referred to as strips.

These strips are reconstructed by selecting the most energetic electromagnetic particle which is within the PF jet cone. It then iteratively searches for other electromagnetic particle hits within a small window around that particle until no other particles are found within that window. After combining all these particles into a single strip object, its four-momentum is calculated, and if its four-momentum is above 2.5 GeV, it is considered as a π^0 candidate for further processing. Then, a new strip is considered, using the next-most energetic electromagnetic particle within the PF jet cone, which has not been used in another strip

reconstruction. This procedure is repeated until all strips which fulfil the requirements are identified.

The size of the search window around the initial electromagnetic particle has been optimized since Run-1. The previous fixed 0.05×0.20 window size in the $\eta - \Phi$ plane sometimes missed strips during reconstruction due to the hits being outside the search window, which happens if the photons created during the π^0 decay undergo multiple conversions and bremsstrahlung. On the other hand, high p_T tau decay products tend to be boosted in the τ_h's flight direction, when the strip is expected to be smaller than in Run-1. The improved algorithm now considers the distance from the hadronic decay products to the strip as well as the momentum of the constituents of the strip when determining the size of the search window. As the strip is assigned a new four-vector every time a new particle is added to it, the calculation is updated accordingly. Therefore, it is referred to as *dynamic strip reconstruction*.

The charged hadron(s) from the tau decay are combined with all possible strips. This procedure also allows for the possibility of no strips being associated with the tau candidate. For each different decay mode of the tau lepton, combinations of the jet and the photon seeds are created. From this plethora of possible combinations, the tau candidate is chosen based on the following criteria:

- First, dependent on the reconstructed decay mode, the candidate is required to be within a mass window around the mass of the intermediate resonance particle as given in Tab. 4.1. For this combination, the mass of all reconstructed $h^{\pm}$ is assumed to be the charged pion mass.
- Next, the combined charge of the tau object has to be $= \pm 1$. For Run2 data taking, the HPS algorithm was extended to reconstruct tau candidates when a charged hadron was missed (or not properly defined as isolated/ its own object). In this case, the charge of the tau object is set to the charge of the charged hadron with the highest momentum. These additional decay modes recover 19% (13%) of the $\tau^- \rightarrow h^-h^+h^-\nu_\tau\,(\pi^0)$ decays, but tend to introduce additional background events. Their use is not recommended at the moment and may be explored in the future.
- In dependence on the reconstructed tau objects momentum, all reconstructed hadrons are required to be within a cone of $\Delta R = 3.0/p_T[\text{GeV}]$ (but limited to $0.05 < \Delta R < 0.1$).
- From all remaining candidates, the highest momentum tau object τ_h candidate is selected. All others are discarded.

The accurateness of this algorithm can be tested in simulation, where it is possible to compare the generated tau decay mode with the reconstructed one. The matching precision varies between 80% accuracy of generated $h^{\pm}$ decay being reconstructed as such to 'only' 36% of $h^{\pm}h^{\mp}h^{\pm}\pi^0$ reconstruction efficiency. In the latter case, 27% are recovered in the $h^{\pm}h^{\mp}h^{\pm}$ decay mode - for such dense jets, it is likely that the neutral pion is not reconstructed / included in the tau object. The HPS algorithm fails to reconstruct the generated decay modes (DMs) in 11% for the $h^{\pm}$ DM, 25% for the $h^{\pm}\pi^0$ DM, 10% for the $h^{\pm}h^{\mp}h^{\pm}$ DM and finally 17% of the % of $h^{\pm}h^{\mp}h^{\pm}\pi^0$ DM. These values were calculated from a $Z \rightarrow \tau\tau$ event sample, where the τ_h is required to fulfil $p_T^{\tau} > 20$ GeV, $|\eta^{\tau}| < 2.3$ [69].

The numbering convention used in CMS to discern different tau decay modes is as follows:

$$DM = 5 * (N_{h^{\pm}} - 1) + N_{strips}$$

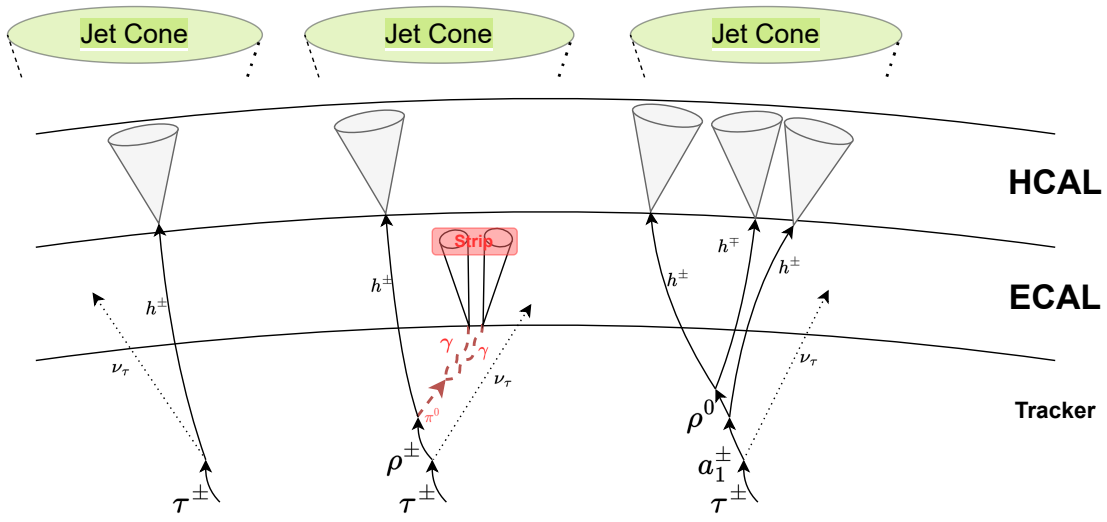

FIGURE 4.1: Simplified sketch of different tau decay modes and their detector signatures. Each incident τ lepton decays very quickly via a representative decay chain, leading to various distinct decay modes. In each case, all decay products have to be within the reconstructed jet cone. The right most part (DM10) can also occur with an associated production of a photon pair, which would lead to an additional strip signature similar to DM1.

where $N_{h^{\pm}}$ is the number of reconstructed charged hadrons associated to the tau and N_{strips} is the number of strips associated to it. E.g. a reconstructed tau object with one charged hadron and one strip gets assigned 'DM1'. A simplified sketch to visualize how different decay modes of the tau are reconstructed is given in Fig. 4.1.

4.2.2 Tau Identification (ID)

It is very important to have a strong differentiation between tau objects from tau lepton decays and quark / gluon jets. Additionally, it is also necessary to discern 'true' taus from electrons and muons, which can create strips close to a jet and be misidentified. In order to achieve this, a strong discriminator tool is needed. In previous years, CMS had developed a multivariate analysis (MVA) based identification method, which provided three discriminator routines - one for each of jets, electrons and muons [69]. During Run2, a new τ_h identification method based on a deep neural network was developed, called the *DeepTauID* [72]. This highly complex algorithm uses classical neural network methods and combines it with image recognition methods in order to provide a four dimensional output, each representing the likelihood score of the analyzed object originating from a hadronic tau decay, a jet, an electron or a muon. The output scores of the non-tau objects are called *vsJet*-discriminator, *vsMuon*-discriminator and *vsElectron*-discriminator.

Inputs: Previous tau identification algorithms only made use of high-level variables - i.e. variables built from combinations of specific reconstruction variables which were promising in terms of discrimination. The DeepTau ID however receives both these high level variables and low-level variables - i.e. all information about all reconstructed particles near the τ_h candidate is directly used as input parameters. This is to accommodate for the possibility that a machine-learning algorithm can use this information to improve overall performance. High-level information known from previous identification methods are still used as input - while the network in principle should be able to learn them on its own, it was shown that

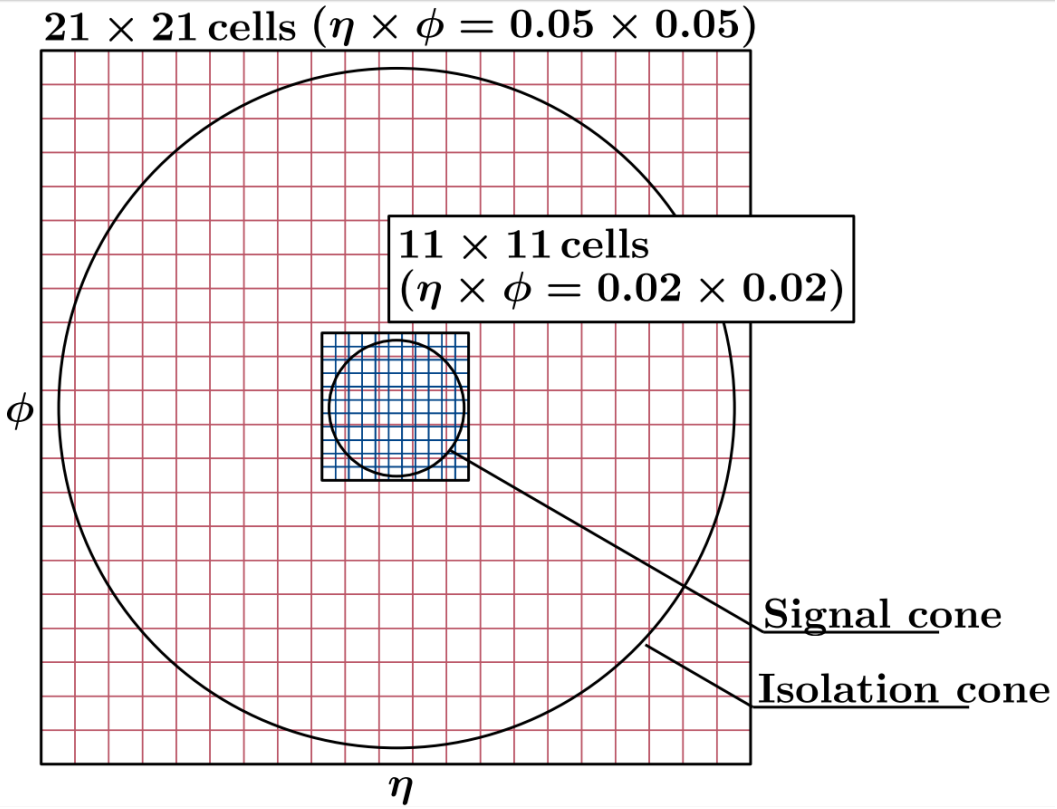

FIGURE 4.2: Input layout of low level input grid structure around the reconstructed τ_h object candidate. The inner grid contains the charged hadron(s) and possible π^0s as reconstructed / required by the HPS algorithm. The large scale grid, which overlays the inner grid is used to extract information on isolation. Image taken from [72].

the number of necessary training events can be lowered and convergence of said training can be improved significantly [73].

Network architecture: These high-level inputs have a different subnetwork processing chain then the low-level inputs, and are combined with the output from the low-level subnetwork only in the last step of the network. The low-level inputs are processed in two separate grids, centered around the τ_h axis. The signal grid is a fine 0.02×0.02 grid with 11×11 cells in the $\eta - \phi$ plane designed to extract information on the collimated hadronic decay products of the tau. This grid always encompasses the signal cone defined from the HPS algorithm. The isolation grid has 21×21 cells with a size of 0.05×0.05 and represents the 'image' of the full isolation cone with $\Delta R = 0.5$. The choice of density and cell number for these grids is given by computational limitations and loss of information - more cells require more computational power while fewer cells reduce resolution power which information could be extracted from. The grid sectioning is shown in Fig. 4.2. This combined input layer is processed with convolutional neural network layers [74], which have been used in image recognition techniques.

Training and validation: This network is trained and validated on selected background simulation samples - specifically Z and W (+jets) training samples, the $t\bar{t}$ background and simulated QCD events. The tau candidates selected for training have to fulfil a minimum requirement of $p_T^\tau > 20$ GeV and $|\eta^\tau| < 2.3$. In total, 140 million events have been used for training and 10 million events have been used to validate the training (i.e. to show, that the parameters determined by the training do indeed give the same / similar result on unknown events).

Discriminatory Power: This identification method has been validated and approved in numerous measurements. An interesting metric for this ID is its comparison in its discriminatory power with previous tau identification methods. This is shown for high transverse momentum τ_h candidates from the W+jets background in Fig. 4.3, where the misidentification rate of QCD jets against the tau identification efficiency can be seen. In this figure, the ideal point would be at 100% τ_h identification effciency (on the x-axis) with a jet-to-τ_h misidentification probability (y-axis) near 0. The previous identification methods had set working point efficiencies while in principle DeepTau outputs a continuous result, depending on the wanted discrimination power. As this line always lies below old MVA based identification methods, the overall discriminatory power of the DeepTau ID is strictly better for the shown tau object candidates with $p_T > 100$ GeV. Different efficiency vs. misidentification rate points are selected, indicating *working points* (WP) of the ID. These points have been selected for further studies w.r.t. scale factor and uncertainty calculations by the CMS tau object group (see Sec. 6.2). The same procedure was applied for discrimination against electrons and muons [72].

In this analysis, the 'Tight' working point is chosen for each discriminator value in order to ensure well isolated tau objects. For discrimination against jets, this corresponds to an estimated misidentification rate (Fig. 4.3) of $\mathcal{O}(10^{-3})$ with a reconstruction efficiency of 75% for high transverse momentum tau objects. Compared to the same working point of the MVA ID, the misidentification rate is approximately a factor of 2 smaller. Discrimination against electrons and muons in this momentum regime is approximately one order of magnitude stronger, while also having a higher τ_h identification efficiency (80% for the vsElectron discriminator, and $>$ 99% for the vsMuon discriminator). This working point and the *VVVLoose* one (which will be relevant for the data-driven technique later) are indicated in Fig. 4.3.

4.3 Reconstruction of missing transverse momentum

Particles and objects which either cannot be detected or are missed by reconstruction methods contribute to missing transverse momentum. This quantity is derived from the principle, that in the initial collision of the two partons, basically no momentum transverse to the direction of travel of the protons exists. This quantity should be the same before and after the collision. Therefore, we can define missing transverse momentum by the negative sum of all momenta of all reconstructed particles [75]. This indirect reconstruction principle works for coordinates in transverse directions - the pseudorapidity η is not assignable for $p_{\mathrm{T}}^{\mathrm{miss}}$.

$$\vec{p}_T^{\,miss} := -\sum_{i \in all} \vec{p}_T^{\,i} \tag{4.1}$$

Due to reconstruction inefficiencies and detector misalignment, this is only an approximation of *true* missing transverse momentum (i.e. the original momentum of the invisible particles). In order to achieve a more precise estimation of this quantity, corrections are applied to $p_{\mathrm{T}}^{\mathrm{miss}}$. The final corrected version of $p_{\mathrm{T}}^{\mathrm{miss}}$ used in this analysis is *Type-1* corrected missing transverse momentum. This quantity takes into account corrections which are applied to the energy scale of jets (JEC) [76]. It is defined as

$$\vec{p}_T^{\,miss} := -\sum_{i \in jets}^{JEC} \vec{p}_T^{\,i} - \sum_{i \in uncl} \vec{p}_T^{\,i} \tag{4.2}$$

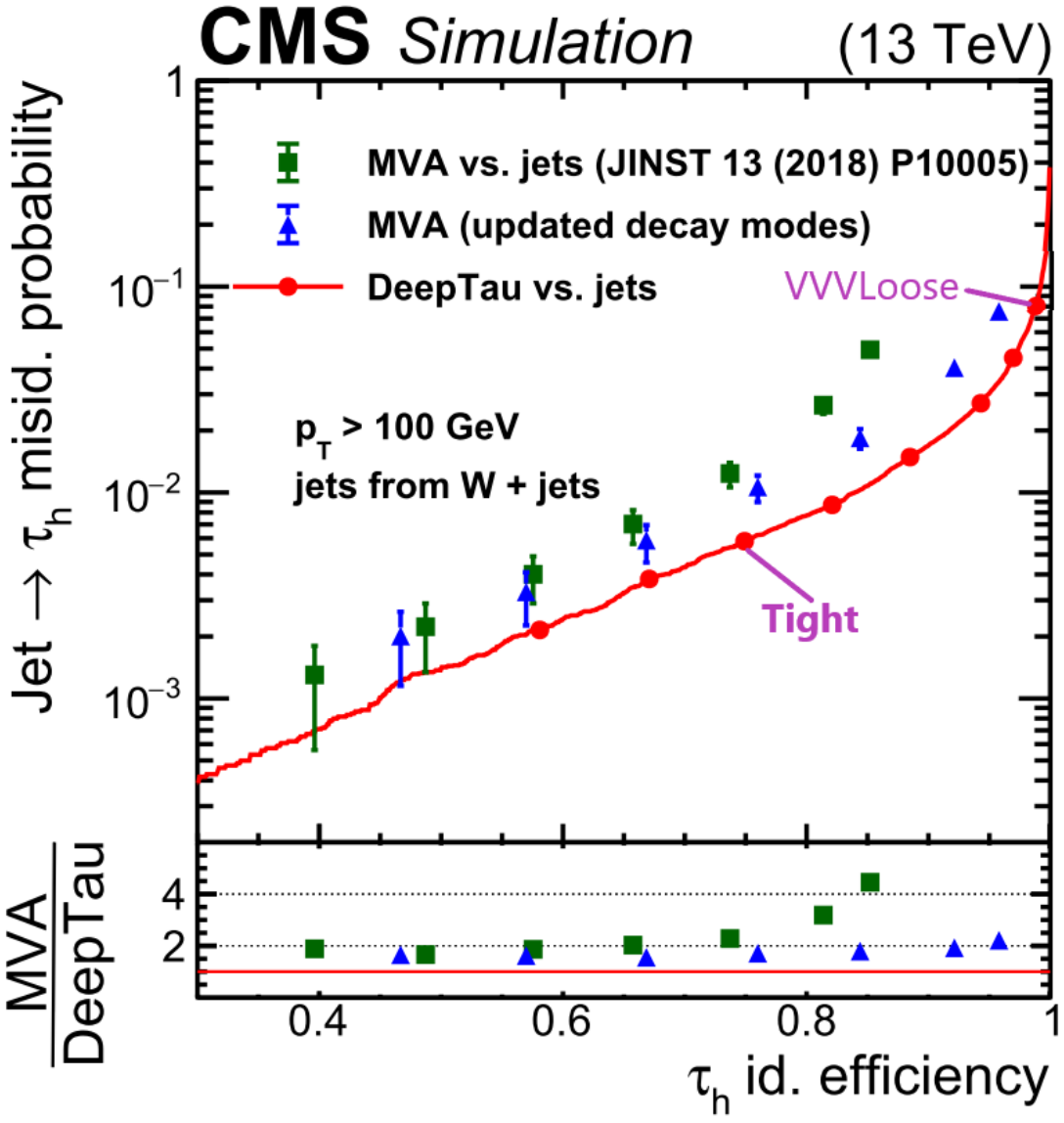

FIGURE 4.3: True τ_h selection efficiency compared with misidentification rate of jets for candidates with $p_T > 100$ GeV estimated from W+jets events. Dots indicate selected working points of the respective IDs. The DeepTau algorithm has overall a higher discriminatory power compared to the MVA identification method. The two working points relevant for this thesis are indicated in purple. Image taken from [72].

Here, the original sum was split between particles which can be clustered as jets and those which could not. For the former, jet energy corrections are applied and substituted for those which were in the sum originally. The latter *unclustered energy* part remains as is. For both contributing sums, uncertainties are applied: The first summand is driven by jet properties - corrections to those are propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$ in order to estimate their effect. This encompasses the propagation of jet energy scale uncertainties as well as jet energy resolution uncertainties. In this analysis, this propagation was performed using a standalone post-processing tool developed and provided by CMS working groups [77].

4.3.1 The m_{T} variable

In some models, we aim to reconstruct the mass of the non-SM particle, e.g. the mass of the W' boson in the SSM. Due to the missing η coordinate of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, it is not possible to reconstruct the exact mass of the original particle or object of the hadronic tau and the missing transverse momentum. Instead, m_{T} is used as a search variable, since it provides good signal to background separation. This quantity is defined as

$$m_T = \sqrt{2 p_T^{\tau_h} p_T^{miss} (1 - \cos(\Delta\Phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})))} \tag{4.3}$$

Here, $p_T^{\tau_h}$ is the transverse momentum of the hadronic tau object, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is the missing transverse momentum vector, $p_{\mathrm{T}}^{\mathrm{miss}}$ is its magnitude, and $\Delta\Phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ is the angle between the τ_h and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane.

4.4 Reconstruction of electrons and muons

Although not of immediate concern for the final state analyzed in this thesis, reconstruction of electrons and muons is necessary to understand for the data driven method used in this analysis. These objects are also vetoed as part of the analysis specific selection of the signal region. This small section gives a brief overview over the basic principles used to reconstruct these objects with the CMS detector. Some variables mentioned in this section are also used in the DeepTau neural network as input variables to discriminate genuine tau candidates from those from misidentification of electrons or muons [72].

4.4.1 Electrons

Electrons are reconstructed at CMS using both information from the tracker and the ECAL. The detector is designed in such a way, that electrons deposit all of their energy within the crystals of the calorimeter, allowing for a reconstruction of their energy and momentum. The charge of an electron can be determined from the bending direction of the track in the magnetic field of the CMS. Details about electron reconstruction used at CMS during Run-2 can be found in Ref. [78].

As electrons transverse the material of the CMS detector, they emit photons via bremsstrahlung, which in turn can convert into electron-positron pairs. In this way, an initial electron might become an electromagnetic shower instead of a single particle, making its reconstruction challenging and interesting. In Run-2, electron reconstruction is based on energy clusters in the ECAL, with minimum energy thresholds of 2-3 times the energy of expected noise. Around this seed cluster, a geometric area is defined, referred to as 'supercluster' in order to account for photon conversions and bremsstrahlung losses. In order to recover the energy and track of the initial particle, these superclusters are used together with information from the tracker as input for a Gaussian sum filter technique [79]. In addition to that, all tracks rebuilt in an event are tested with regard to their compatibility to an electron

trajectory hypothesis. Finally, all of this information (ECAL clusters, supercluster, Gaussian sum filter tracks and 'generic' compatible tracks) is used as input for the PF algorithm to determine electron and photon candidates.

To identify electron candidates, different sets of identification criteria can be used. The two different ones used in this analysis will be described here, namely the cut-based identification method used to veto events in the signal region and the high-energy electron pair identification (HEEP) used for determining side band regions.

The cut-based identification method is a selection approach, where seven variables from the electron candidates have to fulfil certain quality criteria. In this analysis, the *Loose* working point is used to veto events with electrons from the signal region. This working point has a high detection efficiency of 90% to ensure a highly functional event veto. This identification has been optimized for low-momentum ($p_T < 100$ GeV) candidates. A complete list of selection criteria for these types of electron candidates can be found in Tab. 4.2. This table also contains selection criteria for the HEEP identification method, since both identification methods share variables to be selected upon. This second approach has been optimized to identify high energy electrons ($p_T > 35$ GeV). The main difference between HEEP and cut-based identification methods is the use of a sub-detector based isolation method (HEEP) instead of a PF driven one (cut-based).

The specific variables selected upon include the transverse energy E_T and momentum p_T of the electron candidate and might depend on these quantities. The shape of the electromagnetic shower is taken into account via H/E: this is the ratio of energy deposited in the HCAL near the electron candidate relative to the energy deposited in the ECAL. For electrons, this quantity is expected to be small, while hadrons are expected to deposit significant amount of energy in the HCAL. Additionally, the geometric variable $\sigma^{5x5}_{i\eta i\eta}$ quantifies the shape in a 5x5 matrix around the most energetic crystal in the supercluster. This quantity is expected to be small for electron (or photon) showers and broad for neutral meson decays (like the π^0). Geometric variables like $d\eta_{in}$ and $d\Phi_{in}$ quantify the agreement of the initial ECAL seed and the initial tracker track for η and Φ, respectively. Similarly, the difference between the supercluster energy and the momentum of the track is expected to be small. Lastly, electron/ photon conversions are not expected before the detector material, making the number of missing hits in the innermost tracker layer a valuable discriminating variable. Lastly, to discern photons and electrons, a conversion veto is required, which evaluates the first tracker hits for charged particle tracks.

The selection variables used also include requirements on the isolation of the object candidate $I_{combined}$, which is a composite quantity evaluated from the sum of transverse momenta of photons, charged and neutral hadrons inside an isolation cone of $\Delta R = 0.3$ around the electron candidate. Pileup mitigation is also included via the variable ρ, which is the median E_T per unit area in the event. However, the HEEP identification uses sub-detector specific variables, namely shower shape dependent energy deposits in an $n \times m$ crystal grid: $E_{1\times5}$, $E_{2\times5}$ and $E_{5\times5}$. It also evaluates isolation variables based on the sub-detector system for each the tracker ($I_{tracker}$), the ECAL (I_{ECAL}) and the HCAL (I_{HCAL}).

4.4.2 Muons

Muons are, in principle, the only objects which reach the muon chambers of the CMS. They also leave a track in the tracker and have energy deposits in the calorimeters. However, they typically transverse the whole detector and travel outwards of it. Since they are not stopped, momentum measurement of these objects has to be made using the curvature of a muon track, which it inherits due to the large magnetic field of the CMS magnet. This caveat necessitates a high quality track of muon candidates to be able to precisely determine their

Variable	Loose (barrel)	Loose (endcap)
$\sigma^{5x5}_{i\eta i\eta}$	< 0.0112	< 0.0425
$\lvert d\eta_{in}\rvert$	< 0.00377	< 0.00674
$\lvert d\Phi_{in}\rvert$	< 0.0884	< 0.169
H/E	$< 0.05 + 1.16/E_{SC}$ $+0.0324\rho/E_{SC}$	$< 0.0441 + 2.54/E_{SC}$ $+0.183\rho/E_{SC}$
$I_{combined}/E_T$	$< 0.112 + 0.506/p_T$	$< 0.108 + 0.963/p_T$
$\lvert 1/E - 1/p\rvert$	< 0.193	< 0.111
missing inner hits	$<= 1$	$<= 1$
conversion veto	pass	pass
	HEEP (barrel)	HEEP (endcap)
E_T	> 35 GeV	> 35 GeV
η	$\lvert\eta_{SC}\rvert < 1.44$	$1.57 < \lvert\eta_{SC}\rvert < 2.5$
$\sigma^{5x5}_{i\eta i\eta}$	–	< 0.03
$\lvert d\eta_{in}\rvert$	< 0.004	< 0.006
$\lvert d\Phi_{in}\rvert$	< 0.06	< 0.06
H/E	$< 1/E_{SC} + 0.05$	$< 5/E_{SC} + 0.05$[1]
missing inner hits	$<= 1$	$<= 1$
$I_{tracker}$	< 5 GeV	< 5 GeV
$I_{ECAL} + I_{HCAL}$	$< 2 + 0.03E_T + 0.28\rho$	$< 2.5 + 0.28\rho$ $(E_T < 50)$ else $< 2.5 + 0.03(E_T - 50) + 0.28\rho$ [1]
$E_{2\times5}/E_{5\times5}$	> 0.94 OR $E_{1\times5}/E_{5\times5} > 0.83$	–
d_{xy}	< 0.02 cm	< 0.02 cm

TABLE 4.2: Identification criteria for both the cut-based *Loose* electron and the HEEP identification used in this analysis. The former is used to veto events in the signal region which contain an electron candidate, the latter is used to determine a side band region in the data-driven technique. Information taken from [81, 80].

[1] For 2018 data taking, the cuts were slightly adjusted to provide flat scale factors in E_T and η for the endcap region [80]. In 2018, two selections were changed to:
$H/E : 5/E_{SC} \rightarrow (-0.4 + 0.4|\eta|)\rho/E_{SC}$
$I_{ECAL} + I_{HCAL} : 0.28\rho \rightarrow (0.15 + 0.07|\eta|)\rho$

kinematic properties. The description in this chapter follows information from Ref. [64], where all details can be found.

Muon candidate tracks are reconstructed by two approaches. The 'inside-out' reconstruction uses tracks from the inner tracking detector, where every track with $p_T > 0.5$ GeV and $p > 2.5$ GeV is extrapolated outwards towards the muon detectors. If this extrapolation matches at least one muon detector segment, it classifies as a candidate. The 'outside-in' approach instead uses all information from the different muon sub-detectors (see Sec. 3.2.5) to build tracks using a Kalman-filter technique [82]. These are then matched with tracker tracks to find candidates. All these muon candidates are part of the input for the PF algorithm.

In this analysis, two sets of muon identification criteria are used:

- *Loose* identification: All muon candidates selected by the PF algorithm which have been reconstructed by either track reconstruction mentioned above classifies as a Loose muon. This identification method is used to veto events in the signal region. Since no muons are expected in the final state of interest for this analysis, the choice of this set of criteria was made to ensure rejection of all events where a muon could be present.

- *Tight* identification: These candidates have to fulfil additional criteria in addition to those mentioned above. The candidate has to be reconstructed by both track reconstruction methods mentioned above. Additionally, the tracker track has to have at least six layers of inner tracker hits, including at least one in the pixel detector. The track of the candidate has to match at least two segments of the muon system, and a global track fit has to have a $\chi^2/ndof < 10$ while including at least one hit from the muon system. Furthermore, the track has to be compatible with the primary vertex, i.e. it has a required transverse (longitudinal) impact parameter of $d_{xy} < 0.2$ cm ($d_z < 0.5$ cm). This set of criteria is used to ensure a muon is in the event, which will be used in the data driven technique described later.

For high transverse momentum muons ($p_T^{\mu} > 200$ GeV), dedicated reconstruction algorithms can be used [83]. Since this analysis uses muon objects just to define non-signal regions or to veto events and is not specifically interested in its kinematics, these methods are not used here.

Chapter 5

Data Sets and Simulation

5.1 Data Sets

This thesis analyzes the full CMS proton-proton collision dataset taken in the years 2016, 2017 and 2018. The exact data sets used, defined by their data path as described in 3.3 include the *Tau* and the *MET* data sets (which contain events triggered by tau or missing transverse momentum triggers) for the signal region as well as the *SingleMuon* and *SingleElectron/Photon* data sets, which are used for background determination in sideband regions. All used data sets are reconstructed with the *02Apr2020* settings as recommended by the CMS "Phyics Performance and Dataset"–group [84]. This group also provides a so called 'golden JSON' files, which contain data taking periods which have been approved to include events with optimal detector performance. Data events which are not included in these files are not analyzed in this thesis. The remaining amount of data events corresponds to a data set of a total integrated luminosity of ≈ 138 fb^{-1}.

5.1.1 Issues in Run2 data

For each data taking period, either during the data taking itself or after analyzing the reconstructed data initially, unexpected behavior of the detector or the trigger system may be seen. The different problems which occurred in the data taking period analyzed within this thesis will be discussed in this paragraph. A small overview on the affected periods for different problems is given in Tab. 5.1.

Level 1 Pre-firing

During the data taking periods of 2016 and 2017, the gradual timing shift of ECAL was not properly propagated to L1 trigger primitives (TP) [85]. This lead to a significant fraction of high eta TP being mistakenly associated to the previous bunch crossing. Since L1 rules forbid two consecutive bunch crossings to fire, events could veto if a significant amount of ECAL energy is found in the region of $2 < |\eta| < 3$. This effect is not described by the simulations. To accommodate for this effect, event weights have been calculated as functions of p_T and η for both jet and photon trigger objects for each affected year. These

Year	Integrated Luminosity	Known Issues
2016	35.9 fb^{-1}	L1 Pre-firing
2017	41.3 fb^{-1}	L1 Pre-firing, EE noise on p_T^{miss}
2018	60.0 fb^{-1}	HEM 15/16 problem

TABLE 5.1: Overview on the data taking periods, their total integrated luminosity and the occurring phenomena apparent for each corresponding data set.

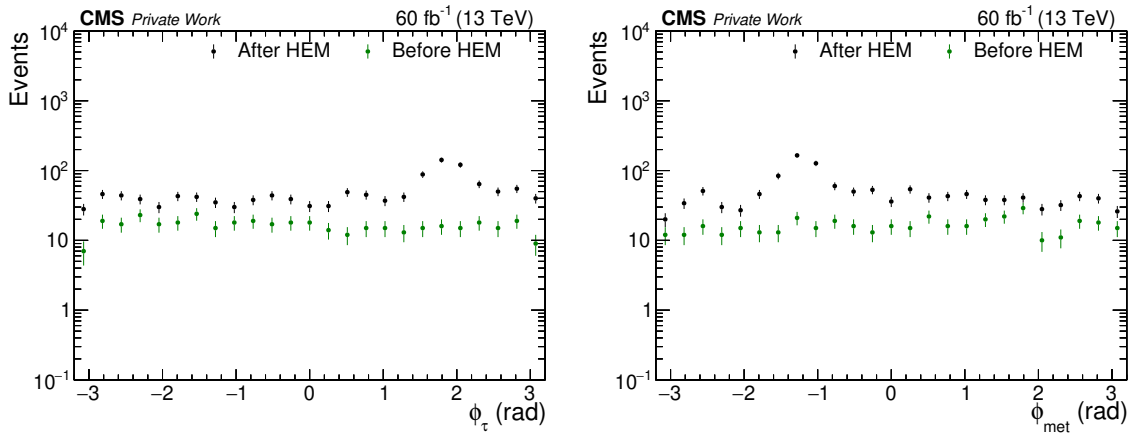

FIGURE 5.1: Distribution of tau (left) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (right) Φ for data events before and after the power outage which disabled HEM sections (HCAL endcap) 15 and 16.

are then applied to simulated events in order to correct for this feature. These weights also have their set of associated uncertainties to account for possible fluctuations.

Effect of EE noise on $p_{\mathrm{T}}^{\mathrm{miss}}$

During 2017 data taking an increased amount of instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ was observed in data / simulation comparisons [86]. The origin was traced to an interplay of various effects, including aging of the ECAL at high values of η and out-of-time pileup. This effect was accounted for after 2017 data taking was completed and corrected for 2018 data taking. For 2017, a recipe was developed to account for this effect: Jets and unclustered objects with a low transverse momentum from this high η region are not considered for the calculation of $p_{\mathrm{T}}^{\mathrm{miss}}$. This recipe is already applied for 2017 samples on production level for this analysis and its uncertainties are incorporated in the uncertainties for $p_{\mathrm{T}}^{\mathrm{miss}}$.

HEM 15/16 failure

During the data taking period of 2018, false fire alarms went of on June 30th 2018, which lead to several power interruptions. Following that incident, the negative endcap sectors HEM15 and HEM16 from the HCAL could not be operated anymore until the end of the 2018 data taking period. The affected area covers a range of $-3.0 < \eta < -1.3$ and $-1.57 < \Phi < -0.87$. This necessarily impacts the reconstruction of objects in that region which is not included in simulated samples produced before 2018 data taking. As the tau object is in its core a jet measured by the HCAL, the impact of this detector failure on these objects is measured specifically for this analysis. Data events recorded before the problem arose are compared to data events taken after the incident. A distribution for that is shown in Fig. 5.1. As can be seen, the affected region is clearly visible in Φ. The effect is shifted by π for missing transverse momentum - jets in the affected HEM region are not fully reconstructed, therefore have less momentum. By the definition of $p_{\mathrm{T}}^{\mathrm{miss}}$ as negative sum of all other particles, the effect is shifted.

Similar difficulties have been shown in multiple analyses which rely on jet or $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction [86]. A recommended solution from studying these problems in detail was applied to this analysis: Events which contain any jet with a transverse momentum of at least 20 GeV in the problematic region, are vetoed. The effect of this solution is shown in

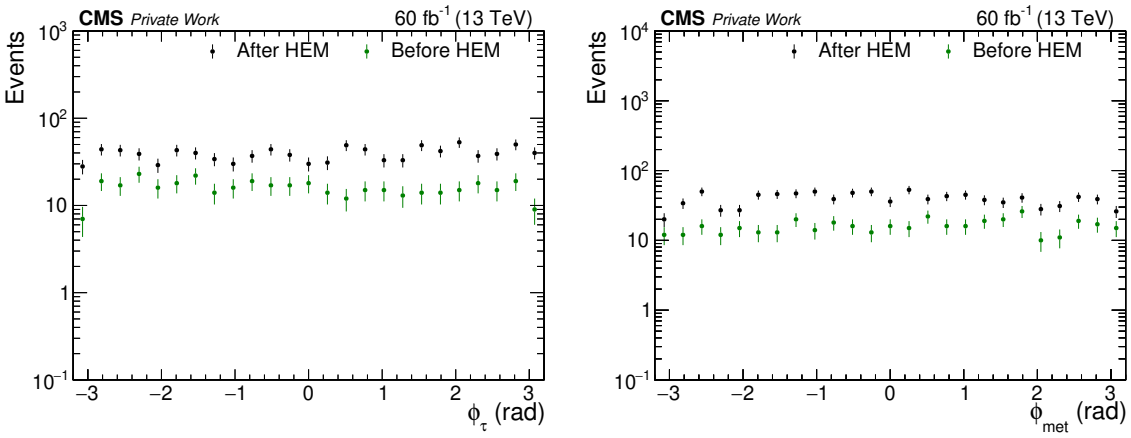

FIGURE 5.2: Distribution of tau (left) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (right) Φ for data events before and after the power outage which disabled HEM sections (HCAL endcap) 15 and 16. The recommended procedure, which vetoes events with jets in the problematic region has been applied.

Fig. 5.2. This can be used as a direct comparison to the previous figure, and it can be seen, that the large, unexpected event spike seen in Φ is effectively removed with this recipe.

5.2 Background simulation

In order to discern new physics which might show up in the measured data, we need a precise background estimation. The background in this search has two fundamentally different types: background with genuine hadronic tau lepton decays and backgrounds with objects misidentified as tau leptons. This section describes the procedures to simulate such backgrounds.

5.3 Simulated Backgrounds

The background comes from known and well studied Standard Model processes, which exhibit similar kinematics as the signal processes. The main background for that type of processes is the W background process. It is kinematically very similar to our signal, except for the mass of the intermediate particle. A Feynman diagram for a simple production of a W boson is given in Fig. 5.3.

Other backgrounds which can produce a similar kinematic final state as the signal are top-processes, the Drell-Yan processes and backgrounds in which two bosons are produced (*Diboson processes*). Exemplary Feynman diagrams for these backgrounds are given in Fig. 5.4. These backgrounds can enter the final selection in special cases. E.g. for $t\bar{t}$ and WW production, two W bosons are produced in the decay chain. If one of these bosons decays hadronically and the other one decays to a tau and a neutrino, the signal final state can be reached. Backgrounds can also enter the final state selection in cases when a particle is missed by the reconstruction - e.g. the DY process can lead to production of two hadronically decaying τ leptons. If one is missed, the event appears in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state.

Another important aspect is the background originating from quark and gluon initiated jets which are misidentified as tau objects. For this analysis, the main SM processes contributing to such backgrounds are QCD processes and the DY process. The latter can mimic the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state when the leptons produced in the decay of the intermediate particle

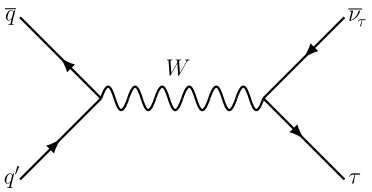

FIGURE 5.3: Feynman diagram of leading order s-channel W production.

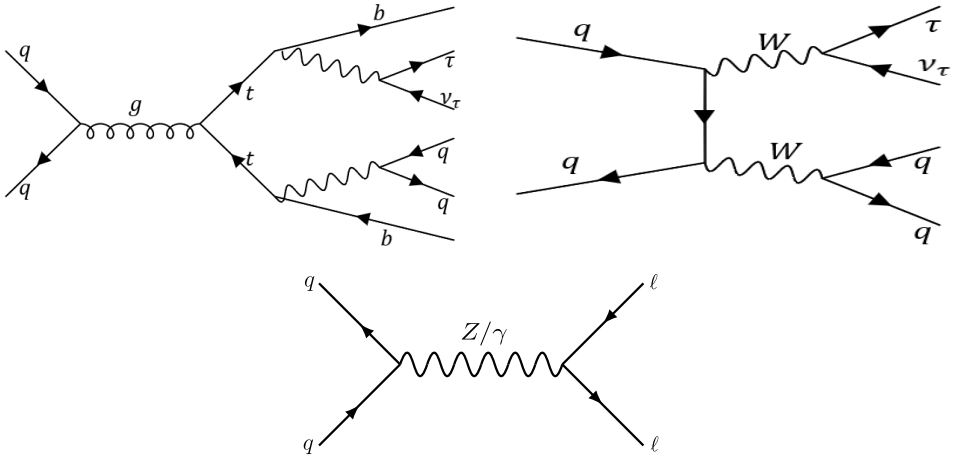

FIGURE 5.4: Exemplary Feynman diagrams for non-leading simulated backgrounds. On the top left, an exemplary diagram of top-pair-production is given. In the top right, the production of two W-bosons is shown and on the bottom, the Drell-Yan process can be seen.

are neutrinos. In this case, a high amount of missing transverse momentum is already apparent in the event, and any initial or final state radiation jet can be misidentified as a tau lepton. Due to the LHC being a hadron collider, many SM processes can contribute to this background, but this will be discussed in detail in Sec. 6.3.

Backgrounds are simulated using *Monte Carlo* (MC) generators and processed through a detector simulation of CMS. This process involves several steps:

- **Generation of events:** First, a MC generator simulates proton-proton collisions and calculates the four-momenta of all outgoing particles created during this collision. Several programs have been designed to simulate such events and can be compared to each other in their prediction of events i.o. to get a precise background estimation. Generators used for background estimation in this analysis are Pythia8 [87], Madgraph_aMC@NLO [88] and PowHeg [89, 90, 91].

- **Parton shower:** Parton showering is performed by Pythia8 for each MC sample. The hard interaction of interest is accompanied by partons from initial state radiation and final state radiation. Other components which need to be taken into account are underlying event conditions like beam-beam remnants (i.e. what remains of the parton scattered from the two protons) and multiple-parton interactions (semi-hard parton scatterings which occur in the same pp collision). These event conditions are controlled in simulation with a set of parameters referred to as *tune*. For 2016 production, the CUETP8M1 tune [92] was used, while 2017–2018 production used the TuneCP5 [93] tune for background simulation. This change between data taking periods is due to updated underlying parton distribution functions and new experimental data obtained for fitting simulation to data in multiple-parton interactions.

- **Detector simulation:** Generated events are input into a precise simulation of the CMS detector. For this, the GEANT4 [94] package is used.
- **Digitization and reconstruction:** Simulated detector signals are digitized and reconstruction algorithms are applied on top, including the simulation of triggering. In this process, pileup is also included in the simulation.
- **Additional algorithms:** The previous step produces files which are already comparable to experimental data. Additional levels of sample post-processing can be applied to these samples. The for this analysis most prominent algorithms applied during this step is the tau identification algorithm and the calculation of jet energy corrections. This analysis uses files in the NanoAOD format [95], a user-friendly data format developed by the CMS community to allow for simple access and analysis.

These steps of event simulation are performed for all SM processes. On-shell W production is simulated at LO with Madgraph_aMC@NLO. The simulated samples include the generic on-shell W, as well as high statistics samples for different ranges of HCAL energy deposits, i.o. to ensure a precise event prediction for the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state. As we are interested in the high-mass tails of this distribution, off-shell W samples are produced with high statistics in this kinematic region. These were simulated with the Pythia8 generator at LO.

Backgrounds from top production (both single and pair production) are simulated using the PowHeg Box generator at NLO. Backgrounds from the DY process are included using NLO samples generated with Madgraph_aMC@NLO. Also for this process, specific background samples with increased statistics in the high-mass region are used. For estimation of misidentified backgrounds, LO QCD background samples generated with Pythia8 are also included. A precise overview on the used simulated samples, their perturbation order and the used cross sections is given in Tab. 5.2. An additional table with background samples which mainly contribute to the background from misidentified tau objects is given in Tab. 5.3. Cross sections may differ slightly between 2016 and 2017–2018 due to different generator tunes and PDF sets.

Overlap between samples which describe the same kinematic region has been removed by preferring those for which a dedicated sample exists. For example, if an event from the W+Jets bulk sample has generated W boson mass $m_W > 200$ GeV, this event is discarded as there exists a dedicated sample for that region.

Background contributions are normalized to the luminosity ($\mathcal{L}$ of the measured data by scaling the respective background with $\frac{\sigma_{MC}\cdot\mathcal{L}}{N_{MC}}$.

5.3.1 Parton Distribution Functions

An important input for event generation is the parton distribution function (PDF) which describes the probability of finding a specific parton (i.e. quarks and gluons) with a longitudinal momentum fraction x at a resolution scale Q^2. The dependency on Q^2 can be evaluated using the DGLAP equations [96]. The x dependency, which describes the inner proton dynamics, lies in the mathematical realm of QCD, where perturbation theory cannot be applied. Therefore, these functions are determined from measurements at various experiment and estimated using effective descriptions. Various groups use different estimation methods to provide descriptions of the measured data and estimations for regions beyond. For high energy searches as presented in this thesis, these functions have little to no data yet, which typically results in a large uncertainty associated to it.

The specific implementation for PDFs in this analysis follows the PDF4LHC guidelines [25, 97] using NNPDF3.1 [24]. For each generated event, a set of replica weights represents the uncertainty originating from both the experimental and the derivation uncertainties from

Sample Name	Generator	Pert. Order	σ[pb] (2016)	σ[pb] (2017/2018)
W background				
WJetsToLNu	aMC@NLO	LO	50260.0	52850.0
- H_T binned				
—70-100 GeV	aMC@NLO	LO	1353.0	1292.0
—100-200 GeV	aMC@NLO	LO	1346.0	1256.0
—200-400 GeV	aMC@NLO	LO	360.1	407.9
—400-600 GeV	aMC@NLO	LO	48.80	57.48
—600-800 GeV	aMC@NLO	LO	12.07	12.87
—800-1200 GeV	aMC@NLO	LO	5.497	5.366
—1200-2500 GeV	aMC@NLO	LO	1.329	1.074
—> 2500 GeV	aMC@NLO	LO	0.0321	0.0080
- m_W binned[1]				
—> 100 GeV	Pythia8	LO	150.1	173.5
—> 200 GeV	Pythia8	LO	6.206	7.259
—> 500 GeV	Pythia8	LO	0.2137	0.2503
—> 1000 GeV	Pythia8	LO	0.01278	0.01497
—> 2000 GeV	Pythia8	LO	4.020×10^{-4}	4.363×10^{-4}
—> 3000 GeV	Pythia8	LO	2.895×10^{-5}	2.787×10^{-5}
—> 4000 GeV	Pythia8	LO	2.694×10^{-6}	2.621×10^{-6}
—> 5000 GeV	Pythia8	LO	2.705×10^{-7}	3.794×10^{-7}
—> 6000 GeV	Pythia8	LO	2.782×10^{-8}	7.349×10^{-8}
$\gamma*/Z$ **background**				
DYJetsToLL $m_Z > 50$	aMC@NLO	NLO	5940.0	6435.0
- 100 GeV$\leq m_Z <$ 200	aMC@NLO	NLO	226.6	247.8
- 200 GeV$\leq m_Z <$ 400	aMC@NLO	NLO	7.770	8.502
- 400 GeV$\leq m_Z <$ 500	aMC@NLO	NLO	0.4259	0.4514
- 500 GeV$\leq m_Z <$ 700	aMC@NLO	NLO	0.2446	0.2558
- 700 GeV$\leq m_Z <$ 800	aMC@NLO	NLO	0.0378	0.0402
- 800 GeV$\leq m_Z <$ 1000	aMC@NLO	NLO	0.0318	0.0341
- 1000 GeV$\leq m_Z <$ 1500	aMC@NLO	NLO	0.0202	0.0183
- 1500 GeV$\leq m_Z <$ 2000	aMC@NLO	NLO	2.286×10^{-3}	2.367×10^{-3}
- 2000 GeV$\leq m_Z <$ 3000	aMC@NLO	NLO	5.375×10^{-4}	5.409×10^{-4}
- 3000 GeV$\leq m_Z$	aMC@NLO	NLO	not available	3.048×10^{-5}
top backgrounds				
top: s-channel leptonic decays	aMC@NLO	NLO	3.365	3.74
top: t-channel antitop	PowHeg	NLO	67.91	67.91
top: t-channel top	PowHeg	NLO	113.3	113.3
$t + W$ production	PowHeg	NLO	38.09	34.91
$\bar{t} + W$ production	PowHeg	NLO	38.06	34.97
$t\bar{t}$: semi-leptonic decays	PowHeg	NLO	320.1	300.9
$t\bar{t}$: leptonic decays	PowHeg	NLO	76.70	72.10
Diboson backgrounds				
WWTo2L2Nu	PowHeg	NLO	10.48	10.48
WWTo4Q	PowHeg	NLO	45.2	45.2
WWToLNuQQ	PowHeg	NLO	45.68	45.68
WZTo1L1Nu2Q	aMC@NLO	NLO	10.73	not available
WZTo1L3Nu	aMC@NLO	NLO	3.054	not available
WZTo2L2Q	aMC@NLO	NLO	5.606	not available
WZTo2Q2Nu	aMC@NLO	NLO	6.317	not available
WZTo3LNu	aMC@NLO	NLO	4.679	not available
WZ	Pythia8	LO	not used	27.59
ZZTo2L2Nu	PowHeg	NLO	0.5644	0.6008
ZZTo2L2Q	aMC@NLO	NLO	3.222	3.703
ZZTo2Q2Nu	aMC@NLO	NLO	4.033	5.033
ZZTo4L	aMC@NLO	NLO	1.204	1.325

TABLE 5.2: Overview on simulated event samples used in this analysis, with their respective generator, perturbation order and cross section (σ). Note, that cross sections differ slightly between 2016 and 2017/18 due to a change in PDF sets and generator tune. The decay specific sample splitting for the WZ process was combined for the 2017/2018 sample production to a general WZ sample.

Sample Name	Generator	Pert. Order	σ[pb] (2016)	σ[pb] (2017/2018)
QCD backgrounds				
p_{T} binned				
50 - 80 GeV	Pythia8	LO	1.910×10^{7}	1.571×10^{7}
80 - 120 GeV	Pythia8	LO	2.735×10^{6}	2.336×10^{6}
120 - 170 GeV	Pythia8	LO	4.689×10^{5}	4.073×10^{5}
170 - 300 GeV	Pythia8	LO	1.178×10^{5}	1.035×10^{5}
300 - 470 GeV	Pythia8	LO	7753.0	6830.0
470 - 600 GeV	Pythia8	LO	642.1	552.1
600 - 800 GeV	Pythia8	LO	185.9	156.5
800 - 1000 GeV	Pythia8	LO	32.05	26.28
1000 - 1400 GeV	Pythia8	LO	9.451	7.470
1400 - 1800 GeV	Pythia8	LO	0.8457	0.6484
1800 - 2400 GeV	Pythia8	LO	0.1124	0.08743
2400 - 3200 GeV	Pythia8	LO	6.752×10^{-3}	5.236×10^{-3}
$>$ 3200	Pythia8	LO	1.626×10^{-4}	1.357×10^{-4}
top: s-channel hadronic decays	aMC@NLO	NLO	not available	11.24
$t\bar{t}$: hadronic decays	PowHeg	NLO	313.9	313.9

TABLE 5.3: List of simulated event samples used for estimating contributions of misidentified tau objects and for cross checking the data driven method in this analysis, with the sample's respective generator, perturbation order and cross section (σ).

the PDF. The two proposed methods to incorporate these uncertainties are either the use of *MC replicas* or Hessian sets. For the former, the uncertainty calculation for the cross section is given by

$$\delta^{PDF}\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\sigma^{i} - \frac{1}{N}\sum_{j=1}^{N}\sigma^{j})^{2}} \tag{5.1}$$

In this equation, $\frac{1}{N}\sum_{j=1}^{N}\sigma^{j}$ is the mean value of the cross section, and σ^{i} are the individual values of each of the N replicas.

For the Hessian set approach, the calculation is based on the 30 stored eigenvectors. The uncertainty then is estimated using

$$\delta^{PDF}\sigma = \sqrt{\sum_{i=1}^{N}(\sigma^{i} - \sigma^{0})^{2}} \tag{5.2}$$

where σ^{i} are the variations and σ^{0} is the nominal cross sections. This results in an uncertainty which is to be understood as a 68% confidence level.

PDF uncertainties for signals are treated differently compared to background simulation. Effects on the normalization originating from PDF uncertainties are neglected for signal processes, which is a common approach in NP searches. This is due to the fact that for new physics processes, theoretical cross section uncertainties are subject to change and expected to improve in the future. Therefore, only the effect of PDF uncertainties on the signal acceptance is taken into account. In case new theoretical cross sections are calculated in the future, the cross sections in this search can therefore be simply rescaled.

5.4 Signal Simulation

The same procedure used for background simulation is also used for signal simulation. Depending on the underlying model, different recommended calculation methods are used to

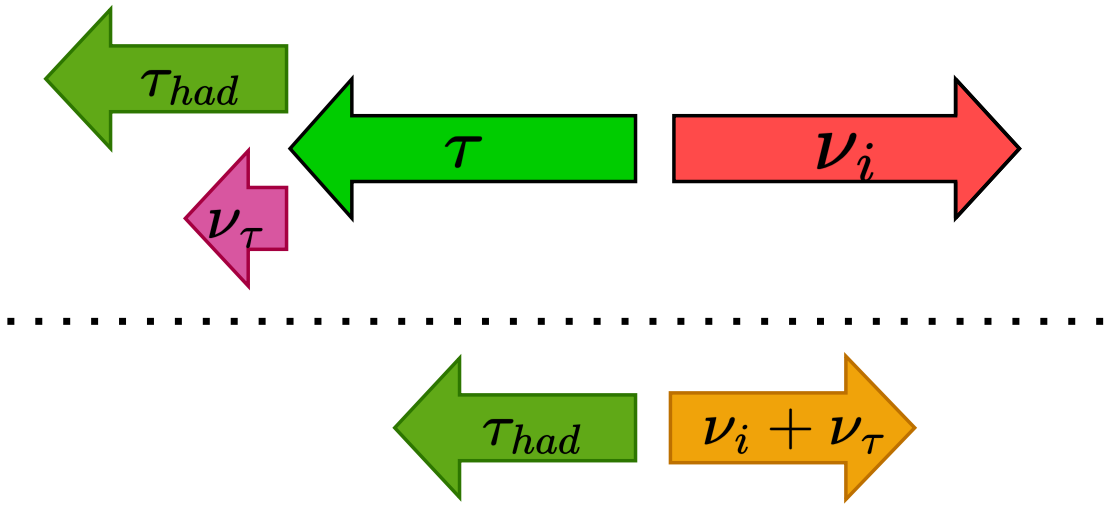

FIGURE 5.5: Simplified sketch of signal kinematics which are similar between all analyzed signals in this thesis. The length of the arrows indicate their transverse momentum. Other coordinates are neglected in this diagram. Above the dotted line, the prompt emission of the τ lepton and the initial neutrino ν_i and the subsequent decay process of the τ lepton to its hadronic components and the ν_τ is visualized, leading to the resulting detectable physical quantities given below the dotted line.

describe the new physics. This section gives a short overview, how all signal processes are generated, as well as an overview on the specific signal kinematics on generator level i.o. to learn discriminating variables to discern them from background.

Distributions shown in this section are shown on generator level - the properties of the object in question are the coordinates without any reconstruction effects. For the tau lepton, there are two 'levels' of generator information, which need to be distinguished. τ_{gen} represents the prompt object, as it might be created in the decay of a SM boson. τ_{gen}^{vis} represents the sum of all visible hadronic particles which originate from a decay of such a tau lepton. The latter corresponds more to what is expected to be measured (assuming perfect reconstruction), while the former can easily be connected to 'pure' kinematics (which do not include additional decays). τ_{gen}^{vis} by construction only exists in events of hadronic tau decays, thus approximately 65% of all signal events. As mentioned in the introduction, leptonic decays from the initial tau decay are neglected in this analysis.

All signal hypotheses analyzed in this thesis have a common kinematic signature, which is visualized in Fig. 5.5. As shown here, some form of new physics produces a τ lepton and a prompt neutrino ν_i, which are emitted back-to-back. The τ lepton subsequently decays, emitting an additional neutrino ν_τ, which travels roughly into the same direction as the original τ lepton. This secondary neutrino typically has much less momentum than the prompt neutrino, therefore only reducing the overall missing transverse momentum. This leads to slightly different kinematic selections for signal extraction comparing to $e/\mu+p_{\mathrm{T}}^{\mathrm{miss}}$ searches [98].

5.4.1 Sequential Standard Model - W'

The SSM assumes the prompt production of a heavy charged vector boson with a certain mass and width. Examples for various mass points are shown in Fig. 5.6. This is a purely theoretical distribution, since the subsequent decay into a tau and a neutrino necessitates the

measurement of $p_{\mathrm{T}}^{\mathrm{miss}}$. Therefore, the approach to reconstruct such a mass is the m_{T} variable, as defined earlier (see Eq. 4.3). This variable, estimated using generator information, is also included in Fig. 5.6. As can be seen, the sharp peak has transformed into a much broader distribution, known as a Jacobian peak. This is characterized by a rising edge towards a peak and a fast subsequent drop in the mass distribution. In the hadronic tau decay channel, an additional neutrino is emitted from the tau decay. This affects the magnitude of $p_{\mathrm{T}}^{\mathrm{miss}}$, again smearing out the mass distribution of the W', indicated by the $m_{\mathrm{T}}^{\mathrm{gen,vis}}$ distributions in Fig. 5.6. These distributions visualize not only the 'evolution' of the reconstructable variable, but also the increasing off-shell production for higher masses of the W' boson. This can be seen in the increasingly large tails towards lower masses in the m^{gen} distributions, which translates towards $m_{\mathrm{T}}^{\mathrm{gen,vis}}$. The shown generator mass m^{gen} is built using matched tau and neutrino four vectors, as is the concurrent $m_{\mathrm{T}}^{\mathrm{gen}}$, which loses information on the neutrino's pseudorapidity. $m_{\mathrm{T}}^{\mathrm{gen,vis}}$ is built from generator combinations of the sum of hadronic decay products and the sum of neutrinos in the event - which is the reconstructable variable. As can be seen, the resonant structure is lost in this process, however the maximum of the $m_{\mathrm{T}}^{\mathrm{gen,vis}}$ variable is very close to the original W' mass.

The effect of the hadronic tau decay on the reconstructable kinematics of the signal can also be visualized very well by comparing the transverse momentum of the generator tau lepton with that of the generator matched hadronic tau lepton decay products, shown for various mass hypotheses of the SSM W' in Fig. 5.7. Here it can be seen that the original peak like structure in transverse τ_{gen} momentum is mostly lost (when looking at τ_{gen}^{vis}) due to the subsequent decay of the tau lepton.

SSM signal simulation was performed using NNPDF2.3 as the PDF set in combination with the generator tune TUNECUETP8M1 for 2016. For 2017–2018, the NNPDF3.1 PDF set was used in combination with TuneCP5.

5.4.2 Coupling Dependent W'

The SSM assumes the same coupling values for the hypothesized W' boson and the SM W boson. However, the model can be expanded to include variations w.r.t. their relative overall coupling strengths in the decay vertex of the W'. In particular, the relative coupling strength $g_{W'}/g_W$ can be investigated upon. This parameter affects the width and the cross section of the new heavy vector boson, as shown in Fig. 5.8. As can be seen, higher ratios lead to an increased width and lower ratios result in a more defined peak. At the same time, the cross section increases with larger coupling ratios, as discussed in Sec. 2.2.1.

Private Signal Sample Generation

In order to analyze the sensitivity to different coupling strengths, signal samples with varying $g_{W'}/g_W$ and therefore varying decay width of the W' are simulated. Since Pythia8 does not include this parameter, signal samples were produced using Madgraph5. For all W' masses, samples were created with coupling ratio values of [0.0001, 0.001, 0.01, 0.1, 1.0, 2.0, 3.0, 5.0], resulting in 240 signal samples in total. PDF sets and tune choices are identical to SSM signal generation (2016: NNPDF2.3 with TuneCUETP8M1, 2017–2018: NNPDF3.1 with TuneCP5).

In order to ensure equivalence between the two generators, we can investigate the generated mass distributions for a coupling ratio value of 1.0, which is the value the SSM assumes. The generator $m_{W'}^{gen}$ distributions for signal mass hypotheses of 1 and 5 TeV is shown in Fig. 5.9. As can be seen, the different generators produce mostly the same result within statistical uncertainties. The only region where the generators predict slightly different event

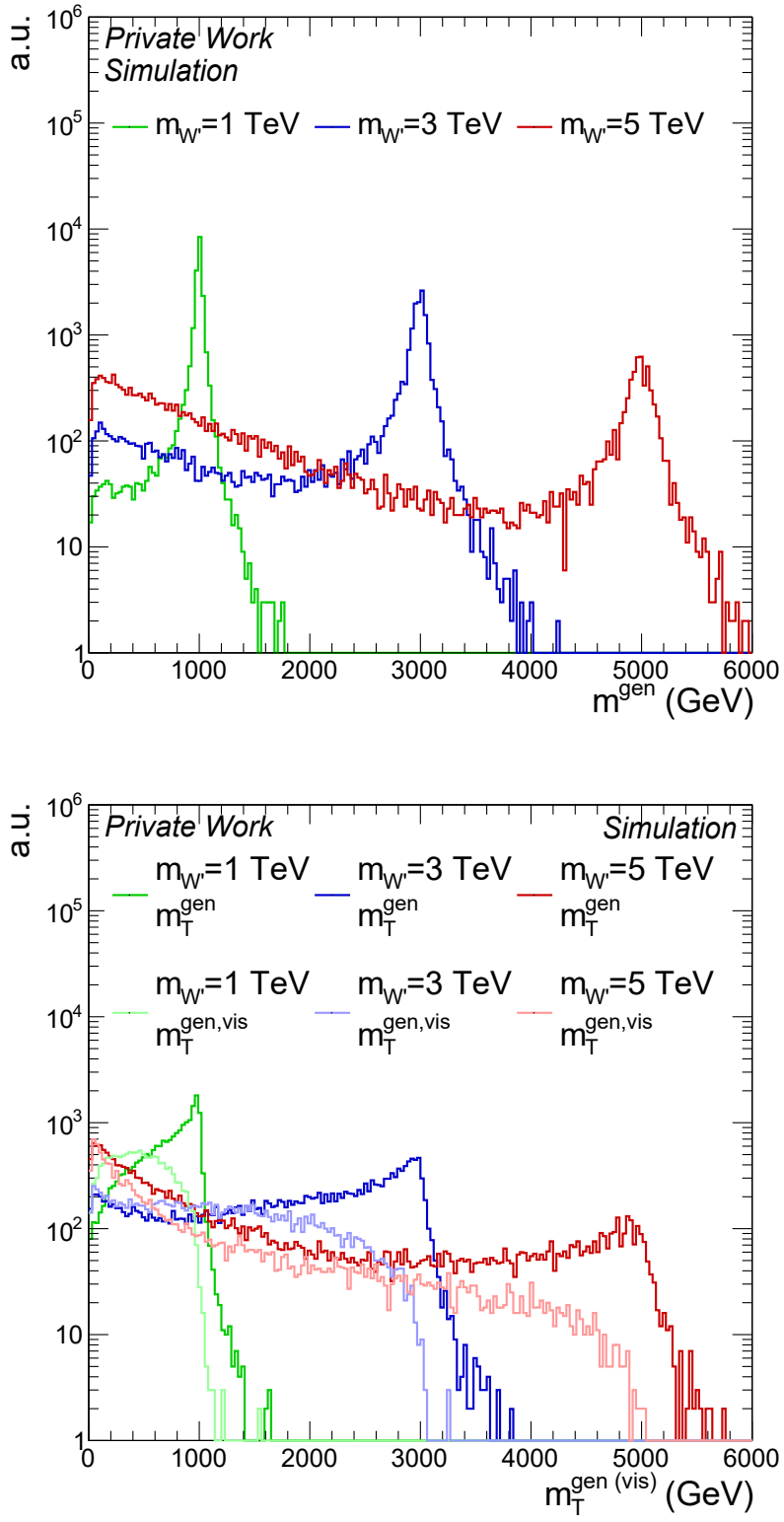

FIGURE 5.6: Depiction of different steps between generated mass and (ideally) reconstructed m_{T} for various mass points for the SSM. Each step loses information on the resonance mass due to the inclusion of additional neutrinos. The generator mass m^{gen} (top) and the transverse mass $m_{\mathrm{T}}^{\mathrm{gen}}$ (bottom) are built using generator matched tau and neutrino four vectors. $m_{\mathrm{T}}^{\mathrm{gen,vis}}$ (bottom) is built from generator combinations of the sum of hadronic decay products and the sum of neutrinos in the event. The resonant structure is lost in this process, however the maximum of the $m_{\mathrm{T}}^{\mathrm{gen,vis}}$ variable is very close to the original W' mass.

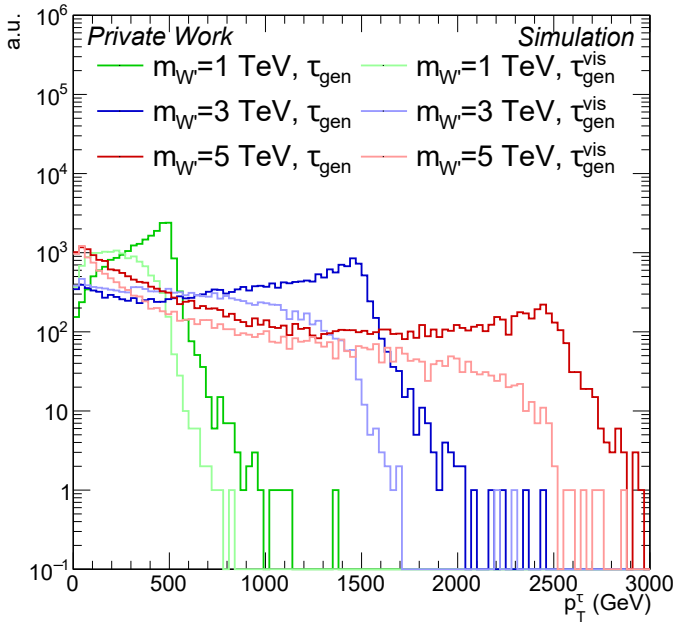

FIGURE 5.7: Comparison of the transverse momentum of τ_{gen} and τ^{vis}_{gen} for various SSM W' signal samples. As can be seen, the characteristic peak structure of the generated lepton is smeared out when accounting for the hadronic decay of that tau lepton due to the loss of information via the emitted tau neutrino. The distributions are normalized to the total number of events in each signal sample.

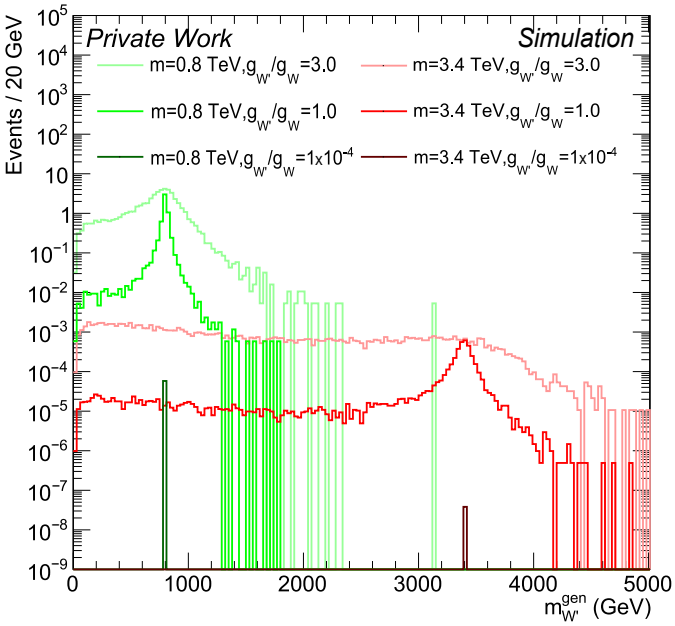

FIGURE 5.8: Distributions of generator $m_{W'}$ for different mass points and different coupling ratio $g_{W'}/g_W$ values. Signal distributions are scaled to their corresponding theoretical cross section to show the impact of the parameters on width and cross section of the signal shape. The low coupling ratio samples ($g_{W'}/g_W = 1 \cdot 10^{-4}$) have been scaled up by a factor of 1000 for better visibility.

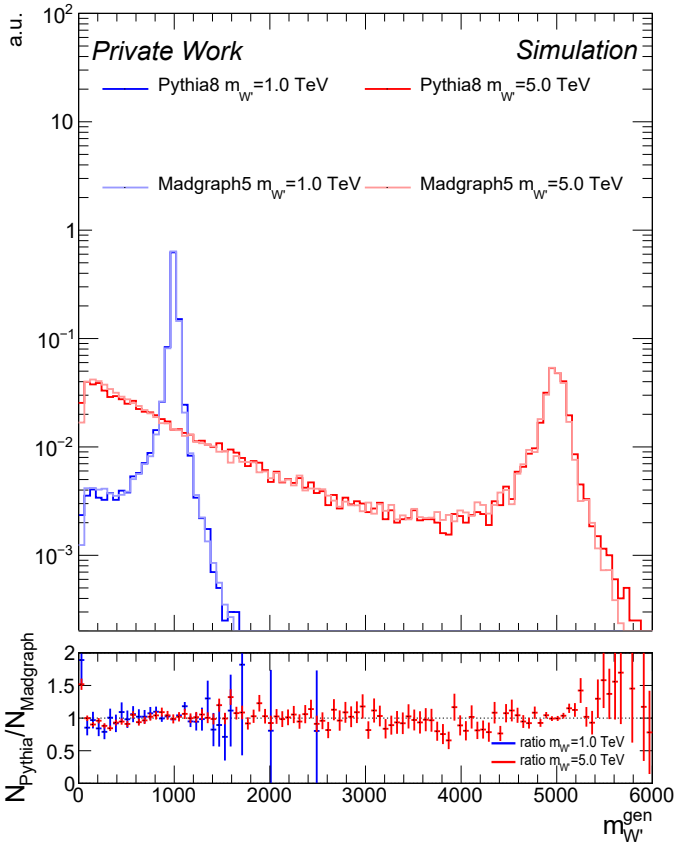

FIGURE 5.9: Comparison of $m_{W'}$ on generator level between Pythia8 and Madgraph5 for mass hypotheses of 1 and 5 TeV, respectively. Both generators produce comparable results within statistical uncertainties. Distributions are normalized to unity.

numbers is the low-mass (off-shell) region[2]. For this analysis however, this is not a significant effect, as events from this region will not pass the high momentum requirements for both the τ_h and the $p_{\mathrm{T}}^{\mathrm{miss}}$.

Previous iterations of this analysis generated these signal samples on generator level only. In order to estimate the signal shape after detector simulation and reconstruction, a reweighing method was used[3]. For this, the reconstructed signal m_{T} of the official signal samples was divided by the generator signal m_{T} of the same generator. This ratio was then applied to generator signal shapes for all coupling values for Madgraph5 signal samples to estimate the effects of simulation and reconstruction. While this method gives a reasonable approximation, for this analysis it was decided to create the signal samples including full detector simulation and reconstruction. The detector simulation was performed by GEANT4 and the decay and hadronization of the generated τ lepton was performed by the Tauola package of Pythia8. Apart from the choice of the generator (Madgraph5 instead of Pythia8), the signal sample processing was kept analogously to the official signal sample production. Samples were fully processed for 2016 and 2018, containing approximately 10^5 events per

[2]Off-shell production in Madgraph (v.2.6.2) was explicitly enabled by changing the BWCUTOFF setting to a large number.

[3]It has to be noted, that the exact procedure to reweigh the signal samples was not available anymore and had to be rebuilt according to the previous analyst's documentation.

sample, resulting in about 24 million events per data taking year[4]. The production was not repeated for 2017, since differences between 2017 and 2018 production are very small. The improvement of this method is shown in Fig. 5.10. Here, the prediction of the reweighing method is compared to the fully reconstructed and analyzed m_T distribution for different mass and coupling hypotheses. As can be seen, the previous method was a decent approximation, but did not describe the full spectrum perfectly, especially at high masses.

5.4.3 Non-Universal Gauge Interaction Model

Of special interest for this analysis is a non-universal gauge interaction model (NUGIM). This model predicts, similar to the SSM, new heavy vector bosons, but allows for the possibility that the coupling of the new particles to different lepton families varies (see Sec. 2.2.2). Specifically, in certain regimes of this model, the branching fraction of tau decays is significantly higher than the branching fraction of light lepton decays. No specific signal generation is necessary for this step, as the coupling dependent analysis can be reinterpreted with regard to this model.

5.4.4 Quantum Black hole

The production of singly positivly charged quantum black holes was simulated using the QBH3.0 generator [99]. The $q\bar{q}$ production of such objects, and their subsequent decay to tau and neutrino are simulated for the assumed number of extra dimensions $n = 4$. The recommended PDF for QBH simulation CTEQ6L1 [100] is used for generation. This PDF set requires the definition of a Q scale, below which QCD is considered non-perturbative and PDFs are specified in a parametrized form. This parameter is taken to be the inverse of the gravitational radius, as in other QBH searches [101]. Mass points created are ranging from 400 GeV at the lowest up to 11 TeV at the highest. For some exemplary mass points, the generator $m_{\text{threshold}}$ and m_T distributions are given in Fig. 5.11. The signal features a minimum threshold mass, at which most events are seen in the m_T distribution. Towards larger values in m_T, the distribution is steeply falling.

5.4.5 The effective field theory approach

The effective field theory model, which translates the measurements of enhanced branching fractions in R_{D*} to a high energy predictive theory is only sensitive to its different coupling types, since no explicit mediator object is assumed. Signal samples for this model were created using Madgraph5 with NNPDF 3.0 PDF set for 2016. For 2017 and 2018 signal production, the NNPDF 3.1 PDF set was used. In both cases, detector simulation was performed using GEANT4 and hadronization was done using Pythia8. For 2016 production the Tune CUETP8M1 was used, for 2017 and 2018 the TuneCP2 [93] was used. This choice of tune was motivated by CMS internal studies which showed problems with non-resonant BSM signal models using TuneCP5.

Instead of a sharp peak this model expects a broad increase in overall event count in the m_T variable. This non-resonant signal shape predicts a slight increase in data in the high-p_T tails of the distribution compared to the SM prediction, and the most sensitive region in m_T at current LHC settings is expected to be around 1-1.5 TeV [55]. The distribution of the signal on generator level is shown in Fig. 5.12. Here, it can be seen that the distribution of the m_T variable on generator level is quite similar between the different coupling types. The 'tensor-like' coupling produces a signal which is slightly more shifted towards lower

[4]The generation of this many events needs a lot of computing resources (approx. one month runtime on local computing cluster)

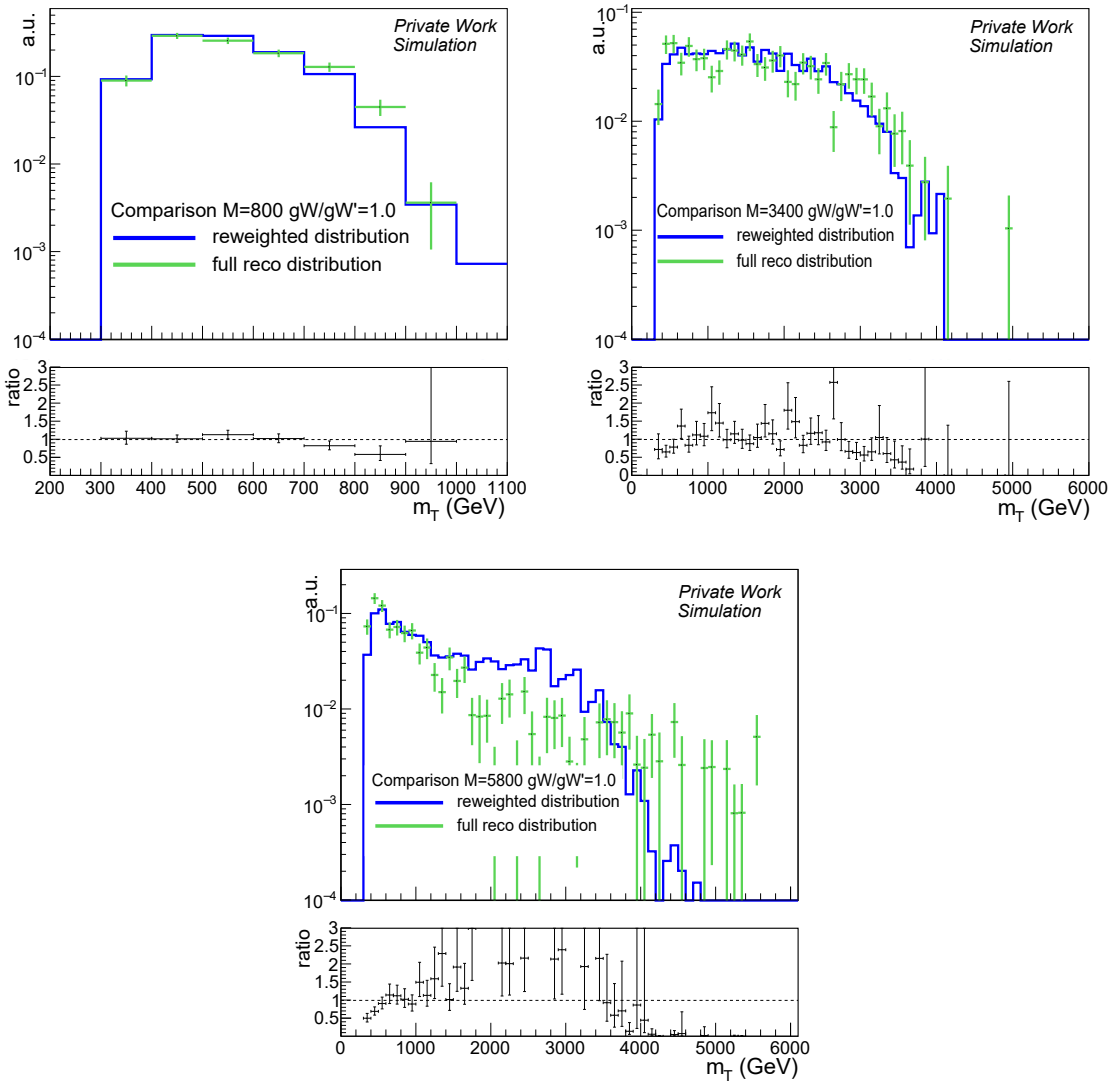

FIGURE 5.10: Comparison between fully reconstructed m_{T} distribution and reweighed m_{T} distribution for signal samples with different masses and couplings. The reweighed signal shapes are obtained by multiplying the generator m_{T} distribution of the signal samples with the ratio of the reconstructed m_{T} distribution for $g_{W'}/g_W = 1$ and its generator m_{T} distribution. The description of the reweighing method is decent, but not as precise as the full reconstruction method.

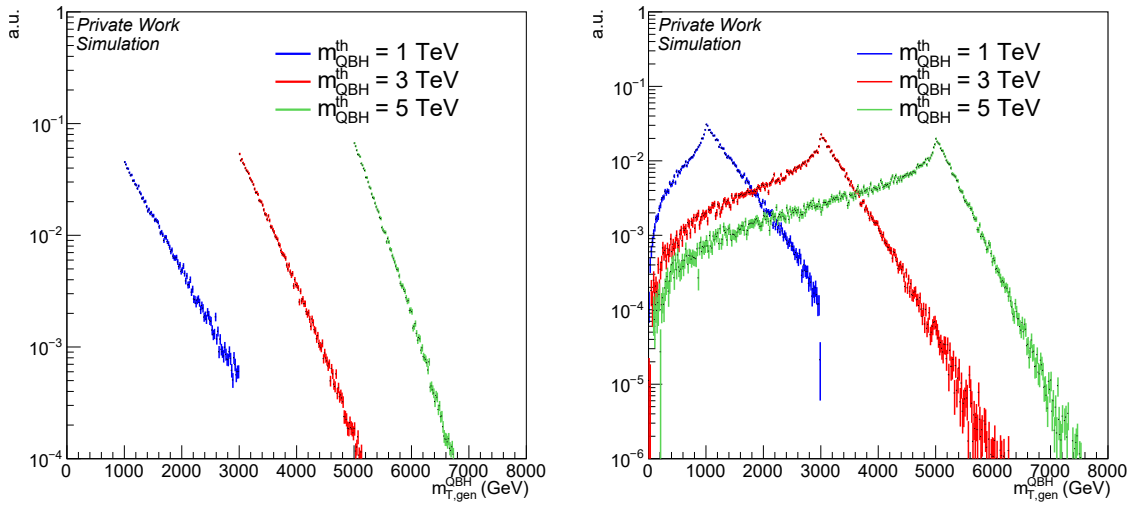

FIGURE 5.11: The m^{gen} (left) and m_T^{gen} (right) distributions for the QBH model on generator level. For m^{gen}, all events lie beyond the generated threshold mass of 1 (blue), 3 (red) and 5 (green) TeV. In m_T^{gen}, most events lie beyond the cutoff mass of the corresponding generated mass point.

values in m_{T} compared to the other coupling types. The signal samples created use the values 0.3, 1.0 and 0.3 for the Wilson coefficients (see Sec. 2.2.4), however, since the event yield (i.e. proportional to the cross section) in this model is directly modified by the square of the coupling, these signal samples can be used to probe a range of coupling values for each coupling type.

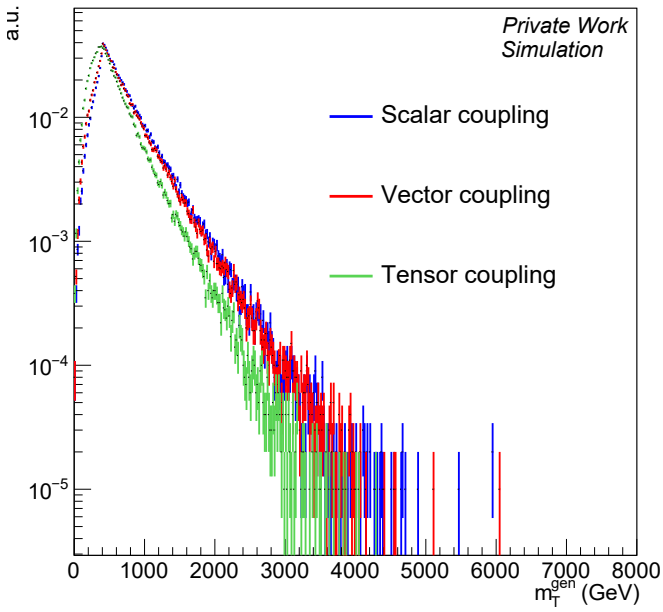

FIGURE 5.12: Transverse mass distribution for EFT model on generator level. Depending on the coupling type, different high mass tails are expected in the τ+MET channel. Note that this distribution will be smeared out more, as at this stage, the tau has not yet decayed. The tau decay is handled in conjunction with hadronization by pythia, and thus, the neutrino coming from tau decay is not accounted for yet.

5.5 Signal Kinematics

All types of signals predict that the difference in the Φ variable of the hadronically decaying tau lepton and the missing transverse momentum $\Delta\Phi(\tau_h, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ is close to π, and that they are balanced in their corresponding transverse momenta. This is shown exemplarily for various signal hypotheses on reconstruction level in Fig. 5.13. As can be seen, all model types exhibit very similar behavior w.r.t. these two variables. The EFT signal shape is slightly broader in $\Delta\Phi$ compared to the other models, however the effect is quite small (the y-axis is logarithmic, so the relative amount of events with $\Delta\Phi$ near π is very high in either case). Therefore, no specific optimization is necessary for these selections between the different models. In the following chapter it will be shown, that these are indeed very potent variables to discern such signals from the background expectation.

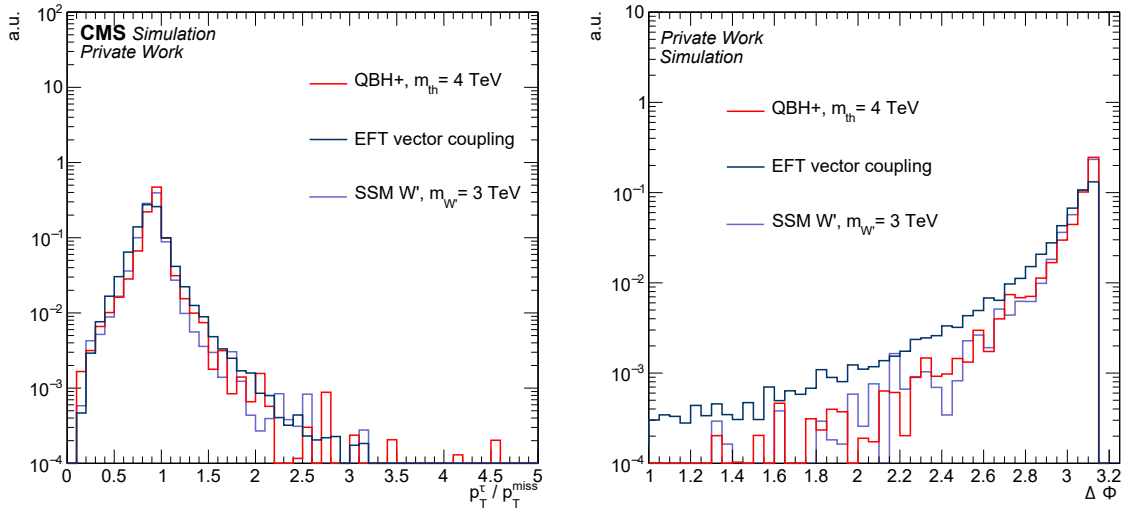

FIGURE 5.13: Significant back-to-back kinematic distributions for the tau and $p_{\mathrm{T}}^{\mathrm{miss}}$ objects on reconstruction level, shown exemplarily for various signal types. As can be seen, the general shape in these variables is very similar between different models.

Chapter 6

Analysis of the Run-2 CMS Dataset

This chapter presents the exact event selections applied in this analysis. The full Run-2 pp-collision data set recorded by the CMS detector corresponding to an integrated luminosity of 138 fb^{-1} is analyzed. Selection criteria are explained and kinematic variables to separate signal from background are analyzed. Additionally, corrections applied to simulated backgrounds are introduced and their application, including uncertainties, is described. The analysis procedure is identical for the analysis of each year (2016, 2017, 2018) except for specially noted selections.

6.1 Event Selection

Events analyzed in the scope of this thesis are required to be triggered by either a missing transverse momentum trigger or a high momentum single hadronic tau trigger[1]. The tau trigger leg requires the candidate object to fulfil isolation requirements, a minimum transverse momentum of 120 (180) GeV for 2016 (2017–2018), and for the tau object to be within $|\eta| < 2.1$. The $p_{\text{T}}^{\text{miss}}$ trigger used requires a certain amount of missing transverse momentum on HLT level ($p_{\text{T}}^{\text{miss}} > 120\,\text{GeV}$). Both trigger legs have been fully active for their corresponding data taking years. In order to ensure a fully efficient trigger, events entering the analysis are required to contain at least one hadronic tau object candidate with 130 (190) GeV of transverse momentum for 2016 ('17–'18) as well as at least 200 GeV of missing transverse momentum. Overlap between the analyzed *Tau* and *MET* datasets is avoided by taking only those events from the *MET* dataset, where the tau trigger has explicitly not fired. The trigger combination was selected in order to ensure the highest sensitivity for the analyzed models (given the possible triggers available in each year). A small discussion with performance of the individual triggers is given in Sec. 6.2.2.

Additional quality requirements are imposed, which are related to missing transverse momentum (*MET filter*). A dedicated CMS work group, the Jet-MET physics object group, has designed multiple event filters in order to suppress events which do not fulfil quality criteria, and are therefore categorized as noise. These filters are [75]:

- **Good Vertex Filter**: Events which have no well reconstructed primary vertices are vetoed.

[1] 2016: HLT_VLooseIsoPFTau120_Trk50_eta2p1
OR HLT_VLooseIsoPFTau140_Trk50_eta2p1
OR HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
2017,2018: HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1
OR HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
OR HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60
OR HLT_PFMETNoMu140_PFMHTNoMu140_IDTight

- **Beam Halo Filter**: Under certain conditions, a small fraction of particles in a particle beam can acquire enough transverse energy from repulsive in-beam forces to create a halo. Events where such a halo is prevalent are vetoed.
- **Dead Cell Trigger Primitive Filter**: Due to radiation, some cells in the ECAL are non-functional anymore (dead). Events which were triggered due to a large amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ in a region with such dead cells are vetoed.
- **Bad PF Muon Filter**: Events with a high energy PF muon candidate, which does not fulfil quality tracking criteria, are vetoed.
- **HBHE noise filters**: Scintillators in the HBHE region have a known issue to produce noise (independent of beam conditions). Events with such noise are vetoed.
- **ECAL endcap supercluster filter**: It was observed, that about 0.7% of the 5×5 ECAL crystals in the endcap do not provide crystal-by-crystal information. Events where such crystals were saturated and contributed to the trigger algorithm are vetoed in experimental data.
- **ECAL bad calibration filter**: In 2017 and 2018, certain crystals in the ECAL rarely showed anomalously large laser calibration corrections. Such events are vetoed. Since this was not observed in 2016 data, no such filter is applied there.

The next requirement imposed on the event is for it to contain at least one well reconstructed tau object. In terms of this analysis, this requires a tau object with at least 130 (190) GeV transverse momentum, a pseudorapidity $|\eta| < 2.1$ and which fulfills the tight working point of each DeepTauID discriminator[2]. The η requirement originates from the tau trigger path. All events fulfilling the above-mentioned requirements are selected. A combined Run-2 distribution of $p_{\mathrm{T}}^{\tau_h}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$ after this basic event selection is given in Fig. 6.1, where background from misidentified τ_h objects is taken from simulation.

In these distributions, the experimental data is indicated as black dots with their respective uncertainties. These dots overlay the sum of all different backgrounds, which are depicted as colored areas. Different background types are indicated by different colors and are stacked on top of each other. Different signal hypotheses are indicated as dashed lines, which would be, if they were real, stacked on top of the sum of the background distributions. In the shown distributions, differences due to an existing signal would be small in the corresponding left (i.e. low) region of the distribution, but significant in the right (high) region.

These distributions show a significant amount of background, composed mostly of off-shell W production and misidentified jets from (mostly) QCD processes. The latter appear significant for the high transverse momentum region of τ_h candidates, but mostly have low amounts of $p_{\mathrm{T}}^{\mathrm{miss}}$. For very high mass signal hypotheses (e.g. the $m_{W'} = 5$ TeV example), the signal could still be extracted quite well from e.g. the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution. However, for low and medium mass hypotheses and non-resonant signal types, the background is very large, and additional selection steps are necessary. At this early stage of the analysis, data and simulation agree mostly with each other, except for regions where simulated background samples are statistically limited (which are mostly backgrounds from misidentified objects).

6.1.1 Signal Selection

In order to discern signals from background, three selection criteria are applied in addition to the object selection. These selections are:

[2] A small study concerning the specific working point selection is given in the Appendix A

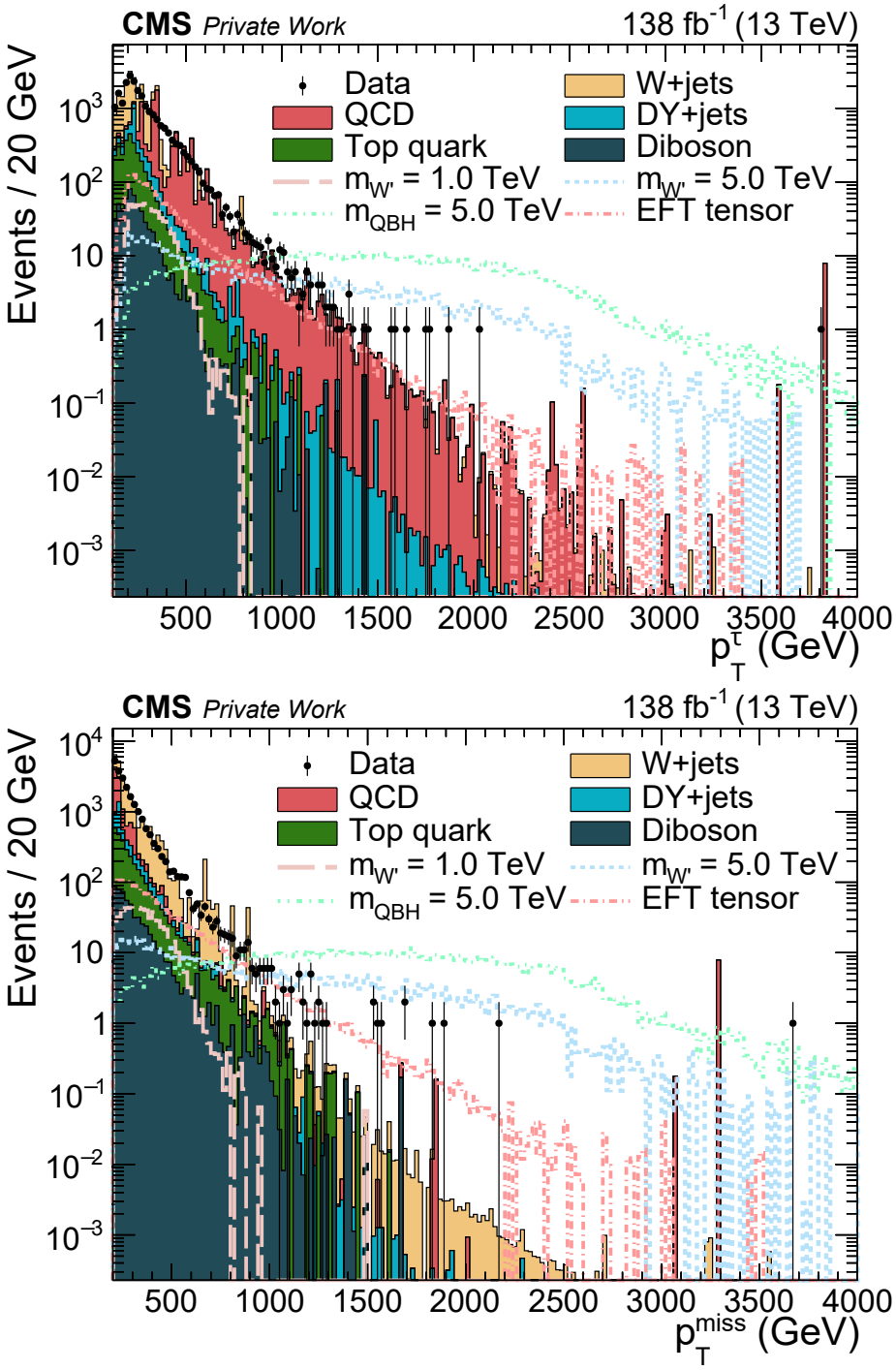

FIGURE 6.1: Preselection distributions after basic object and event selection. No requirements except for trigger, MET filter and the necessity for a high transverse momentum tau object (and sufficient $p_{\mathrm{T}}^{\mathrm{miss}}$) in the event are placed at this point. On the top, the transverse momentum of all tau objects per event is shown. On the bottom, the distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ can be seen. A large amount of background can be seen, making specific signal selections desirable.

- **Second lepton veto**: Since all signal hypotheses do not have a light charged lepton or a secondary tau lepton in the final state, events with such objects are vetoed.
- p_{T} **ratio requirement**: The ratio of $p_{\mathrm{T}}^{\tau_h}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be close to 1. Events

where their ratio is not between 0.7 and 1.3 are vetoed. This selection criterion is well motivated by the amount of QCD events seen in the preselection distributions, where the amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ seems highly unbalanced compared to the τ_h's transverse momentum. This selection should therefore remove many of these background events.

- $\Delta\Phi$ **condition**: The final state objects are expected to be emitted back-to-back. Therefore, events where the absolute difference in ϕ between the tau object and missing transverse momentum is less than 2.4 are discarded.

One of the distributions used to study selection criteria is the "N-1" ("N minus one") distribution. An N-1 distribution is a distribution of a selection criterion without that criterion applied but with all others applied. These distributions are shown for the p_{T} ratio requirement and the $\Delta\Phi$ cut in Fig. 6.2. As can be seen, these selections provide powerful discrimination against background. The requirement on the ratio of the transverse momenta discards events which predominantly originate to misidentified tau objects, but has only minor impact on other background types. The requirement on the difference of the Φ coordinate significantly cleans the selection from events where the tau object and the missing transverse momentum travel into similar directions - a significant amount of SM top production background and recoiling W background[3] can be removed. Both these selections have only minor impact on all signal hypotheses.

[3]These are events, where a SM W boson is produced together with radiated gluon or quark jets [102]. The W then recoils against the jet, resulting in a non-zero transverse momentum W boson. When it decays, its decay products are collimated in the laboratory frame around the travel direction of the W boson, resulting in a lepton and neutrino emitted into similar directions.

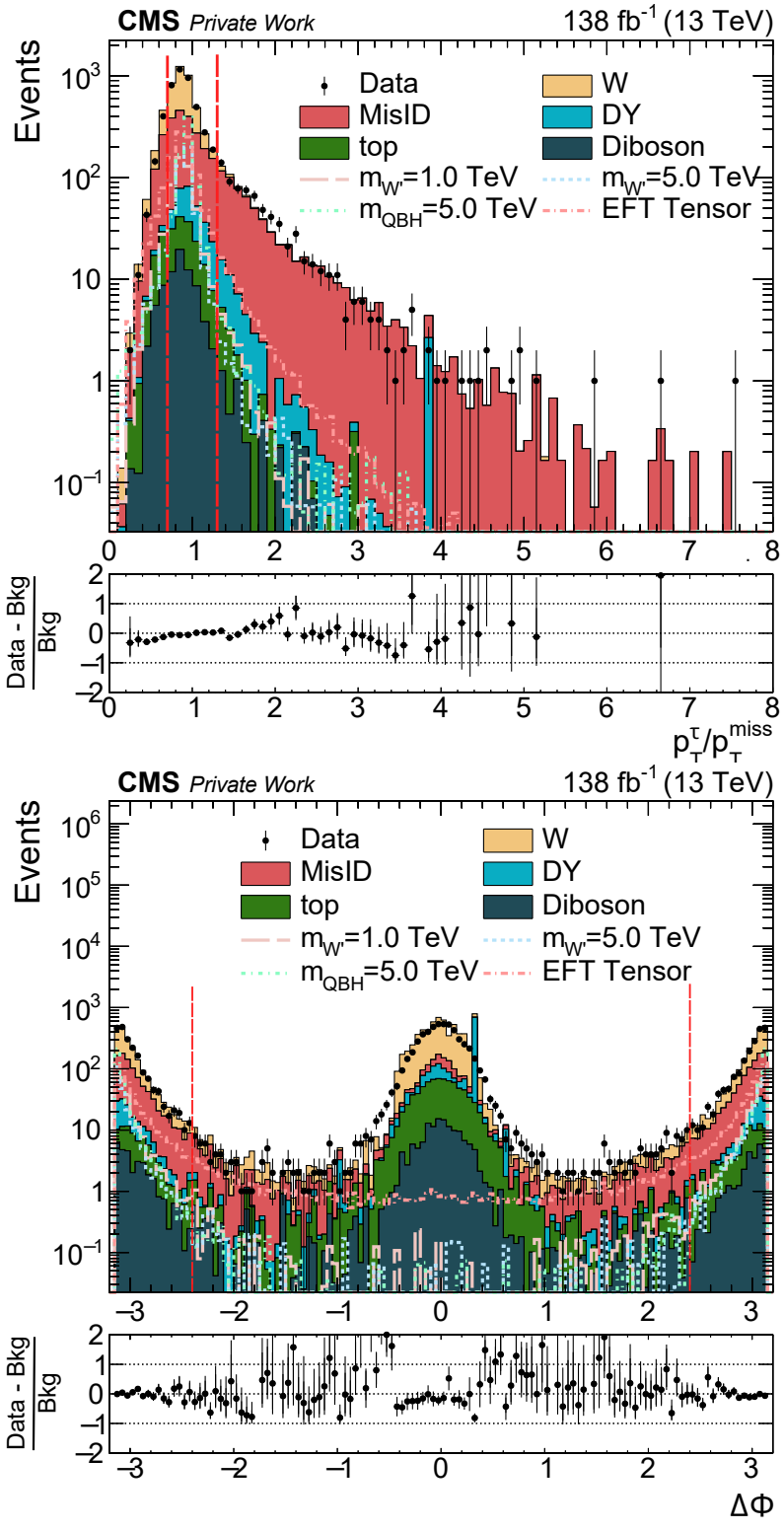

FIGURE 6.2: N-1 distributions for the p_T^τ / p_T^{miss} and $\Delta\Phi$ selection criteria. These distributions show each variable when all other selection criteria are applied, i.e. in the upper one, all selection criteria are applied except for the p_T ratio cut. Analogous requirements are imposed for the N-1 distribution of $\Delta\Phi$ (bottom). As can be seen, each selection removes significant amount of background without reducing signal efficiency significantly. The selection values are indicated in each distribution with red bars.

6.1.2 Signal Efficiency

The effect the different cuts have on the different signal models is shown in Fig. 6.3. The efficiency in each distribution is calculated as the number of events passing the event selection with the required cuts (in order) divided by the total number of events in the signal sample. This means that e.g. detector acceptance $\mathcal{A}$, effects of trigger efficiency and tau decay branching fractions are included (i.e. even with a perfect acceptance and reconstruction the maximum signal efficiency would be around 65% since that is the hadronic tau decay branching fraction). As can be seen, signal efficiency for all model types is not reduced significantly due to the selection criteria.

Signal efficiencies for both the SSM W' signal model and the QBH signal hypotheses follow a steeply rising shape, going into a maximum and then decreasing slowly. This characteristic 'banana-shape' has two main reasons. The fast increase in signal efficiency can be explained by the effect of the transverse momentum threshold, which originate from the trigger selection. Low mass hypotheses have, compared to higher mass hypotheses, a significant amount of events where either the reconstructed tau object or the missing transverse momentum are below the trigger threshold, discarding the event. At very high masses, the effect of signal off-shell production becomes significant (see Fig. 5.6), leading to an increased amount of events with low momentum objects (which then get discarded).

For the coupling dependent W' signal hypotheses, a multidimensional distribution is shown, to illustrate the effects the different coupling ratio and mass values have on the signal efficiency. The bottom distribution in Fig. 6.3 shows the signal efficiency after all selection steps, i.e. after the $\Delta\Phi$ selection cut. Other efficiency selections are not included in this sub-figure to improve visibility. The efficiency can be read by both the color coding and the z-axis of the distribution. As can be seen, the final selection efficiency changes both with mass and coupling ratio values. For large values of $g_{W'}/g_W$, the signal efficiency is quite flat for all masses beyond 1 TeV. The large coupling ratio value leads to large widths, leading to very similar efficiencies for all masses beyond effects due to trigger turn on. For a coupling ratio value of 1, the banana-shape is reproduced. For $g_{W'}/g_W < 1$, the signal distribution becomes sharper (see Fig. 5.8), leading to an efficiency plateau for high mass signal hypotheses.

Signal efficiency is highest for the QBH model overall which originates from the nature of the signal shape - as most events are beyond the generator threshold mass, very little efficiency is lost due to low momentum trigger threshold cuts. It can also be seen, that the second lepton veto, i.e. the requirement that no light lepton or second hadronic tau is in the event, loses up to 5% of the signal efficiency for QBH+ production, increasing with m_{th}^{QBH}.

The signal efficiency for the various coupling types of the EFT model is significantly lower at 3% to 5%, with the vector coupling having the highest efficiency and the tensor coupling the lowest one. The comparably low efficiency is due to the non-resonant signal shape, where most of the events lie at low values of m_{T} (see Fig. 5.12). Therefore, the high trigger thresholds significantly reduce the overall signal efficiency.

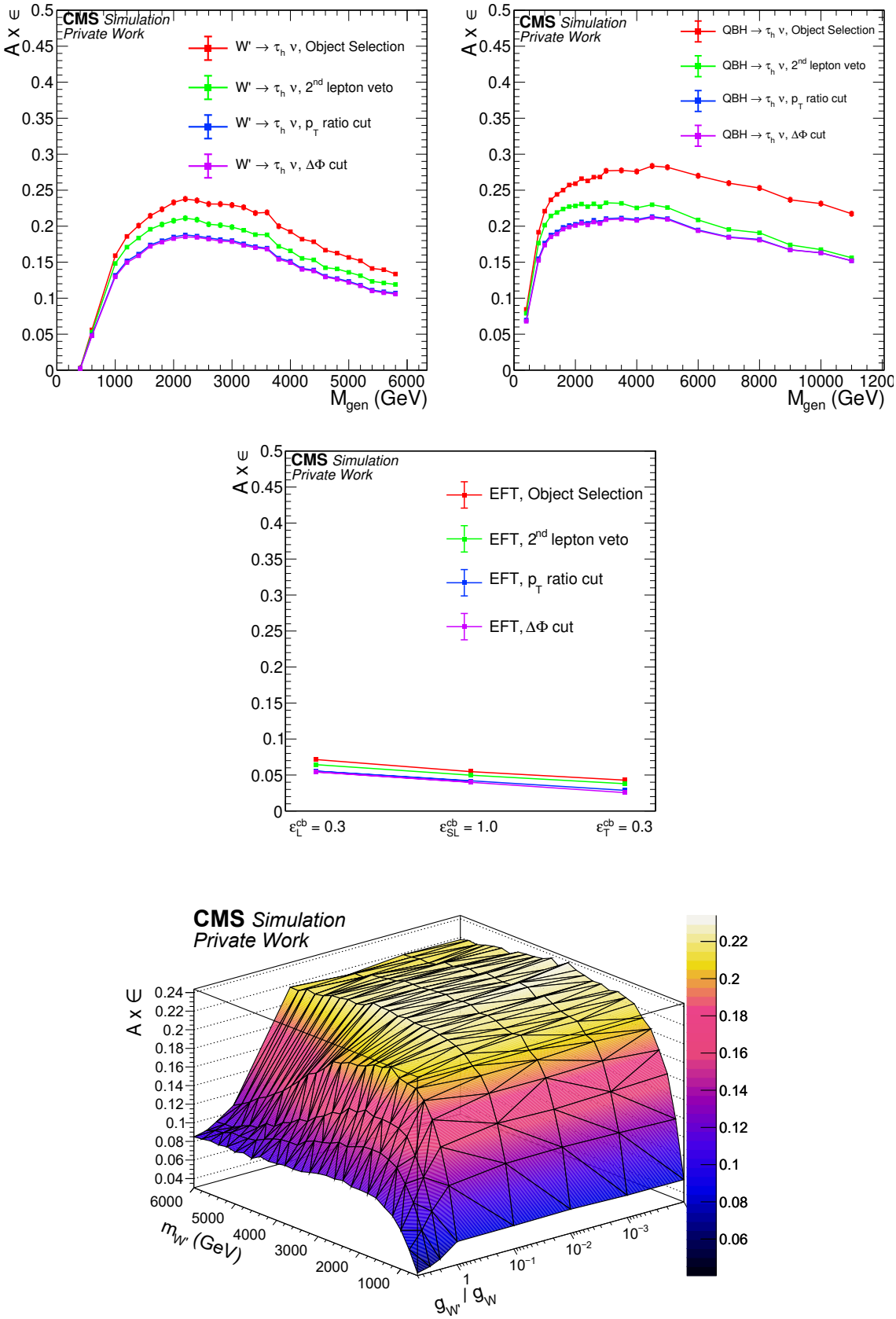

FIGURE 6.3: Signal efficiency, defined as the number of reconstructed events divided by the number of generated events, for various signal hypotheses shown as a function of the parameter of interest for each model, exemplarily for 2017 signal generation. The shown efficiency includes tau decay branching fractions and all reconstruction effects up to this point. The signal efficiency dependent of the variable of interest for each model (SSM top left, QBH top right, EFT middle, coupling dependent W' bottom) is shown.

6.2 Corrections Applied to Simulated Backgrounds

The accuracy of the background description can be increased by applying background corrections. For this, weights or scale factors (SFs) are applied which are typically provided by the different CMS working groups or motivated from higher order calculations from theory. This subsection introduces all the applied corrections and the uncertainties associated to these corrections.

6.2.1 High precision cross section calculation

The dominant background - off-shell W production - is simulated at leading order. As this is expected to be the dominant background in this analysis, it is of high importance to determine the normalization for this background as precisely as possible. This is achieved by using theoretical higher order cross section calculations and introducing an event weight which accounts for the difference between the nominal normalization and the higher order calculation - a so called *k-factor*. The higher order calculations were performed by Klaas Padeken (2016) and Jeongeun Lee (2017, 2018) in dependence on the mass of the generated W boson. Using the FEWZ package [103], QCD LO and NNLO production sections are calculated for this process. For precise determination of the electroweak decay of the W boson, the mcsanc tool [104, 105] was used. Here, NLO and LO electroweak cross sections were calculated in dependence on the mass of the W boson. These calculations depend on the generator tune used to create the off-shell W background samples and are therefore estimated for 2016 and for 2017/18. The results from the production and decay calculations can be combined with different methods: an additive approach and a factorized approach. Both approaches result in similar predictions on the k-factor and their difference is used as an uncertainty in this analysis. Precisely, the k-factor is parametrized with a 4th-order polynomial as a function of the generated W boson mass. The uncertainty is determined by the difference of this polynomial fit and the factorized k-factor. The uncertainties for each mass point originate from statistical fluctuations in the calculation of the k-factor. In order to suppress large fluctuations, especially in the high mass region where the calculations become more difficult, a rolling average over multiple mass bins is used in the calculation of this uncertainty.

The estimated additive and multiplicative k-factors, the parametrization and the uncertainty are shown in Fig. 6.4. The applied uncertainty is given by the continuous blue line, while the uncertainty is depicted as gray bands. The k-factor ranges between 1.3 at low masses and 0.4 at high masses for 2016. For 2017–2018, the k-factor ranges between 1.2 and 0.8 in the same mass region. Differences between the years originate from different PDF sets which were used to generate the signal samples. The resulting normalization (after application of the k-factor) is the same between the years.

6.2.2 Trigger Efficiency Corrections

In order to test the triggering performance, the trigger efficiency has to be measured in data and simulation. In case that they are not aligned - i.e. the trigger efficiency is mis-modeled in MC, a trigger efficiency correction is applied to simulation. These trigger efficiencies are calculated independently for the tau leg and the $p_{\mathrm{T}}^{\mathrm{miss}}$ leg of the trigger setup.

Tau trigger: The tau trigger has been studied in detail by the Tau physics object group of CMS. Its efficiency has been measured in low hadronic activity events to ensure well isolated τ_h objects. Between the years 2016 and 2017/2018, the HLT path was updated with a stricter tau identification on trigger level, which included an increase in the minimum trigger threshold of 120 GeV (2016) to 180 GeV (2017, 2018). The fully efficient tau trigger plateau has been shown to start $\approx$ 10 GeV above the nominal trigger threshold [106]. Small

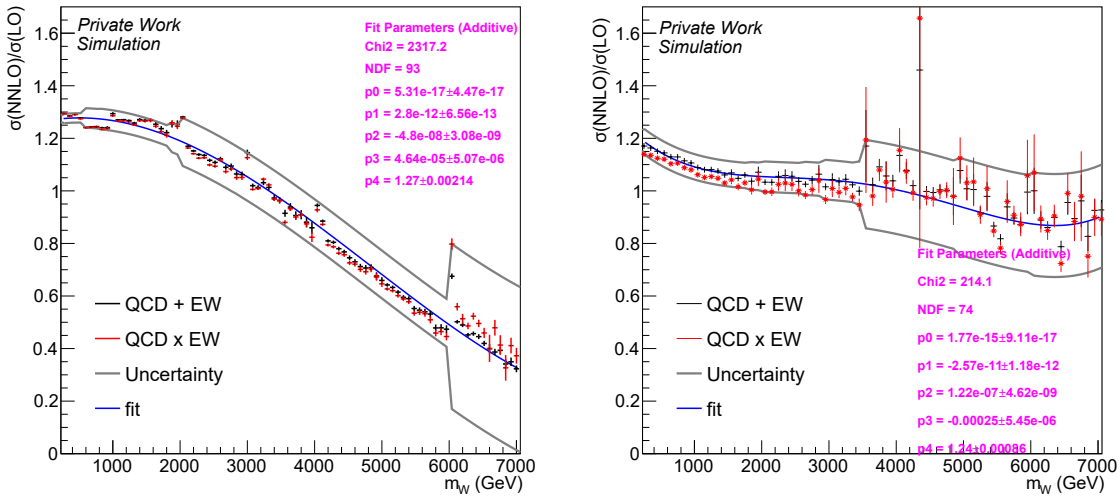

FIGURE 6.4: Depiction of the k-factor calculated for additive and factorial combination of NNLO QCD and NLO EW corrections, which is applied to the SM W boson background for 2016 (left) and 2017/18 (right). The blue line represents a 4-th order polynomial fit to the additive k-factor, the uncertainty band is shown in gray.

differences in trigger performance have been measured in the corresponding years, leading to constant scale factors of 0.88 / 1.08 / 0.87 for 2016 / 2017 / 2018, respectively, measured with $\approx 10\%$ uncertainty in every year. The trigger efficiency is shown exemplarily for 2017 on the left side of Fig. 6.5.

Missing transverse momentum trigger: The $p_{\mathrm{T}}^{\mathrm{miss}}$ branch of the trigger has no provided scale factor calculations from the collaboration side. Therefore, for this trigger branch, the trigger efficiency is measured in data and MC using a W+jets enriched region, which is orthogonal to the signal region. The selection requirements for this sideband region are:

- Events are triggered by a single muon trigger (p_{T} threshold of 27 GeV).
- Application of recommended $p_{\mathrm{T}}^{\mathrm{miss}}$-filters to the events in order to suppress contributions from misreconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$.
- Require exactly one muon candidate, which fulfils tight identification and isolation requirements, which is well above the trigger threshold and within detector acceptance ($p_T^{\mu} > 30$ GeV, $|\eta| < 2.4$).
- In order to ensure event selection from the core W+jets peak and to suppress contamination, an additional $p_T^{\mu} < 100$ GeV requirement is applied.

With this selection, the trigger efficiency can be calculated by taking the ratio of all events and events triggered by the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger. This efficiency is calculated with missing transverse momentum only, ignoring the muon in the event. The result of this is shown exemplarily for 2016/2017 in Fig. 6.5. As can be seen, above the threshold the trigger quickly becomes fully efficient with a trigger plateau close to 100%. To ensure only well triggered events, where differences between triggering in simulation and data is small, a minimum $p_{\mathrm{T}}^{\mathrm{miss}}$ threshold of 200 GeV is applied in the analysis. Differences between simulation and data are corrected with a p_{T} dependent scale factor and its uncertainties. The size of the applied scale factor ranges between 0.96 and 0.995 with relative uncertainties of up to 4%, depending on the data taking year.

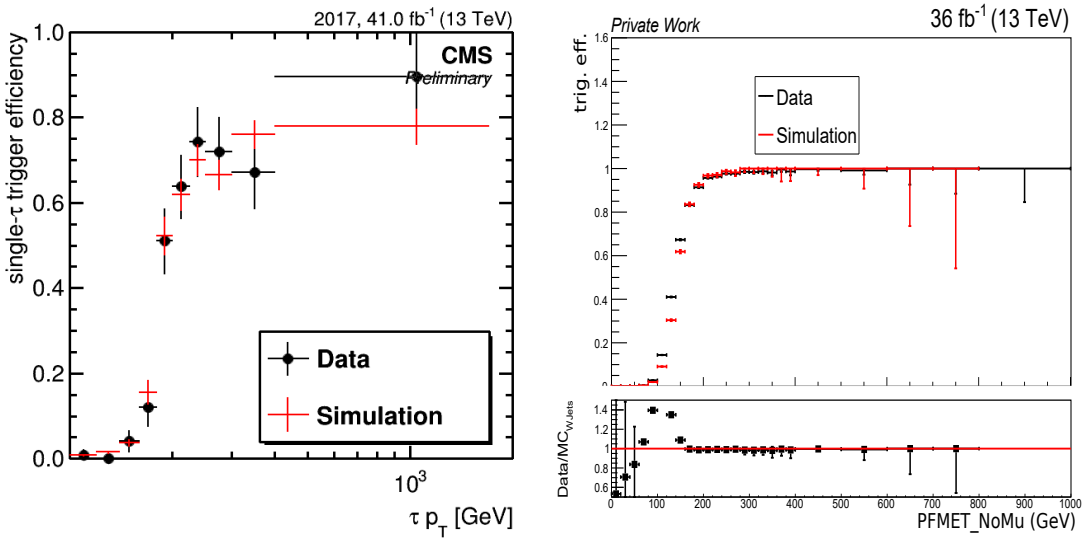

FIGURE 6.5: Trigger efficiency of the different trigger branches for τ_h (left, 2017) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (right, 2016) without application of correction factors. The τ_h trigger efficiency calculation was performed by the TauPOG [106] (image taken from [107]), while the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency was calculated as part of this analysis. Beyond the trigger threshold, the trigger quickly becomes fully efficient with a short turn-on region for each trigger object. Statistical description in the tau trigger leg is difficult, which is incorporated by a large uncertainty.

6.2.3 Hadronic Tau Corrections

Differences in reconstruction between simulation and data also need to be taken into account for hadronic tau objects. There are two different types of corrections for tau objects: one is related to the energy scale determination of tau objects and the other one originates from reconstruction efficiency differences between simulation and data. For each type of correction, the CMS TauPOG provides recommendations, which are followed precisely in this analysis and are described in the following.

Measurement of tau energy scale corrections have been performed in a $Z \to \tau\tau$ region and a $W \to \tau\nu$ region by the CMS Tau group, with the latter correction being designed to provide a better description for high energy taus. However, since this region is exactly the signal region of interest for this analysis, it was chosen to only use the scale factor calculated in the Z boson region. This is necessary in order to avoid a bias. The energy scale correction is applied for each tau object depending on the reconstructed tau decay mode, and is of the order of 1-2%. Uncertainties are of a similar size as the applied correction and are treated as uncorrelated across years and bins in decay mode.

Scale factor corrections originating from the DeepTauID are implemented for each discriminator type (e, μ,jet) separately. Depending on the MC truth matched particle of the reconstructed tau object, these scale factors are applied on an object basis. All these recommended correction factors are provided with a set of up and down shift uncertainties, which are propagated through the analysis to estimate their impact on the final signal region distribution. Scale factors applied due to the vsJet discriminator are applied dependent on the tau objects transverse momentum and range between 5 and 12%. Uncertainties are treated as uncorrelated across years. Scale factors for tau objects from misidentified electrons or muons are applied as a function of $|\eta|$. The corresponding uncertainties originate from statistical uncertainties of the corresponding measurements and are treated as uncorrelated across years and bins in $|\eta|$. The effect of the scale factor and uncertainty due to electron and

muon discriminators is quite small - they are only applied for tau objects originating from electrons or muons, respectively. Due to the strong discriminatory power and the additional light lepton veto applied in this analysis, the resulting impact on this analysis is negligible.

6.2.4 $p_{\mathrm{T}}^{\mathrm{miss}}$ Corrections

As missing transverse momentum is defined via a negative sum of all other objects in an event, its uncertainties are inferred from corrections to those objects. The most significant inferred uncertainty on $p_{\mathrm{T}}^{\mathrm{miss}}$ is the propagation of jet energy scale (JES) and jet energy resolution (JER) uncertainties. By shifting all jets within their recommended uncertainties, and recalculating missing transverse momentum with these shifted objects provides a handle for the inferred uncertainty on $p_{\mathrm{T}}^{\mathrm{miss}}$. A third inferred uncertainty originates from shifts of unclustered energy deposits, which is also propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$ (see Sec. 4.3). These calculations have been centralized for most analyses with the nanoaod-postprocessor tool provided by CMS [77]. Uncertainties from the propagation of JES and JER are treated uncorrelated across years, while uncertainties due to unclustered energy deposits are treated fully correlated across years.

6.2.5 Pileup Correction

In each collision there are typically multiple interactions occurring due to the high instantaneous luminosity. The total number of pp interactions per bunch crossing is called pileup. There are two types of pileup: *In-time pileup* describes multiple interactions which occur in the same bunch crossing as the one which triggered the event, and *Out-of-time pileup* which are interactions which occur in bunch crossings which came before of even after the in-time interaction. These are created due to the integration time of some parts of the CMS detector elements. Detector parts which can lead to out-of-time pileup are the calorimeters and the muon system.

Simulated background samples are generated with an expected profile of pileup interactions, calculated using measurements on the instantaneous luminosity and the total inelastic pp cross section of 69.2 mb [3]. This expected distribution can then be compared against the measured pileup distribution in data. In order to correct for differences between the expected and the observed pileup distributions, a corrective weight is applied to each simulated event based on the number of *true* pileup interactions. The distribution of the number of vertices before and after applying this correction is shown in Fig. 6.6 for 2017 exemplarily. As can be seen, this procedure mainly reduces expectation from simulated backgrounds in the high pileup region. This procedure, as recommended by the CMS luminosity working group, does not improve the overall agreement of experimental data and background expectation significantly. The uncertainty on this method is estimated by applying a $\pm 4.6\%$ shift to the inelastic cross section and recalculating the event weights with this shifted distribution. The uncertainty is treated as fully correlated across data taking years.

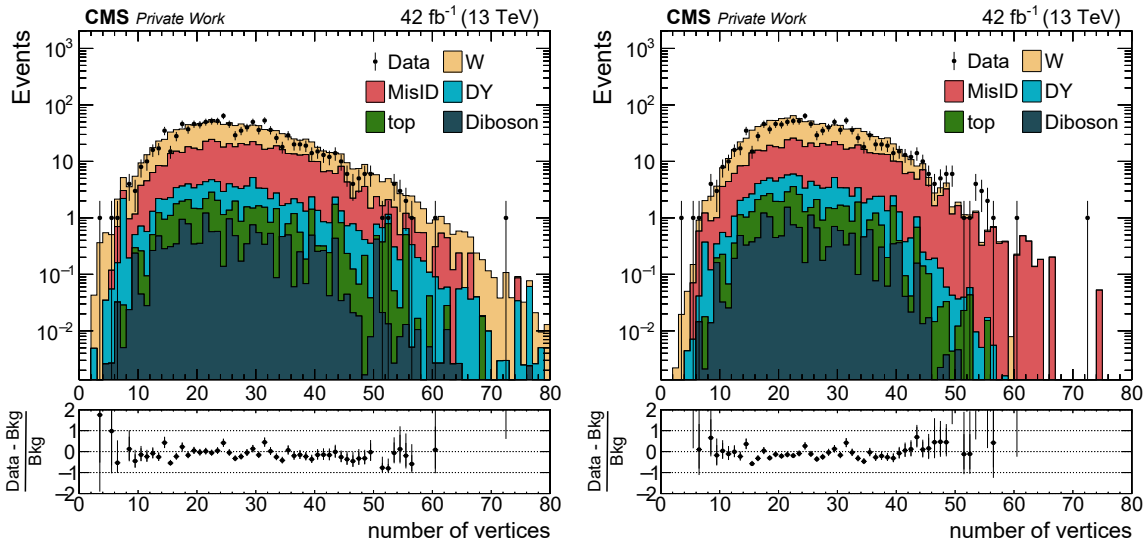

FIGURE 6.6: Number of true pileup interactions before (left) and after (right) applying the pileup reweighing procedure for 2017.

6.3 Backgrounds Estimated from Data

Backgrounds from jets which are misidentified as tau leptons are derived from data. Although the reconstruction and identification algorithms provide a very strong discrimination against such jets, the sheer amount of jets produced at proton-proton collisions makes this a sizeable background in this analysis. These backgrounds are difficult to simulate and have been shown to differ between MC and data by up to 30% [108]. Therefore, a data-driven approach is used in addition to the prediction from simulation i.o. to improve the precision of the description of this particular background. Events from this background mostly come from QCD jets which can be produced in almost every collision due to initial state or final state radiation of quark or gluon jets. Therefore, this background cannot be mapped exactly against MC QCD background predictions but has to be compared to the sum of all simulated backgrounds where the reconstructed tau does not originate from the decay of a tau lepton.

6.3.1 Tight-to-Loose Method

The main assumption this method relies on is, that 'true' tau objects (i.e. tau objects which originate from hadronic tau lepton decays) are well simulated in MC. One can define a sideband region to measure the ratio of events between the *Tight* and *VVVloose* tau ID working points. By subtracting the 'true' tau objects for both these working points, a ratio of misidentified taus between these two working points is acquired. With the second assumption, that this ratio does not change between the sideband region and the signal region, the number of misidentified tau objects can be determined. The application method is sketched in Fig. 6.7. The depicted sideband regions provide template data for the calculation of the ratio of events, which in turn is applied to the application region A, resulting in a prediction for the signal region B.

For the exact method applied in this analysis, two sideband regions are used in order to reduce statistical errors within the method. These regions require either a muon or an electron to be within the event. This requirement makes the regions orthogonal to the signal region, where events with these objects are vetoed. These specific sideband regions are defined as follows:

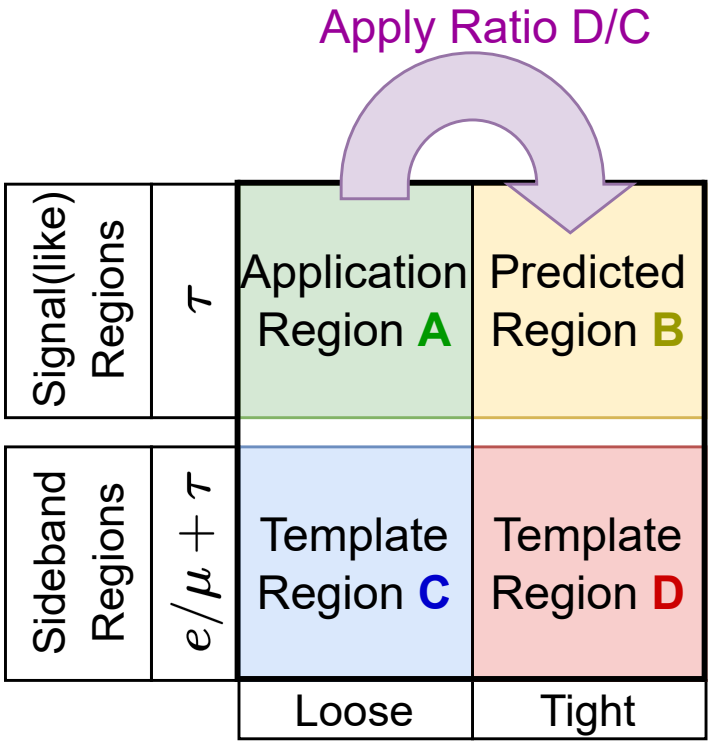

FIGURE 6.7: Simplified visualization of the method used to determine the background for taus being misidentified from QCD jets using data. The goal is to determine the contribution in the signal region B. To achieve this, the sideband regions C and D are determined, the ratio of events calculated and applied to the signal-like region A.

- **Trigger**: For the muon sideband region, the momentum threshold of the used trigger is 50 GeV[4]. For the single electron, the corresponding trigger has a momentum threshold of 27/35/32 GeV[5] for 2016/2017/2018, respectively.

- **Event Requirements**: The events are required to pass all $p_{\mathrm{T}}^{\mathrm{miss}}$ filters as described in Sec. 6.1. The event has to have at least one muon/electron and one tau candidate.

- **Object Selection**: The light lepton needs to fulfil high quality identification criteria. For the muon, these requirements are: $p_T^{\mu} > 53\,\mathrm{GeV}$, $|\eta^{\mu}| < 2.4$, tight identification and isolation criteria [109, 110]. The corresponding selection for the electron sideband is: $p_T^{e} > 35\,\mathrm{GeV}$, $|\eta^{e}| < 2.5$, and the candidate has to fulfil the dedicated high energy electron identification [78]. The event is required to have exactly one light lepton fulfilling these requirements. Events with more than one such lepton are discarded. Additionally, the tau object candidates are required to be within the detector acceptance and fulfil minimum requirements on reconstruction: $p_T^{\tau} > 20\,\mathrm{GeV}$, $|\eta^{\tau}| < 2.3$, a charge of ± 1 and it has to originate from the primary vertex $dz < 0.2$. Furthermore, to ensure selection solely towards misidentified jets, the tau object candidates are required to

[4] HLT_Mu50

[5] 2016: HLT_Ele27_WPTight_Gsf
2017: HLT_Ele35_WPTight_Gsf
2018: HLT_Ele32_WPTight_Gsf

fulfil the Tight working points of both the vsMuon and vsElectron discriminator of the DeepTau identification (see Sec. 4.2.2).

- **Other kinematic requirements**: At this point, the event contains exactly one light lepton and one or more tau object candidates. From these, all lepton+tau combinations are built, and additional requirements are imposed. First, the tau object and the light lepton have to have a distance parameter $\Delta R > 0.5$ in order to ensure that these objects are well reconstructed and not related. Secondly, the light charged lepton and the tau lepton are required to have the same charge. This requirement is imposed in order to suppress contributions from the DY process, which could contaminate the sideband region. If there are still multiple lepton+tau combinations eligible, the combination where the tau object has the highest transverse momentum is selected.

The events selected in this way are then split into two regions, depending on the output of the DeepTau jet discriminator variable for the selected tau object. One region consists of those events where the tau object fulfills the same identification requirement as the tau lepton in the signal region (i.e. the tight working point (or stricter)). The second region is defined by all other events - i.e. those where the tau object fulfills a very loose identification requirement, but explicitly not the tight (or stricter) working point of the ID. In simulation, these regions are further split into those events, where the selected tau object originates from a generator jet and those events that don't. This is necessary, as the tight-to-loose ratio should be estimated from events which originate to jets only. Therefore, the contribution of non-jets is subtracted from data events using this MC truth matching (see first assumption). The muon and the electron region are now combined in their statistics for their corresponding regions. Then the Tight-to-Loose (TtL) ratio is calculated as:

$$R_{TtL} = \frac{N_{data,Tight} - N^{non-jets}_{MC,Tight}}{N_{data,Non-Tight} - N^{non-jets}_{MC,Non-Tight}} \tag{6.1}$$

To visualize this, the distributions of the sideband region with a selected muon and a tau candidate are shown in Fig. 6.8 (Tight ID) and Fig 6.9 (reversed Tight ID). In both Figures, the total contribution is shown in the left panel and the contribution of matched non-jets is shown in the right panel. In the end, the MC contribution from the right distributions is subtracted from the data events in the left distribution. For simplification, this is shown only as a function of the tau object's transverse momentum - in the applied method, this is calculated in a 2-dimensional plane of the tau objects momentum and the decay mode of the tau, as the Tight-to-Loose ratio likely depends on the number of charged hadrons in the tau object. The result of this calculation for each year is given in Fig. 6.10. For 2017, the results of the estimated uncertainty of the Tight-to-Loose ratio is shown exemplarily in Fig. 6.11, where it can be seen that the assigned uncertainty is quite small for the low p_T region and increases towards higher values of the tau candidate's transverse momentum. In order to test that this method is stable, R_{TtL} is calculated for the electron and muon region individually as well. This is shown for 2017 in Fig. 6.12. As can be seen, the ratio is compatible between the two regions, but statistical differences arise in some bins.

6.3.2 Closure Test

In order to validate the method, we can test, if the calculation of R_{TtL} depends on the existence of a light charged lepton in the event or not (assumption 2). This can not be done in data, but as a cross check it can be performed on simulation. For this, the previously calculated Tight-to-Loose ratio is applied to events with signal-like selection in simulation,

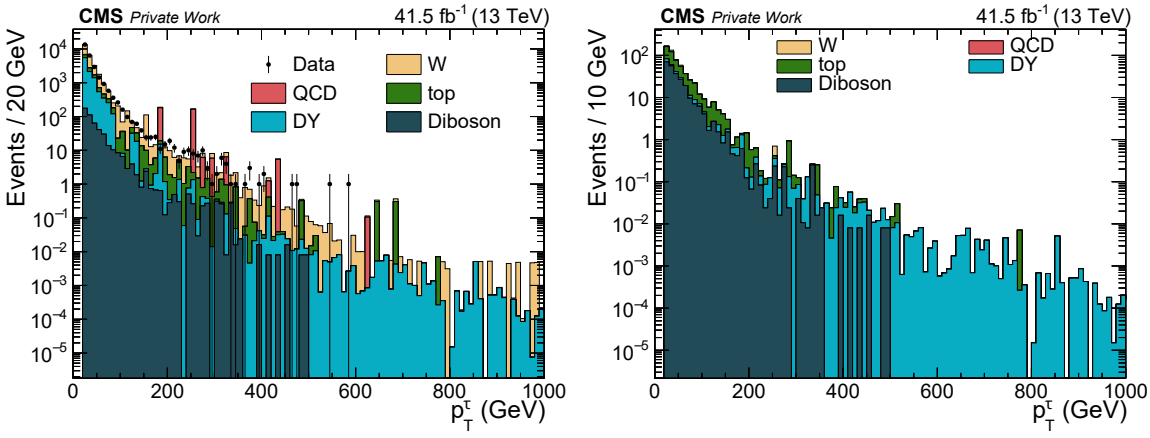

FIGURE 6.8: Tau transverse momentum distributions of the muon sideband region, where additionally a tau object which fulfills the Tight working point of the DeepTau vsJet-discriminator is in the event. On the left, the total data contribution with the total MC prediction is shown. On the right, those events are shown, where the tau object is matched to a non-jet generator particle. Those are subtracted from the data for the calculation of the Tight-to-Loose ratio, which reduces the total event count by $\approx 1\%$.

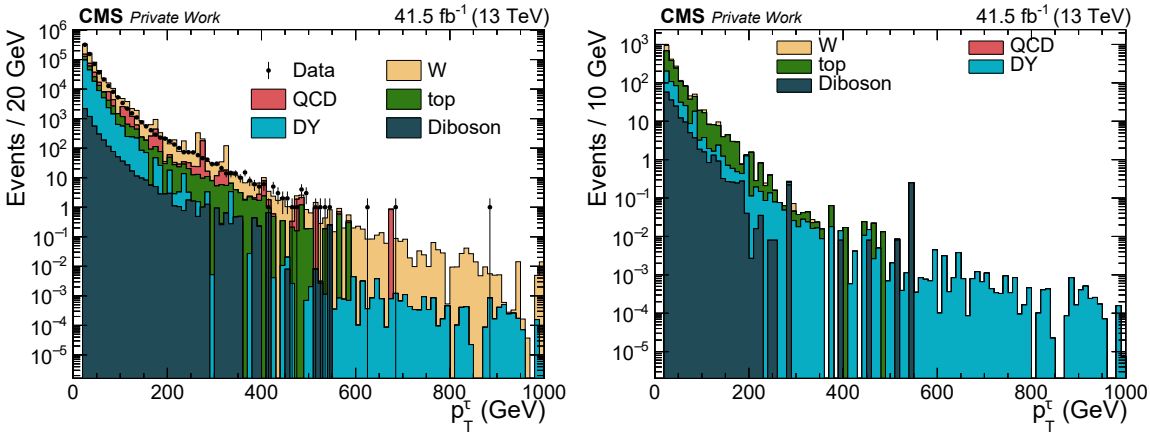

FIGURE 6.9: Tau transverse momentum distributions of the muon sideband region with a tau object which fulfills the loosest working point of the DeepTau vsJet-discriminator but not the Tight working point. On the left, the total data contribution with the total MC prediction is shown. On the right, those events are shown, where the tau object is matched to a non-jet generator particle. Those are subtracted from the data for the calculation of the Tight-to-Loose ratio, which reduces the total event count by $\approx 0.5\%$.

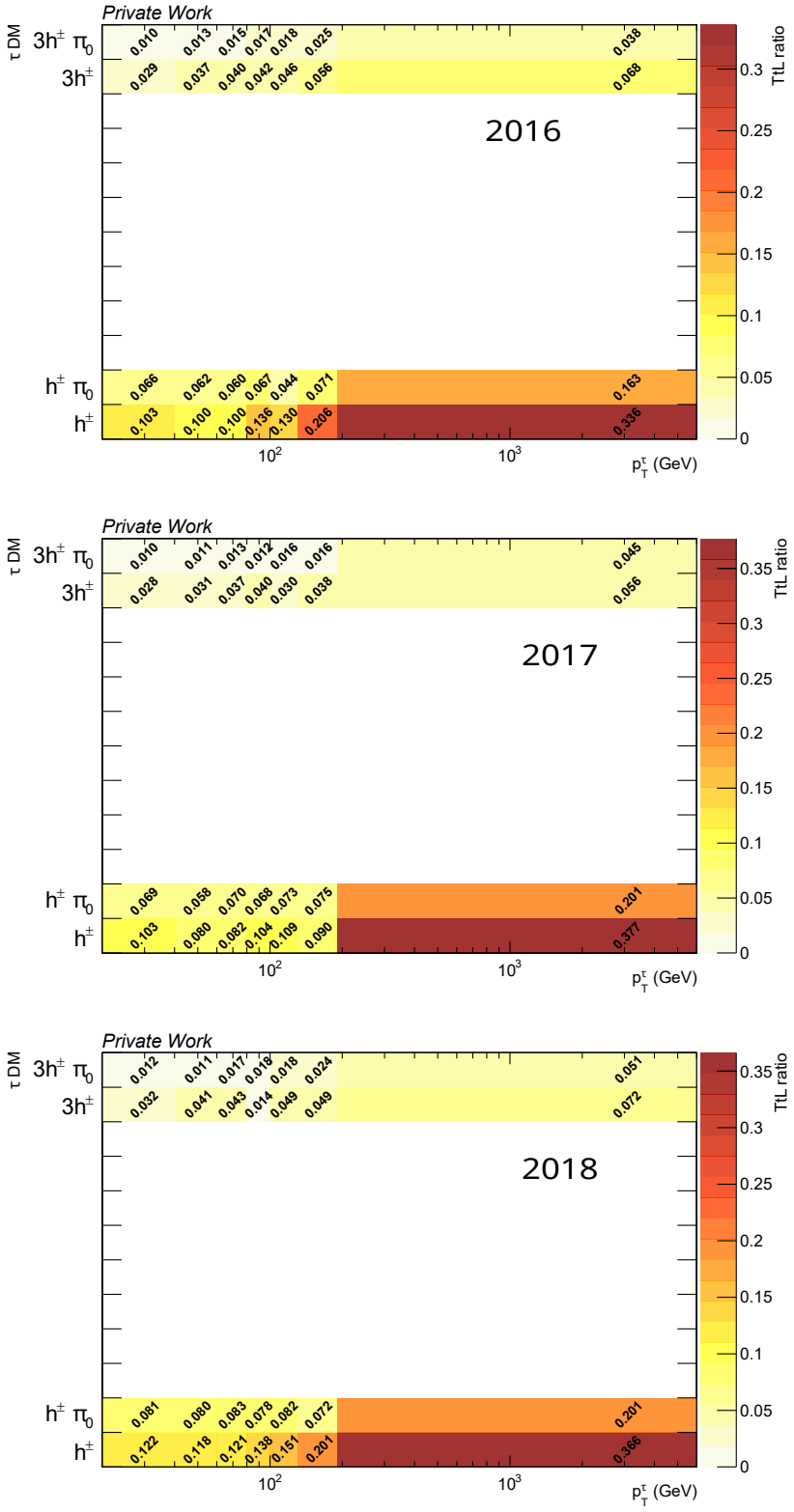

FIGURE 6.10: Calculation of the Tight-to-Loose ratio for 2016 (top), 2017 (middle) and 2018 (bottom). They are estimated as a function of the tau DM and the transverse momentum of the object. As can be seen, this ratio is rising for increasing tau momentum and is significantly higher for low-multiplicity tau candidates (i.e. DM0 and DM1). The tight-to-loose ratio is comparable between years for the full distribution.

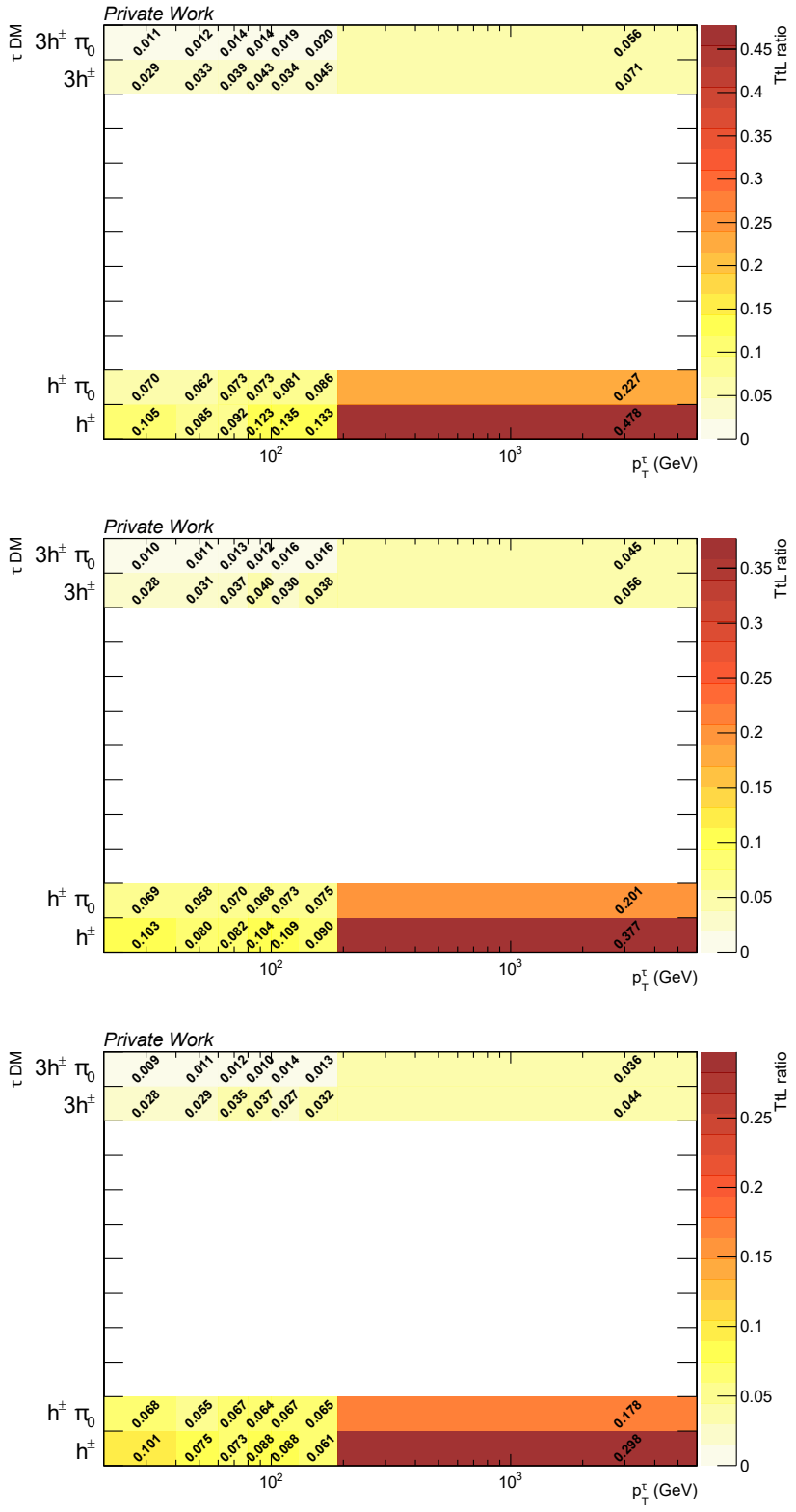

FIGURE 6.11: Uncertainty estimation of the Tight-to-Loose ratio - statistical uncertainties in region C and D are propagated to the ratio estimation for each bin in p_T and DM. The top distribution shows the up shift, the middle distribution the nominal value and the lower distribution the down shift of this calculation for 2017 exemplarily. Even in the combined muon+electron sideband region, the statistical uncertainty is quite significant, especially for high transverse momenta.

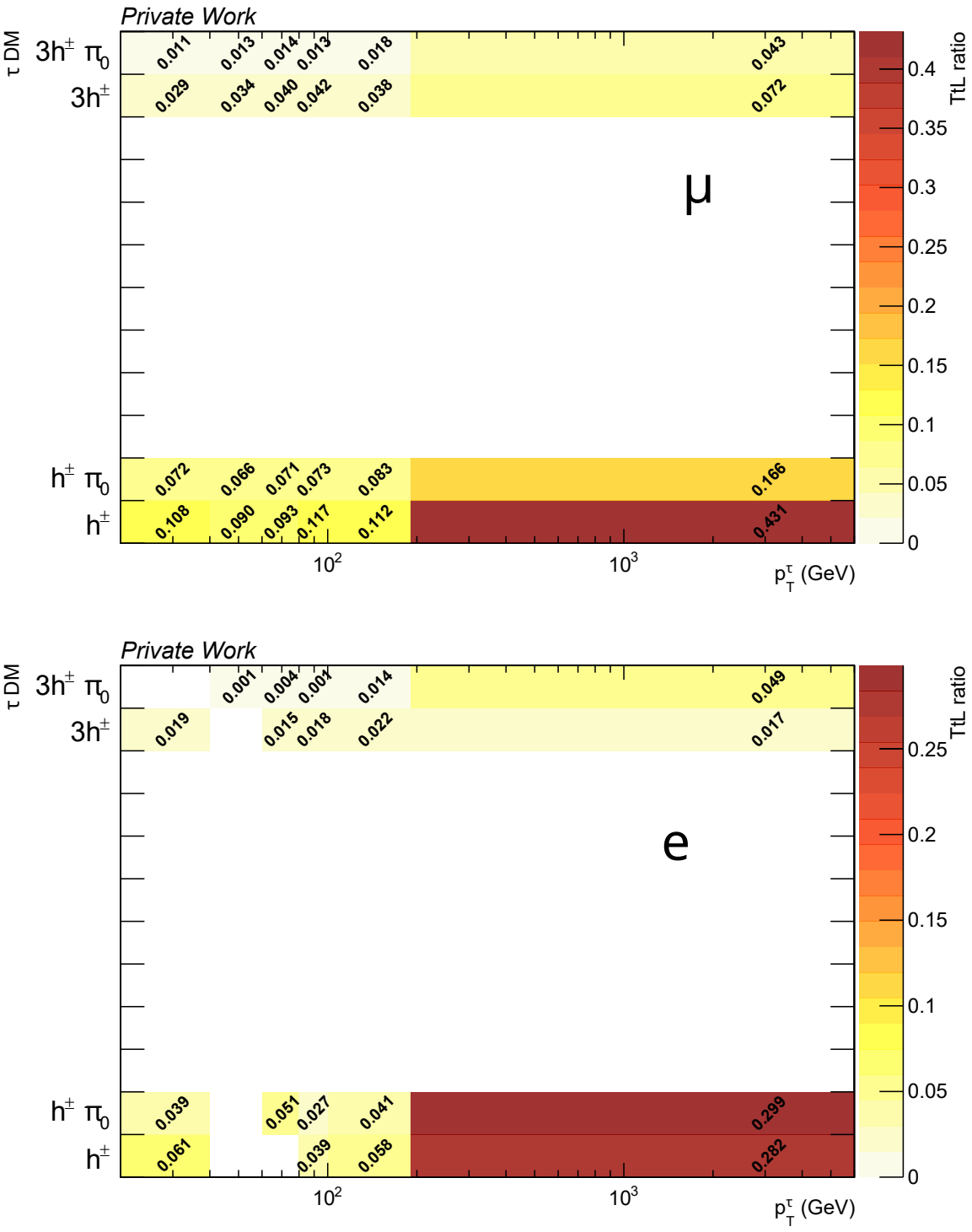

FIGURE 6.12: Tight-to-Loose ratio in 2017 calculated from the muon region (top) and electron region (bottom) individually. As expected, both sidebands lead to similar results, however statistical description becomes difficult especially in the electron sideband region.

where the tau object originates from a generator jet. Comparing the result of this application to the actual prediction of misidentified tau objects in MC, the quality of this method is estimated. This is shown for all three years in Fig. 6.13. As can be seen, the distributions mostly agree quite well except for large statistical fluctuations as indicated by the error bars. These fluctuations originate from limited statistics in simulated background samples, as only a small fraction of simulated events contains jets being misidentified as taus. In order to accommodate for differences in these distributions, a 100% uncertainty is assumed on the normalization of the total background estimated with this method.

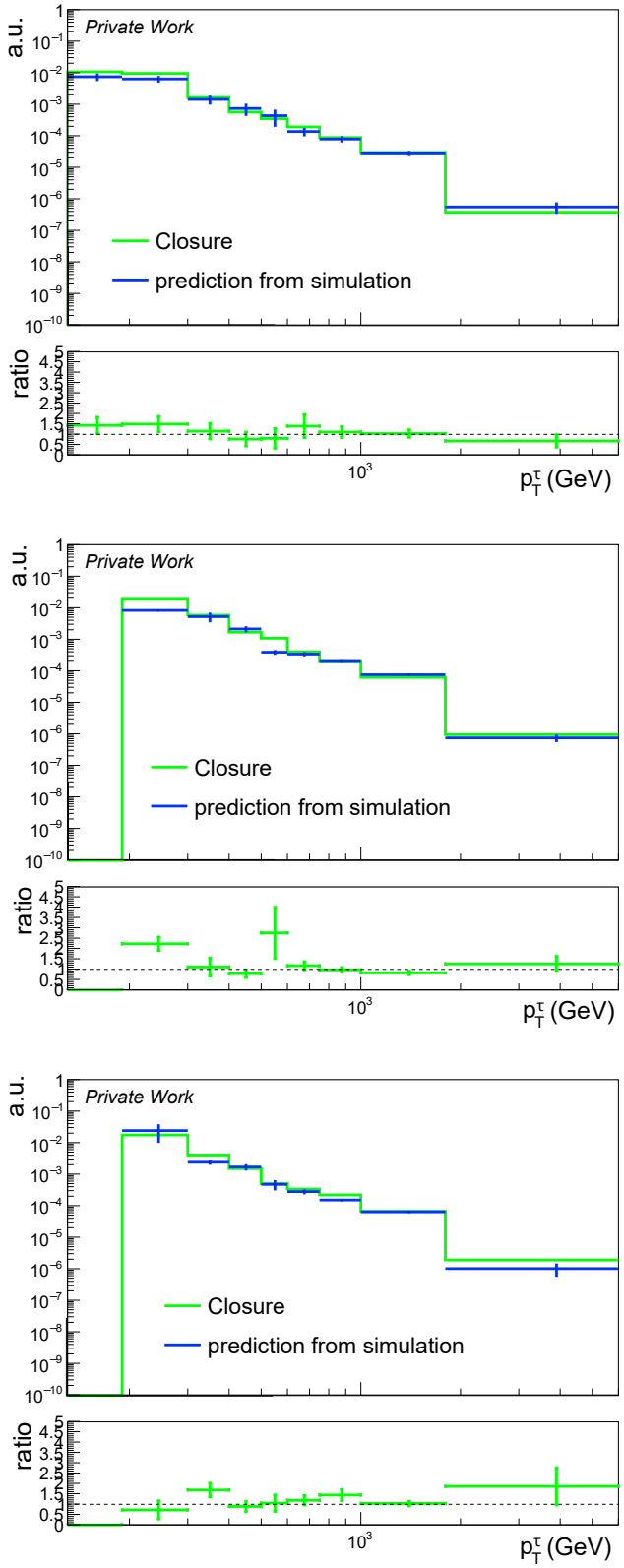

FIGURE 6.13: Closure test on the data-driven method. The tau transverse momentum distributions shown compare event contribution from misidentified taus originating from jets with the event contribution prediction when applying the data-driven method to simulation based event prediction for all three years.

6.4 Systematic uncertainties

While being discussed in detail in their corresponding sections, here a short overview of all systematic uncertainties considered in this analysis is given.

- **Luminosity**: Luminosity measurements have been performed for each data set. Uncertainties on these measurements are propagated to this analysis and amount to 1.0 / 2.0 / 1.5% (uncorrelated) uncertainty. Additionally, an uncertainty of 0.6 / 0.9 / 2.0% correlated between 2016 / 2017 / 2018 is applied. A third component correlated between 2017 and 2018 of 0.6/0.2% is applied, following the recommendations of the CMS luminosity group.

- **JER and JES**: Uncertainties from jet energy scale and jet energy resolution calculations are propagated to the calculation of missing transverse momentum. The analysis chain is reprocessed with the resulting shift on $p_{\mathrm{T}}^{\mathrm{miss}}$ for both up and down shifts of all jets in each event. The result are three uncertainties: one due to JER shifts, one due to JES shifts and one due to shifts of unclustered energy deposits. While the former two are treated as uncorrelated across years, the latter is treated as fully correlated.

- **Tau scale factor**: Application of tau energy scale factors (1-2%) and tau identification scale factors (typically in the range of 5-40% dependent on the discriminator type) come with sets of systematic up and down shifts which are applied on an event by event basis. These uncertainties are typically of the size of the correction (or slightly smaller) and are treated as uncorrelated across years.

- **Data-driven**: Uncertainties for the data-driven estimation include the propagation of the statistical uncertainty from the e/μ sideband regions to the TtL-ratio, as well as a 100% normalization uncertainty for the whole data-driven background. This is a quite conservative approach, and it will be shown later, that this is marginalized in the limit setting procedure towards a much smaller value (see Se. 7).

- **Trigger scale factor**: Due to differences in trigger efficiency between data and simulation, trigger scale factors are applied for both the tau and $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger leg. Each has a set of up and down shifts, which are propagated to the m_{T} distribution as uncertainties. The uncertainties on the corrseponding scale factors of the tau trigger leg are approximately 10%, and those of the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger leg are $< 5\%$.

- **Pileup reweighing**: Events in simulation are reweighed based on pileup differences between measured data and simulation. The minimum bias cross section of 69.2 mb cross section is shifted by $\pm 5\%$ and the resulting final distribution is recalculated to estimate the uncertainty originating from this method.

- **Prefiring**: Corrections applied due to the prefiring effect measured in 2016 and 2017 each have a set of shifts, which are used as event based uncertainty estimations. Not applied in 2018.

- **Cross section**: Theoretical cross section uncertainties for backgrounds are taken into account. For Drell-Yan processes, this uncertainty amounts to 2% normalization uncertainty, for di-boson processes, 4% are applied.

- **W k-factor**: The application of higher order cross section corrections for off-shell SM W boson production comes with a set of uncertainties originating from the differences between additive and factorial combination approaches on calculating these cross sections. They are propagated on an event by event basis, dependent on the generator

matched boson mass for all high mass W background samples. The size of this uncertainty ranges from 3% (5%) at low W boson masses up to 80% (30%) at high masses for 2016 (2017, 2018).

- **PDF uncertainty**: Uncertainties originating from the composition of protons are modeled for each event as a set of 33 Hessian eigenvectors. Calculating the 68 percentiles as both up and down shifts w.r.t. the nominal weight results in the applied PDF uncertainty. This uncertainty affects the event yield in the signal region and increases from a few % at low m_{T} values to up to 50% in the high m_{T} region.

- **Factorization and renormalization scale**: For all backgrounds estimated at NLO, factorization and renormalization scales are each shifted for three values (0.5, 1.0, 2.0), with the combination (1.0, 1.0) being the nominal correction. The resulting event weights (up and down) with the largest difference to the nominal value are taken as systematic uncertainty. In this procedure, the μ_F, μ_R combinations (0.5, 2.0) and (2.0, 0.5) are excluded as they are viewed to be nonphysical.

To visualize the impact each of these uncertainties has on the background event yield in the signal region, the m_{T} distribution and the leading systematic uncertainties for each year are shown in Fig. 6.14 (2016), 6.15 (2017), and 6.16 (2018). The size of the uncertainty is shown per bin, meaning that a 50% uncertainty allows for a shift of 1.5 of the total event yield in the corresponding bin of the m_{T} distribution. The leading uncertainties are very similar for each year. The two most significant uncertainties are the conservatively estimated normalization uncertainty on the data-driven estimated background and the PDF uncertainty, which increases for higher values of m_{T}[6]. The uncertainty on the τ trigger scale factor also contributes significantly to the overall uncertainties. The uncertainty originating from the normalization uncertainty on the Tight-to-Loose ratio has a slightly erratic behaviour in the high-m_{T} region. This originates to the limited statistics in the high-m_{T} region of the application region used in the data-driven method. Uncertainties not shown in these figures impact the event yield of the background prediction by less than the lowest shown uncertainty.

The impact of each uncertainty on the signal yield is shown in Tab. 6.1. The effect of most uncertainties is very low, with the tau trigger and tau identification scale factor uncertainties being the largest ones. Uncertainties not shown in this table affect the signal by less than 0.1%.

6.5 Run-2 m_{T} distribution of τ_h and $p_{\mathrm{T}}^{\mathrm{miss}}$

The m_{T} distribution after full object and event selection combined for all three years is shown in Fig. 6.17. The full dataset of proton-proton collisions taken by the CMS detector from 2016 to 2018, corresponding to an integrated luminosity of 138 fb^{-1} is analyzed. No significant deviations can be seen compared to the SM prediction. There is a small amount of data above the background prediction in the m_{T} region around 1 TeV, however this is fully covered by the uncertainties. A per-year overview on the specific number of events above certain m_{T} thresholds is given in Tab. 6.4. Distributions of kinematic variables for τ_h and $p_{\mathrm{T}}^{\mathrm{miss}}$ objects for each individual year are given in App. B.

[6] This behavior is expected, since the PDF uncertainty should increase the further away the point of estimation is from the point where data on PDFs exists.

	W', $m_{W'} = 1$ TeV	W', $m_{W'} = 3$ TeV	W', $m_{W'} = 5$ TeV
Tau Trigger	14.5%	14.5%	14.5%
MET Trigger	2.4%	2.1%	2.2%
TauID (vsJet)	4.5%	6.6%	6.8%
Tau energy scale	3.5%	1.2%	1.4%
$p_{\mathrm{T}}^{\mathrm{miss}}$ unclustered Energy	0.4%	0.4%	0.4%
$p_{\mathrm{T}}^{\mathrm{miss}}$ jet energy scale	1.0%	0.5%	0.7%
$p_{\mathrm{T}}^{\mathrm{miss}}$ jet energy resolution	0.2%	0.3%	0.4%
Pileup	0.8%	1.2%	1.7%
PDF	0.1%	0.8%	3.8%

TABLE 6.1: Impact of different uncertainties on the signal yield, shown exemplarily for three different mass hypotheses of the SSM W' signal. The leading uncertainties are related to triggering and identification of the tau lepton.

Process	2016 m_T all	2016 $m_T > 1$ TeV	2016 $m_T > 1.5$ TeV
W	593.4±6.5	10.1±0.2	1.18±0.02
MisID	353.7±7.6	2.2±0.7	0.66±0.42
DY	48.5±3.7	0.4±0.1	0.03±0.01
top	27.9±1.3	< 0.01	< 0.01
Diboson	21.1±2.1	0.6±0.3	0.03±0.03
Sum background	1044.6±11.0	13.3±0.8	1.90±0.42
Data	1228.0±35.0	13.0±3.6	1.00±1.00

TABLE 6.2: Overview of integrated event number contribution for each process for 2016 ($\mathcal{L}_{int} = 35.9\text{fb}^{-1}$). Statistical uncertainties only.

Process	2017 m_T all	2017 $m_T > 1$ TeV	2017 $m_T > 1.5$ TeV
W	690.0±6.6	15.8±0.1	1.81±0.02
MisID	402.9±9.3	3.1±0.9	0.20±0.29
DY	56.1±2.3	0.6±0.05	0.05±0.01
top	26.4±3.5	< 0.01	< 0.01
Diboson	14.9±1.6	0.3±0.3	< 0.01
Sum background	1190.2±12.2	19.8±1.0	2.06±0.29
Data	1107.0±33.3	21.0±4.6	2.00±1.41

TABLE 6.3: Overview of integrated event number contribution for each process for 2017 ($\mathcal{L}_{int} = 41.3\text{fb}^{-1}$). Statistical uncertainties only.

Process	2018 m_T all	2018 $m_T > 1$ TeV	2018 $m_T > 1.5$ TeV
W	858.0±22.0	19.1±0.1	2.17±0.03
MisID	576.0±11.0	2.6±0.8	0.42±0.39
DY	62.7±1.9	0.9±0.1	0.07±0.01
top	37.9±2.0	< 0.01	< 0.01
Diboson	22.4±1.7	0.1±0.1	< 0.01
Sum background	1557.0±24.9	22.7±0.8	2.67±0.39
Data	1547.0±39.3	34.0±5.8	3.00±1.73

TABLE 6.4: Overview of integrated event number contribution for each process for 2018 ($\mathcal{L}_{int} = 60.0\text{fb}^{-1}$). Statistical uncertainties only.

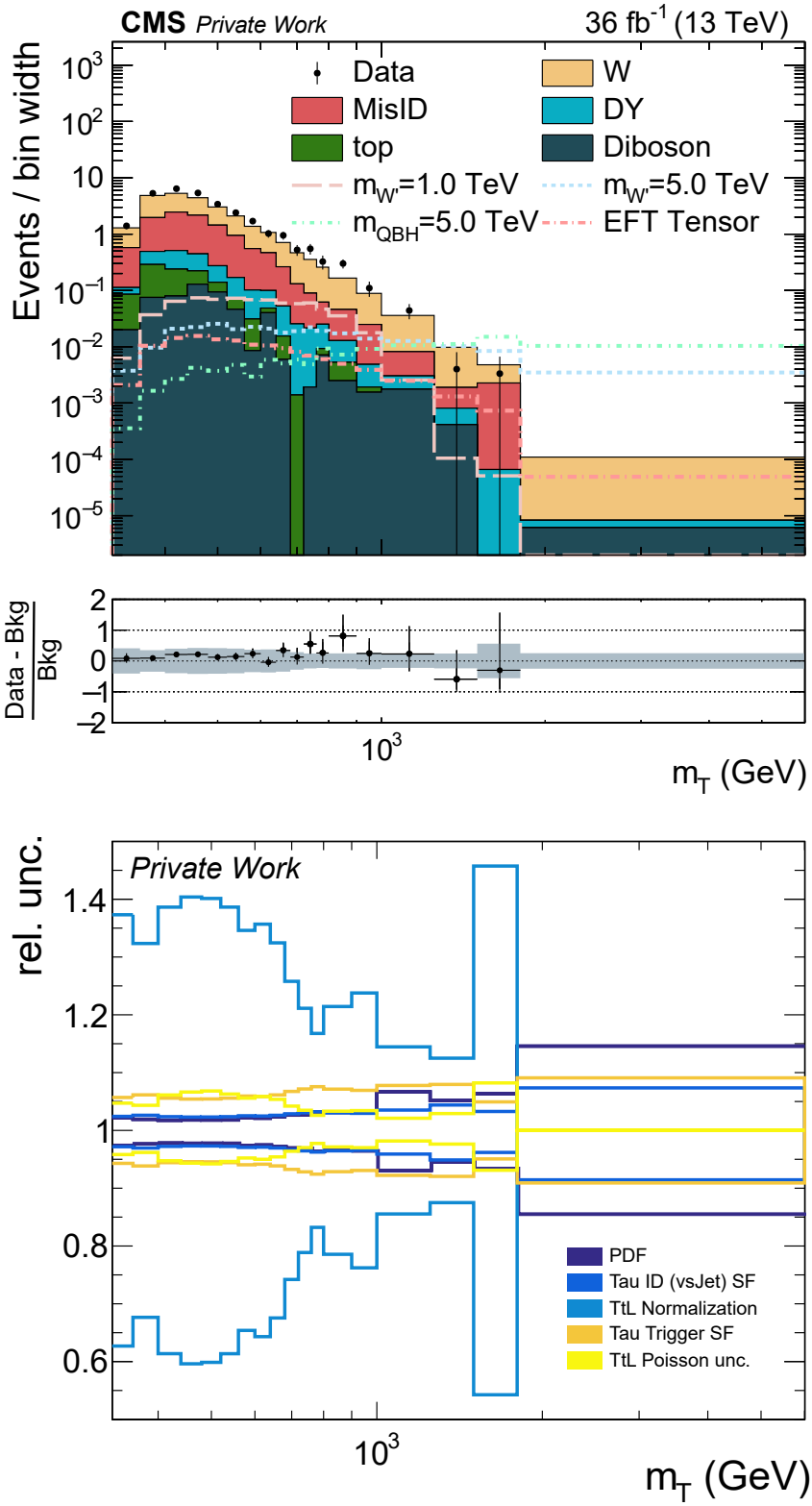

FIGURE 6.14: The m_T distribution (top) after full object and event selection and the leading systematic uncertainties (bottom) for 2016. The lower distribution shows the relative uncertainty per bin in m_T for the leading systematic uncertainties. Here, "TtL" indicates uncertainties related to the data-driven method, namely its normalization uncertainty and the uncertainty originating from the statistical limitations in the sideband regions.

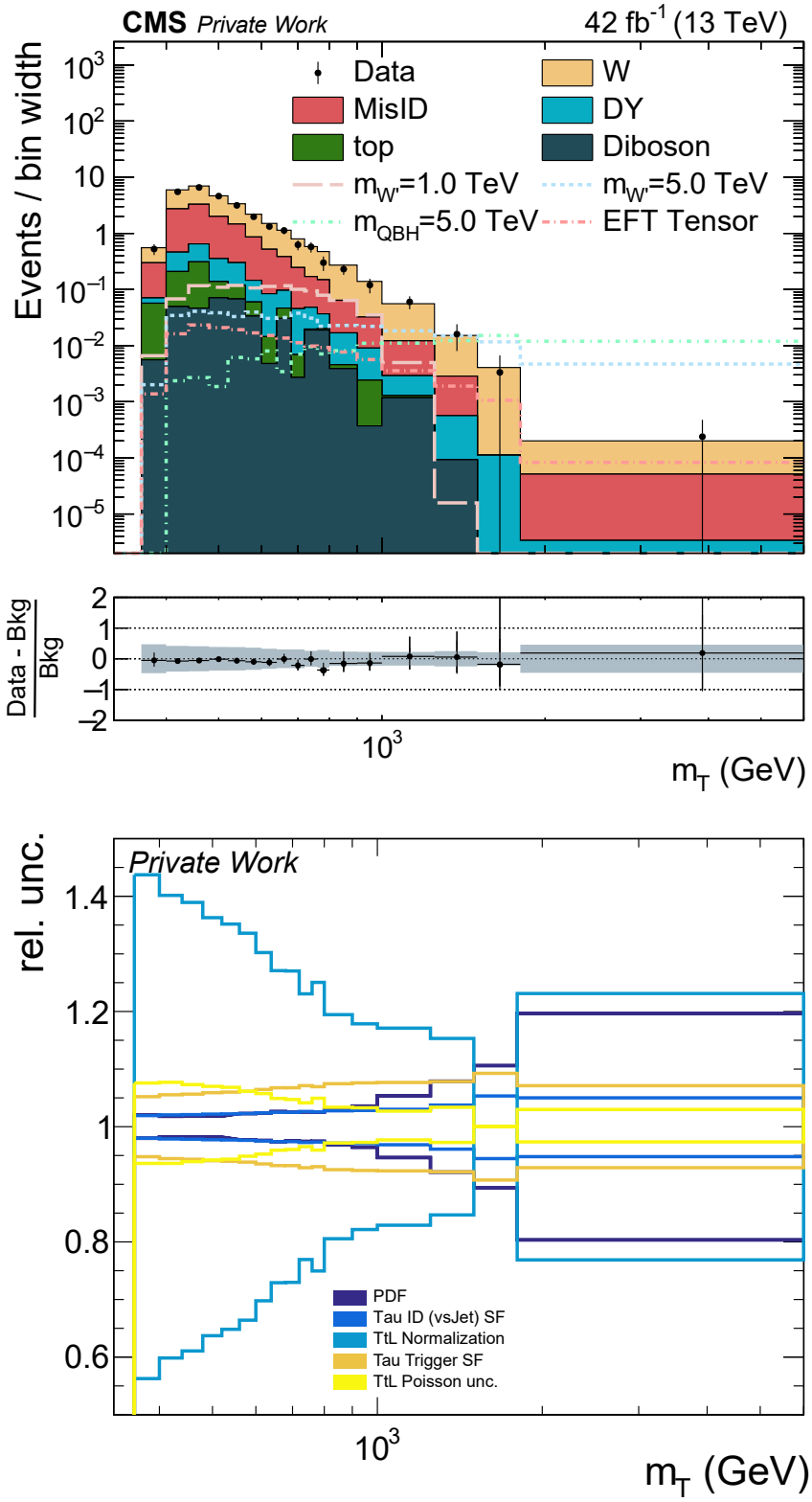

FIGURE 6.15: The m_T distribution (top) after full object and event selection and the leading systematic uncertainties (bottom) for 2017. The lower distribution shows the relative uncertainty per bin in m_T for the leading systematic uncertainties. Here, "TtL" indicates uncertainties related to the data-driven method, namely its normalization uncertainty and the uncertainty originating from the statistical limitations in the sideband regions.

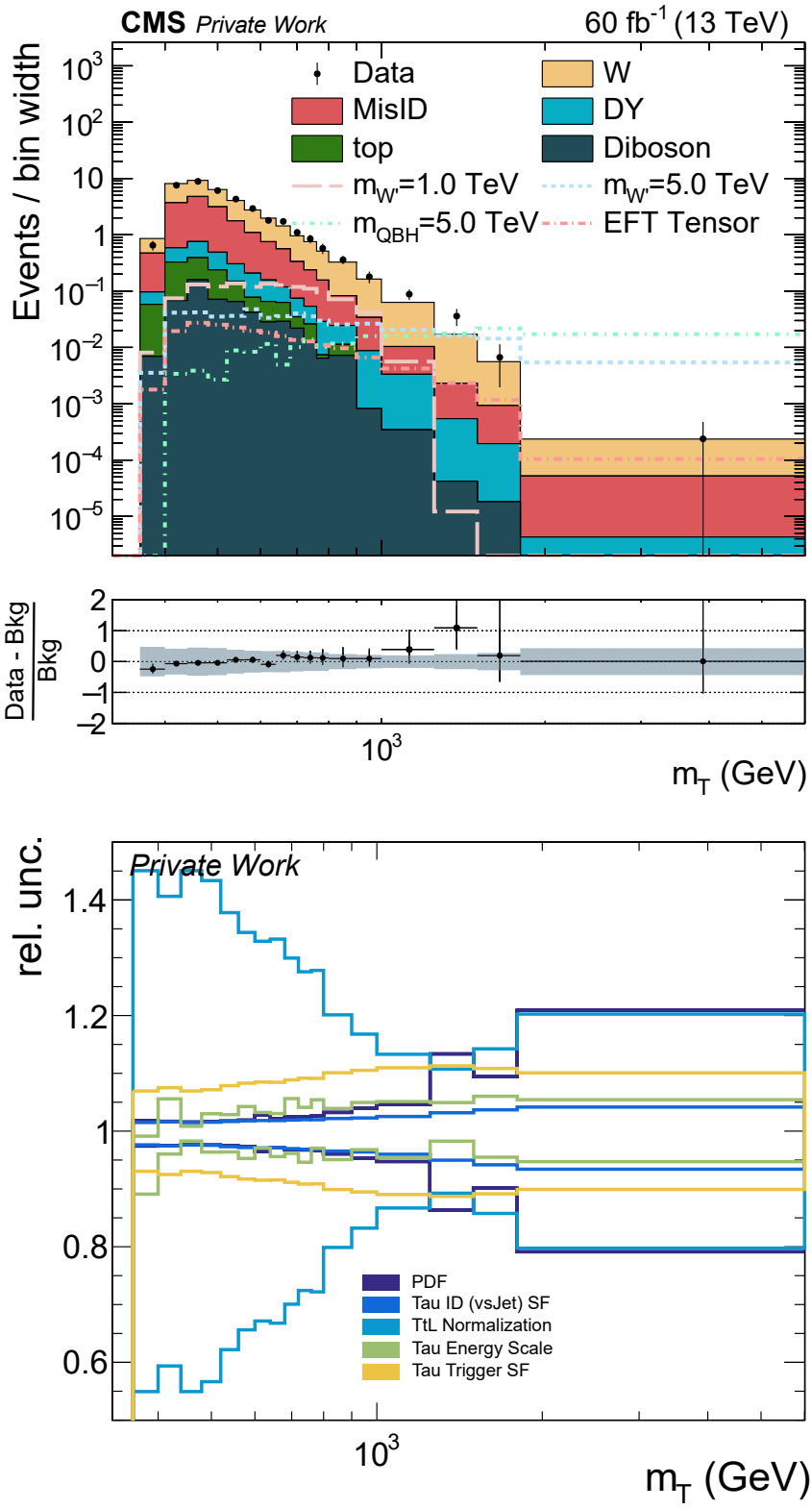

FIGURE 6.16: The m_T distribution (top) after full object and event selection and the leading systematic uncertainties (bottom) for 2018. The lower distribution shows the relative uncertainty per bin in m_T for the leading systematic uncertainties. Here, "TtL" indicates uncertainties related to the data-driven method, namely its normalization uncertainty and the uncertainty originating from the statistical limitations in the sideband regions. Compared to the other years, 2018 has a slightly smaller statistical uncertainty originating from the sideband regions, resulting in the energy scale correction to be of higher relative importance.

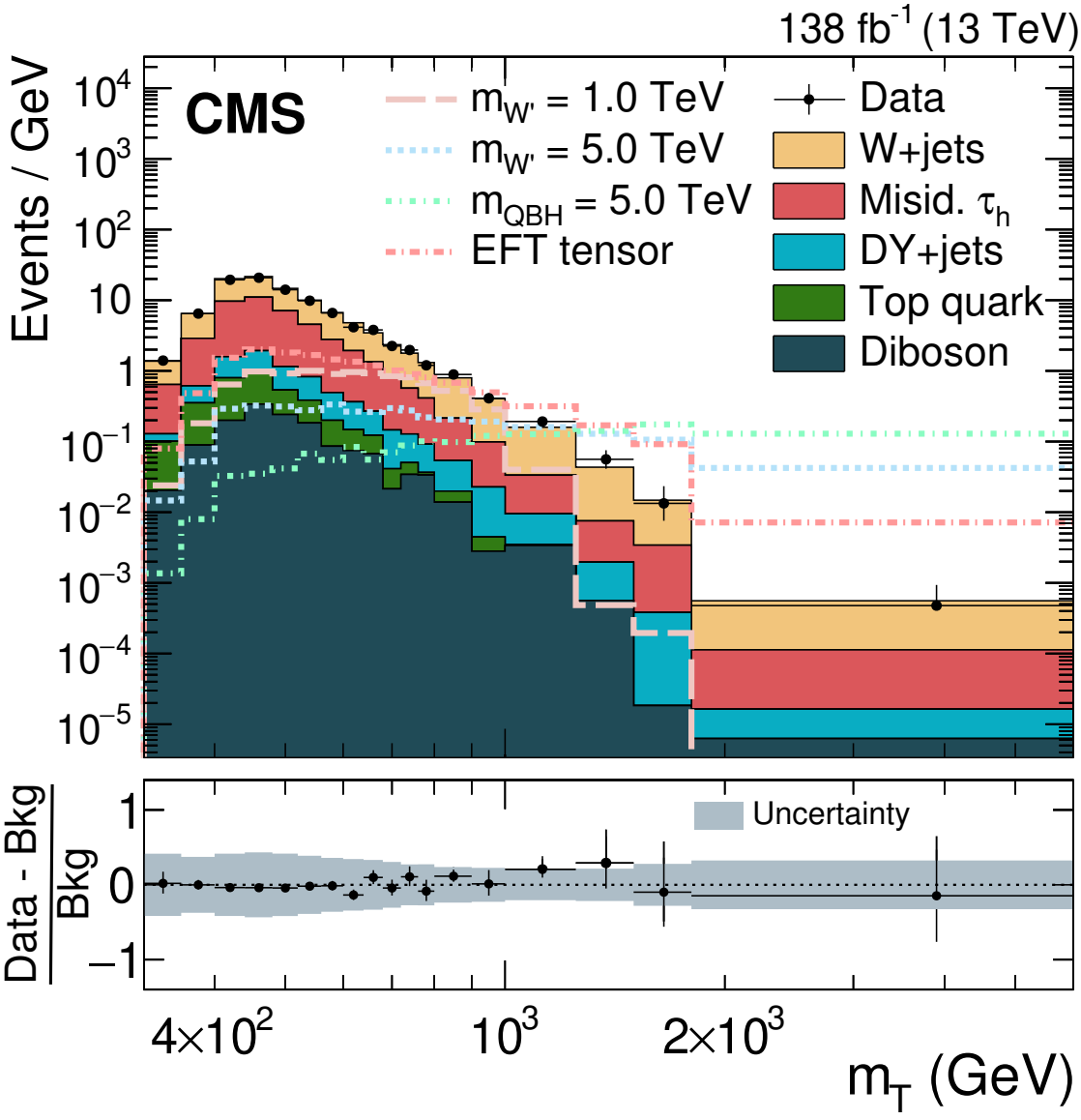

FIGURE 6.17: m_T distribution of events with a hadronically decaying tau lepton and missing transverse momentum, taken with the CMS detector in proton-proton collisions from 2016 to 2018. Backgrounds from quark/ gluon jets being misidentified as taus are taken from the data-driven method. The uncertainty band in the ratio distribution on the bottom shows the quadratic sum of systematic and statistical uncertainties. The data describes the background prediction for most of the distribution with a small overshoot of data in the region around 1 TeV. This abundance is within statistical uncertainties of the data.

In order to visualize reconstructed events in the CMS detector, an exemplary event display with a high m_T event from the 2017 data taking period is shown in Fig. 6.18. The indicated missing transverse momentum is shown in purple while the high energy tau jet is clearly visible by the high energy deposits in the calorimeters. The inner green line represents the information provided by the CMS Tracker. Back-to-back kinematics of the selected objects are clearly visible.

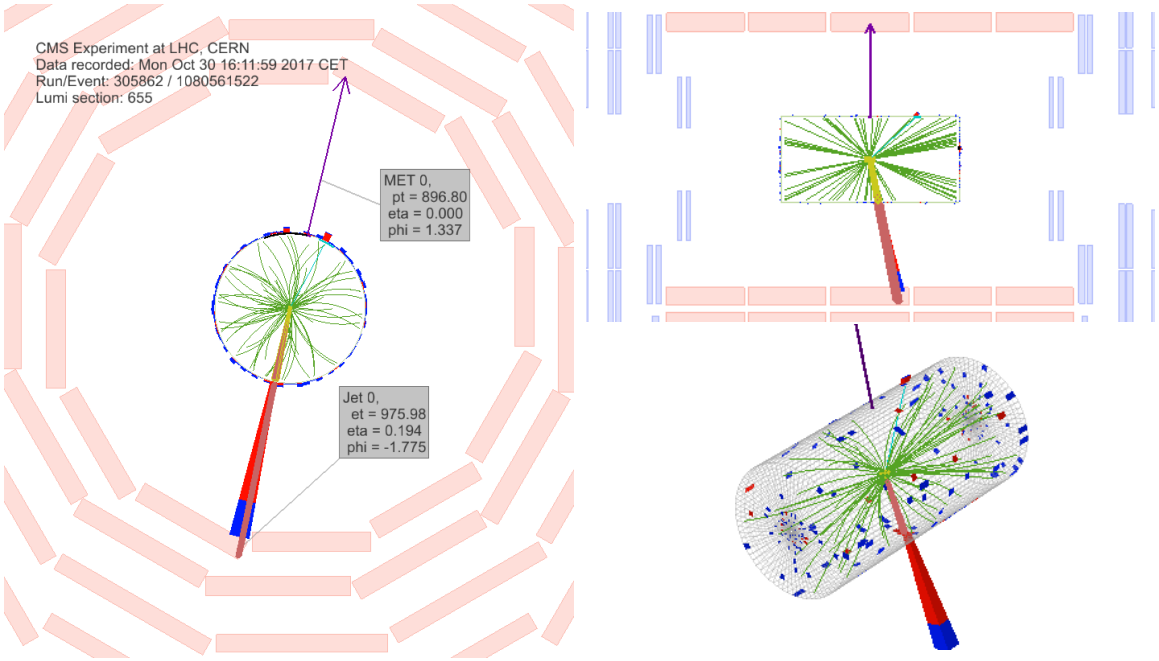

FIGURE 6.18: Exemplary event display from a high m_T selected event in this analysis, from the 2017 data taking period. The back-to-back kinematics are clearly visible. The left panel shows the $\rho - \phi$ view, the top right panel the $\rho - z$ view and the bottom right shows a 3D tower view. All views show tracker tracks in green and energy deposits in the calorimeters in blue and red. The outside blue and red rectangles indicate the muon chambers. The jets, which are reconstructed as tau objects are indicated in the left panel. Date and time of the event are indicated in the top left corner.

Chapter 7

Results

The results seen in the final per year m_{T} distribution (see Figs. 6.14, 6.15, 6.16) are analyzed with regard to deviations from the SM prediction. Since no significant deviations w.r.t. the SM prediction are seen, the results are interpreted in terms of exclusion limits. A Bayesian approach is used to determine the parameters which restrict the compatibility of apparent signals in the observed data [111, 112]. The choice to use the Bayesian approach is based on a previous agreement between the ATLAS and CMS collaborations for W' searches. The statistical methods used in this thesis are the same ones the CMS and ATLAS collaborations agreed upon during the search for the Higgs boson [113], which is based on the RooStats package [114].

This chapter provides a short explanation on the mathematical basics of the limit calculation, before presenting the exclusion limits for all various different model types discussed in Sec. 2.2. Additionally, a model-independent limit is provided near the end of this chapter, which allows for simplified estimations of exclusion limits of untested or even new models for the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state.

7.1 Bayesian statistics

The question to be answered is: "Given the measured data, what are the underlying causes?" In the following sections, different possible signal hypotheses (H_{signal}) are tested as such possible causes. In order to determine their posterior probabilities, *Bayes' theorem* is used:

$$Prob(H_{signal}|data, X) = \frac{Prob(data|H_{signal}, X) \cdot Prob(H_{signal}|X)}{Prob(data|X)} \tag{7.1}$$

The '|' symbol is commonly used in statistics for 'given' - i.e. all comma separated quantities to the right of this symbol are assumed to be true. These probabilities are all dependent on the parameter set X, which denotes the accumulated information already existent, like the information that our signal does indeed have the assumed shape in m_{T}. The denominator in the above equation does not depend on the specified signal hypothesis and (for our cases) is a normalizing factor. $Prob(H_{signal}|X)$ is called the *prior probability* of our signal hypothesis, which represents our state of knowledge about the truth of this hypothesis. This quantity is then tested using the above formula and yields a *posterior probability* $Prob(H_{signal}|data, X)$ representing what we learned about the signal in view of the measured data. In particle physics it is common to set limits with a *confidence level* (CL) of 95%. This means in this case, that for parameters of interest θ like a cross section or signal strength, the upper limit can be extracted from

$$\int_0^{\theta_{95\%CL}} Prob(\theta|data, X)\, dX = 0.95$$

This expresses, that e.g. an observed cross section limit value is estimated to the number, where 95% of the area of the posterior probability function lie below this value.

All other parameters, including background contributions with its uncertainties are treated using marginalization:

$$Prob(H_{signal}, X) = \int_{-\infty}^{\infty} Prob(H_{signal}, Y|X)\, dY \tag{7.2}$$

For any parameter Y, which is relevant to the statistical method but not of special interest for its exact value, the above equation can be used to estimate their effect on the limit calculation. These parameters are called *nuisances* and include all uncertainties on the signal, the background as well as the size of the background contribution itself. For an analysis like this, the method basically becomes a best-fit parameter search in many dimensions, making the best estimation of data with background and resulting in a probability density function for a certain parameter of interest. The parameters of interest here are either the cross section of a given signal model or the signal strength $\mu = \sigma_{signal}/\sigma_{theory}$.

This method is used for all exclusion limits set in this thesis. The prior probability for all signal hypotheses is assumed to be flat - this basically represents the fact that there is no information on the actual truth of the signal. It would be systematically wrong to e.g. put in a biased prior probability density function favoring values which are close to what we would expect it to be from previous analyses, since this would result in posterior probability density functions much sharper than what data actually warrants, giving a wrong sense of accuracy. Except for the model independent limit shown at the end of this section, all exclusion limits are calculated using shape information of data, background and signal obtained from the m_{T} distribution [115]. This is achieved by estimating the likelihood function for each individual bin and combining them after this evaluation.

All limits shown in this section use the full Run-2 CMS data set corresponding to an integrated luminosity of 138 fb^{-1}.

7.1.1 Impacts

In the process of limit setting, the background estimation is shifted within their uncertainties to find the best estimation of the measured data. This is compared to the estimation of background+signal, and the exclusion limit on the signal cross section can be extracted. The nuisance parameters have different impacts on the exclusion limit, as shown in Fig. 7.1. In this figure, different types of uncertainties are shown listed by their relative impact on the exclusion limit. The impact, $\Delta\hat{r}$, is defined as the change in the best-fit value for the parameter of interest $\hat{r}$ when varying a given nuisance parameter θ within its uncertainties. Also shown in the middle column are the best-fit values $\hat{\theta}$ for each nuisance parameter with their uncertainties, compared to their pre-fit values θ_0. E.g. when a value for a given nuisance is larger than 0 and the uncertainty bands are smaller than 1, the best-fit result estimates the mean value of this nuisance to be slightly larger than the pre-fit value and its uncertainties to be smaller than their corresponding pre-fit values (see line 7 of Fig. 7.1). It can be e.g. clearly seen that the deliberately conservatively estimated data driven normalization uncertainty can be reduced significantly for each year in this process. Other uncertainties (e.g. trigger scale factor uncertainties) show very little difference w.r.t. pre- and post-fit behavior.

As can be seen, the most impactful uncertainties are the background PDF uncertainty, followed by the uncertainty originating from the data driven normalization and the uncertainty on the tau trigger scale factor. The chart also includes information about statistical uncertainties for each individual (numbered) bin of the m_{T} distribution and shows that they

are not leading uncertainties. All three normalization uncertainties applied to the background from misidentified tau objects appear to be estimated quite conservatively, as this method evaluates their best-fit solution to be relatively smaller.

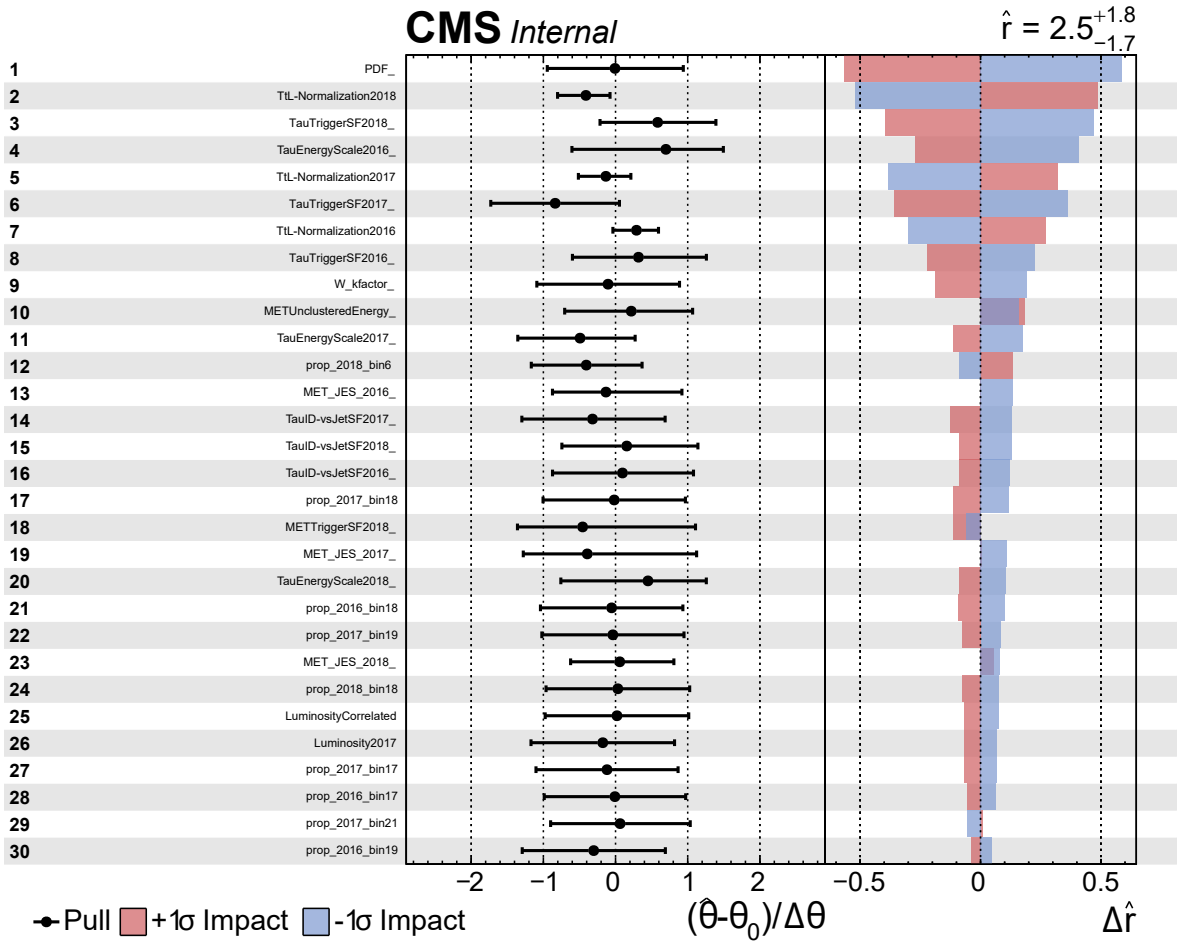

FIGURE 7.1: Impacts of the nuisance parameters on the final limit calculation for an exemplary SSM W' signal with $m_{W'} = 1.4$ TeV for 138fb^{-1}. Shown are only those nuisance parameters with the largest impact (i.e. largest Δr, right column). For each nuisance parameter, the best-fit uncertainty relative to the pre-fit uncertainty is shown (i.e. if it ranges from -1 to +1, the uncertainty is identical pre- and post-fit). Entries labeled with 'prob_X_binY' refer to statistical uncertainties from background sample simulation, where X is the data taking year and Y denotes the Yth bin of the m_T distribution. Nuisances labeled with 'SF' denote a scale factor uncertainty and those starting with 'MET' denote uncertainties related to p_T^{miss}.

7.2 Sensitivity of the Run–2 analysis

7.2.1 Sequential Standard Model W'

For a heavy version of the SM W boson, the W' boson, limits are placed on the mass of such a hypothesized particle, shown in Fig. 7.2. The specific model assumes the decay of such a boson to W+Z to be suppressed, resulting in high branching fractions for diquark or dilepton decays (see Sec. 2.2.1). Signal samples were generated with Pythia8 at LO, but theoretical cross sections have been calculated at NNLO. These are shown as a blue dashed line with a gray band to include the cross section uncertainties coming from PDFs. By determining the intersection of the theoretical cross section prediction and the observed cross section exclusion limit, masses up to

$$m_{W'} = 4.7\,\text{TeV}$$

can be excluded at 95% CL. The expected (dashed black) line is mostly in agreement with the observed (solid black) line, except for the medium mass region between 1 and 2.5 TeV. The exclusion limit in this region is driven by events around 1-1.5 TeV in the m_{T} distribution, where a slightly increased event yield is observed with respect to the SM expectation (see Fig. 6.17). Therefore, the experimental data can not exclude as much parameter space as the expected limit. This is however covered by the one (green) and two (yellow) standard deviation uncertainty bands of the expected exclusion limit, as shown in Fig. 7.2. The lines in this figure represent a linear interpolation between the singular exclusion limits, which have been calculated for all available signal mass hypotheses.

Previous results in this search channel include the CMS (ATLAS) 2016 result quoting an SSM exclusion limit of 4.0 (3.7) TeV at 95% CL [116, 117]. The ATLAS collaboration also published a conference note with an exclusion limit of 5.0 TeV [118]. However, since this model has equal couplings to light leptons, the dedicated search in the combined $e + \mu$ search channels provides an exclusion limit of 5.7 TeV [98]. This analysis therefore significantly improves on previous search results in the specific channel, and complements the more sensitive light charged lepton search channels.

7.2.2 W' Coupling Strength

This approach is expanded, taking into account the relative coupling strengths of the SM W boson and a W' boson. The coupling strength affects the cross section and total decay width of the W' boson, as discussed in Sec. 2.2.1. The scanned parameter space includes coupling ratios $0.0001 \leq g_{W'}/g_W \leq 5.0$ for the same mass range as above. For each mass, the calculated cross section exclusion limit is compared to the LO theoretical prediction in dependency on the coupling ratio[1]. From the intersection in this plane, the limit is extracted. This is exemplarily shown for a W' mass value of 3.2 TeV in Fig. 7.3. This intersection is calculated for both observed limit, expected limit and uncertainty bands for each mass value and the collection of these points is then shown in Fig. 7.4. Coupling ratio values $g_{W'}/g_W > 2$ can be excluded for the full probed mass range. In the low mass region, values of $g_{W'}/g_W > 0.05$ are excluded. Both these limits are significantly weaker than the corresponding combined analyses in the $e + \mu$ channel, which excluded this model up to masses of $m_{W'}^{SSM} < 5.7$ TeV [98], but it complements and completes the full spectrum of this benchmark model investigation.

[1] For the coupling dependent signals, no higher order calculations are available.

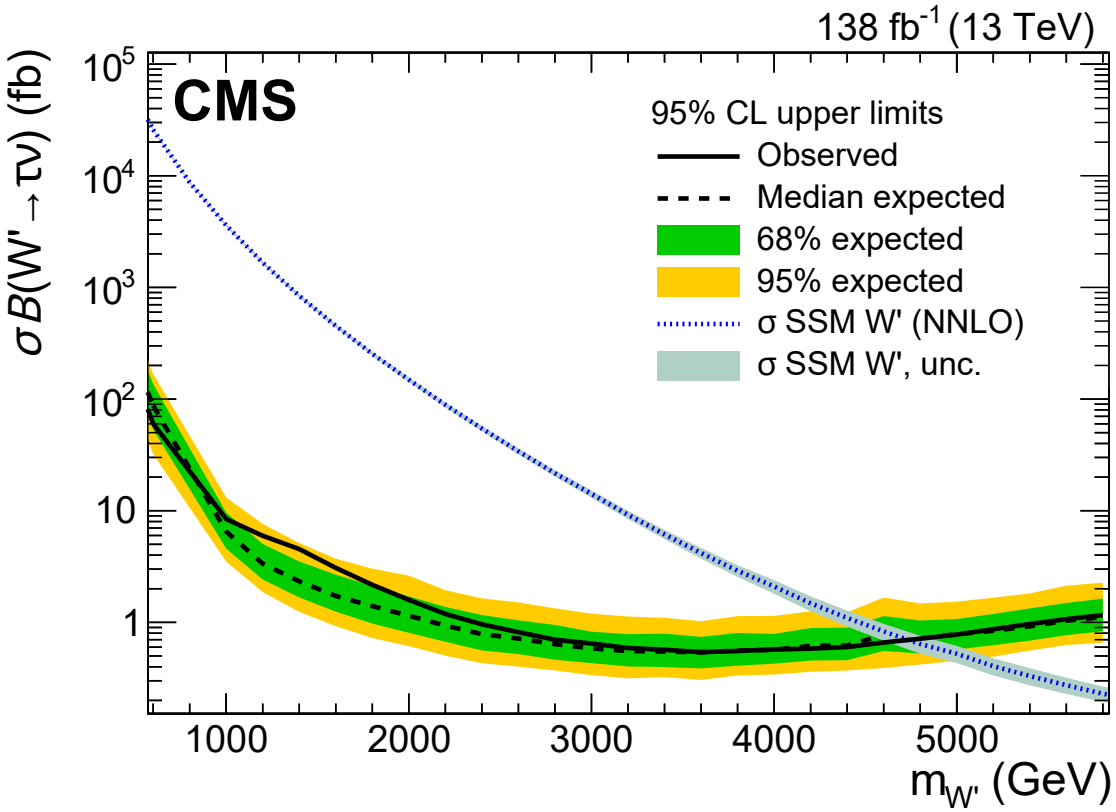

FIGURE 7.2: Upper limit on the cross section times branching fraction for the Sequential Standard Model W′ boson in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ decay channel based on the CMS Run2 data set with 138fb^{-1}. Shown are the observed limit (solid line) based on data and the expected limit based on background prediction (dashed line) with its uncertainty bands (green and yellow). Also included is the theoretical cross section prediction based on NNLO cross section calculations (blue dashed line). Regions where the theoretical prediction is higher than the observed limit are excluded by the data observation, resulting in a mass exclusion limit for the SSM W′ boson of $m_{W'} = 4.7$ TeV at 95% CL.

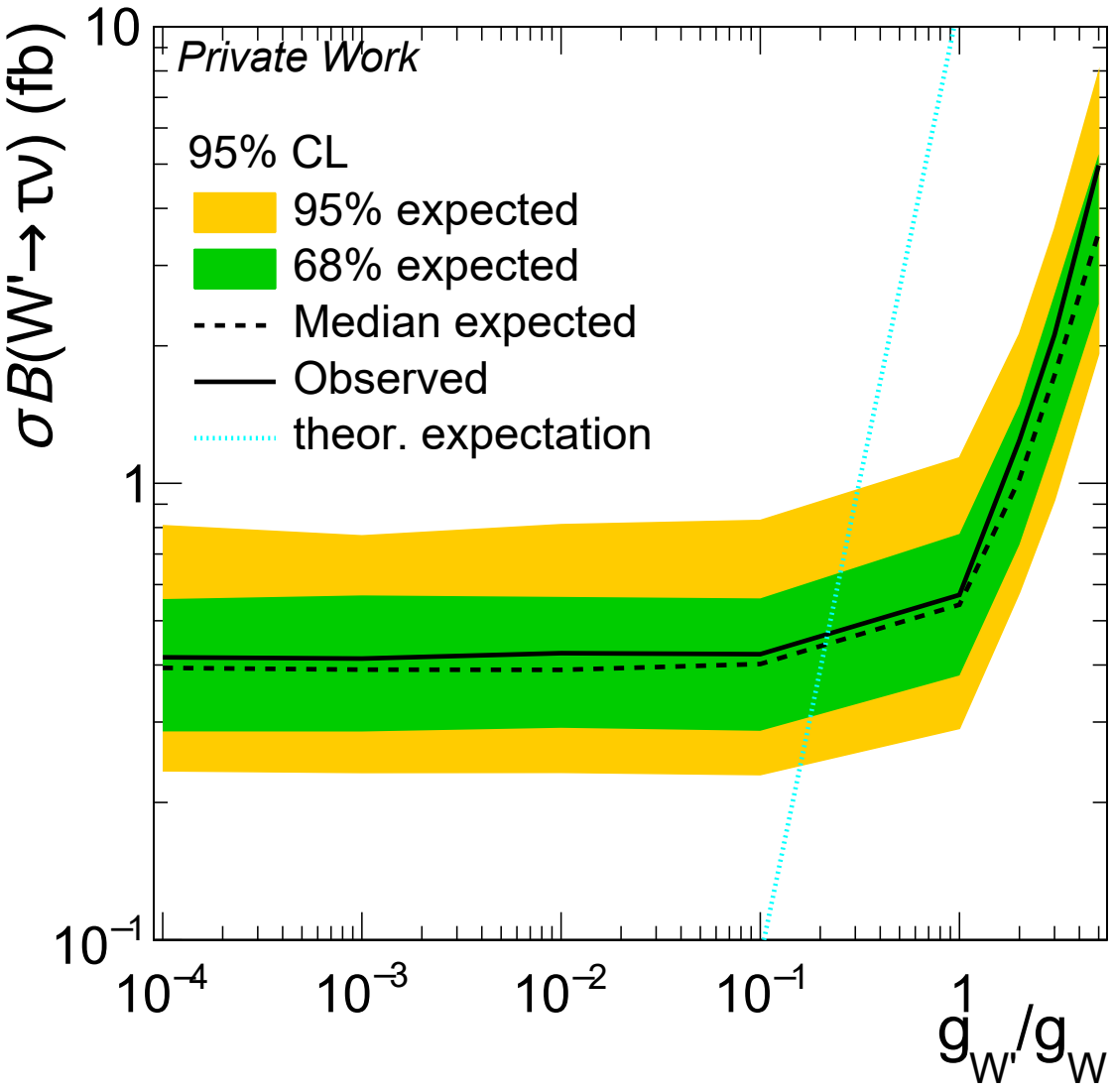

FIGURE 7.3: Example of the 3.2 TeV signal mass point calculation of the coupling limit. The intersections of the theory line and the calculated cross section limits are used as input for the final coupling limit for each mass point.

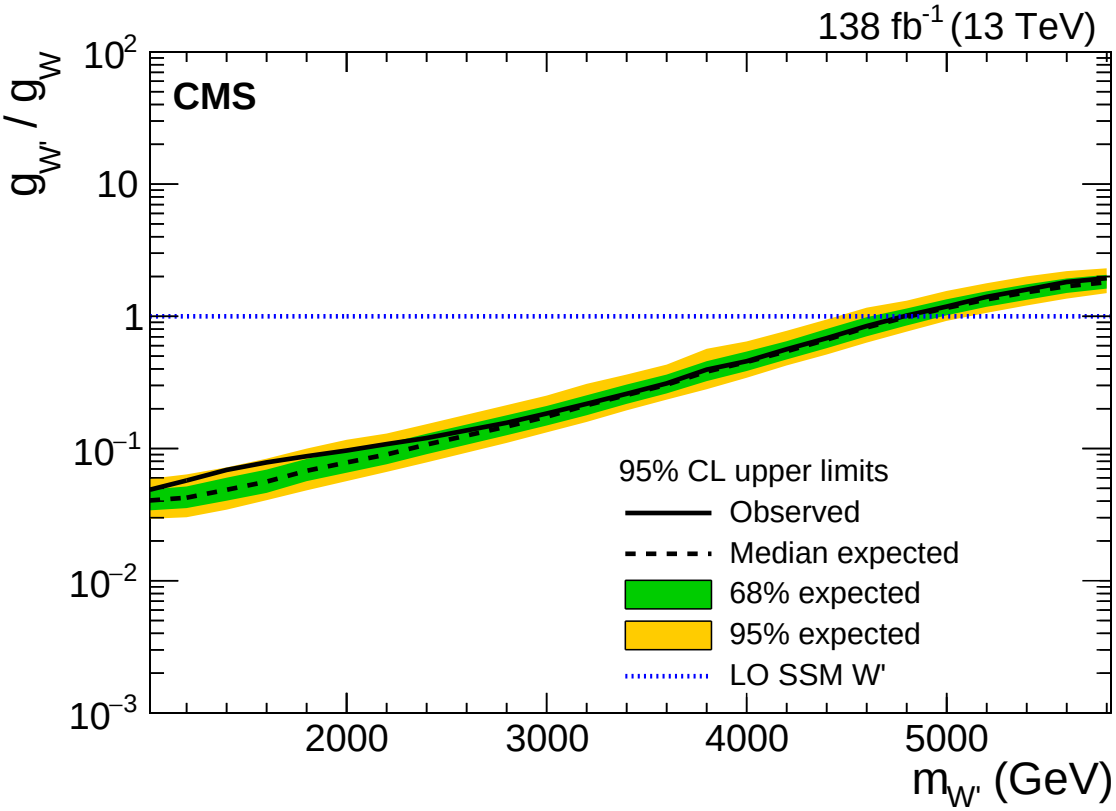

FIGURE 7.4: Coupling dependent limit for a W′ boson in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ decay channel based on the CMS Run2 data set with 138fb^{-1}. Shown are the observed limit (solid line) based on data and the expected limit based on background prediction (dashed line) with its uncertainty bands (green and yellow). Each point is estimated by the intersection of the theoretical LO W′ cross section with the respective expected/observed limit in the $g_{W'}/g_W$-cross section plane. Coupling values above the observed line are excluded.

7.2.3 NUGIM Exclusion Limit

Exclusion limits for the non-universal gauge interaction model are calculated as a function of the mixing angle $\cot(\theta_E)$ and the mediator mass $m_{w'}$. To achieve this, the exclusion limits calculated from the coupling dependent W' signal samples are taken and reinterpreted in terms of this model. This is possible since it is only necessary to adjust cross sections and widths for the purposes of this interpretation. Therefore, similar to the calculation of the coupling dependent model, for each mass hypothesis the cross section exclusion limits are estimated as a function of the width of the heavy mediator. The estimated coupling exclusion limit is translated into a value of $\cot(\theta_E)$; the result is shown in Fig. 7.5. In this distribution, values (bottom) left of the observed line are excluded. Note that values of $\cot(\theta_E) > 5.5$ are theoretically forbidden. This search excludes this model at 95% CL for masses of 2.3 TeV up to 4.8 TeV, depending on the value of $\cot(\theta_E)$. Compared to previous results [116], which excluded masses between 1.7 and 3.9 TeV for the same range of $\cot(\theta_E)$, a significant improvement can be seen.

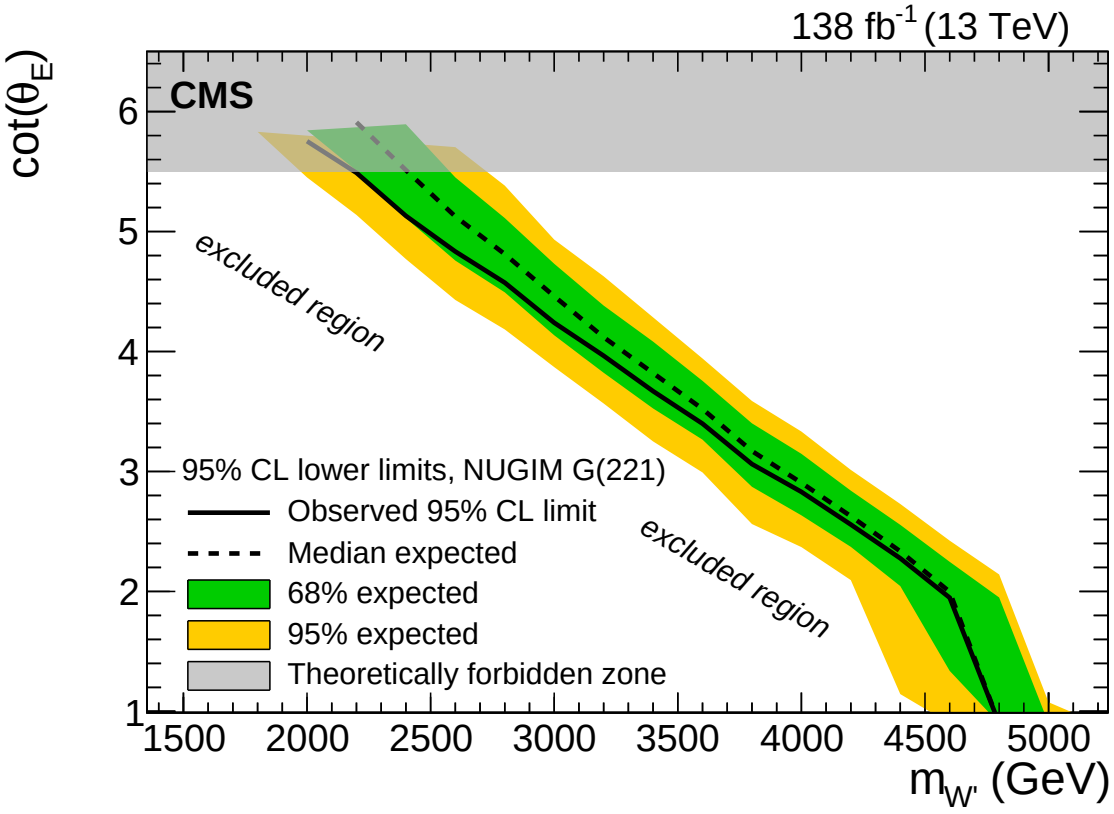

FIGURE 7.5: NUGIM limit in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ decay channel based on the CMS Run2 data set with 138fb^{-1}. Shown are the observed limit (solid line) based on data and the expected limit based on background prediction (dashed line) with its uncertainty bands (green and yellow). Each point is estimated by the intersection of the theoretical NUGIM cross section with the respective expected/observed limit in the $\Gamma_{W'}$-cross section plane. The parameter space left to the observed line is excluded.

7.2.4 Quantum Black Hole

Cross section limits are calculated for production of singly positively charged QBHs, shown in Fig. 7.6. The number of extra dimensions in the underlying ADD model is assumed to be $n = 4$, which is a common reference model. The signal shape typical for quantum black holes allows for strong discrimination against the backgrounds, as most signal events are at very high masses. The shape of the exclusion limit for this model is similar to the shape of the SSM W' exclusion limit, with the low mass dominated by the low signal efficiency due to the trigger thresholds. The production of singly positively charged QBHs can be excluded for masses of up to

$$m_{th} = 6.7\,\text{TeV}$$

at 95% CL. This result adds to the already existing results from other search channels (e.g. Ref. [101]), and further restricts the parameter space of this model. This is the first explicit interpretation of this model in this channel using CMS data.

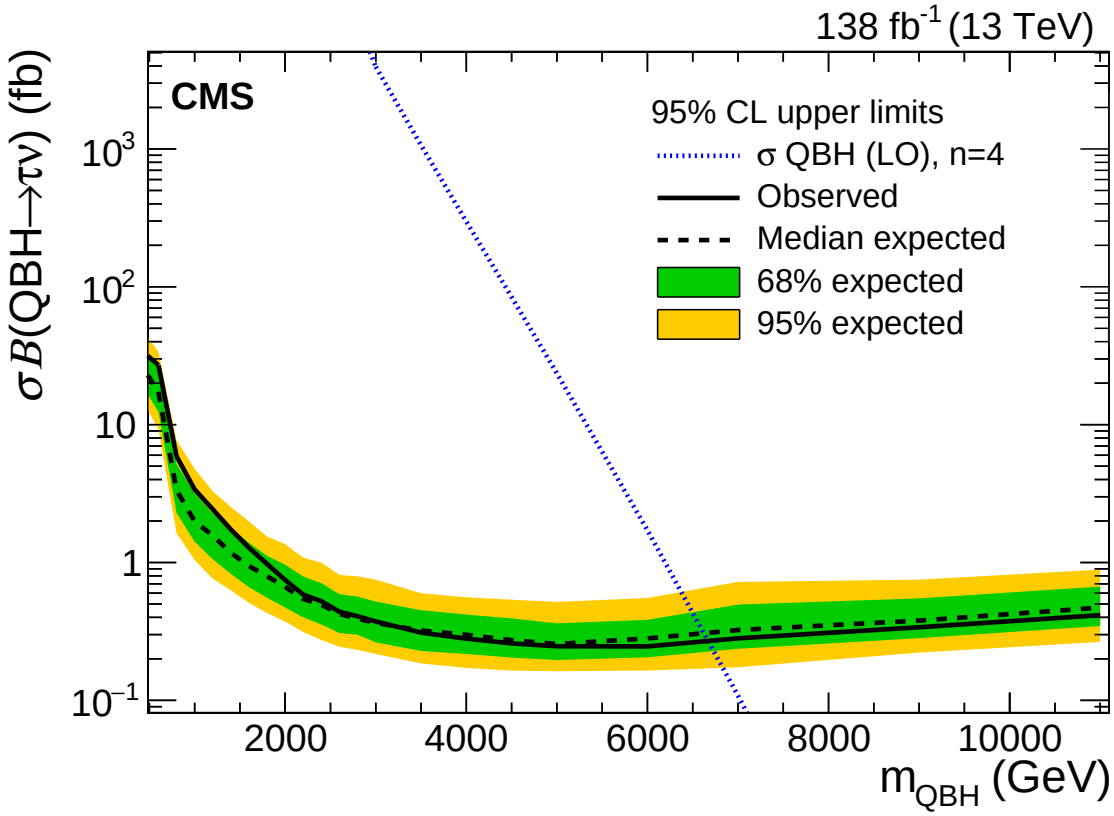

FIGURE 7.6: Upper cross section limit for the production of a positively charged quantum black hole and its decay to $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ based on the CMS Run2 data set with 138fb^{-1}. Shown are the observed limit (solid line) based on data and the expected limit based on background prediction (dashed line) with its uncertainty bands (green and yellow). Also included is the theoretical cross section prediction based on LO cross section calculations (blue dashed line). Regions where the theoretical prediction is higher than the observed limit are excluded by the experimental data, resulting in a mass exclusion limit for the production of a positively charged quantum black hole $m_{QBH^+} = 6.7$ TeV at 95% CL.

7.2.5 Effective Operators

The effective field theory description, which translates the anomalous measurements of the $R_{D^{(*)}}$ ratio of branching fractions into a high energy phenomena search, can be excluded on basis of the Wilson coefficients, which determine the coupling strength to the different coupling types: Scalar-tensor-like coupling, vector-like coupling and tensor-like coupling. Due to the non-resonant signal shape of this interpretation, the exclusion limits obtained here are much more sensitive to the intermediate mass range (1-1.5 TeV in m_{T}), with the exclusion limits shown in Fig. 7.7. Limits on the coupling values are obtained by scanning a coupling value range between 0.05 and 2.0 for each coupling type. The excluded observed (expected) values are

$$\epsilon_{\mathrm{L}}^{\mathrm{cb}} = 0.32(0.27)$$
$$\epsilon_{\mathrm{S_L}}^{\mathrm{cb}} = 0.51(0.41)$$
$$\epsilon_{\mathrm{T}}^{\mathrm{cb}} = 0.27(0.22)$$

As the most sensitive m_{T} region for this model is at the region where a small surplus of data is observed, the observed is slightly weaker than the expected one. The difference is small, however and fully covered by the uncertainties. These are the first limits set on this model's parameters at CMS.

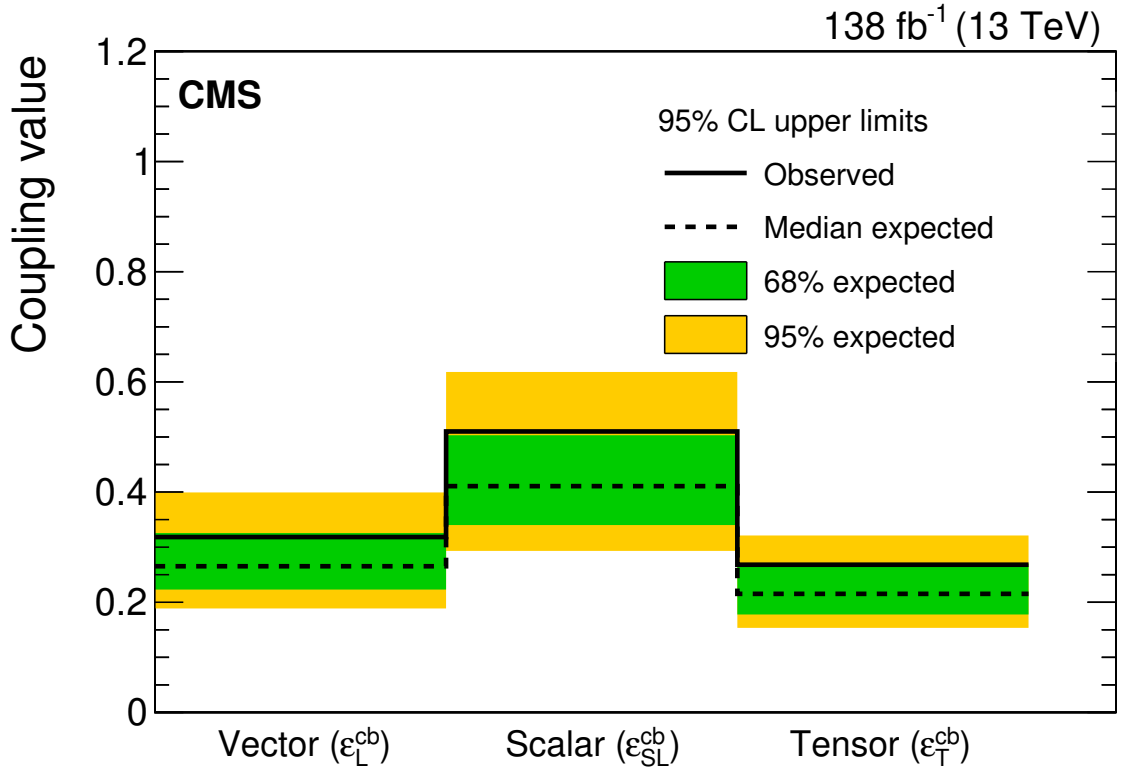

FIGURE 7.7: Bayesian upper exclusion limit on the values of the three different Wilson coefficients proposed in the effective field theory description based on 2016–2018 CMS data. The coupling types represent a LH vector coupling ($\epsilon_{\mathrm{L}}^{\mathrm{cb}}$), a scalar-tensor-like coupling ($\epsilon_{\mathrm{S_L}}^{\mathrm{cb}}$) and a tensor-like coupling ($\epsilon_{\mathrm{T}}^{\mathrm{cb}}$).

7.2.6 Model Independent Limit

All exclusion limits shown so far have some form of significant assumption on the form of the signal shape. And while different signals have been discussed and excluded explicitly, there are many thought and unthought of signals which could be possible in this final state. In order to make the analysis fruitful for any possible new physics scenario leading to an increased event yield in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state, a model independent limit is provided. In order to calculate such a limit, it is by construction necessary to have no specific signal shape when calculating the exclusion limit. This is achieved by changing the limit setting approach. Instead of a shape-based interpretation of the m_{T} distribution, the results are statistically analyzed using a counting-experiment approach. To achieve this, a variable m_T^{min} is defined as a lower threshold on the m_{T} distribution. Integrating all events in both background and data above this limit threshold results in a single-bin of events, where background can be compared to data.

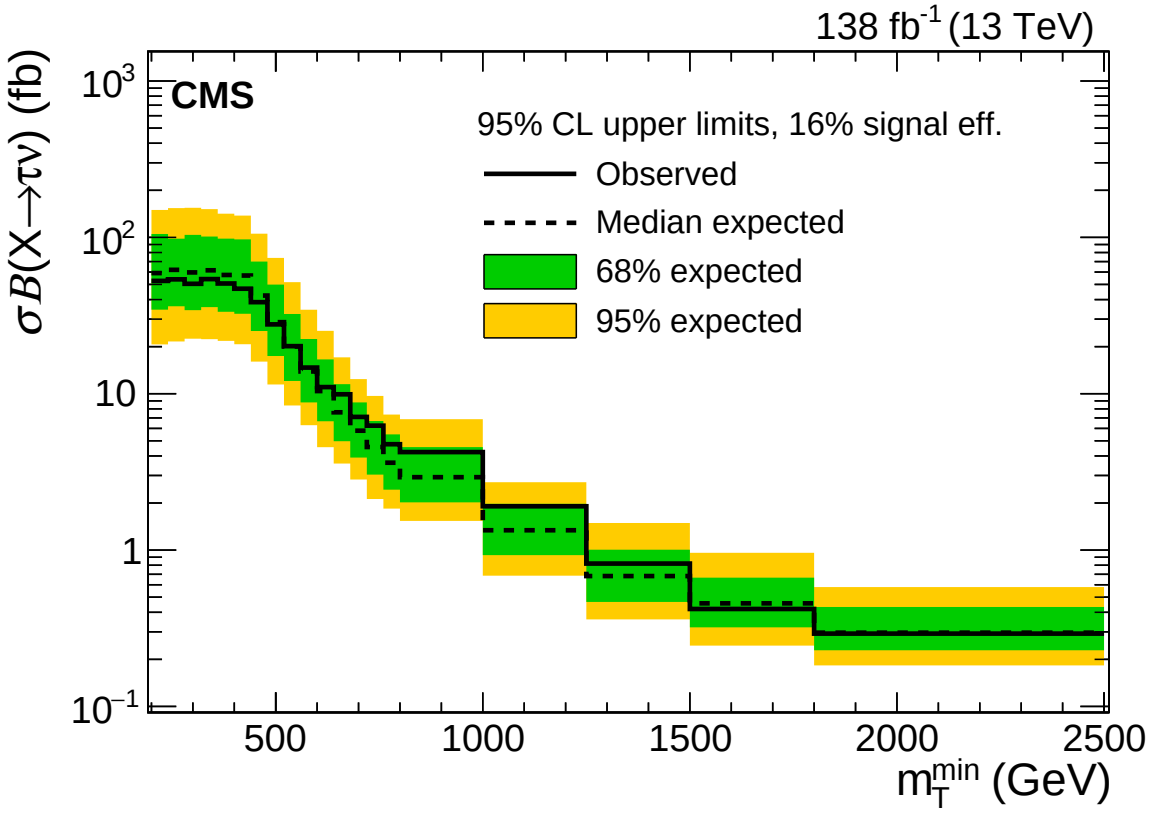

FIGURE 7.8: Model independent upper limit on cross section times branching fraction in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ final state. Each cross section value solely depends on the number of integrated data and expected background events based on the minimum m_{T} threshold m_T^{min}. For each value of m_T^{min}, the same constant signal efficiency given in the figure has been used.

In order to calculate an exclusion limit, some assumption has to be made about the signal efficiency. In this analysis, this number was chosen to be 16%, which is close to a typical signal efficiency for high mass searches (see Fig. 6.3) and includes the hadronic decay branching fraction of the tau lepton. Note, that no information on signal shape enters this assumption. This procedure (single-bin limit with an assumed constant signal efficiency) is repeated for various thresholds of m_T^{min} and the results of calculating such limits is shown in Fig. 7.8. The step structure of this limit originates from the binning from the m_{T} distribution. Opposite to previous shown exclusion limits, this representation of the model independent exclusion

was chosen, since linear approximations between the different m_T^{min} points of exclusion limit calculation need to be avoided.

This exclusion limit can now be used for an easy reinterpretation of the results. If someone proposes a new physics model, which would lead to an increased event yield compared to SM expectation in the final state with a hadronically decaying tau lepton, this person can estimate a cross section limit on this process by evaluating, for each value of m_T^{min}, his signal efficiency $\epsilon = \mathrm{N}_{m_T > m_T^{min}} / \mathrm{N}_{total}$, where N stands for the number of signal events. This can then be translated into a cross section exclusion limit for $(\sigma\mathcal{B})_{\text{new model}}(m_T > m_T^{min}) = \sigma_{\text{total}} \cdot \epsilon_{\text{signal}} > \sigma_{\text{excluded}}$. The excluded cross section limit is then the minimum value from all individual cross section limits obtained from scanning m_T^{min}.

Since reconstructed and generated m_{T} differ only slightly, the provided cross section limit can in this way be used to estimate an exclusion limit on any model with a τ lepton and missing transverse momentum using generator information. This model independent limit could be e.g. used to calculate upper cross section limits on the direct production of tau leptons and dark matter (which would be an invisible object like the neutrino).

7.3 High-luminosity sensitivity study

In addition to the shown results, a study was performed to estimate the sensitivity of this search at high-luminosity LHC (HL-LHC) [119]. In the next 20 years, it is expected to have much higher instantaneous luminosity as well as a center of mass energy of 14 TeV at the LHC - 3000 fb^{-1} of p-p collision data are expected to be delivered at the end of this running period (see Fig 7.9). To understand what insights into physics can be gained from such efforts CERN scientists have collaborated to create the *Yellow Report* [120]. This document combines many studies of experimental sensitivity obtainable at HL-LHC, including (some of the) results shown in this section. This analysis part was published as a physics analysis summary by the CMS Collaboration [121].

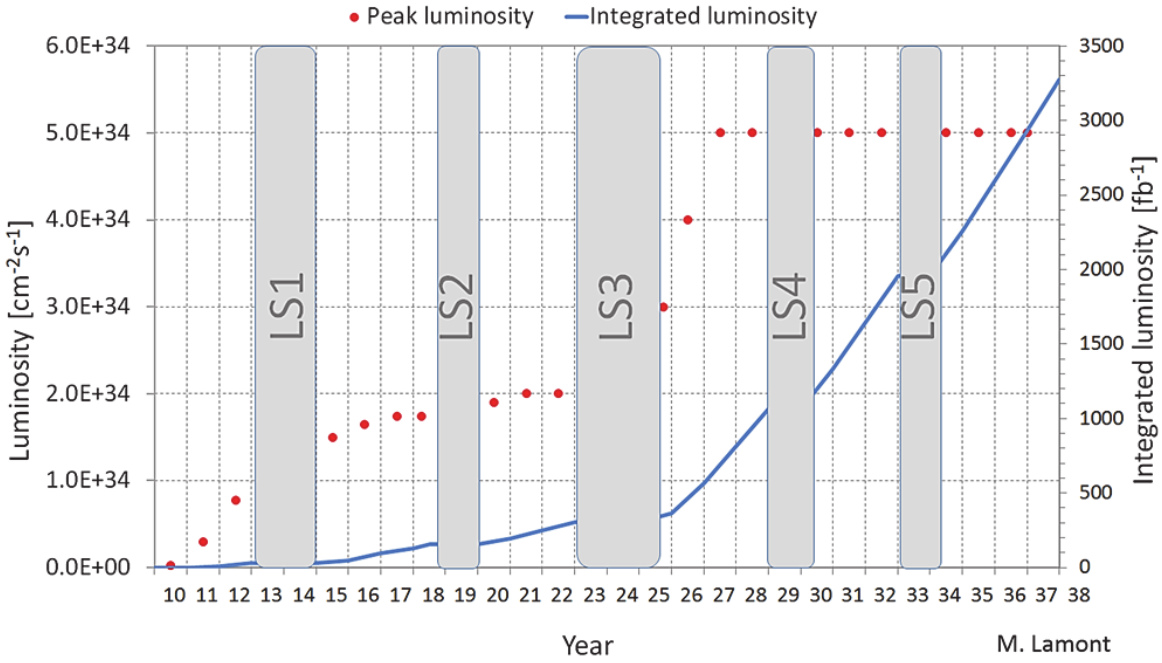

FIGURE 7.9: Estimated luminosity delivered by the luminosity at the end of the high-luminosity era until the end of 2037[2]. Image taken at the time of this study (Fall 2018) from [122].

The CMS detector will be substantially upgraded towards HL-LHC in order to exploit the physics potential of the HL-LHC and to be able to handle the expected harsh operational conditions. Expected improvements include a higher granularity tracking system, extending to $|\eta| = 4$, an overall upgrade of the combined ECAL+HCAL system, as well as an improved muon system with upgraded back-end electronics and an extended coverage up to $|\eta| = 2.8$. Additionally, a new timing detector is envisioned, capable of 4-dimensional vertex reconstruction, which would greatly reduce performance degradation at high pileup. Much more information on the planned CMS upgrades is given in Refs. [124, 125, 126, 127, 128]. A detailed overview on object performance at the upgraded CMS detector can be found in [129].

This study is a projection and performance of the Phase-2 CMS detector is not yet considered. Only simulation is used as an input for this analysis. Contributions to the $\tau_h + p_{\mathrm{T}}^{\mathrm{miss}}$ final state are estimated using simulated events processed with DELPHES 3.4.1 [130]. Unlike a full detector simulation, DELPHES uses generator input and parametrizes the detector

[2]This timeline has been adjusted slightly after the analysis presented in this chapter was performed, see [123]. The expectation to obtain at least 3000 fb^{-1} before 2040 remained the same.

Object definition	
Tau object	$W_\tau^{DELPHES} > 0.1$
MisID jet object	$W_\tau^{DELPHES} <= 0.1$, apply fake rate ($< 0.12\%$)
Object selection	
momentum threshold	$p_T^{object} > 80$ GeV
calorimeter acceptance	$\eta^{object} < 2.8$
$p_{\mathrm{T}}^{\mathrm{miss}}$ threshold	$p_T^{miss} > 200$ GeV
Kinematic Selection	
Veto events with light lepton	$p_T^l > 25$ GeV
Back-to-back requirement	$\Delta\Phi(\tau, p_T^{miss}) > 2.4$
Balanced momentum	$0.7 < p_T^\tau / p_T^{miss} < 1.3$

TABLE 7.1: Overview on the event selection criteria used in the HL-LHC upgrade study. These selections are applied to events processed with DELPHES for an average pileup scenario of 200.

performance based on independent detector studies. An average pileup of 200 was used to estimate the effect of the high number of primary vertices expected at each collision at the HL-LHC.

Dedicated identification algorithms of hadronically decaying tau leptons are not included in DELPHES. Therefore, it is necessary to parametrize both the tau identification rate and the misidentification rate from jets. The tool assigns a likelihood $W_\tau^{DELPHES}$ for PUPPI jets to originate from hadronically decaying taus. If this value is > 0.1, the jet is considered to be a tau object - the reconstruction efficiency is assumed to be a constant 45%. The misidentification rate is parametrized for all jets which do not fulfil the tau likelihood, dependent on the jet's transverse momentum. For high momentum jets $p_T^{jet} > 190$ GeV, the applied misidentification rate becomes constant around 0.1%. These reconstruction and misidentification probabilities are assumed to be within a factor of 2 compared to the Run-2 MVA based identification method.

Events which contain exactly one 'reconstructed' τ_h with $p_T^\tau > 80$ GeV and $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV are selected, assuming similar trigger performance as in 2016[3]. Events with a light lepton with sufficient transverse momentum $p_T^l > 25$ GeV are vetoed. The same kinematic selections $0.7 < p_T^\tau / p_T^{miss} < 1.3$ and $\Delta\Phi(\tau, p_T^{miss}) > 2.4$ are required. An overview of the selections is given in Tab. 7.1.

For this study, background samples were centrally produced at $\sqrt{s} = 14$ TeV for various backgrounds. The dominant W+Jets background is binned in number of jets (1, 2, 3, 4+). While these samples describe the on-shell spectrum very well, its statistical description in the high-m_{T} region is very sparse. In order to address this problem, a W+Jets sample, inclusive in number of jets, was produced where a requirement was set on the generated W boson mass $m_W > 400$ GeV. Overlap with the official W+Jets sample is removed based on the mass of the generated W boson. Signal samples for this study were also created for 8 different mass points ranging from 1 to 8 TeV in W' mass at LO using Madgraph. The estimated cross sections are given in Tab. 7.2.

The resulting m_{T} distribution for the DELPHES study is shown in Fig 7.10. It can be seen, that the relative contribution of background is expected to be increased for top based backgrounds in the sub-TeV region. The broad binning shown is chosen due to small statistics in the simulated background samples, which are already significantly improved by the privately produced high mass W sample. This is still a good estimation of the expected sensitivity towards W' searches, since little to no background is expected in the very high

[3] Although that specific trigger does not exist at the time of writing this thesis.

Mass $m_{W'}^{gen}$	$\sigma\mathcal{B}$ (pb)
1000 GeV	2807.2
2000 GeV	132.98
3000 GeV	14.690
4000 GeV	2.3607
5000 GeV	0.5131
6000 GeV	0.1606
7000 GeV	0.0705
8000 GeV	0.0369

TABLE 7.2: Cross sections times branching fraction at generated mass points of LO SSM W' production at $\sqrt{s} = 14$ TeV.

Type of Systematic Uncertainty	Value
Luminosity	1%
Jet Energy Scale	2.5%
Jet Energy Resolution	3%
$p_{\mathrm{T}}^{\mathrm{miss}}$ Energy Scale	2.5%
$p_{\mathrm{T}}^{\mathrm{miss}}$ Energy Resolution	3%
PDF	2.4%
Tau ID	2.5%
Tau Energy Scale	3%

TABLE 7.3: Systematic uncertainties assigned in the DELPHES based upgrade study, corresponding to the best estimate of what can be achieved at HL-LHC.

mass region. It can be seen, that even with the high amount of pileup, the contribution of misidentified backgrounds is rather small[4]. It has to be kept in mind, however, that these are taken from simulation and use an optimistic approach for estimation of the misidentification rate.

Systematic uncertainties originating from performance of the detector are difficult to estimate, since no comparisons with data can be made, especially under the high expected pileup conditions at HL-LHC. Therefore, a simplified approach is used, following a common recipe used between analyzes contributing to the Yellow Report [131]. Flat systematic uncertainties are assigned to the final m_{T} distribution, as given in Tab. 7.3. These uncertainties are reduced by 50% compared to current systematic uncertainties. This is effectively the best estimate of what can be achieved at HL-LHC. Compared to the uncertainty originating from limited samples size, these uncertainties are less significant.

Under these assumptions, the expected sensitivity and exclusion limits can be calculated with the same procedure as described in the beginning of this chapter 7.1. The results are shown in Fig. 7.11. In case that there exists a heavy charged vector boson with SSM like properties, it could be discovered with 5 σ for masses of up to 6.4 TeV in the $\tau + p_{\mathrm{T}}^{\mathrm{miss}}$ channel, while the 3000fb^{-1} would be enough to measure evidential significance (3σ) for SSM-like W' bosons with masses of up to 6.9 TeV. In case no significant difference from data can be found, the expected exclusion limit for such particles would improve to 7 TeV at 95% confidence level. For comparison, the exclusion limits are also provided for an integrated luminosity of 300fb^{-1}, which corresponds to the expected amount of data taken at LHC before the HL-LHC era. The sensitivity can be improved by roughly 1 TeV in W' mass, proving that the HL-LHC can provide sensitivity for high mass searches like this.

[4] Assuming that the reconstruction algorithms perform as described above

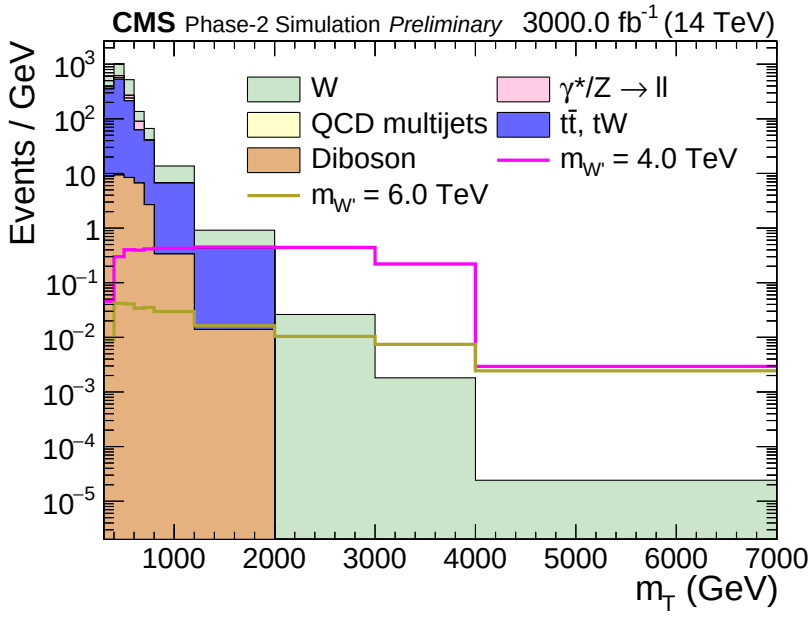

FIGURE 7.10: Distribution of m_T for the Yellow Report upgrade study performed with the detector parametrization tool DELPHES in the $\tau+p_T^{miss}$ final state for an expected average pileup of 200 and an integrated luminosity of 3 ab^{-1}. As in the Run-2 analysis, the dominant background originates from off-shell W production. Binning is chosen to accommodate the lack of statistics in the very high m_T region.

This sensitivity check can be expanded upon for the $g_{W\prime}/g_W$ coupling dependent case. Since this outlook analysis was made before the Run-2 analysis, the reweighing method was used to estimate the impact of a change in this coupling ratio parameter, instead of processing all signal samples through DELPHES. This was done for coupling values of [0.0001, 0.001, 0.01, 0.1, 1.0, 2.0, 3.0, 5.0]. The expected median cross section limit as well as the corresponding coupling strength dependent limit are shown in Fig 7.12. The two dimensional cross section distribution shows that the upper cross section exclusion limit roughly follows the shape of the signal efficiency, as shown in Fig. 6.3. The erratic behavior of the uncertainty bands originates from the low statistics with which this sensitivity study had to be performed.

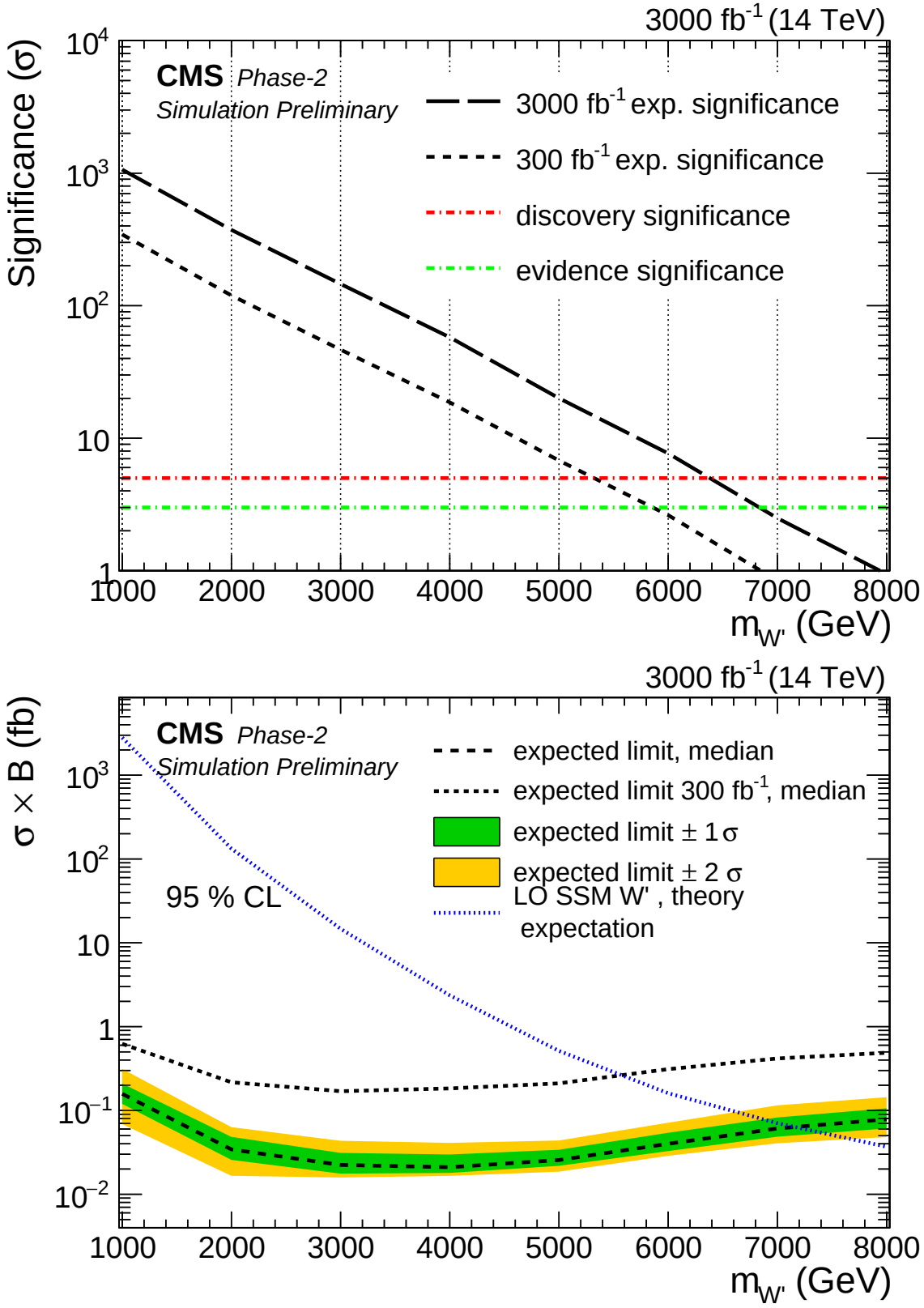

FIGURE 7.11: Expected sensitivity (top) and exclusion limit (bottom) possible in the $\tau+p_{\mathrm{T}}^{\mathrm{miss}}$ channel for the SSM W' model with 3000fb^{-1} of data at $\sqrt{s} = 14$ TeV. The expected amount of data would allow for a 5σ discovery of SSW like W' bosons with masses up to 6.4 TeV in this channel, given the aforementioned assumptions hold. In case no difference to the measured data is seen, the model will be excluded for W' masses up to 7 TeV at 95% CL.

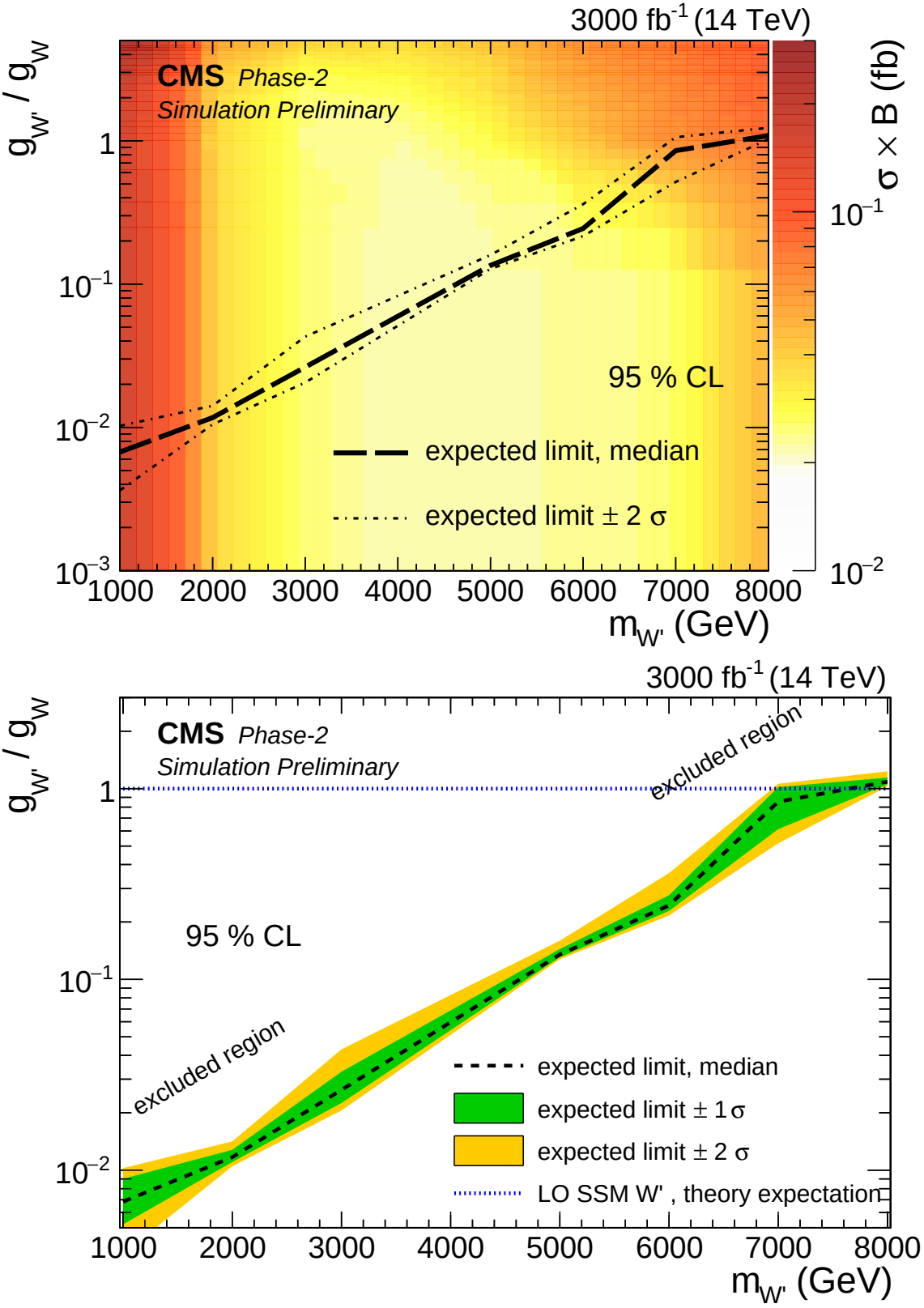

FIGURE 7.12: Expected exclusion limit for coupling dependent SSM-like production of a W' for the 3 ab^{-1} projection. On the top, the median expected cross section limit is shown in a color coded distribution, where lighter colors indicate a higher (i.e. worse) excluded cross section limit. On the bottom, the expected coupling dependent limit is shown, estimated in the same way as the Run-2 result.

Chapter 8

Summary

A search for new physics in events produced by proton-proton collisions has been presented. Events which contained a single hadronically decaying tau lepton and a significant amount of missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ were selected and analyzed with respect to different models. The two objects have been required to be balanced in their respective momenta and needed to be traveling into opposite directions. The expected background from Standard Model processes creating prompt τ_h was estimated using simulation, while background originating from misidentified tau objects has been identified using a data driven technique. Systematic uncertainties are taken into account - the leading uncertainty originates from uncertainties to parton distribution functions. This analysis has been performed for the three Run-2 $\sqrt{s} = 13$ TeV data sets recorded by the CMS collaboration in the years 2016, 2017 and 2018. The combined data set correspond to a total integrated luminosity of 138fb^{-1}. The analysis revealed no evidence for new physics. Therefore, for various models, exclusion limits have been calculated. Some of these models have no previous exclusion limits calculated in this search channel. A sequential standard model W', a heavy charged vector boson with analogous parameters to the standard model W boson has been excluded for masses between 0.6 and 4.7 TeV. A limit has been set on possible coupling ratio values of a W' w.r.t. the W boson couplings. In the non-universal interaction model, possible different branching fractions of a W' for decays to third generation leptons are possible. This model was excluded dependent on the mixing parameter $\cot(\theta_E)$, where masses of 2.2 TeV (4.8 TeV) have been excluded for values of $\cot(\theta_E) = 5.5$ ($\cot(\theta_E) = 1.0$). A lower exclusion limit has been provided for a quantum black hole model, excluding minimum threshold masses of singly positively charged quantum black holes of 6.6 TeV. A limit on an effective field theory, which translates the anomalous measurements of $R_D^{(*)}$ into a new physics hypotheses which would manifest at the TeV scale, is also provided dependent on the three different coupling types. To make reinterpretation of the presented results for other models which predict non-SM event rates in the $\tau + p_{\mathrm{T}}^{\mathrm{miss}}$ final state, a model-independent limit has also been provided.

In the last part, a projection of the analysis towards the high-luminosity run era of the LHC was performed. An analysis performed for an integrated luminosity of 3000fb^{-1} would be able to measure SSM like W' boson with masses of up to 6 TeV, assuming that reconstruction and misidenfication probabilities of hadronically decaying tau leptons are similar to current values. In case no difference will be found, masses of up to 7 TeV can be excluded at 95% confidence level in this best case scenario.

Appendix A

Working point selection study

In this analysis, tau obejcts are selected based on the output of the DNN based DeepTauID. Since this is a new algorithm, a small study was performed in order to estimate the differences between different working points of the ID - in this case Loose and Tight working points are compared. This study was performed using background simulation from 2018 as input and corresponding expected sensitivites are calculated for the SSM W' signal hypothesis. The results are shown in Fig. A.1. Only very small differences can be seen in this comparison. Therefore, the Tight working point of the ID was chosen, in order to ensure reduced misidentified background contribution even though this reduces overall signal efficiency slightly.

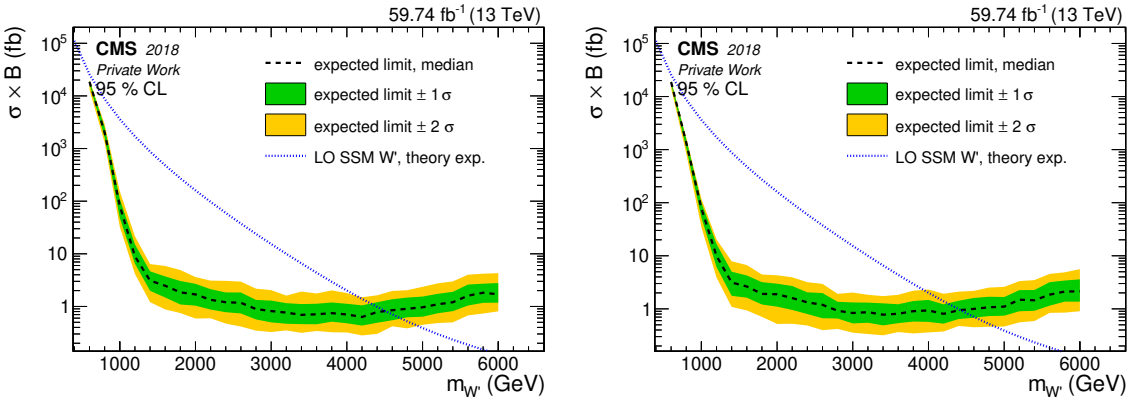

FIGURE A.1: Expected sensitivity for the SSM W' for 'Loose' (left) and 'Tight' (right) working points of the DeepTau ID for 2018. Only minor differences are visible.

Appendix B

Object kinematics for each year

This appendix shows kinematic variables of the τ_h and $p_{\mathrm{T}}^{\mathrm{miss}}$ objects for each individual year after full object and kinematic selections applied in the analysis. Distributions shown here show backgrounds from misidentified τ_h candidates estimated via the data-driven method. The kinematic variables can be checked for any hints of noticeable disagreements between experimental data and background expectation. The broad binning chosen for the m_{T} distribution is not analogously applied here. Therefore, the effect of statistical relics can be seen in some of the high momentum regions of the transverse momentum distributions of both τ_h (Figs. B.1,B.3,B.5) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (Figs. B.2,B.4,B.6). The one notable thing to mention is, that for 2016 a small overall normalization offset between experimental data and expected background can be seen. Taking into account the assigned uncertainties however, they are all within acceptable agreement. As discussed in Sec. 7.1, this normalization effect can be accounted for during the statistical evaluation of the results.

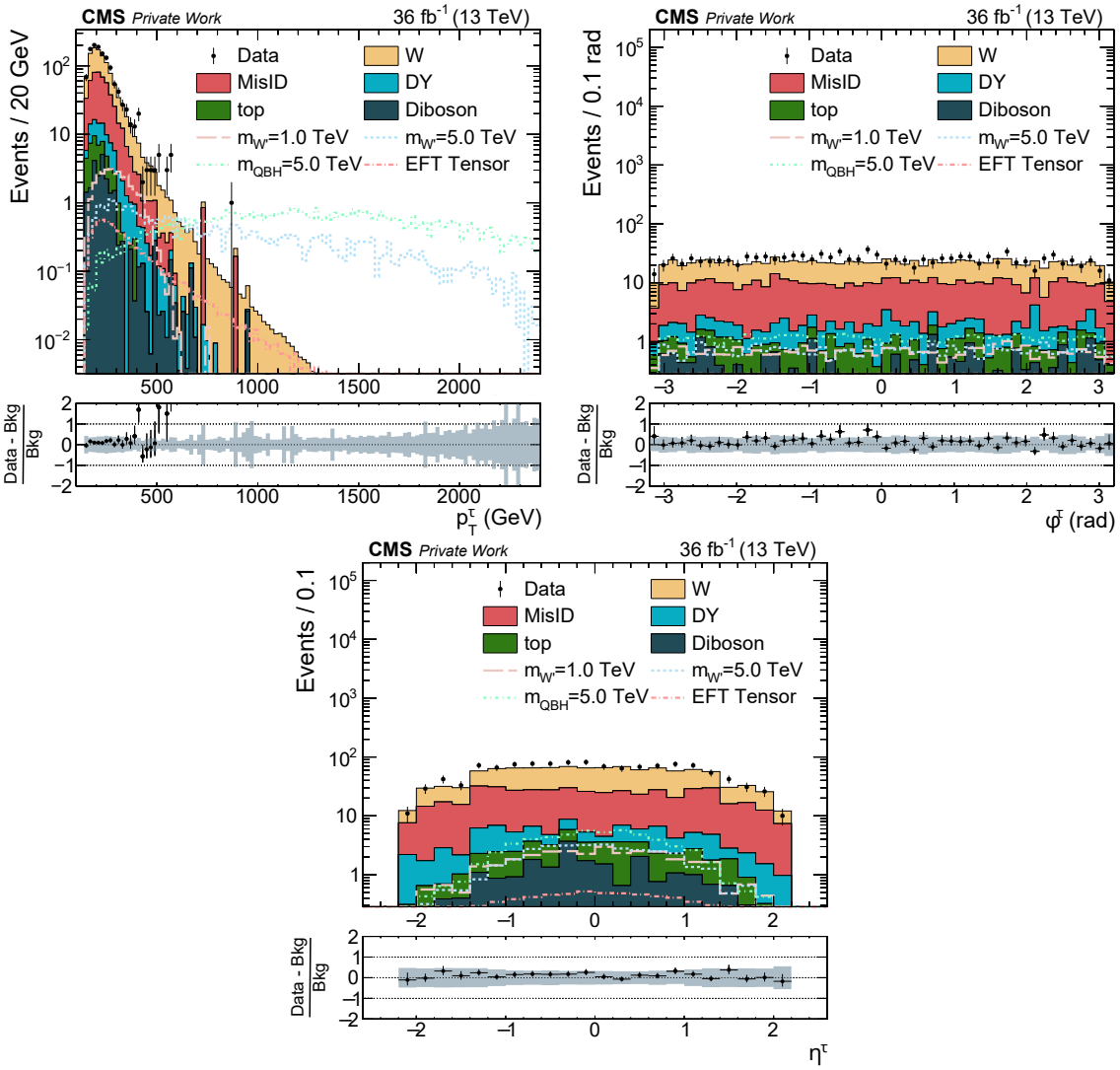

FIGURE B.1: Kinematic variables for τ_h candidates after the full event selection for 2016. The top left panel shows the transverse momentum of the candidate, the top right panel shows the Φ coordinate. On the bottom, the η coordinate is shown.

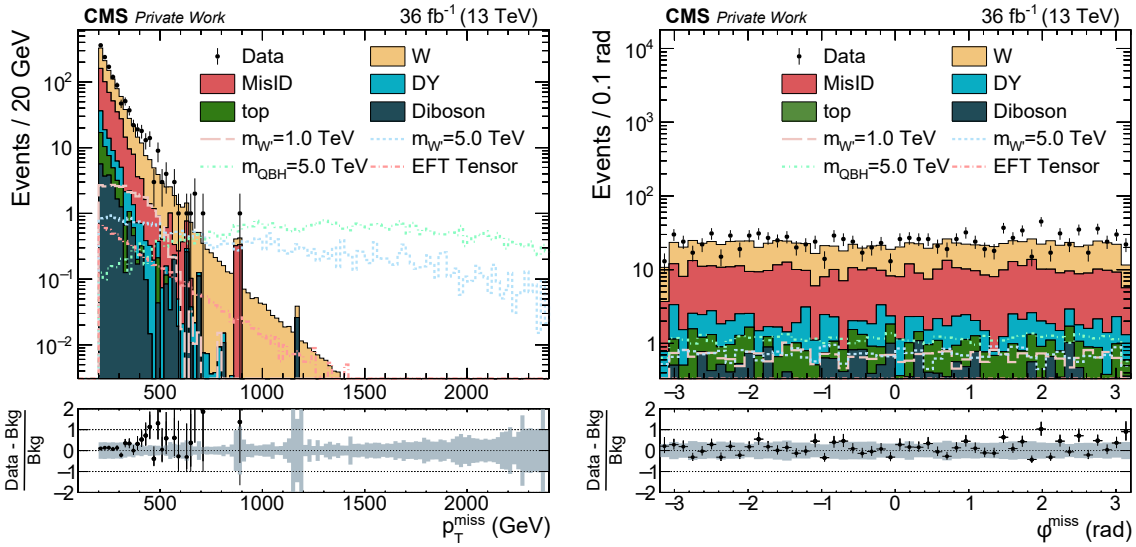

FIGURE B.2: Kinematic variables for $p_{\mathrm{T}}^{\mathrm{miss}}$ after the full event selection for 2016. The left panel shows the magnitude of the missing transverse momentum, the right panel shows its Φ coordinate.

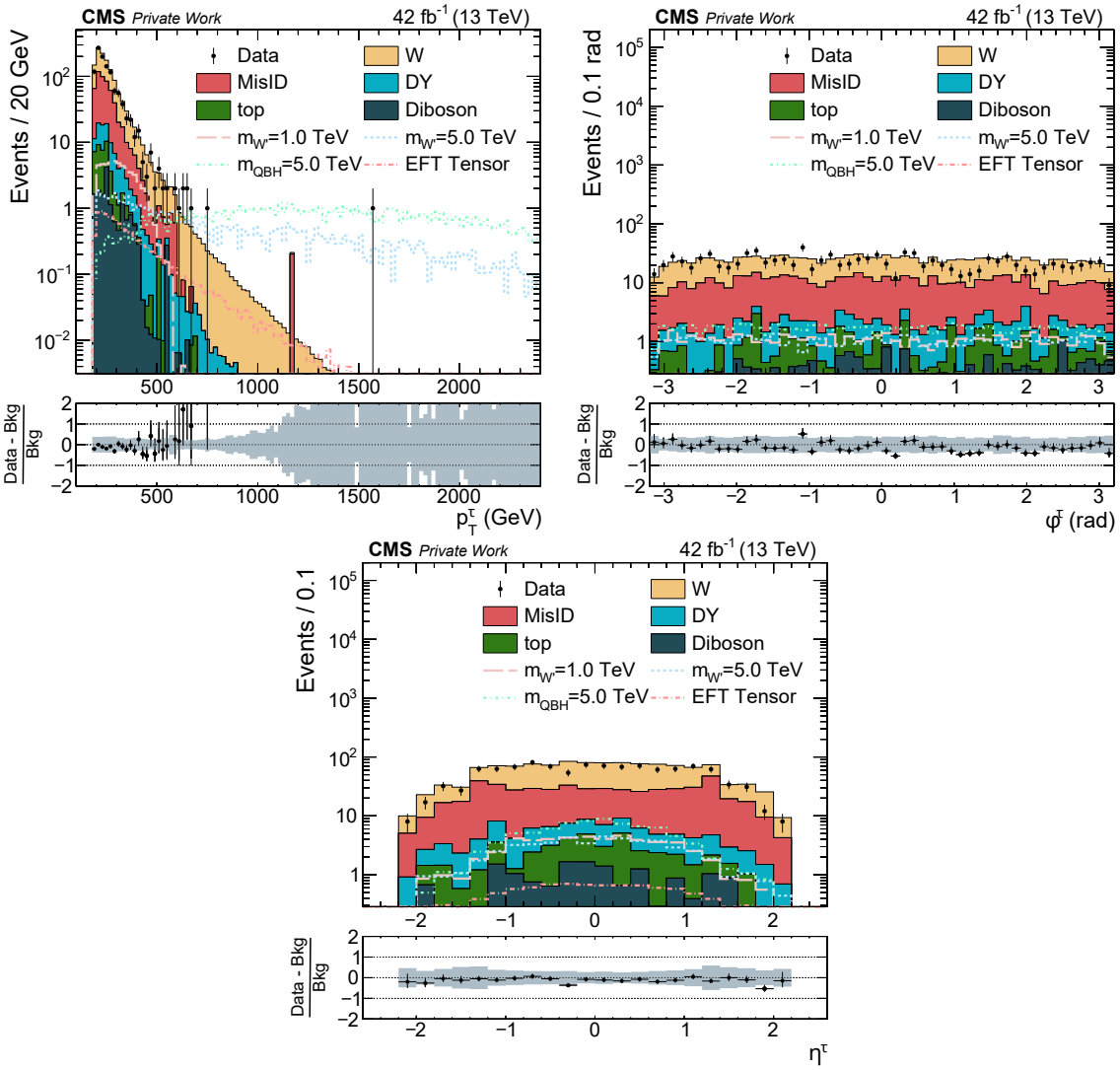

FIGURE B.3: Kinematic variables for τ_h candidates after the full event selection for 2017. The top left panel shows the transverse momentum of the candidate, the top right panel shows the Φ coordinate. On the bottom, the η coordinate is shown.

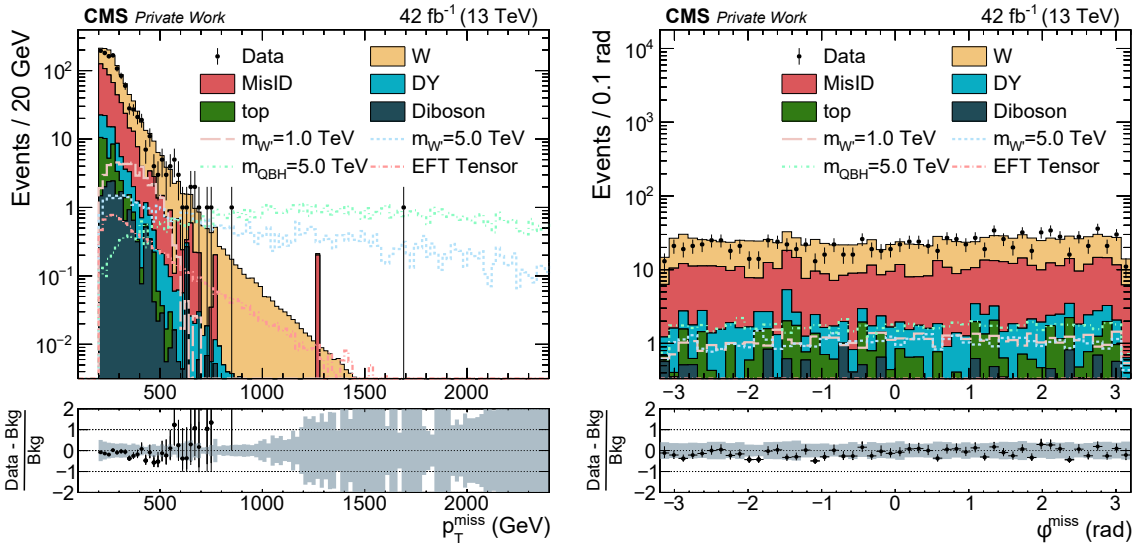

FIGURE B.4: Kinematic variables for $p_{\mathrm{T}}^{\mathrm{miss}}$ after the full event selection for 2017. The left panel shows the magnitude of the missing transverse momentum, the right panel shows its Φ coordinate.

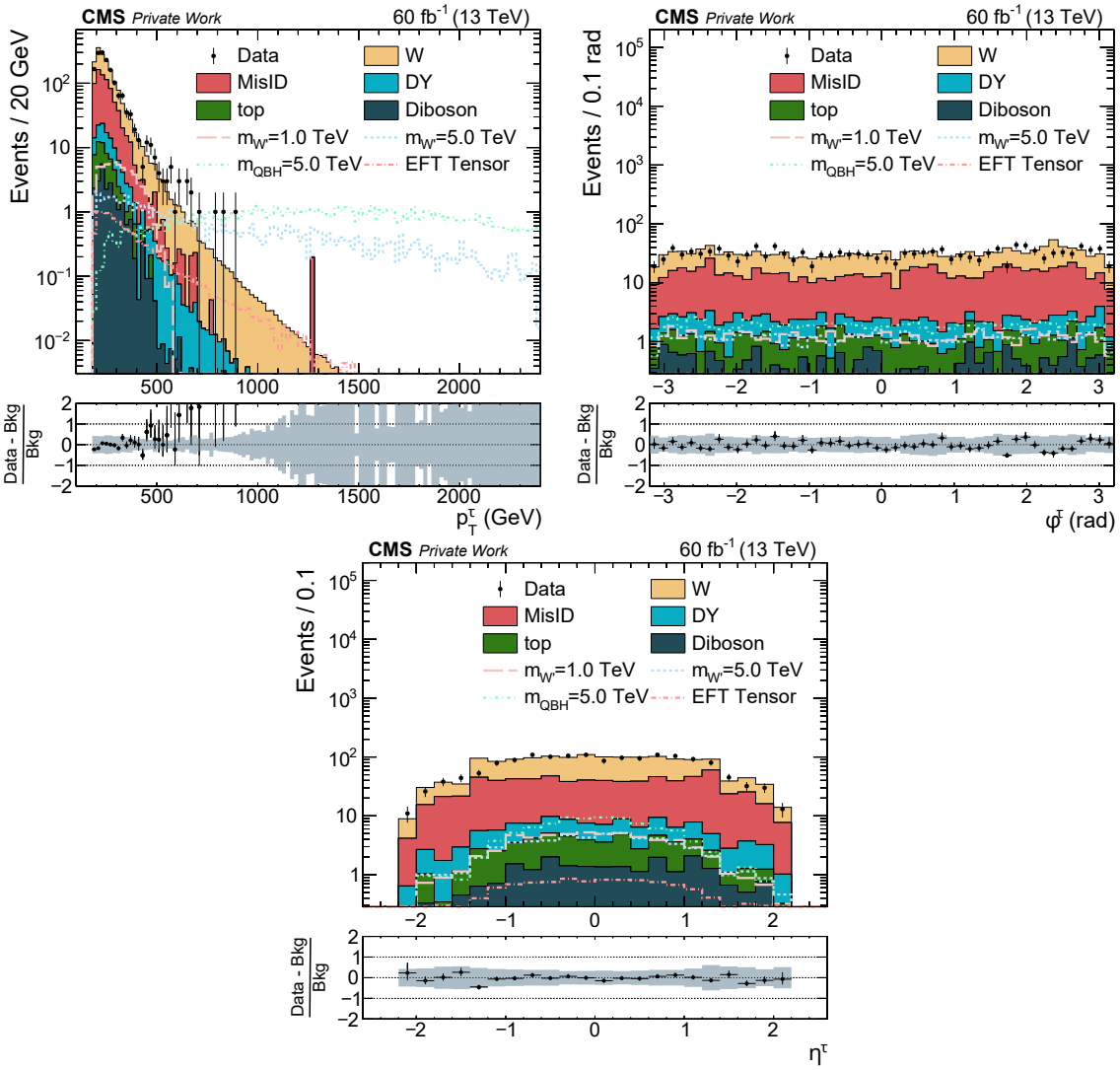

FIGURE B.5: Kinematic variables for τ_h candidates after the full event selection for 2018. The top left panel shows the transverse momentum of the candidate, the top right panel shows the Φ coordinate. On the bottom, the η coordinate is shown.

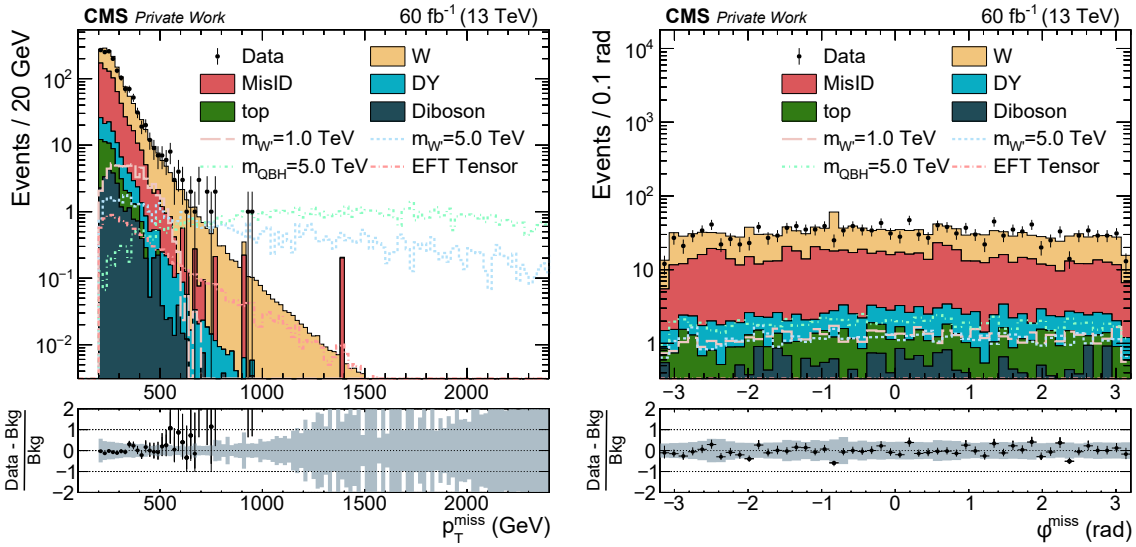

FIGURE B.6: Kinematic variables for $p_{\mathrm{T}}^{\mathrm{miss}}$ after the full event selection for 2018. The left panel shows the magnitude of the missing transverse momentum, the right panel shows its Φ coordinate.

Bibliography

[1] Georges Aad et al. 'Search for new phenomena in three- or four-lepton events in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector'. In: *Phys. Lett. B* 824 (2022), p. 136832. DOI: 10.1016/j.physletb.2021.136832. arXiv: 2107.00404 [hep-ex].

[2] Albert M Sirunyan et al. 'MUSiC: a model-unspecific search for new physics in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$'. In: *Eur. Phys. J. C* 81.7 (2021), p. 629. DOI: 10.1140/epjc/s10052-021-09236-z. arXiv: 2010.02984 [hep-ex].

[3] CMS Lumi POG. *CMS luminosity - public resulsts*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. 2022.

[4] G Ranucci et al. 'Overview and accomplishments of the Borexino experiment'. In: *Journal of Physics: Conference Series* 675.1 (Feb. 2016), p. 012036. DOI: 10.1088/1742-6596/675/1/012036. URL: https://doi.org/10.1088/1742-6596/675/1/012036.

[5] Christopher W. Walter. 'The Super-Kamiokande Experiment'. In: *Neutrino Oscillations* (Mar. 2008), pp. 19–43. DOI: 10.1142/9789812771971_0002. URL: http://dx.doi.org/10.1142/9789812771971_0002.

[6] J. P. Lees et al. 'Evidence for an Excess of $\bar{B} \to D(*)^+ + \tau^- \bar{\nu}_\tau$ Decays'. In: *Physical Review Letters* 109.10 (Sept. 2012). ISSN: 1079-7114. DOI: 10.1103/physrevlett.109.101802. URL: http://dx.doi.org/10.1103/PhysRevLett.109.101802.

[7] J. P. Lees et al. 'Measurement of an excess of $\bar{B} \to D(*)^+ \tau^- \bar{\nu}_\tau$ decays and implications for charged Higgs bosons'. In: *Physical Review D* 88.7 (Oct. 2013). ISSN: 1550-2368. DOI: 10.1103/physrevd.88.072012. URL: http://dx.doi.org/10.1103/PhysRevD.88.072012.

[8] M. Huschle et al. 'Measurement of the branching ratio of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle'. In: *Phys. Rev. D* 92.7 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. arXiv: 1507.03233 [hep-ex].

[9] R. Aaij et al. 'Measurement of the Ratio of Branching Fractions $\mathcal{B}(\bar{B}^0 \to D(*)^+ \tau^- \bar{\nu}_\tau)$ $/\mathcal{B}(\bar{B}^0 \to D(*)^+ + \mu^- \bar{\nu}_\mu)$'. In: *Physical Review Letters* 115.11 (Sept. 2015). ISSN: 1079-7114. DOI: 10.1103/physrevlett.115.111803. URL: http://dx.doi.org/10.1103/PhysRevLett.115.111803.

[10] S. Aoki et al. 'Review of lattice results concerning low-energy particle physics'. In: *Eur. Phys. J. C* 77.2 (2017), p. 112. DOI: 10.1140/epjc/s10052-016-4509-7. arXiv: 1607.00299 [hep-lat].

[11] P.A. Zyla et al. 'Review of Particle Physics'. In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[12] A. Tumasyan et al. 'Search for new physics in the τ lepton plus missing transverse momentum final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: (Dec. 2022). arXiv: 2212.12604 [hep-ex].

[13] Gudrun Hiller, Dennis Loose and Ivan Nišandžić. 'Flavorful leptoquarks at hadron colliders'. In: *Phys. Rev. D* 97.7 (2018), p. 075004. DOI: 10.1103/PhysRevD.97.075004. arXiv: 1801.09399 [hep-ph].

[14] Michael J. Baker et al. 'High- p_T signatures in vector–leptoquark models'. In: *Eur. Phys. J. C* 79.4 (2019), p. 334. DOI: 10.1140/epjc/s10052-019-6853-x. arXiv: 1901.10480 [hep-ph].

[15] Lorenzo Calibbi, Andreas Crivellin and Tianjun Li. 'Model of vector leptoquarks in view of the B-physics anomalies'. In: *Phys. Rev. D* 98.11 (2018), p. 115002. DOI: 10.1103/PhysRevD.98.115002. arXiv: 1709.00692 [hep-ph].

[16] P.W. Higgs. 'Broken symmetries, massless particles and gauge fields'. In: *Physics Letters* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)91136-9. URL: https://www.sciencedirect.com/science/article/pii/0031916364911369.

[17] F. Englert and R. Brout. 'Broken Symmetry and the Mass of Gauge Vector Mesons'. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[18] CMS Collaboration. 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC'. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[19] ATLAS Collaboration. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[20] G. W. Bennett et al. 'Final report of the E821 muon anomalous magnetic moment measurement at BNL'. In: *Physical Review D* 73.7 (Apr. 2006). DOI: 10.1103/physrevd.73.072003. URL: https://doi.org/10.1103%2Fphysrevd.73.072003.

[21] B. Abi et al. 'Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm'. In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 141801. DOI: 10.1103/PhysRevLett.126.141801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.141801.

[22] Andy Buckley et al. 'General-purpose event generators for LHC physics'. In: *Physics Reports* 504.5 (July 2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. URL: https://doi.org/10.1016%2Fj.physrep.2011.03.005.

[23] H. Abramowicz et al. 'Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data'. In: *Eur. Phys. J. C* 75.12 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[24] Richard D. Ball et al. 'Parton distributions from high-precision collider data'. In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[25] Jon Butterworth et al. 'PDF4LHC recommendations for LHC Run II'. In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[26] SNO Collaboration. 'Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory'. In: *Physical Review Letters* 89.1 (June 2002). ISSN: 1079-7114. DOI: 10.1103/physrevlett.89.011301. URL: http://dx.doi.org/10.1103/PhysRevLett.89.011301.

[27] J. P. Lees et al. 'Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays'. In: *Phys. Rev. Lett.* 109 (2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802. arXiv: 1205.5442 [hep-ex].

[28] J. P. Lees et al. 'Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons'. In: *Phys. Rev. D* 88.7 (2013), p. 072012. DOI: 10.1103/PhysRevD.88.072012. arXiv: 1303.0571 [hep-ex].

[29] Roel Aaij et al. 'Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)$ $/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$'. In: *Phys. Rev. Lett.* 115.11 (2015). [Erratum: Phys.Rev.Lett. 115, 159901 (2015)], p. 111803. DOI: 10.1103/PhysRevLett.115.111803. arXiv: 1506.08614 [hep-ex].

[30] R. Aaij et al. 'Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong τ-lepton decays'. In: *Phys. Rev. Lett.* 120.17 (2018), p. 171802. DOI: 10.1103/PhysRevLett.120.171802. arXiv: 1708.08856 [hep-ex].

[31] R. Aaij et al. 'Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays'. In: *Phys. Rev. D* 97.7 (2018), p. 072013. DOI: 10.1103/PhysRevD.97.072013. arXiv: 1711.02505 [hep-ex].

[32] Roel Aaij et al. 'Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays'. In: *Phys. Rev. Lett.* 122.19 (2019), p. 191801. DOI: 10.1103/PhysRevLett.122.191801. arXiv: 1903.09252 [hep-ex].

[33] R. Aaij et al. 'Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays'. In: *JHEP* 08 (2017), p. 055. DOI: 10.1007/JHEP08(2017)055. arXiv: 1705.05802 [hep-ex].

[34] Roel Aaij et al. 'Test of lepton universality in beauty-quark decays'. In: *Nature Phys.* 18.3 (2022), pp. 277–282. DOI: 10.1038/s41567-021-01478-8. arXiv: 2103.11769 [hep-ex].

[35] Roel Aaij et al. 'Tests of lepton universality using $B^0 \to K^0_S\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays'. In: *Phys. Rev. Lett.* 128.19 (2022), p. 191802. DOI: 10.1103/PhysRevLett.128.191802. arXiv: 2110.09501 [hep-ex].

[36] S. Hirose et al. 'Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$'. In: *Phys. Rev. Lett.* 118.21 (2017), p. 211801. DOI: 10.1103/PhysRevLett.118.211801. arXiv: 1612.00529 [hep-ex].

[37] S. Hirose et al. 'Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle'. In: *Phys. Rev. D* 97.1 (2018), p. 012004. DOI: 10.1103/PhysRevD.97.012004. arXiv: 1709.00129 [hep-ex].

[38] F. Zwicky. 'Die Rotverschiebung von extragalaktischen Nebeln'. In: *Helvetica Physica Acta* 6 (Jan. 1933), pp. 110–127.

[39] Yoshiaki Sofue and Vera Rubin. 'Rotation Curves of Spiral Galaxies'. In: *Annual Review of Astronomy and Astrophysics* 39.1 (Sept. 2001), pp. 137–174. ISSN: 1545-4282. DOI: 10.1146/annurev.astro.39.1.137. URL: http://dx.doi.org/10.1146/annurev.astro.39.1.137.

[40] Guido Altarelli, Barbara Melé and M Ruiz-Altaba. 'Searching for new heavy vector bosons in $p\overline{p}$ colliders'. In: *Z. Phys. C* 45 (Mar. 1989), 109–121. 46 p. DOI: 10.1007/BF01552335. URL: https://cds.cern.ch/record/197730.

[41] Elena Accomando et al. 'Interference effects in heavy W'-boson searches at the LHC'. In: *Physical Review D* 85.11 (June 2012). DOI: 10.1103/physrevd.85.115017. URL: https://doi.org/10.1103%2Fphysrevd.85.115017.

[42] E. Boos et al. 'Interference between W' and W in single-top quark production processes'. In: *Physics Letters B* 655.5-6 (Nov. 2007), pp. 245–250. DOI: 10.1016/j.physletb.2007.03.064. URL: https://doi.org/10.1016%2Fj.physletb.2007.03.064.

[43] Vardan Khachatryan et al. 'Search for physics beyond the standard model in final states with a lepton and missing transverse energy in proton-proton collisions at sqrt(s) = 8 TeV'. In: *Phys. Rev. D* 91.9 (2015), p. 092005. DOI: 10.1103/PhysRevD.91.092005. arXiv: 1408.2745 [hep-ex].

[44] K. Padeken. 'Search for new physics in the tau plus missing energy final states at CMS'. In: *Publication Server of RWTH Aachen University* (2017). DOI: 10.18154/RWTH-2017-04198.

[45] Cheng-Wei Chiang et al. 'The Family $SU(2)_l$ x $SU(2)_h$ x $U(1)$ Model'. In: *Phys. Rev. D* 81 (2010), p. 015006. DOI: 10.1103/PhysRevD.81.015006. arXiv: 0911.1480 [hep-ph].

[46] Lisa Edelhäuser and Alexander Knochel. 'Observing nonstandard W' and Z' through the third generation and Higgs lens'. In: (Aug. 2014). arXiv: 1408.0914 [hep-ph].

[47] Nima Arkani-Hamed et al. 'Infinitely Large New Dimensions'. In: *Physical Review Letters* 84.4 (Jan. 2000), pp. 586–589. DOI: 10.1103/physrevlett.84.586. URL: https://doi.org/10.1103%2Fphysrevlett.84.586.

[48] Savas Dimopoulos and Greg L. Landsberg. 'Black holes at the LHC'. In: *Phys. Rev. Lett.* 87 (2001), p. 161602. DOI: 10.1103/PhysRevLett.87.161602. arXiv: hep-ph/0106295.

[49] Steven B. Giddings and Scott D. Thomas. 'High-energy colliders as black hole factories: The End of short distance physics'. In: *Phys. Rev. D* 65 (2002), p. 056010. DOI: 10.1103/PhysRevD.65.056010. arXiv: hep-ph/0106219.

[50] V. M. et al. Abazov. 'Search for Large Extra Dimensions via Single Photon plus Missing Energy Final States at $\sqrt{s} = 1.96$ TeV'. In: *Phys. Rev. Lett.* 101 (1 June 2008), p. 011601. DOI: 10.1103/PhysRevLett.101.011601. URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.011601.

[51] Luis A. Anchordoqui et al. 'Updated limits on TeV-scale gravity from the absence of neutrino cosmic ray showers mediated by black holes'. In: *Phys. Rev. D* 68 (10 Nov. 2003), p. 104025. DOI: 10.1103/PhysRevD.68.104025. URL: https://link.aps.org/doi/10.1103/PhysRevD.68.104025.

[52] Serguei Chatrchyan et al. 'Search for Large Extra Dimensions in the Diphoton Final State at the Large Hadron Collider'. In: *JHEP* 05 (2011), p. 085. DOI: 10.1007/JHEP05(2011)085. arXiv: 1103.4279 [hep-ex].

[53] Patrick Meade and Lisa Randall. 'Black Holes and Quantum Gravity at the LHC'. In: *JHEP* 05 (2008), p. 003. DOI: 10.1088/1126-6708/2008/05/003. arXiv: 0708.3017 [hep-ph].

[54] K.S. Thorne. 'Nonspherical Gravitational Collapse–A Short Review'. In: *Magic Without Magic: John Archibald Wheeler*. Klauder, John R., 1972, p. 231.

[55] Admir Greljo, Jorge Martin Camalich and José David Ruiz-Álvarez. 'Mono-τ Signatures at the LHC Constrain Explanations of B-decay Anomalies'. In: *Phys. Rev. Lett.* 122.13 (2019), p. 131803. DOI: 10.1103/PhysRevLett.122.131803. arXiv: 1811.07920 [hep-ph].

[56] Lyndon Evans and Philip Bryant. 'LHC Machine'. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088/1748-0221/3/08/s08001.

[57] CMS Collaboration. 'The CMS experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[58] CMS Collaboration. *CMS detector web page, Image taken from https://cms.cern/news/cms-detector-design on 11.01.22*. 2022.

[59] The Tracker Group of the CMS Collaboration. *The CMS Phase-1 Pixel Detector Upgrade*. 2020. arXiv: 2012.14304 [physics.ins-det].

[60] 'The CMS electromagnetic calorimeter project: Technical Design Report'. In: (1997). URL: https://cds.cern.ch/record/349375.

[61] CMS Collaboration. 'Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data'. In: *Journal of Instrumentation* 5.03 (Mar. 2010), T03012–T03012. DOI: 10.1088/1748-0221/5/03/t03012. URL: https://doi.org/10.1088%2F1748-0221%2F5%2F03%2Ft03012.

[62] A.M. Sirunyan et al. 'Particle-flow reconstruction and global event description with the CMS detector'. In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/10/p10003. URL: http://dx.doi.org/10.1088/1748-0221/12/10/P10003.

[63] 'CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter'. In: (Sept. 2012). Ed. by J. Mans et al. DOI: 10.2172/1151651.

[64] CMS Collaboration. 'Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV'. In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. DOI: 10.1088/1748-0221/13/06/p06015. URL: https://doi.org/10.1088%2F1748-0221%2F13%2F06%2Fp06015.

[65] CMS Collaboration. 'The CMS trigger system'. In: *Journal of Instrumentation* 12.01 (Jan. 2017), P01020–P01020. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/01/p01020. URL: http://dx.doi.org/10.1088/1748-0221/12/01/P01020.

[66] 'CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger'. In: (Dec. 2002). Ed. by P. Sphicas. URL: http://cds.cern.ch/record/578006.

[67] Matteo Cacciari, Gavin P Salam and Gregory Soyez. 'The anti-ktjet clustering algorithm'. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[68] CMS Collaboration. 'Performance of τ-lepton reconstruction and identification in CMS'. In: *Journal of Instrumentation* 7.01 (Jan. 2012), P01001–P01001. DOI: 10.1088/1748-0221/7/01/p01001. URL: https://doi.org/10.1088/1748-0221/7/01/p01001.

[69] 'Reconstruction and identification of τ lepton decays to hadrons and $\nu\tau$ at CMS'. In: *Journal of Instrumentation* 11.01 (Jan. 2016), P01019–P01019. DOI: 10.1088/1748-0221/11/01/p01019. URL: https://doi.org/10.1088/1748-0221/11/01/p01019.

[70] 'Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2'. In: (2016). URL: https://cds.cern.ch/record/2196972.

[71] CMS Collaboration. 'Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s}$=13 TeV'. In: *Journal of Instrumentation* 13.10 (Oct. 2018), P10005–P10005. DOI: 10.1088/1748-0221/13/10/p10005. URL: https://doi.org/10.1088/1748-0221/13/10/p10005.

[72] Armen Tumasyan et al. 'Identification of hadronic tau lepton decays using a deep neural network'. In: (Jan. 2022). arXiv: 2201.08458 [hep-ex].

[73] E. Bols et al. 'Jet flavour classification using DeepJet'. In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012. URL: http://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[74] Y. Lecun et al. 'Gradient-based learning applied to document recognition'. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

[75] CMS Collaboration. 'Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 13 TeV using the CMS detector'. In: *Journal of Instrumentation* 14.07 (July 2019), P07004–P07004. DOI: 10.1088/1748-0221/14/07/p07004. URL: https://doi.org/10.1088%2F1748-0221%2F14%2F07%2Fp07004.

[76] CMS Collaboration. 'Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV'. In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. DOI: 10.1088/1748-0221/12/02/p02014. URL: https://doi.org/10.1088%2F1748-0221%2F12%2F02%2Fp02014.

[77] CMS Members. *cms-nanoAOD*. https://github.com/cms-nanoAOD/nanoAOD-tools. 2022.

[78] Albert M Sirunyan et al. 'Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC'. In: *JINST* 16.05 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[79] W Adam et al. 'Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC'. In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (July 2005), N9–N20. DOI: 10.1088/0954-3899/31/9/n01. URL: https://doi.org/10.1088%2F0954-3899%2F31%2F9%2Fn01.

[80] CMS Members. *HEEP ID Run-2*. https://twiki.cern.ch/twiki/bin/view/CMS/HEEPElectronIdentificationRun2. CMS Internal. 2022.

[81] CMS Members. *Cut based Electron Identification Run-2*. https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2. CMS Internal. 2022.

[82] R. Frühwirth. 'Application of Kalman filtering to track and vertex fitting'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: https://www.sciencedirect.com/science/article/pii/0168900287908874.

[83] CMS Collaboration. 'Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: *Journal of Instrumentation* 15.02 (Feb. 2020), P02027. DOI: 10.1088/1748-0221/15/02/P02027. URL: https://dx.doi.org/10.1088/1748-0221/15/02/P02027.

[84] Physics Performance and Dataset Group. *New "Physics Performance and Dataset" Group at CMS*. https://cms.cern/news/new-physics-performance-and-dataset-group-cms. 2012. URL: https://cms-info.web.cern.ch/coordination/physics-performance-datasets-ppd/.

[85] Albert M Sirunyan et al. 'Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: *JINST* 15.10 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[86] Armen Tumasyan et al. 'Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: *JHEP* 11 (2021), p. 153. DOI: 10.1007/JHEP11(2021)153. arXiv: 2107.13021 [hep-ex].

[87] Torbjörn Sjöstrand et al. 'An introduction to PYTHIA 8.2'. In: *Comp. Phys. Comm.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[88] Johan Alwall et al. 'MadGraph 5: going beyond'. In: *JHEP* 06 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[89] Paolo Nason. 'A New method for combining NLO QCD with shower Monte Carlo algorithms'. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[90] Stefano Frixione, Paolo Nason and Carlo Oleari. 'Matching NLO QCD computations with Parton Shower simulations: the POWHEG method'. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[91] Simone Alioli et al. 'A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX'. In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[92] Peter Skands, Stefano Carrazza and Juan Rojo. 'Tuning PYTHIA 8.1: the Monash 2013 Tune'. In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].

[93] Albert M Sirunyan et al. 'Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements'. In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[94] S. Agostinelli et al. 'Geant4—a simulation toolkit'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[95] CMS Members. *cms-nanoAOD*. https://github.com/cms-nanoAOD. 2022.

[96] G. Altarelli and G. Parisi. 'Asymptotic freedom in parton language'. In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(77)90384-4. URL: https://www.sciencedirect.com/science/article/pii/0550321377903844.

[97] A. Accardi et al. 'A Critical Appraisal and Evaluation of Modern PDFs'. In: *Eur. Phys. J. C* 76.8 (2016), p. 471. DOI: 10.1140/epjc/s10052-016-4285-4. arXiv: 1603.08906 [hep-ph].

[98] CMS Collaboration. *Search for new physics in the lepton plus missing transverse momentum final state in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*. 2022. DOI: 10.48550/ARXIV.2202.06075. URL: https://arxiv.org/abs/2202.06075.

[99] Douglas M. Gingrich. 'Quantum black holes with charge, colour, and spin at the LHC'. In: *J. Phys.* G37 (2010), p. 105008. DOI: 10.1088/0954-3899/37/10/105008. arXiv: 0912.0826 [hep-ph].

[100] J. Pumplin et al. 'New generation of parton distributions with uncertainties from global QCD analysis'. In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195.

[101] CMS Collaboration. *Search for heavy resonances and quantum black holes in* $e\mu$*,* $e\tau$*, and* $\mu\tau$ *final states in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*. 2022. DOI: 10.48550/ARXIV.2205.06709. URL: https://arxiv.org/abs/2205.06709.

[102] V.M. Abazov et al. 'A novel method for modeling the recoil in W boson events at hadron colliders'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 609.2 (2009), pp. 250–262. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.08.056. URL: https://www.sciencedirect.com/science/article/pii/S0168900209016623.

[103] Ryan Gavin et al. *W physics at the LHC with FEWZ 2.1*. 2012. arXiv: 1201.5896 [hep-ph].

[104] Sergey G. Bondarenko and Andrey A. Sapronov. 'NLO EW and QCD proton-proton cross section calculations with mcsanc-v1.01'. In: *Comput. Phys. Commun.* 184 (2013), pp. 2343–2350. DOI: 10.1016/j.cpc.2013.05.010. arXiv: 1301.3687 [hep-ph].

[105] A. Arbuzov et al. 'Update of the MCSANC Monte Carlo integrator, v. 1.20'. In: *JETP Letters* 103.2 (Jan. 2016), pp. 131–136. ISSN: 1090-6487. DOI: 10.1134/s0021364016020041. URL: http://dx.doi.org/10.1134/S0021364016020041.

[106] CMS Tau POG. *Tau Trigger Scale Factor Github*. https://github.com/cms-tau-pog/TauTriggerSFs. 2022.

[107] M Meyer and A. Raspereza. *Measurement of the high p_T single-tau trigger efficiency with W*→ $\tau\nu$ events*. https://indico.cern.ch/event/982713/contributions/4138615/attachments/2162407/3653661/TauId_SingleTauTrigger.pdf. CMS Internal. 2020.

[108] 'Jet→Tau Misidentification Probability with 2016 pp collision data at $\sqrt{s}$=13 TeV'. In: (Aug. 2017). URL: https://cds.cern.ch/record/2282910.

[109] The CMS collaboration. 'Performance of CMS muon reconstruction in pp collision events at $\sqrt{s}$ = 7 TeV'. In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002–P10002. ISSN: 1748-0221. DOI: 10.1088/1748-0221/7/10/p10002. URL: http://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[110] CMS Collaboration. 'Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: *Journal of Instrumentation* 15.02 (Feb. 2020), P02027–P02027. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/02/p02027. URL: http://dx.doi.org/10.1088/1748-0221/15/02/P02027.

[111] D. S. Sivia and J. Skilling. *Data Analysis - A Bayesian Tutorial*. 2nd. Oxford Science Publications. Oxford University Press, 2006.

[112] G. Cowan, "Statistics", Ch. 39 in Particle Data Group, C. Patrignani et al. 'Review of particle physics'. In: *Chin. Phys. C* 40 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001. URL: http://pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf.

[113] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11. 2011. URL: https://cds.cern.ch/record/1379837.

[114] Lorenzo Moneta et al. *The RooStats Project*. 2010. DOI: 10.48550/ARXIV.1009.1003. URL: https://arxiv.org/abs/1009.1003.

[115] J. S. Conway. 'Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra'. In: *PHYSTAT 2011*. 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[116] CMS Collaboration. 'Search for a W′ boson decaying to a τ lepton and a neutrino in proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: *Physics Letters B* 792 (May 2019), pp. 107–131. DOI: 10.1016/j.physletb.2019.01.069. URL: https://doi.org/10.1016%2Fj.physletb.2019.01.069.

[117] ATLAS Collaboration. 'Search for High-Mass Resonances Decaying to $\tau\nu$ in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector'. In: *Physical Review Letters* 120.16 (Apr. 2018). DOI: 10.1103/physrevlett.120.161802. URL: https://doi.org/10.1103%2Fphysrevlett.120.161802.

[118] 'Search for high-mass resonances in final states with a tau lepton and missing transverse momentum with the ATLAS detector'. In: (June 2021). URL: https://cds.cern.ch/record/2773301.

[119] 'High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1'. In: 4/2017 (2017). Ed. by G. Apollinari et al. DOI: 10.23731/CYRM-2017-004.

[120] Andrea Dainese et al., eds. *Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC*. Vol. 7/2019. CERN Yellow Reports: Monographs. Geneva, Switzerland: CERN, 2019. ISBN: 978-92-9083-549-3. DOI: 10.23731/CYRM-2019-007.

[121] CMS Collaboration. 'Sensitivity study for a heavy gauge boson W' in the decay channel with a tau lepton and a neutrino at the High-Luminosity LHC'. In: (2019). URL: http://cds.cern.ch/record/2655312.

[122] L Rossi and O Brüning. 'Introduction to the HL-LHC Project'. In: *Adv. Ser. Dir. High Energy Phys.* 24 (2015), pp. 1–17. DOI: 10.1142/9789814675475_0001. URL: https://cds.cern.ch/record/2130736.

[123] 'High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report'. In: 10/2020 (2020). Ed. by I. Béjar Alonso et al. DOI: 10.23731/CYRM-2020-0010.

[124] D Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. Geneva, June 2015. DOI: 10.17181/CERN.VU8I.D59J. URL: https://cds.cern.ch/record/2020886.

[125] Armen Tumasyan et al. 'The Phase-2 Upgrade of the CMS Tracker'. In: (June 2017). Ed. by K. Klein. DOI: 10.17181/CERN.QZ28.FLHW.

[126] Thomas Hebbeker and Andrey Korytov. 'The Phase-2 Upgrade of the CMS Muon Detectors'. In: (Sept. 2017). URL: https://cds.cern.ch/record/2283189.

[127] *The Phase-2 Upgrade of the CMS Barrel Calorimeters*. Tech. rep. This is the final version, approved by the LHCC. Geneva: CERN, Sept. 2017. URL: https://cds.cern.ch/record/2283187.

[128] 'The Phase-2 Upgrade of the CMS Endcap Calorimeter'. In: (2017). URL: https://cds.cern.ch/record/2293646.

[129] The CMS Collaboration. *Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC*. Tech. rep. Geneva: CERN, Dec. 2018. URL: https://cds.cern.ch/record/2650976.

[130] J. de Favereau et al. 'DELPHES 3: a modular framework for fast simulation of a generic collider experiment'. In: *Journal of High Energy Physics* 2014.2 (Feb. 2014). DOI: 10.1007/jhep02(2014)057. URL: https://doi.org/10.1007%2Fjhep02%282014%29057.

[131] 'Projected performance of Higgs analyses at the HL-LHC for ECFA 2016'. In: (2017). URL: https://cds.cern.ch/record/2266165.