Karlsruher Institut für Technologie

$H \rightarrow \tau^+\tau^-$ cross section measurement at high p_T with CMS data of proton-proton collisions at the Large Hadron Collider at CERN

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

angenommene
Dissertation

von
M.Sc. Olha Lavoryk
aus Rivne (Ukraine)

Tag der mündlichen Prüfung: 31. Oktober 2025
1. Referent: Prof. Dr. Markus Klute
2. Referent: Priv. Doz. Dr. Roger Wolf

Declaration of Originality

I hereby declare that the work presented in this thesis is my own original work. All sources and references that have been used or quoted are explicitly indicated in the text and listed in the bibliography.

I further confirm that I have not used any other sources or aids apart from those referenced. Large Language Models (LLMs) have been used solely to improve the grammar, style, and clarity of the text, but not to generate original scientific content.

Karlsruhe, 22.09.2025

Olha Lavoryk

Abstract

The discovery of the Higgs boson has opened a new era in precision tests of the Standard Model, with particular focus on its couplings to fermions. Among these, the coupling to the τ lepton plays a key role due to its large Yukawa interaction strength. The boosted $H \rightarrow \tau\tau$ final state provides a unique opportunity to probe Higgs boson properties in the high-momentum regime, but also presents significant experimental challenges in terms of its reconstruction.

This thesis presents an analysis of the boosted $H \rightarrow \tau\tau$ channel using proton-proton collision data collected by the CMS detector during the data-taking year of 2018 of the LHC Run 2 at $\sqrt{s} = 13$ TeV. A machine-learning-based approach, ParticleNet, was employed for the identification of hadronically decaying τ leptons in boosted topologies, treating them as substructure of a large-radius jets. To reconstruct the di-τ mass, the collinear mass approximation was applied, providing a discriminating observable in the presence of invisible neutrinos from the τ decays.

The performance of ParticleNet was validated against collision data and demonstrated satisfying agreement, significantly improving identification of boosted τ. The collinear mass approximation provided strong discriminating power between signal and background, though reconstructed mass peaks were observed to shift relative to their true values. A conservative 10% uncertainty on boosted τ identification efficiency was assumed, leaving room for improvement through more precise calibrations.

The methodologies developed in this work are broadly applicable to other searches involving boosted τ leptons, including $Z' \rightarrow \tau\tau$ and Higgs-sector extensions such as the MSSM and NMSSM. With the larger datasets from Run 3 and the High-Luminosity LHC, along with advancements in detector technology and machine-learning techniques, these methods are expected to play a central role in precision Higgs physics and searches for physics beyond the Standard Model.

Zusammenfassung

Die Entdeckung des Higgs-Bosons hat eine neue Ära präziser Tests des Standardmodells eröffnet, wobei insbesondere seine Kopplungen an Fermionen im Fokus stehen. Unter diesen spielt die Kopplung an das τ-Lepton eine zentrale Rolle, da es aufgrund seiner großen Yukawa-Wechselwirkung von besonderer Bedeutung ist. Der Zerfall $H \to \tau\tau$ im geboosteten Regime bietet eine einzigartige Möglichkeit, die Eigenschaften des Higgs-Bosons bei hohen Impulsen zu untersuchen, stellt jedoch zugleich erhebliche experimentelle Herausforderungen an die Rekonstruktion.

In dieser Arbeit wird eine Analyse des geboosteten $H \to \tau\tau$-Kanals vorgestellt, basierend auf Proton-Proton-Kollisionsdaten, die mit dem CMS Detektor während Run 2 des LHC bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV aufgezeichnet wurden. Zur Identifikation hadronisch zerfallender τ-Leptonen in geboosteten Topologien wurde ein maschinelles Lernverfahren, *ParticleNet*, eingesetzt, das τ-Leptonen als Jet-Unterstruktur behandelt. Zur Rekonstruktion der Di-τ-Masse wurde die kollineare Massenapproximation angewendet, als diskriminierende Observable mit Anwesenheit unsichtbarer Neutrinos aus τ-Zerfällen angewendet.

Die Trennkraft von ParticleNet wurde mit Kollisionsdaten validiert und Identifikation geboosteter τ-Leptonen deutlich verbesserte. Auch die kollineare Massenapproximation erwies sich als nützliches Werkzeug zur Unterscheidung von Signal und Untergrund, wenngleich die rekonstruierten Massenpeaks gegenüber ihren wahren Werten verschoben erscheinen. Für diese Arbeit wurde eine konservative Unsicherheit von 10% auf die Effizienz der geboosteten τ-Identifikation angenommen, wodurch Potenzial für Verbesserungen durch präzisere Kalibrierungen verbleibt.

Die in dieser Arbeit entwickelten Methoden sind nicht nur für die Untersuchung des Higgs-Bosons von können angewendet werden, sondern auch für andere Analysen mit geboosteten τ-Leptonen, wie etwa $Z' \to \tau\tau$ oder Erweiterungen des Higgs-Sektors wie das MSSM und NMSSM. Mit den größeren

Datensätzen von Run 3 und dem High-Luminosity LHC sowie den Fortschritten in der Detektortechnologie und im maschinellen Lernen wird erwartet, dass diese Methoden eine zentrale Rolle sowohl in der präzisen Higgs-Physik als auch in der Suche nach neuer Physik jenseits des Standardmodells spielen werden.

Contents

List of Abbreviations

$t\bar{t}H$	associated production with a top quark.
ASIC	Application-Specific Integrated Circuits.
ATLAS	A Toroidal LHC Apparatus.
AUC	Area Under the ROC Curve.
BSM	beyond the Standard Model.
CA8	CA jet with ΔR =0.8.
CMS	Compact Muon Solenoid.
CSCs	Cathode Strip Chambers.
DAQ	Data Acquisition.
DGCNN	Dynamic Graph Convolutional Neural Network.
DM	Decay Mode.
DTs	Drift Tubes.
ECAL	Electromagnetic Calorimeter.
Edge-Conv	Edge Convolution.
EFT	Effective Field Theory.
FCC	Future Circular Collider.
FPGA	Field-Programmable Gate Arrays.

ggF gluon-gluon fusion.

GNN Graph Neural Network.

HCAL Hadronic Calorimeter.

HF Forward Hadronic Calorimeter.

HLT High Level Trigger.

HPS Hadron-Plus-Strip.

JEC Jet Energy Corrections.

JES Jet energy scale.

Jet ID Jet Identification.

KF Kalman Filtering.

L1 Level-1.

LHC Large Hadron Collider.

LO leading order.

MC Monte Carlo.

MET Missing Transverse energy.

MLE Maximum Likelihood Estimation.

MSSM Minimal Supersymmetric Standard Model.

MVA Multivariate Analysis.

NLO next-to-leading order.

NMSSM Next-to-Minimal Supersymmetric Standard Model.

PDF Parton Distribution Functions.

PF Particle-Flow.

PUPPI	Pile-up per particle identification.
PV	Primary Vertex.
QCD	Quantum Chromodynamics.
QED	Quantum Electrodynamics.
QFT	Quantum Field Theory.
ROC	Receiver Operating Characteristic.
RPCs	Resistive Plate Chambers.
S&M	Shuffle and Merge.
SM	Standard Model.
SMEFT	Standard Model Effective Field Theory.
SS	same-sign.
SSB	Spontaneous Symmetry Breaking.
SUSY	Supersymmetry.
VBF	vector boson fusion.
VEV	Vacuum Expectation Value.
WP	Working Point.
WPs	Working Points.

List of Figures

List of Tables

Part I.

Introduction

1. Introduction to particle physics

1.1. Standard Model of particle physics

The Standard Model (SM) is a theoretical framework in particle physics that classifies all the particles responsible for forming matter, based on their intrinsic properties [1].

A schematic representation of the SM is summarized in Figure 1.1.

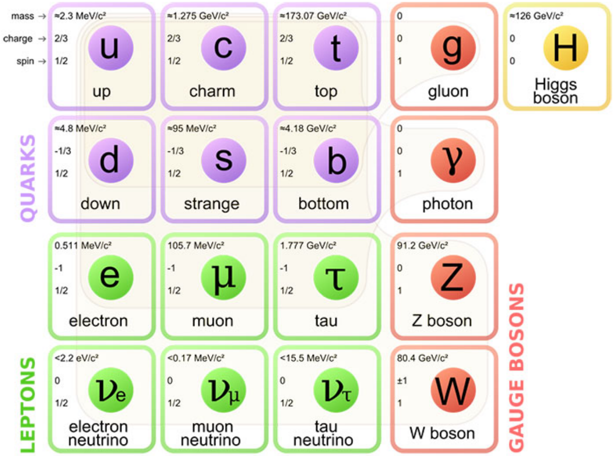


Figure 1.1.: The schematic representation of the SM highlights its key components: quarks, leptons (fermions), and mediators (gauge bosons). Fermions are categorized into three distinct generations. In the illustration, the masses, charges, and spins of the particles are depicted. The heavier a particle is, the higher the generation it belongs to [2].

According to this classification, elementary particles fall into three groups: leptons and quarks (both are fermions), and bosons. Most bosons serve as mediators of the fundamental interactions (the photon, $W^{\pm}$, Z, and gluons), while the Higgs boson plays a different role by providing masses through the Higgs mechanism [3]. Quarks and leptons are organized into three generations, with each subsequent generation containing heavier particles than the previous one.

Quarks cannot exist in a free state; instead, they form composite particles known as *hadrons*. Typical examples of hadrons are the proton and neutron. The proton consists of two up quarks and one down quark (uud), while the neutron is composed of one up quark and two down quarks (udd). Hadrons are categorized into two subgroups: *baryons*, which consist of an odd number of quarks (typically three), and *mesons*, which contain an equl number of quarks and antiquarks. Quarks possess spin $\frac{1}{2}$ and carry an electric charge of either $\pm\frac{1}{3}$ or $\pm\frac{2}{3}$ times the elementary charge. Among the six known quark flavours, the up quark is one of the lightest, while the top quark is the heaviest.

Leptons, like quarks, also have spin $\frac{1}{2}$ but, in contrast to quarks, can exist as free particles. They carry an electric charge of either ± 1 or 0 (for neutrinos).

All gauge bosons have spin 1. Among them, only the $W^{\pm}$ bosons carry electric charge ± 1; the Z boson is electrically neutral. The W and Z bosons are massive, whereas the remaining gauge bosons (such as the photon and gluon) are massless.

The Higgs boson is a scalar (spin 0) particle and is electrically neutral [4].

In addition to mass, spin, and electric charge, particles in the SM are also characterized by further quantum numbers associated with the gauge symmetries of the theory. Quarks carry color charge under $SU(3)_C$ and are classified by flavour (up, down, strange, charm, bottom, top). Leptons and quarks transform differently under the electroweak group $SU(2)_L \times U(1)_Y$, where left-handed fermions form weak isospin doublets while right-handed fermions are singlets. The corresponding weak isospin (T) and weak hypercharge (Y) determine the electric charge via the Gell-Mann–Nishijima relation [5]

$$Q = T_3 + \frac{1}{2}Y \tag{1.1}$$

Approximate global quantum numbers such as baryon number and lepton number are conserved in all SM interactions, while flavour quantum numbers (e.g. strangeness, charm, bottomness) are conserved by strong and electromagnetic interactions but can change in weak processes.

1.2. Forces of nature and their carriers

1.2.1. Quantum Field Theory Picture

In the framework of Quantum Field Theory (QFT), particles are described as excitations of underlying quantum fields [6, 7]. Interactions arise from specific terms in the Lagrangian of the theory. Each interaction term corresponds to a vertex in a Feynman diagram, with the coupling constant determining the strength of the interaction. The propagators represent the exchange of virtual particles, i.e. the mediators of the interaction. In this way, the intuitive picture of “forces with carrier particles” is rigorously formulated in terms of fields, vertices, and propagators. This framework provides the foundation for the description of the electromagnetic, weak, and strong interactions.

1.2.2. Electromagnetic Interaction

The electromagnetic interaction is the simplest among them and is described by Quantum Electrodynamics (QED) [8, 9]. It explains electromagnetic interactions mediated by photons. A typical QED process is electron-positron annihilation, as illustrated in Figure 1.2.

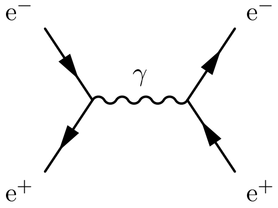
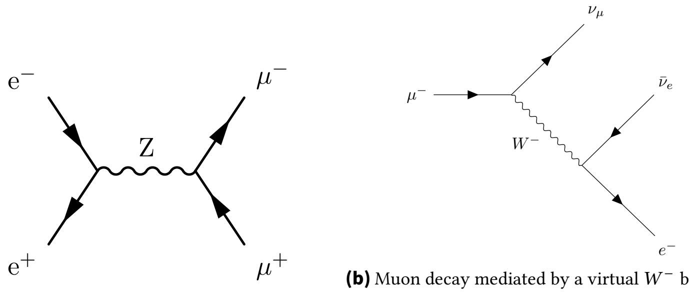


Figure 1.2.: Electron-positron annihilation via a virtual photon γ, resulting in a fermion-antifermion pair.

1.2.3. Weak interaction

The weak interaction can be divided into two types: charged interactions (mediated by $W^\pm$ bosons) with the charge transition involved and neutral interactions (mediated by the Z boson) without charge transition involved [10]. A typical example of such an interaction is demonstrated in Figure 1.3.



(a) Electron-positron annihilation via a virtual Z boson, resulting in a fermion-antifermion pair.

(b) Muon decay mediated by a virtual W^- boson, producing a muon neutrino, electron, and electron antineutrino.

Figure 1.3.: Feynman diagrams of processes mediated by the weak interaction: (a) electron-positron annihilation via Z boson exchange, and (b) muon decay via W^- boson exchange.

1.2.4. Electroweak Unification and Weak Isospin

The electromagnetic and weak interactions are unified in the framework of the electroweak theory, first formulated by Glashow, Weinberg, and Salam [11, 12]. The underlying gauge symmetry is

$$SU(2)_L \times U(1)_Y,$$

where $SU(2)_L$ corresponds to the weak isospin and $U(1)_Y$ to the weak hypercharge. The weak isospin groups left-handed fermions into doublets, while right-handed fermions appear as singlets under $SU(2)_L$.

After spontaneous symmetry breaking via the Higgs mechanism, the gauge fields mix to form the physical bosons: the photon (γ), the charged $W^\pm$ bosons, and the neutral Z boson. In this way, the electroweak theory explains both the similarities and the differences between electromagnetic and weak processes, and provides a unified description of two of the fundamental interactions of nature.

1.2.5. Strong interaction

The strong force is described by Quantum Chromodynamics (QCD) and explains interaction mediated by gluons [13, 14]. An illustrative example of a QCD process, quark-antiquark annihilation, is presented in Figure 1.4.

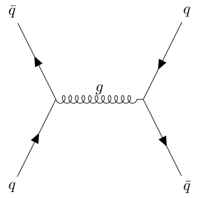


Figure 1.4.: Feynman diagram of quark-antiquark annihilation via a virtual gluon g, resulting in a quark-antiquark pair of the same or different flavor.

1.3. Local gauge invariance

Local gauge invariance is a property of the Lagrangian that it is invariant under local transformation. Unlike the global transformation, the local transformation requires a dependence on the spatial coordinate. For example, Equation 1.2 illustrates the global transformation, while Equation 1.3 illustrates the local invariance [15].

$$\psi \longrightarrow e^{i\theta}\psi \tag{1.2}$$

$$\psi \longrightarrow e^{i\theta(x)}\psi \tag{1.3}$$

Considering the Dirac Lagrangian [16], which describes relativistic massive particles with spin $\frac{1}{2}$

$$\mathcal{L} = \hbar i c\overline{\psi}\gamma^{\mu}\partial_{\mu}\psi - mc^{2}\psi\overline{\psi} \tag{1.4}$$

and is invariant under global gauge invariance, but not under local gauge invariance. If we require the local gauge invariance, the Lagrangian transforms to

$$\mathcal{L} = \hbar i c \overline{\psi} \gamma^\mu \partial_\mu \psi - mc^2 \psi \overline{\psi} - \frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} - (q \overline{\psi} \gamma^\mu \psi) A_\mu \tag{1.5}$$

where A_μ is a new massless field which has a corresponding massless carrier - a photon.

This is a Lagrangian for quantum electrodynamics - the description of electrons interacting with the photon field.

The Equation 1.2 can be represented as a matrix multiplication $\psi \longrightarrow U\psi$, where $U^\dagger U = 1$. The group of such matrices is U(1).

Since the Lagrangian 1.5 lacks a term describing photon self-coupling, this implies that photons cannot interact with one another directly. An electron can interact with another electron indirectly through the exchange of a virtual photon.

We apply the same procedure to the QCD theory. According to SM, quarks have three color charges - red, blue, green.

Equations 1.6 define the color structure of a quark field in QCD. The quark field ψ is represented as a color triplet, consisting of components ψ_r, ψ_b, and ψ_g, which correspond to the red, blue, and green color charges, respectively. Its Dirac adjoint, denoted by $\overline{\psi}$, is similarly expressed as a triplet with components $\overline{\psi_r}$, $\overline{\psi_b}$, and $\overline{\psi_g}$. This representation reflects the fact that quarks transform under the fundamental representation of the SU(3) color gauge group

$$\begin{cases} \psi = (\psi_r,\ \psi_b,\ \psi_g), \\ \overline{\psi} = (\overline{\psi_r},\ \overline{\psi_b},\ \overline{\psi_g}) \end{cases} \tag{1.6}$$

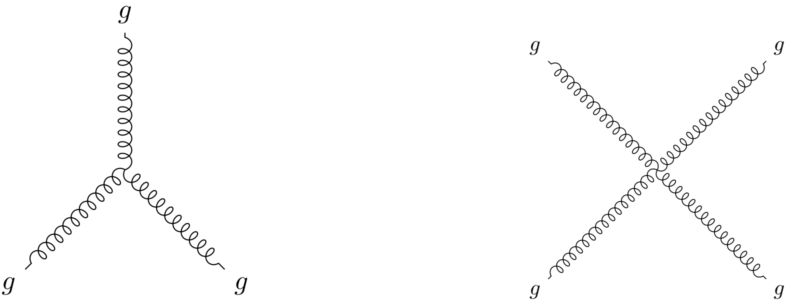
The QCD Lagrangian [17] is

$$\mathcal{L} = \frac{1}{4\pi} F_a^{\mu\nu} F_{\mu\nu}^a + \sum_f \overline{\psi}_i^{(f)} (iD_{ij} - m_f \delta_{ij}) \psi_j^{(f)} \tag{1.7}$$

with

$$
\begin{aligned}
D^{\mu}_{ij} &= \partial^{\mu}\delta_{ij} + i g^{s} t^{a}_{ij} A^{\mu}_{a} \quad \text{and} \\
F^{a}_{\mu\nu} &= \partial_{\mu} A^{a}_{\nu} - \partial_{\nu} A^{a}_{\mu} - g_{s} f_{abc} A^{b}_{\mu} A^{c}_{\nu}
\end{aligned}
\tag{1.8}
$$

The resulting term A contains eight gauge fields, each associated with one of the eight massless gluons, the mediators of the strong interaction. In contrast to QED, QCD permits self-interactions among its gauge bosons, which is manifested through terms like $A^{b}_{\mu} A^{c}_{\nu}$. The corresponding process is illustrated in Figure 1.5.



(a) Three gluon vertex

(b) Four gluon vertex

Figure 1.5.: Feynman diagrams representing gluon self-interactions in QCD. (a) Three-gluon vertex. (b) Four-gluon vertex. These interactions arise from the non-Abelian nature of the strong force.

Weak force is more complicated case. If we analogously require SU(2) gauge symmetry, it leads to the existence of three massless gauge bosons: two charged and one uncharged. However, experimental data confirms that the Z and $\mathrm{W}^{\pm}$ bosons have mass [18, 19]. To describe the mass generation of these gauge bosons, the Higgs mechanism is used, which is more precisely described in Chapter 2.

SM of particle physics provides a unified theoretical framework that describes three of the four fundamental forces in nature: the electromagnetic, weak, and strong interactions. These forces are mediated by gauge bosons: photons, W and Z bosons, and gluons, which arise naturally from requiring local gauge invariance of the Lagrangian under the symmetry groups U(1), SU(2), and SU(3), respectively. Matter fields in the SM consist of quarks and leptons, which interact via these gauge bosons according to their respective charges and representations.

Local gauge invariance serves as a guiding principle in the formulation of interaction terms. By promoting global symmetries to local ones, one introduces gauge fields whose dynamics and couplings are constrained by the symmetry structure. For example, the requirement of SU(3) local gauge invariance leads to the introduction of gluons and their self-interactions, a hallmark of QCD. Similarly, the electroweak sector emerges from the unification of SU(2) and U(1) gauge groups, with spontaneous symmetry breaking giving rise to massive W and Z bosons.

Overall, the SM successfully explains a wide range of observed phenomena in high-energy physics, although it remains incomplete due to its exclusion of gravity and open questions such as neutrino masses and dark matter.

2. Higgs Physics: theoretical and experimental aspects

2.1. Higgs mechanism

The SM of particle physics is a remarkably successful theoretical framework that has been extensively validated by experimental results over the past decades [6]. One of its central components is the Higgs boson, a fundamental scalar particle associated with the Higgs mechanism [20, 21], which endows elementary particles with mass through spontaneous electroweak symmetry breaking [11, 12].

For many years, the Higgs boson remained the only element of the SM whose existence had not been experimentally confirmed. This changed in 2012 with its discovery at the Large Hadron Collider (LHC), marking a major milestone in high-energy physics [22, 23]. However, it is important to note that the SM does not predict the mass of the Higgs boson. Consequently, the experimental effort was not only aimed at confirming the existence of the Higgs field but also at precisely measuring the mass of the associated scalar particle.

Prior to the LHC era, constraints on the Higgs boson mass had already been established by earlier experiments at LEP and Tevatron [24, 25], which excluded large regions of the mass spectrum and narrowed down the viable mass range. These indirect limits meant that the Higgs mass was not an entirely unconstrained parameter.

A crucial theoretical motivation for the Higgs boson's existence is encapsulated in what is often referred to as the *no-lose theorem* [26]. This argument states that, in the absence of the Higgs boson (or an alternative mechanism of electroweak symmetry breaking), scattering amplitudes involving longitudinally polarized vector bosons (such as $W_L W_L \to W_L W_L$) would grow with energy and eventually violate unitarity and yielding probabilities greater than one, which is a clear sign of a breakdown in physical consistency. The introduction of the Higgs boson restores unitarity in these processes by cancelling the divergences in the amplitude at high energies. Thus, the presence of some new physics, whether in the form of a Higgs boson or another mechanism, is theoretically required to preserve unitarity in vector boson scattering.

These theoretical arguments, together with experimental constraints, provided strong motivation for the search for the Higgs boson. A key theoretical concept is Spontaneous Symmetry Breaking (SSB) [27], which occurs when the ground state of a system does not respect the symmetry of its Lagrangian. In the SM, this mechanism is realized through a scalar potential with a nontrivial vacuum structure, often visualized as the so-called “Mexican hat” potential shown in Figure 2.1.

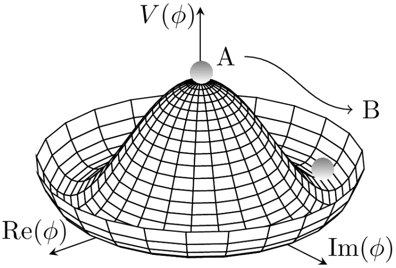


Figure 2.1.: The Higgs potential, often referred to as the "Mexican hat" [28]. It illustrates the concept of SSB, where the ground state of the theory is not invariant under the original symmetry.

The Higgs mechanism is the process by which gauge bosons acquire mass through *SSB* of a gauge symmetry, while preserving the *renormalizability* and *gauge invariance* of the theory [29, 30]. It plays a central role in the SM, providing masses to the weak gauge bosons ($W^{\pm}$, Z) and to fermions.

2.1.1. Electroweak Gauge Symmetry

The electroweak sector of the SM is based on the gauge group

$$SU(2)_L \times U(1)_Y \tag{2.1}$$

where $SU(2)_L$ is the weak isospin symmetry and $U(1)_Y$ corresponds to weak hypercharge [11, 12]. Gauge invariance forbids explicit mass terms for the gauge bosons in the Lagrangian density, as such terms would violate local gauge symmetry.

2.1.2. Spontaneous Symmetry Breaking

To generate particle masses dynamically, a complex scalar field known as the Higgs field is introduced in the SM [3, 21, 20]. This field transforms as a doublet under $SU(2)_L$ and carries hypercharge $Y = +1$ [11, 12]. Its dynamics are governed by the scalar potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{2.2}$$

as described in standard gauge theory texts [6, 30, 29]. The shape of the potential depends on the sign of the parameter μ^2, which determines whether the electroweak symmetry remains unbroken ($\mu^2 > 0$) or is spontaneously broken ($\mu^2 < 0$).

In the broken phase, the neutral component ϕ^0 acquires a nonzero Vacuum Expectation Value (VEV),

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v = \sqrt{-\mu^2/\lambda} \tag{2.3}$$

which breaks the gauge symmetry

$$SU(2)_L \times U(1)_Y \longrightarrow U(1)_{\text{em}} \tag{2.4}$$

The value $v \approx 246$ GeV is determined from low-energy measurements of the Fermi constant [31].

2.1.3. Mass Generation for Gauge Bosons

The interaction between the Higgs field and the gauge bosons is introduced via the covariant derivative, which ensures local gauge invariance under the transformation of electroweak symmetry group $SU(2)_L \times U(1)_Y$. The covariant derivative acting on the Higgs doublet field ϕ is defined as

$$D_\mu \phi = \left(\partial_\mu - i\frac{g}{2} \vec{\tau} \cdot \vec{W}_\mu - i\frac{g'}{2} B_\mu \right) \phi \tag{2.5}$$

Here, ∂_μ is the usual spacetime derivative, g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge couplings, respectively, $\vec{\tau}$ are the Pauli matrices, $\vec{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ are the weak gauge fields, and B_μ is the hypercharge gauge field.

Upon SSB, the kinetic term of the Higgs field,

$$(D_\mu\phi)^\dagger(D^\mu\phi), \tag{2.6}$$

generates mass terms for the gauge bosons when expanded around the Higgs VEV. The resulting mass terms Lagrangian density are

$$\mathcal{L}_{\text{mass}} = \frac{1}{2}v^2\left[g^2 W_\mu^+ W^{-\mu} + \frac{1}{2}(g^2 + g'^2)Z_\mu Z^\mu\right] \tag{2.7}$$

The physical mass eigenstates are obtained by diagonalizing the gauge boson mass matrix

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2), \quad m_W = \frac{1}{2}gv \tag{2.8}$$

$$Z = \cos\theta_W W^3 - \sin\theta_W B, \quad m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}\,v \tag{2.9}$$

$$A = \sin\theta_W W^3 + \cos\theta_W B \quad \text{(massless photon)} \tag{2.10}$$

The angle θ_W, known as the Weinberg angle, is defined by the ratio of coupling constants

$$\tan\theta_W = \frac{g'}{g} \tag{2.11}$$

The Higgs mechanism [3, 20, 21] not only gives mass to the $W^\pm$ and Z bosons while leaving the photon massless, but also predicts a scalar particle, the Higgs boson, with mass

$$m_H = \sqrt{2\lambda}\,v \tag{2.12}$$

where λ is the Higgs self-coupling constant.

Out of the four real degrees of freedom in the complex Higgs doublet, three are absorbed as longitudinal components of the massive vector bosons via the Higgs mechanism [11, 12], leaving one observable scalar field. This mechanism ensures the correct number of polarization states for the massive W and Z bosons and maintains renormalizability.

Moreover, the presence of the Higgs boson is crucial for preserving unitarity in high-energy longitudinal vector boson scattering processes [26].

Without it (or an alternative mechanism), scattering amplitudes would grow unboundedly, violating probabilistic interpretation.

These theoretical foundations are discussed in detail in standard texts such as [6, 29, 30].

2.1.4. Physical Higgs Boson

Out of the four real degrees of freedom in the complex Higgs doublet, three are absorbed as the longitudinal polarization components of the massive gauge bosons $W^\pm$ and Z, these are the so-called *would-be Goldstone bosons* [3, 20, 21]. The remaining degree of freedom manifests as a physical scalar particle: the Higgs boson.

The Higgs field can be expanded around the VEV as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.13}$$

where $h(x)$ is the physical Higgs boson field, representing quantum excitations above the vacuum.

The mass of the physical Higgs boson is determined by the second derivative (curvature) of the Higgs potential evaluated at the vacuum, leading to

$$m_H^2 = 2\lambda v^2 \tag{2.14}$$

where λ is the self-coupling parameter appearing in the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{2.15}$$

as introduced in the SM scalar sector [6, 29]. This relationship links the mass of the Higgs boson directly to the strength of the scalar self-interaction and the symmetry-breaking scale.

2.1.5. Fermion Masses via Yukawa Couplings

Fermions acquire mass in the SM through their interaction with the Higgs field via so-called *Yukawa couplings* [6, 11]. These interactions are described by the Lagrangian term

$$\mathcal{L}_Y = -y_f \, \bar{\psi}_L \phi \psi_R + \text{h.c.} \tag{2.16}$$

where y_f is the Yukawa coupling for a fermion of flavor f, ψ_L and ψ_R are the left- and right-handed fermion fields, and ϕ is the Higgs doublet.

After electroweak symmetry breaking, the Higgs field acquires a VEV, and the Yukawa interaction term generates a fermion mass

$$m_f = \frac{y_f v}{\sqrt{2}} \tag{2.17}$$

Each fermion mass is therefore directly proportional to its Yukawa coupling y_f, which remains a free parameter within the SM. This formulation explains the hierarchical pattern of fermion masses but does not predict their absolute values, making the Yukawa sector one of the least constrained parts of the model.

The structure of Yukawa interactions is critical for both the generation of mass and the flavor structure of the SM, and plays a central role in Higgs boson decay modes.

2.2. SM problems in Higgs sector

Although the SM has been confirmed experimentally with extraordinary precision [31], it does not provide answers to all fundamental questions in particle physics. One of the major open problems is the *mass hierarchy problem* [32]. It remains unclear why the different fundamental forces exhibit such vastly different strengths. For instance, the strong interaction is approximately 10^{14} times stronger than gravity and several orders of magnitude stronger than the weak interaction.

The Higgs boson, responsible for electroweak symmetry breaking, was long expected to have a mass significantly larger than 125 GeV [29]. However, the SM does not predict the mass of the Higgs boson, it remains a free parameter. Historically, similar theoretical inconsistencies have led to the prediction of new particles. A notable example is the positron, the antimatter counterpart of the electron, predicted by Paul Dirac in 1928 [16]. Dirac formulated a relativistic wave equation for spin-$\frac{1}{2}$ particles, whose solutions included states with negative energy, a result incompatible with classical physics. To resolve this paradox, Dirac postulated the existence of a new particle identical to the electron but with opposite charge. This groundbreaking prediction of antimatter was soon experimentally confirmed [33].

Analogously, many physicists anticipate that solving the shortcomings of the SM may again require the introduction of new particles. In the case of the Higgs sector, the mass hierarchy problem becomes particularly severe. In the SM, the Higgs boson mass receives quantum corrections that grow quadratically with the energy scale. Without a protective symmetry, these corrections must be finely tuned to cancel out, a situation considered unnatural and theoretically unsatisfying [32].

One elegant solution to this problem is provided by Supersymmetry (SUSY) [34]. SUSY postulates a symmetry between bosons and fermions, predicting for every SM particle a heavier superpartner with opposite spin statistics. These superpartners naturally cancel the large quantum corrections to the Higgs mass, stabilizing it without fine-tuning. While no SUSY particles have been observed so far, the framework remains one of the most compelling extensions of the SM.

2.2.1. Supersymmetry and its application to problems in Higgs sector

The SM radiative corrections to the Higgs mass arising from loop diagrams, such as those involving the top quark, take the following form

$$\delta m_H^2 \approx -\frac{|y_t|^2}{8\pi^2}\Lambda^2 + \cdots \tag{2.18}$$

where y_t is the top Yukawa coupling, and Λ is the ultraviolet cutoff scale, typically assumed to be as high as the Planck scale ($\sim 10^{19}$ GeV). This implies that, in the absence of fine-tuning, the physical Higgs mass would naturally be pushed up to the cutoff scale. The observed Higgs mass of 125 GeVseems to be unnatural according to this assumption, unless it receives some additional terms for cancellation [34, 35].

SUSY offers a natural solution by introducing superpartners for all SM particles. In an unbroken supersymmetric theory, the quadratic divergences in the Higgs mass cancel between particles and their superpartners

$$\delta m_H^2\big|_{\text{boson}} + \delta m_H^2\big|_{\text{fermion}} = 0 \tag{2.19}$$

This cancellation occurs because bosons and fermions contribute with opposite signs to loop corrections, and SUSY enforces equal coupling strengths and masses between them [34, 30].

As a result, the Higgs mass is stabilized against large quantum corrections, making the electroweak scale technically natural.

It is postulated that for every known particle, there exists a corresponding *superpartner* with spin differing by half a unit:

- Fermions (spin-$\frac{1}{2}$) → bosonic superpartners (spin-0), called *sfermions* (e.g., selectron, squark)
- Bosons (spin-1 or spin-0) → fermionic superpartners (spin-$\frac{1}{2}$), called *gauginos* and *higgsinos*

2.3. MSSM and NMSSM models

Supersymmetric extensions of the SM, such as the Minimal Supersymmetric Standard Model (MSSM) and the Next-to-Minimal Supersymmetric Standard Model (NMSSM), offer elegant solutions to some of the theoretical limitations of the SM. In particular, they address the hierarchy problem by stabilizing the Higgs boson mass against large radiative corrections through cancellations between divergent quantum contributions from SM particles and their superpartners [34, 36, 37].

Contrary to the SM, which contains a single complex scalar Higgs doublet, the MSSM introduces two Higgs doublets, denoted H_u and H_d. This extension is necessary to ensure anomaly cancellation and to generate masses for both up-type and down-type fermions [34, 36].

After electroweak symmetry breaking, the two complex Higgs doublets yield five physical Higgs bosons:

- Two CP-even neutral scalars: the lighter h^0 and the heavier H^0
- One CP-odd neutral pseudoscalar: A^0
- Two charged Higgs bosons: $H^\pm$

At tree-level, the Higgs sector of the MSSM can be described using two parameters

$$\tan\beta = \frac{v_u}{v_d}, \qquad m_A \tag{2.20}$$

where v_u and v_d are the VEV of the up-type and down-type Higgs doublets, and m_A is the mass of the CP-odd pseudoscalar A^0.

A notable prediction of the MSSM is an upper bound on the tree-level mass of the lightest CP-even Higgs boson

$$m_{h^0}^{\text{tree}} \leq M_Z|\cos 2\beta| \leq M_Z \tag{2.21}$$

where M_Z is the mass of the Z boson. This is below the observed Higgs mass. However, large radiative corrections, especially those involving top quark and top squark (stop) loops, can raise the Higgs mass significantly. The leading one-loop correction takes the form [38]

$$\Delta m_{h^0}^2 \approx \frac{3m_t^4}{2\pi^2 v^2} \ln\left(\frac{M_S^2}{m_t^2}\right) \tag{2.22}$$

where m_t is the top quark mass, $v = \sqrt{v_u^2 + v_d^2} \approx 246\,\text{GeV}$ is the electroweak VEV, and M_S is the geometric mean of the two stop mass eigenstates.

The NMSSM extends the MSSM by introducing a gauge-singlet chiral superfield S. This addition addresses the so-called μ-problem of the MSSM, where the supersymmetric Higgs mass parameter μ must be of the electroweak scale without a natural origin [37]. In the NMSSM, the superpotential relevant to the Higgs sector is given by

$$W_{\text{NMSSM}} = \lambda S H_u H_d + \frac{\kappa}{3} S^3 \tag{2.23}$$

where λ and κ are dimensionless couplings. After the singlet scalar field acquires a VEV, $\langle S\rangle$, an effective μ-term is dynamically generated

$$\mu_{\text{eff}} = \lambda\langle S\rangle \tag{2.24}$$

The scalar potential in the NMSSM contains three scalar fields: H_u, H_d, and S. After electroweak symmetry breaking, the model predicts seven physical Higgs bosons

- Three CP-even neutral scalars: h_1, h_2, h_3
- Two CP-odd neutral pseudoscalars: a_1, a_2
- Two charged Higgs bosons: $H^\pm$

An important improvement of the NMSSM over the MSSM is the modified tree-level upper bound on the lightest Higgs mass

$$m_{h_1}^2 \leq M_Z^2 \cos^2 2\beta + \lambda^2 v^2 \sin^2 2\beta \tag{2.25}$$

which includes a positive contribution proportional to λ^2. This allows the model to accommodate a Higgs boson with mass around 125 GeVmore naturally, without requiring extremely heavy stop masses or large loop corrections [39].

Furthermore, the mixing between the Higgs doublets and the singlet scalar field can give rise to distinctive Higgs phenomenology. In particular, the observed 125 GeVHiggs boson may exhibit suppressed couplings to SM particles due to a singlet admixture. The NMSSM also allows for exotic Higgs decays, such as

$$h_1 \longrightarrow a_1 a_1 \longrightarrow 4\tau,\ 4b,\ \text{or}\ 2\mu 2\tau \tag{2.26}$$

These novel signatures are actively searched for in current LHC analyses.

2.4. SM Higgs boson properties

The Higgs boson is an electrically neutral scalar gauge boson with a mass of approximately 125.38 GeVand spin 0 [40].

Within the SM, the τ lepton lifetime is predicted [41] to be

$$\tau = (2.903 \pm 0.005) \times 10^{-13}\ \text{s},$$

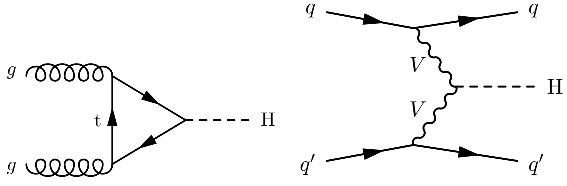
which corresponds to a decay width of

$$\Gamma = \frac{\hbar}{\tau} = 4.14 \pm 0.02\ \text{MeV}.$$

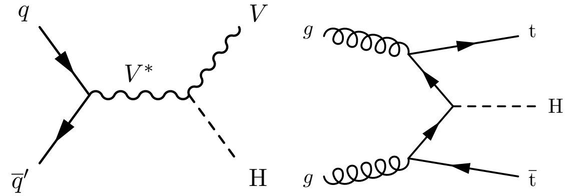
In proton-proton collisions, there are four primary production mechanisms for the Higgs boson:

1. gluon-gluon fusion (ggF),
2. vector boson fusion (VBF),
3. Higgs-Strahlung (associated production with a W or Z boson),
4. associated production with a top quark ($t\bar{t}H$).

Figure 2.2 shows representative Feynman diagrams for the four main Higgs boson production mechanisms in proton-proton collisions. At the LHC, the dominant production mode is ggF, followed by VBF, Higgs-Strahlung, and $t\bar{t}H$ [40].



(a) ggF **(b)** VBF



(c) Higgs-Strahlung **(d)** $t\bar{t}H$

Figure 2.2.: Representative Feynman diagrams for the four main Higgs boson production mechanisms in proton-proton collisions: (a) ggF, (b) VBF, (c) Higgs-Strahlung, and (d) $t\bar{t}H$

Figure 2.3 shows the predicted production cross sections of the SM Higgs boson in proton-proton collisions as a function of the center-of-mass energy $\sqrt{s}$. The four dominant production mechanisms are displayed. As the energy increases, the cross sections for all channels rise due to the increase in parton luminosities. Among them, ggF remains the dominant process across all energy scales, with VBF contributing significantly at higher energies. These predictions play a crucial role in determining the expected event rates for different Higgs production modes at the LHC.

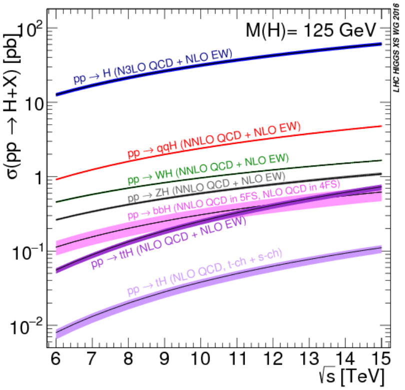


Figure 2.3.: SM predictions for Higgs boson production cross sections in proton-proton collisions as a function of the center-of-mass energy $\sqrt{s}$. The four main production modes are shown: ggF, VBF, associated production with a vector boson (Higgs-Strahlung), and $t\bar{t}H$. ggF dominates across the entire energy range, while the relative contribution of other mechanisms increases with energy [42]

The Higgs boson has an extremely short lifetime, which prevents it from traveling a measurable distance within the detector before decaying. As a result, the Higgs boson cannot be directly reconstructed; instead, its presence must be inferred by reconstructing its decay products.

Within the SM, the Higgs boson can couple to leptons, quarks, and electroweak gauge bosons. The strength of these couplings is not arbitrary: it is predicted to be proportional to the mass of the particle. Specifically, the coupling to fermions is proportional to their mass, while the coupling to vector bosons is proportional to the square of their mass

$$y_{H\to f\bar{f}} = \frac{m_f}{v}, \qquad y_{H\to VV} = \frac{2m_V^2}{v}, \tag{2.27}$$

where m_f is the mass of the fermion, m_V is the mass of the vector boson (W or Z). These relationships are central predictions of the SM and are tested through precision measurements of Higgs boson decays [40].

Due to the energy and momentum conservation, the Higgs boson cannot decay into two on-shell W or Z bosons: its mass is 125 GeV, while the masses of the W and Z bosons are 81 GeVand 91 GeV, respectively. As a result, at least one of the vector bosons in such decays must be off-shell.

Figure 2.4 shows the measured strengths of Higgs boson couplings to various fermions and vector bosons as a function of their mass. This plot illustrates the remarkable predictive power of the SM: the coupling strength increases proportionally with particle mass, as expected from theory.

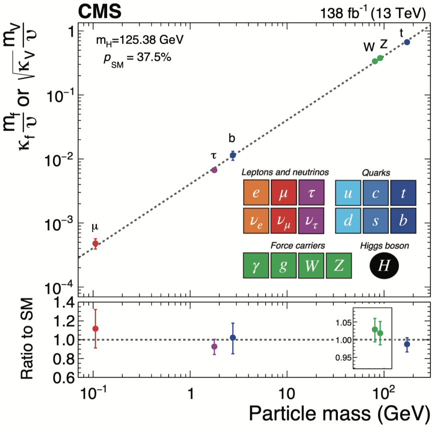


Figure 2.4.: Measured Higgs boson couplings to fermions and vector bosons as a function of particle mass. The results confirm the SM prediction that the coupling strength is proportional to the particle mass for fermions and to the mass squared for vector bosons [43].

The smallest predicted coupling is for the decay $H \longrightarrow e^+e^-$; however, this channel has not yet been observed at the LHC due to its extremely small branching ratio and overwhelming background. Nonetheless, it remains a compelling physics case for future high-luminosity colliders such as the Future Circular Collider (FCC) [44].

The decay $H \longrightarrow \mu^+\mu^-$ was successfully observed during Run 2 of the LHC [43]. While this channel benefits from a clean experimental signature, it remains challenging due to its low signal-to-background ratio.

The decay $H \rightarrow \tau^+\tau^-$ was also confirmed in Run 2 [45]. It has the highest branching ratio among all leptonic Higgs decays, but the identification of τ leptons is complicated by their frequent hadronic decays, which can be misidentified as jets.

The $H \rightarrow WW^*$ channel features a relatively large branching ratio. However, the presence of neutrinos in the final state degrades the invariant mass resolution, and the analysis suffers from considerable background contributions from non-resonant WW production [46].

In contrast, the decay $H \rightarrow ZZ^* \rightarrow 4\ell$ is widely known as the "golden channel" due to its exceptionally clean experimental signature and excellent mass resolution, making it ideal for precision measurements.

2.5. Boosted Higgs boson: motivation and theoretical basics

Since the discovery of the Higgs boson in 2012 [23, 22], the study of its properties has become one of the central pursuits in experimental particle physics. Among the key observables, the differential cross section as a function of the Higgs boson's transverse momentum (p_T) plays a particularly important role.

Inclusive Higgs boson production is dominated by the low-p_T region (p_T <200 GeV), which is well described by SM predictions and largely constrained by existing measurements [47]. In this regime, the production kinematics are governed by soft and collinear QCD emissions, and the sensitivity to possible effects from physics beyond the Standard Model (BSM) is significantly diluted.

In contrast, the high-p_T region ($p_\mathrm{T} > 200$ GeV) probes energy scales where new physics contributions may become relevant [41]. This part of the spectrum can be influenced by loop-level effects from heavy BSM particles, such as vector-like quarks or additional Higgs states, or by modifications of effective couplings in the context of Standard Model Effective Field Theory (SMEFT) [48]. As a result, precision measurements of the high-p_T tail of the Higgs spectrum serve as a sensitive probe for deviations from SM expectations and provide a complementary approach to direct BSM searches.

Historically, probing the high-p_{T} regime of Higgs production ($p_{\mathrm{T}} > 200$ GeV) was extremely challenging due to limitations in detector resolution, object identification, and reconstruction algorithms, especially in final states with highly Lorentz-boosted topologies. As a result, this corner of phase space remained largely unexplored in early LHC runs.

In contrast, the inclusive production at low transverse momentum ($p_{\mathrm{T}} <$ 200 GeV) has been well measured and is in good agreement with SM predictions.

Thanks to recent advancements in reconstruction techniques, including boosted object tagging, substructure-based identification, and improved tracking and calorimetry, experiments can now access the high-p_{T} tail of the Higgs spectrum with sufficient precision.

An enhancement in the Higgs transverse momentum spectrum at high p_T^H may signal the presence of new physics. As discussed earlier in this chapter, addressing open questions in the Higgs sector often requires the introduction of new heavy BSM particles. While such particles may lie beyond the direct reach of current colliders, their effects can manifest indirectly through higher-dimensional operators in the framework of an Effective Field Theory (EFT) [49].

In this formalism, the low-energy Lagrangian is extended by a series of dimension-six operators $\mathcal{O}_i$, suppressed by powers of the new physics scale Λ

$$\mathcal{L}_{\mathrm{EFT}} \supset \frac{c_i}{\Lambda^2}\mathcal{O}_i, \tag{2.28}$$

where c_i are dimensionless Wilson coefficients, $\mathcal{O}_i$ are gauge-invariant operators constructed from SM fields, and Λ represents the mass scale of the heavy new physics.

The contribution of such operators to physical observables is energy-dependent. In particular, the cross section receives corrections that scale quadratically with the momentum transfer

$$\Delta\sigma \sim \frac{p_T^2}{\Lambda^2}, \tag{2.29}$$

where p_{T} denotes the transverse momentum of the Higgs boson. This scaling implies that deviations from SM predictions become more pronounced at higher p_{T}, making the high-p_T^H region especially sensitive to possible effects from BSM physics [50, 51].

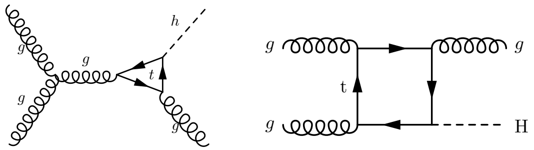
Another reason for that is the Yukawa coupling of the Higgs boson to the top quark where the top effects (in particularly top mass) start playing a significant role. At high pT, the SM prediction must use the full top-mass-dependent amplitude, as the heavy top approximation is no longer valid [52].

In the SM, the dominant Higgs boson production mechanism at the LHC is ggF. This process proceeds through a quantum loop involving top quarks, as gluons do not couple directly to color neutral scalar particles like the Higgs boson. At leading order, the interaction is mediated by a loop of virtual top quarks, which effectively induces a coupling between the gluons and the Higgs field

$$g + g \rightarrow H \tag{2.30}$$

The Feynman diagram of such process could be found on the Figure 2.2 a). At low Higgs p_{T}, the kinematic properties are such that the energy scales involved are well below the top quark mass, $p_T^H \ll m_t$ [53].

At high additional gluon radiation starts playing more important role. The corresponding Feynman diagrams are shown in the Figure 2.5.



(a) Triangle diagram for Higgs boson production via ggF at high transverse momentum. The Higgs is produced through a top-quark loop, with the recoil balanced by an additional hard gluon. This configuration becomes relevant when $p_T^H \gtrsim m_t$, where the full top-mass dependence must be accounted for.

(b) Box diagram contributing to Higgs boson production at high p_T via ggF. The process involves two incoming gluons and one outgoing gluon, with the Higgs produced through a closed top-quark loop. This topology encodes additional kinematic structure and contributes significantly to the high-p_T tail of the spectrum.

Figure 2.5.: Representative Feynman diagrams for Higgs boson production via ggF in the SM at high transverse momentum. (a) Triangle diagram in which the Higgs boson is produced through a top-quark loop. (b) Box diagram with the Higgs boson produced via a closed top-quark loop. Both topologies contribute significantly to the high-p_T tail of the spectrum and require full top-mass dependence for accurate theoretical predictions, as the effective field theory approximation breaks down in this regime.

In the high-p_T regime (typically above ~200–300 GeV), the Higgs boson recoils against a highly energetic object resulting from the radiation of a quark or gluon. Sensitivity to the internal structure of the top-quark loop arises from matrix-element terms involving the ratio p_T^H/m_t

$$\left|\mathcal{M}_{Hj(j)}\right|^2 \;\propto\; m_t^4\, \log^4\!\left(\frac{(p_T^H)^2}{m_t^2}\right). \tag{2.31}$$

The effect of the different partonic sub-processes is shown on the Figure 2.6. The visible top mass effects appear at p_T^H >250 GeV [41].

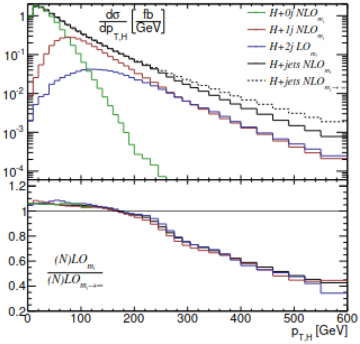


Figure 2.6.: Transverse momentum distribution $p_{T,H}$ for $H \to W^+W^-$ + jets production with Sherpa at NLO (left panel). Distributions are shown for exclusive and merged jet samples, including both finite top-mass effects ($m_t = 173$ GeV) and the low-energy approximation ($m_t \to \infty$). The Y-axis spans several orders of magnitude, reflecting the rapidly falling cross section at high p_T. [54]

The study of Higgs boson production at high transverse p_T, the so-called boosted regime, has become increasingly important in the search for BSM physics. Overall, the boosted Higgs regime provides a powerful window into the structure of Higgs interactions and remains a promising frontier for future discoveries in both direct BSM searches and precision measurements.

Part II.

Objects reconstruction in context of CMS experiment at CERN

3. Relevant subdetectors and reconstruction algorithms

3.1. The CMS Experiment

Compact Muon Solenoid (CMS) is a general-purpose detector at the LHC, the world's largest proton-proton collider. Together with its counterpart, the A Toroidal LHC Apparatus (ATLAS) experiment, located diametrically opposite to CMS, it played a central role in the discovery of the Higgs boson in 2012 [23, 22]. The LHC has a circumference of 26.7 km and provides collisions at the highest centre-of-mass energies ever achieved in a laboratory.

As of 2025, Run 3 of the LHC is ongoing. For both CMS and ATLAS, the centre-of-mass energy is $\sqrt{s} = 13.6\,\text{TeV}$. During the current data-taking period (2022-2025), CMS has recorded an integrated luminosity of 220.3 fb^{-1}, approximately twice the dataset collected in Run 2 (2016-2018) at $\sqrt{s} = 13\,\text{TeV}$. Run 3 is characterised by challenging pile-up conditions, with an average of about 60 simultaneous interactions per bunch crossing, compared to ~35 during Run 2.

The physics programme of CMS in Run 3 covers a broad range of topics:

- precision measurements of Higgs boson properties, including its couplings and self-coupling,
- precision tests of the SM, including QCD, electroweak, and top-quark physics, as well as searches for BSM phenomena,
- heavy-ion physics, in particular studies of the quark-gluon plasma and the equation of state,
- flavour physics, with emphasis on the properties and rare decays of B mesons.

The versatile design of the CMS detector enables this wide and diverse physics programme, providing sensitivity to a broad spectrum of signatures ranging from precision measurements to searches for entirely new phenomena.

3.2. The structure and geometry of CMS experiment

3.2.1. Brief description of the experiment

CMS is a cylindrical detector with a diameter of 15 m, a length of 21 m, and a total weight of about 125,000 tonnes. At its centre lies the beam pipe, with a diameter of 4.34 cm, which is surrounded by the silicon tracking system. This tracker consists of pixel and strip detectors that provide precise measurements of charged-particle trajectories as well as primary and secondary vertices. The tracking system is surrounded by the Electromagnetic Calorimeter (ECAL), designed to measure the energy of photons and electrons, and the Hadronic Calorimeter (HCAL), which measures the energy of hadrons. These calorimeters are enclosed within a steel return yoke that also hosts the muon detection system. A defining feature of CMS is its powerful superconducting solenoid, which provides a magnetic field of 4 T, enabling highly precise momentum measurements of charged particles [55]. A schematic view of the detector layout is shown in Figure 3.1.

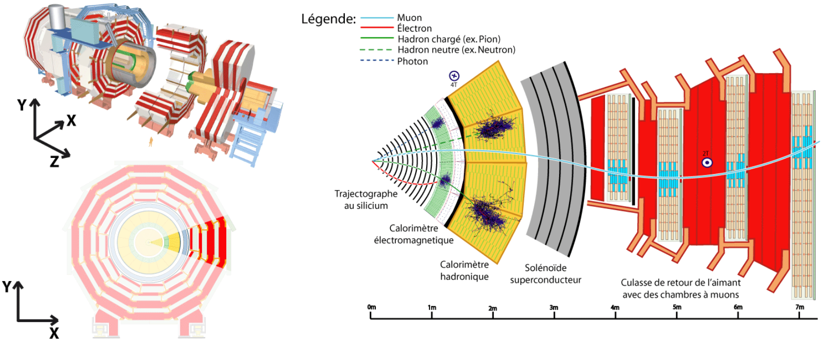


Figure 3.1.: Schematic slice of the CMS detector [56].

3.2.2. Coordinate system

Due to the cylindrical geometry of the detector, the use of a Cartesian coordinate system is not optimal. Instead, experimentalists employ a coordinate system based on the following variables:

- pseudorapidity η is a spatial coordinate that describes the angle of a particle relative to the beam axis. It is defined as

  $$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \tag{3.1}$$

  where θ is the polar angle with respect to the beam axis.

  From a geometrical perspective, η provides a convenient way to characterise detector coverage. Values of $\eta \approx 0$ correspond to particles produced at large polar angles (close to perpendicular to the beam axis, i.e. the central or "barrel" region of the detector), while large $|\eta|$ values correspond to small polar angles, i.e. particles emitted close to the beam direction (the "endcap" or "forward" regions). For example, a polar angle of $\theta = 90^\circ$ corresponds to $\eta = 0$, while $\theta = 10^\circ$ corresponds to $\eta \approx 2.4$.

  Thus, the pseudorapidity provides both a boost-invariant description of particle kinematics and a natural coordinate for describing detector acceptance in collider experiments.

- The azimuthal angle ϕ describes the angle of a particle's momentum in the plane transverse to the beam axis. It is defined as

  $$\phi = \arctan\left(\frac{p_y}{p_x}\right), \tag{3.2}$$

  where p_x and p_y are the momentum components in the transverse plane.

  By convention, ϕ is measured around the beam axis (z-axis) from the x-axis towards the y-axis, and it takes values in the range

  $$-\pi < \phi \leq \pi. \tag{3.3}$$

  Together with the pseudorapidity η, the coordinate ϕ provides a natural parametrisation of detector geometry in cylindrical coordinates (η, ϕ), which is widely used in collider physics to describe the direction and separation of particles.

The angular distance between two particles in this space is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}\,, \tag{3.4}$$

which is a Lorentz-invariant measure under boosts along the beam axis.

The CMS coordinate system is illustrated in Figure 3.2.

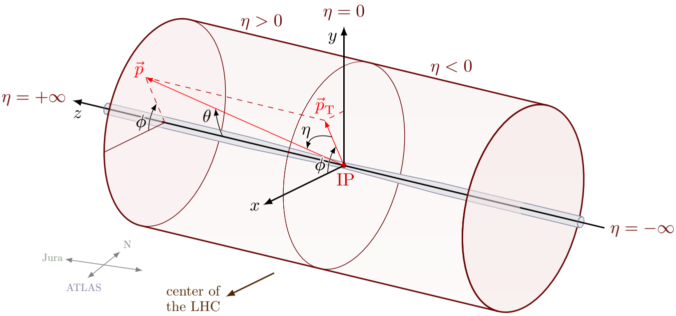


Figure 3.2.: Schematic view of the CMS coordinate system.

To summarize, the detector's coordinate system, based on pseudorapidity and azimuthal angle, offers a natural description of particle kinematics in a collider environment.

3.2.3. Division into barrel and endcaps regions

The CMS detector is divided into a central barrel region and two endcap regions, which together provide nearly hermetic coverage around the interaction point. The barrel corresponds to the region at small pseudorapidities ($|\eta| \lesssim 1.3$), where the detector layers form cylindrical shells around the beam axis. In this region, the geometry is approximately uniform in η and ϕ, which simplifies calibration and leads to the best detector resolution.

The endcap regions extend the coverage, up to $|\eta| \sim 3$ for the calorimeters and even beyond for the forward detectors. The endcaps are disk-shaped structures placed perpendicularly to the beam axis on both sides of the barrel. They allow the experiment to detect particles emitted at small polar angles with respect to the beam line, which are crucial for studies of boosted objects, forward jets, and heavy-ion collisions.

The distinction between barrel and endcaps is particularly important for performance studies: while the barrel generally offers the highest measurement precision, the endcaps provide the extended acceptance necessary for a comprehensive physics program [55].

Table 3.1 summarizes a pseudoraidity coverage of the main CMS subdetectors.

Table 3.1.: Approximate pseudorapidity coverage of the main CMS subdetectors in the barrel and endcap regions [57].

Subsystem	Barrel coverage	Endcap coverage
Tracker	$\|\eta\| < 1.2$	up to $\|\eta\| \approx 2.5$
ECAL	$\|\eta\| < 1.48$	$1.48 < \|\eta\| < 3.0$
HCAL	$\|\eta\| < 1.3$	$1.3 < \|\eta\| < 3.0$
Muon system	$\|\eta\| < 1.2$	$0.9 < \|\eta\| < 2.4$
Forward calorimeters (HF)	-	$3.0 < \|\eta\| < 5.2$

3.3. Overview of subdetectors

3.3.1. Superconducting magnet

The CMS detector is equipped with a large superconducting solenoid magnet [55], which provides an axial and nearly uniform magnetic field of 3.8 T over a length of 12.5 m and a free bore radius of 3.15 m. This geometry is sufficient to enclose both the ECAL and the HCAL, while minimising the amount of material in front of the calorimeters, thus preserving their energy resolution.

Under nominal operating conditions, the solenoid provides a bending power of about 4.9 Tm at the inner surface of the calorimeter system. This strong magnetic field enables precise curvature measurements of charged-particle tracks, which are essential for accurate momentum determination and for distinguishing between charged and neutral particle deposits. For example, a charged particle with transverse momentum $p_T = 20\,\mathrm{GeV}$is deflected in the transverse plane by approximately 5 cm when reaching the ECAL, a separation that is sufficient to distinguish it from a photon shower [56].

In combination with the finely segmented tracking and calorimeter systems, the solenoid magnet is a key feature that underpins the excellent momentum resolution and particle identification capabilities of CMS.

3.3.2. Yoke

Outside the superconducting solenoid, a massive steel return yoke is installed to suppress and guide the strong magnetic field produced by the magnet, thereby preventing uncontrolled fringe fields in the outer detector volume. The yoke also serves as an absorber for the muon system, providing both mechanical support and a magnetic field environment for bending muon tracks.

The yoke is composed of 11 large elements: 5 barrel wheels and 6 endcap disks (3 on each side). Each element is built from construction steel with strong ferromagnetic properties, allowing it to channel the magnetic flux lines back around the solenoid and thus close the magnetic circuit. The individual components are extremely massive, with weights ranging from about 400 t for the lightest endcap disks up to nearly 1920 t for the central barrel wheel, which houses the solenoid coil and its cryostat [57, 58].

In addition to providing flux return, the layered structure of the yoke is instrumented with muon detectors, making it an integral part of the overall muon system. This dual role - as a magnetic flux return path and as a host for muon chambers - is a defining feature of the CMS design and contributes to its excellent muon detection and momentum resolution capabilities.

3.3.3. Tracking system

The inner silicon tracking system is a cylindrical detector with an outer radius of about 1.2 m and a length of 5.6 m [59, 60]. It consists of a barrel section and two endcaps, providing coverage up to $|\eta| < 2.5$. The barrel is composed of three layers of pixel detectors surrounded by ten layers of silicon micro-strip detectors, while each endcap contains two pixel disks followed by twelve micro-strip disks. The pixel detector, with a typical granularity of $100 \times 150\ \mu\text{m}^2$ per pixel, provides excellent spatial resolution close to the interaction point and is crucial for efficient vertex reconstruction and b-tagging. The outer strip detectors extend the lever arm for momentum measurements and improve the overall track reconstruction efficiency.

Owing to its fine granularity and large number of readout channels, the tracker achieves excellent precision in reconstructing charged-particle trajectories. Under standard operating conditions, the transverse momentum of charged hadrons with $p_T < 20$ GeVcan be measured with a resolution of about 1% [56]. At higher p_T values, the relative resolution gradually decreases due to the reduced curvature of the tracks in the 3.8 T magnetic field, but the system still provides sufficient accuracy for reconstructing high-momentum particles produced in proton-proton collisions.

3.3.4. The electromagnetic calorimeter

The ECAL is a hermetic and homogeneous calorimeter constructed from lead tungstate ($PbWO_4$) crystals. The barrel crystals have a length of 22 cm, corresponding to about 25.8 radiation lengths (X_0), while the endcap crystals are also 22 cm long, corresponding to about 24.7 X_0. This depth is sufficient to absorb more than 98% of the energy of photons and electrons with energies up to 1 TeV.

The $PbWO_4$ material amounts to roughly one nuclear interaction length, which leads to about two thirds of charged hadrons initiating showers in the ECAL before reaching the HCAL. The very fine granularity of the calorimeter, with crystal cross sections of approximately $2.2 \times 2.2\ \text{cm}^2$ in the barrel (corresponding to $\Delta\eta \times \Delta\phi \approx 0.0174 \times 0.0174$), enables the separation of nearby electromagnetic showers, for example distinguishing photon deposits from hadronic activity within distances of order 5 cm [61, 62].

The relative energy resolution of the ECAL for electrons and photons can be parameterised as

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E\ [\text{GeV}]}} \oplus \frac{12\%}{E\ [\text{GeV}]} \oplus 0.3\%, \tag{3.5}$$

where the three terms correspond to the stochastic contribution, the noise term, and the constant term, respectively. The small stochastic term provides excellent energy resolution, particularly in the range $E = 1\text{-}50\,\text{GeV}$, which is typical for electrons from hadronic jets and for photons [56].

Overall, the ECAL is a key component of the CMS detector, providing precise measurements of electromagnetic objects such as photons and electrons, as well as contributing to the reconstruction of jets and the calculation of the missing transverse momentum.

3.3.5. The hadron calorimeter

The HCAL is a hermetic sampling calorimeter, composed of alternating layers of brass absorber and plastic scintillator tiles. It is located directly outside the ECAL and is divided into a central barrel section and two endcap disks, providing coverage up to $|\eta| \lesssim 3.0$. The scintillator tiles are read out via wavelength-shifting fibres coupled to hybrid photodiodes, which convert the light signal into electronic pulses for energy reconstruction.

In the barrel region, the absorber thickness corresponds to about six nuclear interaction lengths, increasing to over ten interaction lengths at larger pseudorapidities. This ensures that the majority of hadronic showers are fully contained within the calorimeter system. Beyond the endcaps, the Forward Hadronic Calorimeter (HF) calorimeter, based on Cherenkov light detection in quartz fibres, extends the coverage up to $|\eta| \approx 5$, providing sensitivity to very forward jets and particles [63].

The HCAL, together with the ECAL, forms the complete calorimetric system of CMS, which is crucial for the measurement of jet energies, missing transverse momentum, and for distinguishing hadronic from electromagnetic activity.

3.3.6. The muon detectors

The muon system of CMS is located outside the superconducting solenoid and is embedded in the massive steel return yoke, which guides the magnetic flux and prevents a strong, uncontrolled magnetic field in the outer detector volume. The yoke is segmented into three barrel wheels and four endcap disks on each side, providing space for four layers of muon stations arranged concentrically around the solenoid [58, 64].

The muon stations host three different detector technologies, optimised for complementary pseudorapidity regions:

- Drift Tubes (DTs): used in the barrel region ($|\eta| < 1.2$), providing precise position measurements in a low-background environment.
- Cathode Strip Chambers (CSCs): employed in the endcap regions ($0.9 < |\eta| < 2.4$), where the particle flux and radiation levels are higher, offering good spatial resolution and fast response.
- Resistive Plate Chambers (RPCs): installed in both barrel and endcaps up to $|\eta| < 1.6$, providing excellent time resolution and redundancy for the trigger system.

This combination of technologies ensures robust muon detection across a wide pseudorapidity range, allowing for efficient triggering, precise momentum measurements at high transverse momentum, and unambiguous muon identification, which is essential for many physics analyses.

3.3.7. The trigger and DAQ

The CMS data acquisition (Data Acquisition (DAQ)) system employs a two-level trigger architecture designed to select events that are potentially interesting for physics analyses [65]. The first stage, the Level-1 (L1) Trigger, is implemented in custom hardware based on fast Field-Programmable Gate Arrays (FPGA) and Application-Specific Integrated Circuits (ASIC). It receives information from the calorimeters (ECAL and HCAL) and the muon system and must reach a decision within about 4 μs, corresponding to the maximum available latency of the front-end electronics. The L1 Trigger reduces the event rate from the LHC bunch crossing frequency of 40 MHz to approximately 100 kHz, a rate at which full detector readout becomes feasible.

The second stage is the High Level Trigger (HLT), a software-based trigger system running on a large computing farm. Unlike the L1 trigger, the HLT has access to the full event information from all detector subsystems and performs a fast but sophisticated reconstruction of physics objects such as electrons, muons, jets, and missing transverse momentum. The HLT reduces the event rate further from $\sim$ 100 kHz to about 1 kHz, a rate that can be stored permanently for offline analysis. The entire trigger system thus ensures that rare and interesting events are efficiently selected while maintaining manageable data volumes for storage and subsequent physics studies.

3.4. Summary of detector performance

The CMS detector combines a high-granularity silicon tracker, a homogeneous crystal ECAL, a sampling HCAL, and a large-area muon system embedded in the return yoke of a 3.8 T superconducting solenoid. Together, these subsystems provide nearly hermetic coverage and complementary measurements, allowing the reconstruction and identification of a wide variety of final-state particles with high precision.

Table 3.2 summarizes the typical performance of CMS for key reconstructed objects. The excellent spatial resolution of the silicon tracker ensures precise momentum measurements of charged hadrons and muons, while the fine granularity and high light yield of the $PbWO_4$ crystal ECAL provide superior energy resolution for photons and electrons. The combined calorimetric system enables reliable jet reconstruction and missing transverse momentum measurements, while the muon detectors deliver robust identification and tracking in a wide pseudorapidity range.

Table 3.2.: Summary of the typical performance of the CMS detector for key physics objects in Run 2/3 [57, 56].

Physics object	Typical performance
Electrons/photons	Energy resolution $\sim$ 1% for $E \gtrsim$ 50 GeV(barrel)
Muons	p_T resolution $\sim$ 1% at $p_T = 50$ GeV(barrel)
Charged hadrons	p_T resolution $\sim$ 1% for $p_T < 20$ GeV(tracker-dominated)
Jets	Energy resolution $\sim$ 10% at $p_T = 100$ GeV
Missing E_T	Resolution $\sim$ 10-20 GeVfor low pile-up events

In summary, the CMS detector achieves excellent performance across a broad spectrum of final states. Its precise tracking, superior ECAL, reliable HCAL, and unique muon system make it a powerful general-purpose experiment, enabling both precision measurements within the SM and sensitive BSM searches.

3.5. Objects reconstruction in CMS

The excellent resolutions of the individual CMS subdetectors allow for precise measurements of different aspects of a particle's trajectory and energy deposition. To exploit this information in an optimal way, CMS employs a global event reconstruction strategy based on the Particle-Flow (PF) algorithm [56].

The core idea of the PF approach is to combine, or "link," detector signals from the tracking system, the calorimeters, and the muon detectors into a consistent set of reconstructed particles, referred to as PF candidates. This method takes advantage of the excellent momentum resolution of the silicon tracker for charged hadrons, the precise energy measurement of the ECAL for photons and electrons, and the large coverage of the muon system.

The geometrical linking of tracks and calorimeter deposits in the PF algorithm is illustrated in Figure 3.3.

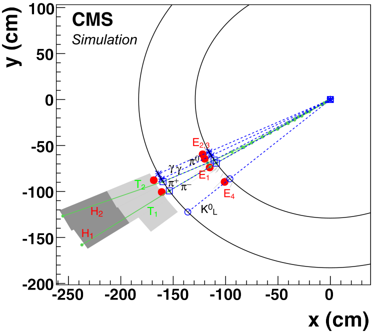


Figure 3.3.: Schematic view of the PF geometrical linking between a track, calorimeter deposits, and the absence of deposits in the muon system [56].

Compared to calorimeter-only reconstruction, the PF approach significantly improves the resolution of jets and missing transverse momentum, while maintaining excellent performance for electrons, photons, and muons. The resulting list of PF candidates serves as the common input for higher-level physics objects such as jets, τ leptons, and $\vec{p}_T^{\,\text{miss}}$, making PF the cornerstone of event reconstruction in CMS.

3.5.1. Tracking and Vertexing

Every charged particle traversing the inner detector leaves hits in the silicon tracking system. To reconstruct tracks from these hits, a Kalman Filtering (KF) [66] algorithm is employed. The procedure begins with the creation of track *seeds*, formed from a few hits that are geometrically compatible with the trajectory of a charged particle in the 3.8 T magnetic field. From these seeds, candidate trajectories are propagated through the tracker layers, where additional hits are collected to extend the track. Finally, a fitting step is performed to refine the track parameters and determine the transverse momentum, direction, and point of closest approach to the beam spot, which is used to reconstruct the Primary Vertex (PV).

To ensure high-quality tracks, selection criteria are applied. Tracks are typically required to have at least two hits in the pixel detector and a minimum of eight hits in total, with at most one missing hit along the expected trajectory. Only tracks that satisfy these requirements are retained for subsequent reconstruction steps and physics analyses.

3.5.2. Muons

Muons are reconstructed in CMS by combining information from the inner silicon tracker and the outer muon system, which is embedded in the steel return yoke of the solenoid and provides coverage up to $|\eta| \sim 2.4$.

Several reconstruction strategies are defined:

- **Standalone muons**: tracks built only from hits in the muon chambers.
- **Tracker muons**: tracks reconstructed in the silicon tracker that are matched to at least one muon segment in the muon chambers. This approach is efficient at low p_T where the muon may not leave enough hits in multiple stations of the outing tracking system.
- **Global muons**: tracks reconstructed by combining hits from both the tracker and the muon chambers, requiring track segments in at least two muon stations. This reconstruction provides excellent momentum resolution at higher p_T.

Muon identification builds upon these reconstruction categories. Quality requirements such as the number of hits in the tracker and muon chambers, the χ^2 of the global track fit, and compatibility with the primary vertex are used to suppress backgrounds from hadron decays in flight or punch-through. In addition, isolation requirements, based on the absence of significant additional activity around the muon direction, are commonly applied in analyses.

The performance of muon reconstruction in Run 2 was very good: the reconstruction efficiency exceeded 95% over a broad range of p_T and η, and the transverse momentum resolution was about 1% at $p_T = 50$ GeVin the barrel region, increasing to a few percent in the endcaps [64].

3.5.3. Photons and Electrons

Electrons and photons are reconstructed primarily in the ECAL. Since photons are electrically neutral, no matching track is expected in the tracker; they are identified as clusters of energy deposits in the ECAL without associated tracks. Electrons, on the other hand, are reconstructed by matching ECAL superclusters with tracks from the silicon tracker. The energy is measured in the ECAL, the momentum is obtained from the track, and the two are combined to achieve the best resolution.

To distinguish electrons and photons from hadronic backgrounds, Multivariate Analysis (MVA) techniques are applied. For electrons, the discriminating variables include shower-shape observables, track–cluster matching quantities, and isolation with respect to nearby activity. For photons, the identification relies on shower-shape variables and isolation from hadronic activity. Different identification Working Points (WPs) (loose, medium, tight) are defined to optimise efficiency and purity depending on the requirements of individual physics analyses.

Typical identification efficiencies are around 80-90% for electrons [67] and above 70% for photons [67].

3.5.4. Jets

Jets are collimated sprays of hadrons and other particles produced in the hadronisation of quarks and gluons. On average, about 65% of the jet energy is carried by charged hadrons, 25% by photons (mostly from neutral pion decays), and 10% by neutral hadrons [56].

In CMS, jets are reconstructed from PF candidates. The reconstruction is performed using sequential recombination algorithms, which iteratively merge pairs of particles according to a distance measure defined in the rapidity–azimuth plane:

$$d_{ij} = \min\left(p_{T,i}^{2\alpha}, p_{T,j}^{2\alpha}\right) \frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = p_{T,i}^{2\alpha}, \tag{3.6}$$

where $\Delta R_{ij} = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2}$, R is the jet radius parameter (commonly $R = 0.4$ or $R = 0.8$), and α defines the clustering scheme [68].

- $\alpha = 1$: k_T algorithm [69, 70]. Clustering begins with the softest particles, merging them iteratively into larger objects. This algorithm is useful for jet substructure studies but tends to form irregular jet shapes.
- $\alpha = 0$: Cambridge–Aachen (CA) algorithm [71, 72]. Clustering proceeds purely based on angular distance, independent of p_T. This algorithm is often used for boosted-object tagging.
- $\alpha = -1$: anti-k_T algorithm [68]. Clustering starts effectively from the hardest particles, producing nearly conical jets with well-defined boundaries. This is the default jet algorithm in CMS.

In addition to clustering, jets require calibration and identification. *Jet Energy Corrections (JEC)* [73] are applied to account for the detector response, pile-up, and residual differences between data and simulation. Furthermore, a dedicated *Jet Identification (Jet ID)* [74] is applied to suppress spurious jets originating from calorimeter noise, beam halo, or pile-up interactions. After these corrections and quality selections, jets provide a reliable representation of the underlying parton kinematics and serve as essential inputs for many physics analyses.

3.5.5. Missing transverse energy

By definition, the missing transverse momentum (Missing Transverse energy (MET)) is the negative vectorial sum of the transverse momenta of all reconstructed particles in the event:

$$\vec{p}_T^{\,\text{miss}} = -\sum_i \vec{p}_{T,i}. \tag{3.7}$$

The observable is especially important for SUSY and other BSM searches, where dark matter candidates or other weakly interacting particles may leave no signal in the detector, appearing only as an imbalance in the transverse momentum. In CMS, MET is typically reconstructed from particle-flow (PF) candidates, which combine information from all subdetectors to provide the most accurate event description.

Pile-up interactions can degrade the MET resolution by contributing additional, uncorrelated transverse momentum. To mitigate this effect, advanced algorithms such as Pile-up per particle identification (PUPPI) [75] are employed to suppress or downweight contributions from particles not associated with the PV, thereby improving the MET resolution under high pile-up conditions.

4. τ_h lepton reconstruction and identification

4.1. τ_h lepton reconstruciton via HPS algorithm

4.1.1. τ_h lepton properties

The τ lepton is the heaviest lepton in the SM, with a mass of 1776.86 ± 0.12 MeV/c^2 [76]. Among the leptons it couples most strongly to the Higgs boson through the Yukawa interaction, making the $H \longrightarrow \tau^+\tau^-$ channel the most sensitive purely leptonic probe of the Higgs sector.

Due to its short lifetime of $t = 2.9 \times 10^{-13}$ s, corresponding to a decay length of $c\tau \approx 87$ μm, the τ lepton decays well before reaching the inner layers of the CMS tracking system (beam pipe radius $\sim$ 3 cm). As a consequence, τ leptons cannot be reconstructed directly but are instead identified through their visible decay products. At higher energies, the relativistic time dilation extends the mean decay length: for example, at $E \approx 200$ GeVthe decay length increases to about $c\tau \approx 10.2$ mm, though still short compared to the tracker dimensions.

In the $H \longrightarrow \tau\tau$ process, the subsequent τ decays retain the spin information of the τ leptons. This property makes it an excellent probe of the spin and *CP* structure of the parent Higgs boson. In the following, the discussion will focus on hadronically decaying τ leptons (τ_h), which represent the majority of τ decays and provide distinctive experimental signatures.

The main τ decay modes and their branching fractions are summarized in Table 4.1. Approximately 35% of τ decays are leptonic, producing an electron or a muon together with two neutrinos. The remaining ~65% are hadronic decays, in which the τ decays into one or more charged hadrons (predominantly pions) accompanied by neutrinos. The hadronic decays are dominated by so-called "1-prong" channels, in which a single charged track is reconstructed, and "3-prong" channels, which yield three charged tracks, often from the decay of an intermediate $a_1^\pm$ resonance. Rare multi-prong modes contribute only a small fraction of the branching fraction.

Table 4.1.: Main τ lepton decay modes and their branching fractions [76].

Decay mode	**Branching fraction (%)**
Leptonic decays ($\tau \to e\,\nu_e\,\nu_\tau$)	17.8
Leptonic decays ($\tau \to \mu\,\nu_\mu\,\nu_\tau$)	17.4
Hadronic decays (1-prong, mostly $\pi^\pm\nu_\tau$, $\rho^\pm\nu_\tau$)	~49.5
Hadronic decays (3-prong, mostly $a_1^\pm\nu_\tau$)	~15.2
Other hadronic decays (rare multi-prong modes)	~0.1
Total hadronic decays	**~65**
Total leptonic decays	**~35**

Experimentally, the leptonic τ decays are relatively clean, since the final-state electron or muon can be reconstructed with high purity. In contrast, the hadronic τ decays produce narrow, low-multiplicity jets that must be distinguished from the abundant QCD jet background. This makes τ_{had} identification both experimentally challenging and particularly important, since hadronic decays constitute the majority of all τ decays.

4.1.2. HPS reconstruction algorithm

Hadronic τ reconstruction in CMS is performed using the Hadron-Plus-Strip (HPS) algorithm [77].

First, the reconstruction starts from jet seeds. The seeding jets are clustered from all particles reconstructed by the PF algorithm using the anti-k_T algorithm with a distance parameter of $R = 0.4$. For the subsequent steps of τ_h reconstruction, all particles located within a cone of radius $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.5$ around the jet axis are considered.

In the next step, the algorithm searches for neutral pion (π^0) constituents. Neutral pions decay into two photons, $\pi^0 \to \gamma\gamma$, which have a high probability of converting into e^+e^- pairs in the detector material. The algorithm clusters electrons and photons with $p_\text{T} > 0.5$ GeVinto strips. During Run 1, a fixed strip size of 0.05×0.2 in $\eta \times \phi$ was originally used. Later in Run 2, a dynamic strip reconstruction method was adopted, in which the strip size depends on the energy of its constituents. The four-momenta of electrons and photons are added.

After forming the strips, the algorithm combines them with the charged constituents of the jet, i.e., charged pions $\pi^{\pm}$. The four-momenta of τ_h candidates are formed the following way:

$$p_{\mu}^{\tau_h} = \sum_{i \in \text{charged}} p_{\mu}^{i} + \sum_{j \in \text{strips}} p_{\mu}^{j}, \tag{4.1}$$

where the four-momenta of the τ_h candidate, $p_{\mu}^{\tau_h}$, are obtained by summing the four-momenta of the charged hadrons and the strips included in the reconstruction according to the selected Decay Mode (DM).

Then HPS algorithm rejects all candidates whose masses are not compatible with those of mesonic resonances expected in the τ decay chain, namely the $\rho(770)$ and $a_1(1260)$ mesons [78, 79].

Figure 4.1 shows the distribution of the reconstructed visible invariant mass of τ_h candidates.

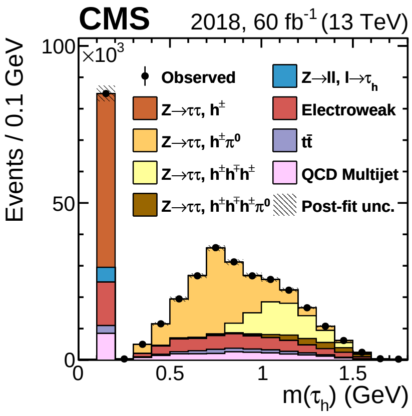


Figure 4.1.: Distribution of the reconstructed visible invariant mass of τ_h candidates for various DMs. For DM0, the τ_h mass corresponds to that of a charged pion, resulting in a peak around 1.4 GeV(close to the charged pion mass of approximately 1.39 GeV) [80].

Depending on the number of strips and charged hadrons, the HPS algorithm reconstructs the following DM:

- **DM0**: one charged hadron and no strip candidates, with $0 < m_{\tau_h} < 1\,\text{GeV}$
- **DM1**: one charged hadron and one strip, with

  $$0.3\text{ GeV} - \Delta m_{\tau_h} < m_{\tau_h} < 1.3\text{ GeV} \cdot \sqrt{p_T^{\tau_h}/(100\text{ GeV})} + \Delta m_{\tau_h}$$

  for the $h^{\pm}\pi^0$ mode. The mass window is enlarged for high-p_{T} τ_h candidates to account for resolution effects. The upper limit is constrained to lie between 1.3 and 4.2 GeV.
- **DM2**: one charged hadron and two strips, with

  $$0.4\text{ GeV} - \Delta m_{\tau_h} < m_{\tau_h} < 1.2\text{ GeV} \cdot \sqrt{p_T^{\tau_h}/(100\text{ GeV})} + \Delta m_{\tau_h}$$

  for the $h^{\pm}\pi^0\pi^0$ mode. The upper limit is restricted to lie between 1.2 and 4.0 GeV.
- **DM10**: three charged hadrons and no strips, with $0.8 < m_{\tau_h} < 1.5\,\text{GeV}$ for the $h^{\pm}h^{\mp}h^{\pm}$ mode.
- **DM11**: three charged hadrons and one strip, with

  $$0.9\text{ GeV} - \Delta m_{\tau_h} < m_{\tau_h} < 1.6\text{ GeV} + \Delta m_{\tau_h}$$

  for the $h^{\pm}h^{\mp}h^{\pm}\pi^0$ mode.

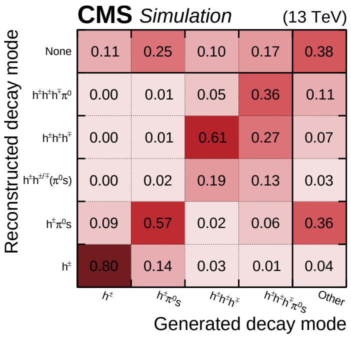
The DM number is defined by the formula:

$$\mathrm{DM} = 5 \cdot (N_{\text{prong}} - 1) + N_{\text{strip}}, \tag{4.2}$$

where N_{prong} is the number of charged hadrons and N_{strip} is the number of reconstructed neutral pions (π^0).

In practice, DM2 is often merged into DM1, as their reconstructed masses are close to each other, making a reliable separation difficult.

Figure 4.2 shows the confusion matrix representing the reconstruction efficiency of τ_h candidates across different DMs.



	$h^{\pm}$	$h^{\pm}\pi^0$s	$h^{\pm}h^{\pm}h^{\mp}$	$h^{\pm}h^{\pm}h^{\mp}\pi^0$s	Other
None	0.11	0.25	0.10	0.17	0.38
$h^{\pm}h^{\pm}h^{\mp}\pi^0$	0.00	0.01	0.05	0.36	0.11
$h^{\pm}h^{\pm}h^{\mp}$	0.00	0.01	0.61	0.27	0.07
$h^{\pm}h^{\pm/\mp}(\pi^0$s)	0.00	0.02	0.19	0.13	0.03
$h^{\pm}\pi^0$s	0.09	0.57	0.02	0.06	0.36
$h^{\pm}$	0.80	0.14	0.03	0.01	0.04

Figure 4.2.: Distribution of the reconstructed visible invariant mass of τ_h candidates for various DMs. For DM0, the τ_h mass corresponds to that of a charged pion, resulting in a peak around 1.4 GeV(close to the charged pion mass of approximately 1.39 GeV) [80].

The fraction of non-reconstructable τ_h candidates ranges from 11% for DM0 to 25% for DM1. The overall reconstruction efficiency is primarily limited by the ability to reconstruct tracks from charged hadrons, which is approximately 90%. The best-identified DM is DM0, with a reconstruction efficiency of 80%, due to its simple topology: a single charged pion track and no neutral pions. It is followed by DM1, consisting of one charged pion and one neutral pion, with an identification efficiency of 57%.

DM10, characterized by a clear signature of three charged pion tracks, reaches an efficiency of 61%, while DM11, which includes three charged pion tracks and one neutral pion, has a lower efficiency of 36%. In both DM10 and DM11, there is a possibility that one of the charged pion tracks is not reconstructed due to inefficiencies in the PF algorithm. As a result, the object may be misidentified as a two-prong τ_h candidate instead of a three-prong one.

These misreconstructed two-prong τ_h candidates contribute to the two-prong row of the confusion matrix. However, such DMs were generally not considered in Run 2 analyses due to poor charge assignment and the ambiguity associated with these configurations [80].

This chapter described the reconstruction of hadronically decaying τ leptons (τ_h) in CMS using the HPS algorithm. The reconstruction performance varies across DMs, with efficiency primarily limited by charged track reconstruction. Misidentification, particularly in complex modes like DM11, is illustrated through a confusion matrix, highlighting the appearance of two-prong candidates due to track loss. Overall, the HPS algorithm achieves a good balance between efficiency and background rejection and plays a central role in τ_h identification at CMS.

4.1.3. τ_h identification using convolutional neural network

Following the reconstruction of τ_h candidates with the HPS algorithm, the identification of genuine hadronic tau decays is performed using the DeepTau algorithm [80]. DeepTau is a deep convolutional neural network trained to distinguish real τ_h decays from background objects such as quark and gluon jets, electrons, and muons. It performs multiclass classification using both low-level and high-level input features.

Low-level inputs include information about individual PF candidates,such as tracks and energy deposits,within a fixed-size $\eta \times \phi$ grid centered on the reconstructed τ_h direction. These inputs are processed through convolutional layers that exploit the spatial correlations between PF objects in the vicinity of the τ_h. High-level variables, such as the transverse momentum, energy, and pseudorapidity of the τ_h candidate and its constituents, are passed through a set of fully connected layers.

Figure 4.3 illustrates the architecture of the DeepTau algorithm. The input consists of low-level PF candidates,such as charged hadrons ($\pi^{\pm}$), neutral pions (π^0), photons (γ), and electrons ($e^{\pm}$),organized in an η–ϕ grid centered on the HPS candidate. These features are passed through a series of convolutional blocks that learn spatial correlations and patterns specific to genuine τ_h decays. Simultaneously, high-level features like the transverse momentum and total energy (E) of the HPS candidate are processed through fully connected layers. The network outputs class probabilities, which are later used to construct final discriminants against background objects.

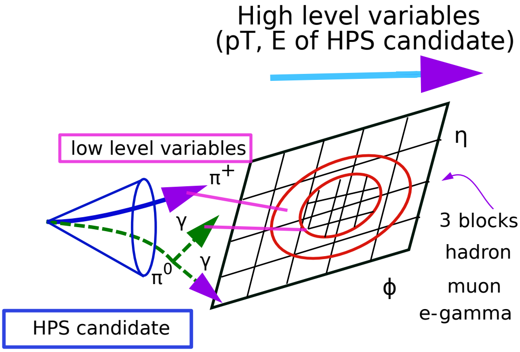


Figure 4.3.: Schematic representation of the DeepTau algorithm used for τ_h identification in CMS. The input to the network consists of low-level PF information distributed in a fixed-size η–ϕ grid around the HPS candidate, including charged hadrons, photons, and electrons. This spatial information is processed through convolutional blocks. In parallel, high-level variables such as the transverse momentum and energy of the HPS candidate are passed through fully connected layers. The outputs are class scores used to distinguish genuine τ_h decays from jets, electrons, and muons.

The output of the network is a set of probabilities,

$$y_{\text{prob}} = (y_{\tau_h}, y_{\text{jet}}, y_\mu, y_e),$$

representing the likelihood that the input corresponds to a genuine τ_h, jet, muon, or electron. These values are used to construct binary discriminants against each background class. For instance, the discriminant against jets is defined as

$$D_\alpha(y) = \frac{y_{\tau_h}}{y_{\tau_h} + y_\alpha}, \tag{4.3}$$

where $\alpha \in \{\text{jet}, \mu, e\}$.

WPs (e.g., loose, medium, tight) are defined by applying thresholds to these discriminants. The τ_h candidates are required to pass particular Working Point (WP) of the `vsJet`, `vsMu`, `vsEle` discriminant to suppress jet, muon and electron misidentification respectively while retaining high signal efficiency.

4.1.4. DeepTau v2.1 vs DeepTau v2.5

DeepTau v2.5 is an improved version of its predecessor DeepTau v2.1.

The most important change - is an introduction of domain adaptation techniques - a process of training the network on a mixture of Monte Carlo (MC) simulated samples and detector data to mitigate data/MC mismodelling. This is implemented by an introduction of a parallel neural network that distinguishes data from simulation while the main part classifies τ_h candidates.

The DeepTau architecture, illustrated in Figure 4.4, is designed to optimally use both low-level and high-level features for τ_h identification. Low-level inputs are taken from PF candidates in two concentric grids around the HPS τ_h axis: an 11×11 "inner" grid and a 21×21 "outer" grid in η–ϕ space. Each cell contains categorized inputs from electrons/photons, muons, and hadrons. These grids are processed separately by convolutional layers to extract spatial features.

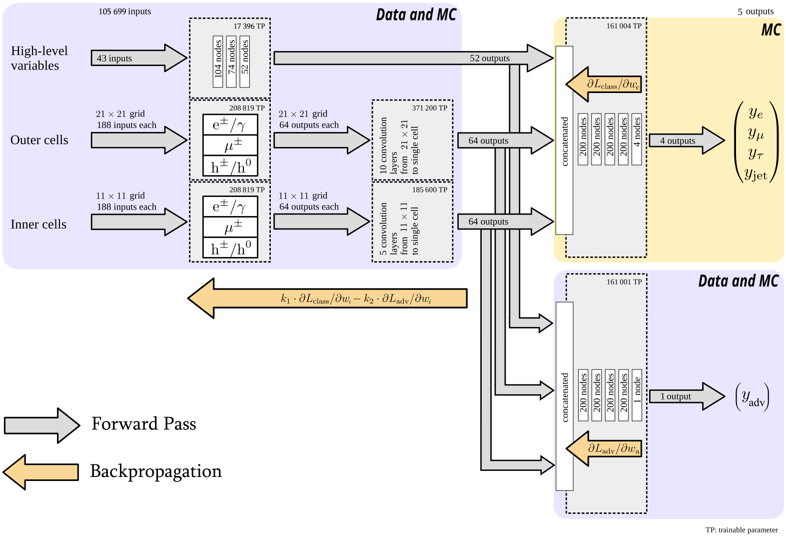


Figure 4.4.: Architecture of the DeepTau algorithm used in CMS for τ_h identification. The network combines multiple inputs: high-level variables, as well as low-level PF information from inner and outer cells centered around the HPS τ_h candidate. The low-level inputs are arranged into $\eta \times \phi$ grids (11×11 for inner cells, 21×21 for outer cells), containing PF candidates categorized as $e^{\pm}/\gamma$, $\mu^{\pm}$, and $h^{\pm}/h^{0}$. These are processed by convolutional layers and combined with high-level features (such as p_{T}, η, and energy). The resulting outputs are passed through dense (fully connected) layers to produce final class scores for electron (y_e), muon (y_μ), tau (y_τ), and jet (y_{jet}) hypotheses. An additional adversarial output y_{adv} is used to improve robustness against simulation mismodeling. Both forward propagation (gray arrows) and backpropagation (orange arrows) are shown. [81]

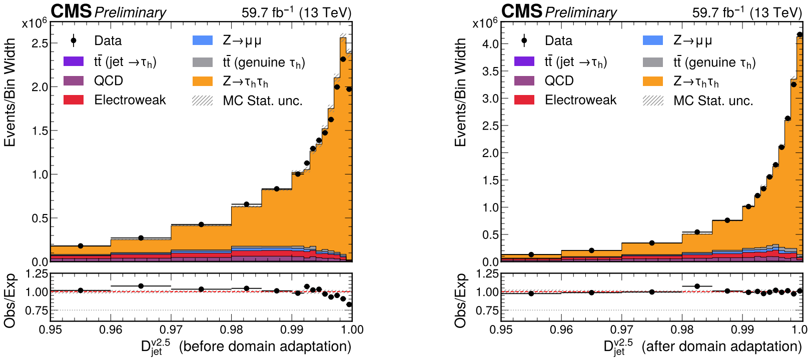
In parallel, high-level features such as the τ_h candidate's transverse momentum, energy, and isolation variables are processed as a separate input branch. The outputs of the convolutional and high-level branches are concatenated and passed through fully connected layers. The network produces four class probabilities corresponding to e, μ, τ_h, and jet hypotheses. Additionally, an adversarial branch predicts a simulation-specific output y_{adv}, trained with a separate loss to reduce dependence on simulation features and improve agreement with data. Forward and backward passes are indicated by gray and orange arrows, respectively.

Mathematically, domain adaptation in DeepTau is implemented via a modified loss function:

$$L_{\text{tot}} = k_1 \cdot L_{\text{class}} - k_2 \cdot L_{\text{adv}}, \tag{4.4}$$

where k_1 and k_2 are hyperparameters that control the contributions of the classification and adversarial components, respectively. In the final training configuration, these are set to $k_1 = 1$ and $k_2 = 10$. The negative sign in front of L_{adv} ensures that the main network is optimized to *maximize* the adversarial loss, thereby preventing the adversarial branch from effectively distinguishing between data and simulation. This mechanism encourages the network to learn features that are insensitive to domain-specific discrepancies, thus improving generalization to real data.

The impact of domain adaptation can be observed in Figure 4.5.



(a) Data/MC agreement before domain adaptation.

(b) Data/MC agreement after domain adaptation.

Figure 4.5.: Comparison of distributions before and after domain adaptation. An improved agreement between data and simulation is particularly visible in the last bin, as the neural network was trained to focus on high `vsJet` discriminant values [81].

An additional improvement introduced in comparison to DeepTau v2.1 is the use of the *Shuffle and Merge (S&M)* procedure. Since some training samples are significantly larger than others, training directly on the full dataset would result in a biased model, potentially leading to overfitting or poor generalization.

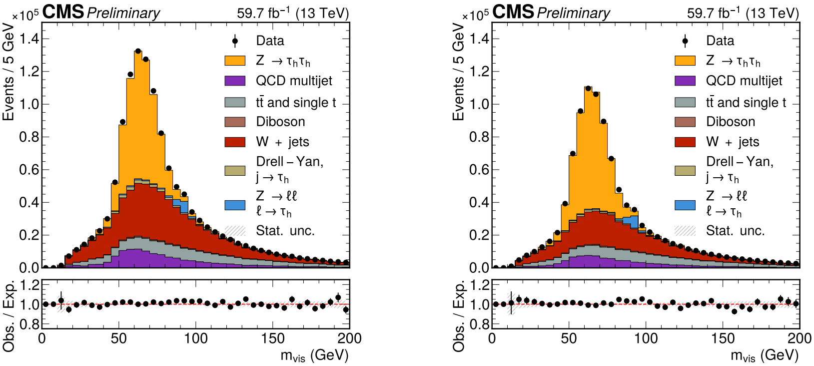
To mitigate this, a S&M strategy is applied. The method assigns a sampling probability to each τ_h candidate, proportional to:

$$p(\tau_h) \sim \frac{P_{\text{target}}(x_s(\tau_h))}{N(\text{bin, group})}, \tag{4.5}$$

where $P_{\text{target}}(x_s(\tau_h))$ represents the desired frequency of the event in the target distribution, and $N(\text{bin, group})$ is the number of events in the corresponding bin and group of the input data. This weighting ensures that rare events and underrepresented classes are sufficiently sampled and included in the final training dataset.

Another crucial step is the *feature standardization* process, which is applied to suppress numerical artifacts. All input features are scaled to lie within the range $[-1, 1]$, ensuring consistent feature magnitudes and improved numerical stability during training.

Overall, the introduction of these improvements results in a significant reduction of the fake τ_h background, by approximately 30% for a given τ_h identification efficiency . This effect is clearly visible in Figure 4.6.



(a) DeepTau v2.1

(b) DeepTau v2.5

Figure 4.6.: Comparison of m_{vis} distributions using DeepTau v2.1 and v2.5 as discriminants. The fake τ_h rate is reduced by approximately 30% with DeepTau v2.5, particularly in the W+jets background [81].

The improvements demonstrate enhanced robustness and generalization of the τ_h identification performance with the updated version of the algorithm.

4.2. Limitations of the standard reconstruction and identification in a boosted regime

Although the HPS reconstruction algorithm and the DeepTau identification technique demonstrate high efficiency under standard conditions, their performance becomes limited in scenarios involving two boosted hadronically decaying tau leptons (τ_h).

Let us consider the process illustrated in Figure 2.5, where a jet recoils against a Higgs boson. In such a configuration, the Higgs boson acquires a significant transverse momentum, which leads to a boost of its decay products. As a result, the two tau leptons produced in the Higgs decay are highly collimated , their separation in angular space (ΔR) becomes small due to the Lorentz boost of the mother particle.

A schematic view of this boosted topology is shown in Figure 4.7.

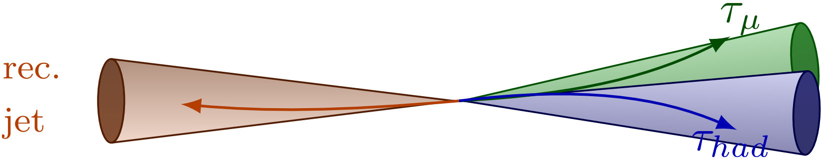


Figure 4.7.: Illustration of a boosted Higgs boson decaying into two tau leptons , one decaying hadronically (τ_h), the other leptonically into a muon (μ), in the presence of a jet recoiling against the Higgs boson.

In this topology, the decay products of one tau lepton may spatially overlap with those of the second tau lepton. This overlap poses a challenge for the standard τ_h reconstruction algorithms, such as HPS, which are designed to identify isolated and well-separated decay products. As a result, in boosted regimes, where isolation criteria are not fulfilled and multiple tau candidates share the same jet cone, the efficiency of standard algorithms is significantly reduced.

To improve reconstruction in such complex environments, dedicated techniques that account for tau overlap and boosted decay topologies are required, often involving jet substructure analysis or machine learning–based taggers.

4.3. Boosted τ_h tagging using ParticleNet algorithm

ParticleNet is a Graph Neural Network (GNN)-based algorithm developed for jet tagging, i.e., identifying the origin of jets (for example, distinguishing jets initiated by quarks or gluons) [82]. The key idea behind the algorithm is to take jets reconstructed using standard anti-k_T algorithms (AK4 or AK8) and predict their origin based on the internal structure of their constituent particles.

There are several advantages to using ParticleNet over traditional algorithms, particularly for tagging boosted topologies such as highly energetic τ leptons:

- In boosted regimes (e.g., decays of high-p_{T} Higgs bosons), resolving the internal substructure of jets is crucial for distinguishing signal from background.
- ParticleNet's graph-based architecture is specifically optimized to exploit the spatial and relational information among jet constituents.

The network architecture is based on the Dynamic Graph Convolutional Neural Network (DGCNN) framework [83] and employs Edge Convolution (Edge-Conv) layers to capture relational information between particles. A schematic illustration of the Edge-Conv operation is shown in Figure 4.8.

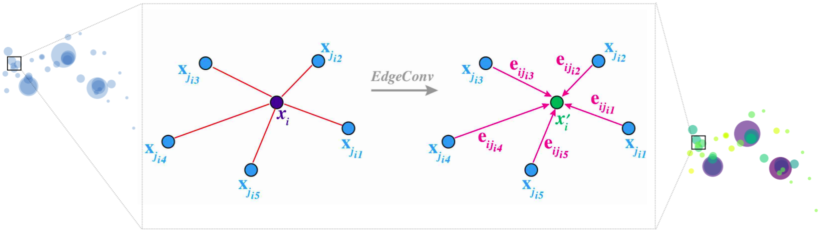


Figure 4.8.: Illustration of the Edge-Conv operation on a particle jet [82].

The algorithm treats each jet as a set of unordered particles in space,referred to as a "particle cloud",where each particle is described by a vector of features, such as transverse momentum (p_{T}), pseudorapidity η, and azimuthal angle ϕ. The input to the network consists of the PF constituents of the selected jet.

For each particle, the DGCNN constructs a graph by identifying its k-nearest neighbors in feature space and applies the Edge-Conv operation to define an "edge feature" between the central particle and its neighbors:

$$h_\Theta(x_i, x_{ij}) = \bar{h}_\Theta(x_i, x_{ij} - x_i), \tag{4.6}$$

where x_i denotes the feature vector of the central node (e.g., a particle in the jet), x_{ij} is the feature vector of its neighbor in the graph, h_Θ is a learnable function defining how edge features are computed, and $\bar{h}_\Theta$ is a reformulated function that operates on both the central particle's features and their relative difference.

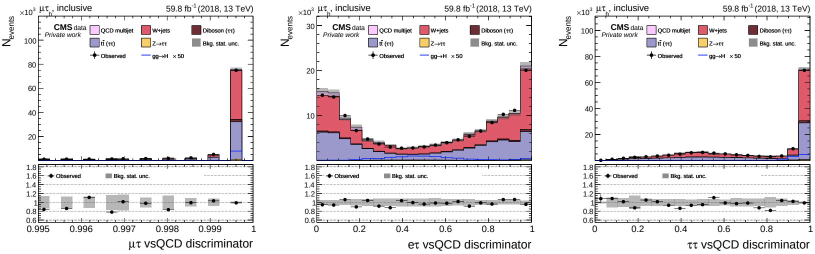
To ensure permutation invariance, the edge features from neighboring particles are aggregated using a symmetric function, such as summation, mean, or maximum. This aggregation guarantees that the learned representation is independent of the ordering of particles within the jet, which is required given the unordered nature of particle clouds.

Furthermore, ParticleNet is designed to capture both local and global structures within jets. Local structures,i.e., fine-grained correlations among nearby constituents,are learned through successive Edge-Conv layers. Global features are extracted through pooling operations that combine information from all particles, enabling the network to form a comprehensive representation of the jet. This combination of local and global feature learning significantly improves the network's ability to classify complex jet substructures.

4.4. Properties of Boosted τ_h via ParticleNet

4.4.1. ParticleNet discriminants

In the boosted regime, ParticleNet provides three discriminants corresponding to the final states $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$. These `vsJet` discriminants are specifically trained to distinguish genuine boosted di-τ signatures from generic QCD jets. They are designed to identify the visible DMs of a highly boosted di-τ system within a single AK8 jet, where one or both τ leptons decay visibly. The distributions of these discriminants are shown in Figure 4.9.



(a) Distribution of the $\mu\tau_h$ discriminant.

(b) Distribution of the $e\tau_h$ discriminant.

(c) Distribution of the $\tau_h\tau_h$ discriminant.

Figure 4.9.: Distributions of the ParticleNet `vsJet` discriminants for tagging boosted di-τ final states: $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$. The contribution from the $H \rightarrow \tau\tau$ process is enhanced by a factor of 50. A cut of ParticleNet score > 0.995 is applied to the $\mu\tau_h$ discriminant, while no cuts are applied to the remaining two.

The $\mu\tau_h$ `vsJet` discriminant shows a sharp rise near unity, which is attributed to numerical effects at extremely high selection efficiencies. A large fraction of events accumulates in the final histogram bins, indicating that this score is highly discriminative for signal-like boosted $\mu\tau_h$ topologies.

These discriminants are trained to distinguish genuine τ-lepton signatures from QCD-like jets using both the particle-level structure and correlations captured by the graph neural network. The availability of separate discriminants allows for dedicated selection strategies for different decay channels, enabling improved signal sensitivity in analyses targeting various τ final states.

As discussed above, ParticleNet operates on reconstructed AK8 jets,commonly referred to as *fatjets*,and attempts to determine whether such jets are consistent with the collimated decay products of a di-τ system. This tagging is particularly relevant in scenarios where the Higgs boson is produced with high transverse momentum, resulting in the visible τ decay products being contained within a single large-radius jet.

It is worth noting that although the discriminants target distinct final states, some overlap between them may occur due to similarities in the jet substructure of $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$ decays, especially at low resolution or in presence of detector effects. Therefore, careful tuning of thresholds and cross-validation with generator-level information is essential when applying these scores in analysis selections for event categorization.

To assess the idealized performance of the tagger independent of reconstruction effects, a Receiver Operating Characteristic (ROC) curve is constructed based on generator-level matching. The ROC curve is a graphical representation of the classifier's ability to discriminate between signal and background by plotting the true positive rate (signal efficiency) against the false positive rate (background misidentification rate) as the discriminant threshold is varied. The Area Under the ROC Curve (AUC) quantifies the overall discriminating power of the classifier, with an AUC of 1.0 corresponding to a perfect separation of signal and background.

We focus on the $\mu\tau_h$ final state, where a hadronically decaying tau is reconstructed as a fatjet and the accompanying muon originates from the same Higgs boson decay. Signal fatjets are defined as those from Higgs $\rightarrow \tau\tau$ simulated samples that are matched to a generator-level muon stemming from a Higgs boson decay via an intermediate tau lepton, within an angular distance of $\Delta R < 0.8$ from that muon, and are positively tagged by the ParticleNet `vsJet` $\mu\tau_h$ discriminant, defined as having a discriminant score greater than 0. Background fatjets are taken from simulated W+jets events and are defined as those not matched to such muons but tagged by ParticleNet.

4.4.2. ParticleNet Tagging

While ParticleNet is a powerful tagging tool, it is important to consider its limitations. For instance, Higgs or Z bosons are often produced in association with recoiling jets. In some cases, such recoiling jets may mimic a di-τ topology and be incorrectly tagged by the discriminant.

To study this behavior, it is useful to examine the angular separation, $\Delta\phi$, between the tagged fatjet and the generator-level $\mu\tau_h$ system originating from a common mother particle.

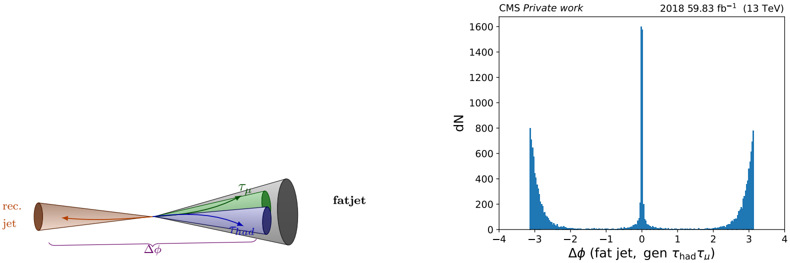
This observable is added as the angular separation $\Delta\phi$. It is defined as the absolute difference in azimuthal angle between the reconstructed fatjet and the generator-level $\mu\tau_h$ system:

$$\Delta\phi = \left|\phi_{\text{fatjet}} - \phi_{\mu\tau}^{\text{gen}}\right|, \quad \Delta\phi \in [0, \pi]. \tag{4.7}$$

Here, ϕ_{fatjet} denotes the azimuthal angle of the reconstructed AK8 jet tagged by the ParticleNet `vsJet` $\mu\tau_h$ discriminant, and $\phi_{\mu\tau}^{\text{gen}}$ refers to the azimuthal angle of the visible system formed from the generator-level muon and hadronic tau originating from the Higgs boson decay. The absolute value ensures a non-negative result, and the range is wrapped into $[0, \pi]$ due to the periodic nature of angular coordinates. A value of $\Delta\phi \approx 0$ corresponds to a collinear configuration, which is expected when the tagged fatjet correctly matches the direction of the boosted $\mu\tau_h$ system. Conversely, a value near $\Delta\phi \approx \pi$ indicates that the tagged fatjet is nearly opposite in azimuth to the generator-level $\mu\tau_h$ system, suggesting a possible misidentification of a recoiling jet as signal.

For this purpose, simulated Higgs boson events with $p_{\text{T}} > 200$ GeVare used, where one τ decays hadronically and the other into a muon.

Figure 4.10 shows both a sketch of the $\Delta\phi$ observable (left) and the corresponding distribution between the fatjet (tagged by the ParticleNet `vsJet` $\mu\tau_h$ discriminant) and the generator-level $\mu\tau_h$ system (right).



(a) Illustration of $\Delta\phi$ between the tagged fatjet and the generator-level $\mu\tau_h$ pair.

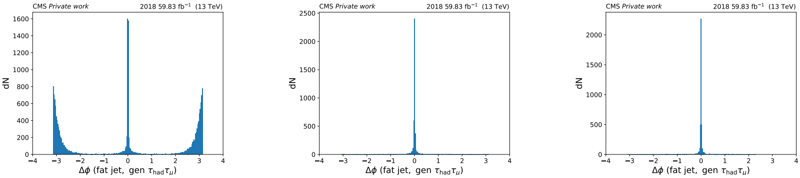
(b) Distribution of $\Delta\phi$ between tagged fatjet and gen-level $\mu\tau_h$ pair.

Figure 4.10.: Comparison of the tagged fatjet and generator-level $\mu\tau_h$ system. Left: schematic illustration of the $\Delta\phi$ observable. Right: distribution of $\Delta\phi$ between the fatjet tagged by the $\mu\tau_h$ discriminant and the $\mu\tau_h$ pair from Higgs boson decays. The central peak corresponds to correct tagging, while the side peaks are due to recoiling jets being misidentified.

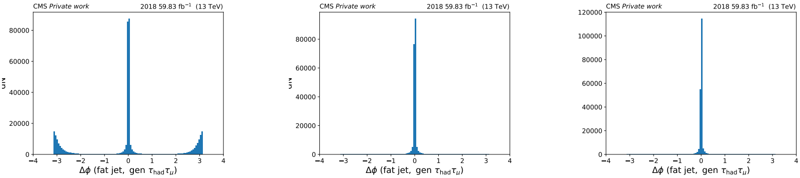
The distribution exhibits a pronounced peak at $\Delta\phi = 0$, corresponding to correctly tagged Higgs boson decays where the fatjet and the $\mu\tau_h$ system are collinear. In contrast, the side peaks near $\Delta\phi = \pi$ arise from cases where the tagged fatjet is opposite in direction to the true di-τ system, indicating that a recoiling jet was mistakenly tagged.

The distributions shown above are obtained without applying any ParticleNet discriminant. It is therefore reasonable to assume that imposing a tighter ParticleNet selection would reduce the number of incorrectly tagged jets.

Figure 4.11 illustrates the effect of the ParticleNet `vsJet` $\mu\tau_h$ discriminant on the suppression of recoiling jets that were falsely identified as boosted $\mu\tau_h$ pairs.



(a) No ParticleNet cut **(b)** $\mu\tau_h$ ParticleNet > 0.15 **(c)** $\mu\tau_h$ ParticleNet > 0.9



(d) No ParticleNet cut **(e)** $\mu\tau_h$ ParticleNet > 0.15 **(f)** $\mu\tau_h$ ParticleNet > 0.9

Figure 4.11.: Distributions of the angular separation $\Delta\phi$ between the fatjet tagged by the ParticleNet `vsJet` $\mu\tau_h$ discriminant and the generator-level $\mu\tau_h$ pair, shown for different discriminant thresholds. The upper row corresponds to Higgs boson events with $p_T > 200$ GeV, and the lower row to Z boson events with $p_T > 100$ GeV. The plots illustrate how tighter WPs suppress recoiling jets that are falsely identified as boosted $\mu\tau_h$ candidates.

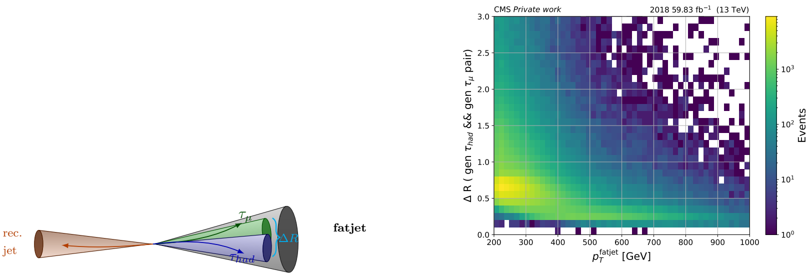
As expected, applying a stricter WP significantly reduces the number of misidentified candidates. The upper row of the figure shows the distributions of the angular separation $\Delta\phi$ for Higgs boson events with $p_T > 200$ GeV, while the lower row corresponds to Z boson events with $p_T > 100$ GeV.

For the Higgs boson sample, applying a $\mu\tau_h$ discriminant cut of > 0.15 removes approximately 77% of falsely tagged jets, and a tighter cut of > 0.9 removes about 78%. In the case of the Z boson sample, a cut of > 0.15 eliminates 44% of falsely identified jets, while a cut of > 0.9 removes 45%. This behavior suggests that a moderately tight WP already provides effective suppression, while more aggressive cuts yield diminishing returns in terms of additional background rejection.

In this context, falsely tagged jets are defined as those not matched to a generator-level $\mu\tau_h$ system within a reasonable angular distance. The chosen discriminant thresholds,0.15 as a loose WP and 0.9 as a tight one,are used to study the trade-off between background rejection and signal retention. Although stricter cuts reduce the background more effectively, they may also suppress signal events. Hence, selecting the optimal WP requires balancing tagging efficiency against background contamination.

4.4.3. Properties of boosted di-τ pair

The angular separation (ΔR) between the visible decay products of a di-τ system specifically, a muon and a hadronically decaying τ lepton, decreases as the transverse momentum of the parent particle increases. This behavior is a direct consequence of Lorentz boosting, which causes the decay products to become more collimated at higher energies. The correlation between $\Delta R(\tau_\mu, \tau_h)$ and the transverse momentum of the reconstructed fatjet is illustrated in Figure 4.12. The fatjet serves as a proxy for the parent boson's boost, and is tagged using the ParticleNet `vsJet` $\mu\tau_h$ discriminant.



(a) Schematic illustration of the Lorentz boost effect, where the decay products of a high-p_{T} parent particle become increasingly collimated.

(b) Scatter plot showing the angular distance $\Delta R(\tau_\mu, \tau_h)$ as a function of the transverse momentum of the tagged fatjet.

Figure 4.12.: Angular separation $\Delta R(\tau_\mu, \tau_h)$ as a function of the transverse momentum p_{T} of the fatjet tagged by the ParticleNet `vsJet` $\mu\tau_h$ discriminant. Events are selected from DY+jets samples with generator-level $p_{\mathrm{T}} > 100$ GeV. The figure demonstrates the expected collimation of decay products at higher p_{T}.

This observation confirms that highly boosted pairs of τ leptons become collimated, a feature that can be exploited in the analysis. One direct consequence of this collimation is seen in the angular separation between the fatjet and the missing transverse energy (MET) in the event, as shown in Figure 4.13.

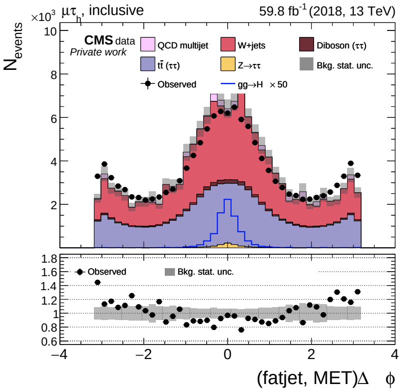


Figure 4.13.: Distribution of the angular separation $\Delta\phi(\text{fatjet}, MET)$ for signal (Higgs) and background events. The fatjet is tagged by the ParticleNet $\mu\tau_h$ discriminant.

Since τ leptons decay into neutrinos that escape detection, the missing transverse energy (MET) is assumed to be primarily due to these neutrinos. In signal events, where both visible τ decay products are collimated and boosted, the resulting MET tends to align with the fatjet direction. This is reflected by the peak near $\Delta\phi \approx 0$ in Figure 4.13 for the signal (Higgs) sample. In contrast, background events exhibit a broader distribution due to the lack of such kinematic correlation.

This property forms the basis of the *collinear approximation* [84], a technique used to reconstruct the invariant mass of the di-τ system by assuming that the neutrinos are emitted in the same direction as the visible τ decay products.

Part III.

Analysis

5. H $\longrightarrow \tau\,\tau$ analysis

5.1. Brief overview of previous boosted H$\longrightarrow \tau\,\tau$ analysis

The CMS Collaboration has recently released a paper [85] reporting a measurement of the Lorentz-boosted production cross section of the $H \longrightarrow \tau^+\tau^-$ process. In this analysis, a different strategy for identifying boosted di-τ systems was employed compared to the one used in this work. The method relies on declustering a large-radius CA jet with ΔR =0.8 (CA8) into two subjets, each expected to contain a hadronically decaying τ lepton reconstructed with the HPS algorithm. The τ reconstruction proceeds by applying the standard HPS DM identification to the constituents of the subjets. Since the subjets are built from non-overlapping collections of particles, the requirement that two τ candidates do not share any common PF candidates is automatically satisfied by construction. Following the reconstruction, additional discriminators are applied to suppress backgrounds and select genuine τ leptons.

The fiducial inclusive production cross section in the boosted phase space was measured to be

$$\sigma^{\text{fid}}_{H\rightarrow\tau\tau} = 3.88^{+1.69}_{-1.35}\ \text{fb},$$

which is compatible with the SM prediction of

$$\sigma^{\text{SM}}_{H\rightarrow\tau\tau} = 2.36 \pm 0.51\ \text{fb}.$$

No significant deviation from the SM expectation was observed.

5.2. Event selection

The purpose of the event selection in this analysis is to identify high-quality fatjets in the $\mu\tau$ final state. This channel is chosen because it offers the cleanest reconstruction among all $\tau\tau$ final states, primarily due to the high efficiency and reliability of muon reconstruction.

Selected fatjets are required to be tagged by the ParticleNet `vsJet` $\mu\tau$ tagger and must satisfy a set of minimal quality criteria. These include:

- Transverse momentum $p_T > 175$ GeV,
- Pseudorapidity $|\eta| < 4.7$,
- Passing the Tight Jet ID,
- Failing the Tight Lepton Veto.

Since the $\mu\tau$ final state is not explicitly labeled in the samples, it is defined based on reconstructed objects. Specifically, a fatjet is considered to belong to the $\mu\tau$ final state if exactly one good muon and no good electrons are geometrically matched to the fatjet within $\Delta R < 0.8$. This requirement ensures a clean identification of the topology corresponding to a $\mu\tau$ decay, while rejecting final states like $Z \rightarrow \mu\mu$ or $Z \rightarrow ee$, which typically feature two reconstructed leptons.

A good muon or electron is defined as a lepton that satisfies a set of baseline selection criteria, summarized in Table 5.1:

Selection criterion	Good muon	Good electron
p_T	>10 GeV	>10 GeV
$\|\eta\|$	<2.4	<2.4
Identification	loose ID	mvaNoIso WP80 ID
Impact parameters	$\|d_{xy}\| < 0.05$ cm, $\|d_z\| < 0.2$ cm	$\|d_{xy}\| < 0.05$ cm, $\|d_z\| < 0.2$ cm

Table 5.1.: Selection criteria applied to good muons and electrons.

The selection criteria listed in Table 5.1 are designed to ensure a high reconstruction efficiency for genuine leptons while suppressing backgrounds from misidentified objects and pile-up interactions.

The minimum transverse momentum requirement of $p_\mathrm{T} > 10$ GeVfor both muons and electrons ensures that the leptons are efficiently reconstructed and identified within the tracker and calorimeter systems. This threshold also reduces the contribution from soft leptons originating from hadron decays in flight or low-energy processes [86, 67].

The pseudorapidity selection $|\eta| < 2.4$ reflects the geometrical acceptance of the tracker and muon systems. Beyond this range, the reconstruction efficiency and momentum resolution degrade significantly due to limited detector coverage [87].

For particle identification, a loose identification working point is applied for muons, while electrons are required to pass the `mvaNoIso` WP80 working point. These working points are optimized by CMS experiment to provide high efficiency for genuine leptons while effectively suppressing misidentified hadrons and photon conversions [86, 67].

Finally, the impact parameter requirements, $|\mathrm{d_{xy}}| < 0.05$ cm and $|\mathrm{d_z}| < 0.2$ cm, ensure that the selected leptons originate from the primary interaction vertex rather than from pile-up vertices or displaced secondary decays. This cut significantly improves the purity of the lepton sample [87].

These requirements are consistent with standard CMS analysis practices and provide a good balance between signal efficiency and background suppression.

Furthermore, the selected muon must be matched to a trigger object within $\Delta R < 0.8$ and must satisfy the requirements of the single-muon trigger path `HLT_Mu50`, which targets high-p_T muons. This ensures a consistent and efficient trigger selection in the high-momentum regime relevant for the boosted Higgs and Z boson decays under consideration.

In addition to the ParticleNet `vsJet` $\mu\tau$ tagger, each fatjet is also evaluated using the `vsJet` $e\tau$ and `vsJet` $\tau\tau$ taggers.

To enhance the purity of the selected $\mu\tau$ final state, a requirement is imposed that the ParticleNet `vsJet` $\mu\tau$ score must be greater than both the $e\tau$ and $\tau\tau$ scores for the same fatjet:

$$\texttt{vsJet}\,\mu\tau > \texttt{vsJet}\,e\tau, \quad \texttt{vsJet}\,\mu\tau > \texttt{vsJet}\,\tau\tau$$

This ensures that the tagging algorithm assigns the highest likelihood to the $\mu\tau$ hypothesis, thereby reducing contamination from events with electrons or multiple hadronic taus that may otherwise mimic the signal topology.

The resulting Data/MC agreement after applying the event selection described in this chapter is shown in Figure 5.1.

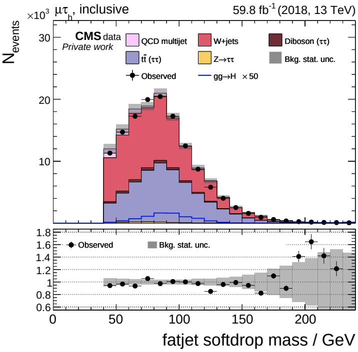


Figure 5.1.: Distribution of the soft-drop mass of the selected fatjets after the full event selection. Data points are compared to the stacked MC prediction.

A generally good Data/MC agreement is observed across most of the phase space, indicating that the selection is well-modeled by the simulation. Deviations are visible only in the tails of the distribution, where the statistical uncertainties are larger due to the limited number of events. The overall agreement validates the applied selection criteria and provides confidence in the modeling of the boosted $\mu\tau$ topology.

5.3. Backgrounds

- W**+jets:** At the LHC, this process is produced at leading order (LO) predominantly via quark–antiquark fusion. The W boson decays leptonically via $W \to \mu\nu$, providing a genuine muon, while one of the accompanying jets is misidentified as a hadronically decaying tau (τ_h). Due to its large production cross section and relatively high probability for jet misidentification, this process constitutes one of the dominant backgrounds in the $\mu\tau$ final state.

- $t\bar{t}$ **production:** At the LHC, this process is primarily produced at LO via ggF. Each top quark decays as $t \rightarrow bW$, and the subsequent W bosons may decay either leptonically or hadronically. Several configurations can mimic the $\mu\tau$ final state:
    - One W decays to a muon and a neutrino ($W \rightarrow \mu\nu$), while a jet from the hadronic W decay or a b-quark jet is misidentified as a τ_h.
    - One W decays to a tau lepton that subsequently decays hadronically, and the other W decays to a muon.

  The large cross section, presence of real muons, and abundance of jets make $t\bar{t}$ an important background in τ-based analyses.

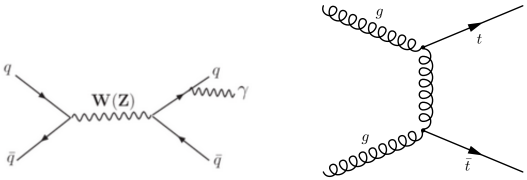
- **QCD multijet events:** Although QCD processes do not contain genuine leptons, they can enter the signal region through double misidentification:
    - One jet is misidentified as a muon,
    - Another jet is misidentified as a τ_h,
    - In addition, jets may contain genuine non-prompt muons from semileptonic decays, typically with lower transverse momentum.

  Despite the low probability for such double fakes, the extremely high production rate of QCD events makes this background non-negligible, particularly in regions with relaxed identification criteria.

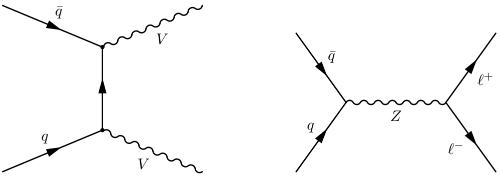
- **Diboson production (WW, WZ, ZZ):** These events can resemble the signal if one of the bosons decays leptonically (e.g., $W \rightarrow \mu\nu$) and the other decays hadronically, with a jet being misidentified as a τ_h. Additionally, real taus may be produced via decays such as $W \rightarrow \tau\nu$ or $Z \rightarrow \tau\tau$, contributing to the background in a more signal-like manner.

- **Drell–Yan process $Z/\gamma^* \rightarrow \tau^+\tau^-$:** This is a background that is very complicated to supress, as it produces the same final state as the signal: two genuine tau leptons, with one decaying into a muon and the other hadronically. Although the topology is identical to $H \rightarrow \tau\tau$, the kinematic properties, such as the invariant mass and transverse momentum of the $\tau\tau$ system, differ significantly from the Higgs boson signal. These differences can be exploited in the analysis to enhance signal-to-background separation.

The main background processes that can mimic the Higgs boson signal at the LHC are illustrated in Figure 5.2. These include W+jets production, top quark pair ($t\bar{t}$) production, diboson production, and Drell–Yan (Z boson) production.



(a) W+jets production **(b)** Top quark pair ($t\bar{t}$) production



(c) Diboson production **(d)** Drell–Yan (Z boson) production

Figure 5.2.: Representative Feynman diagrams for background processes that can mimic the Higgs-boson signal at the LHC: (a) W+jets production; (b) top-quark pair ($t\bar{t}$) production; (c) diboson (WW/WZ/ZZ) production; (d) Drell–Yan ($Z/\gamma^* \to \tau^+\tau^-$) production.

Despite the efficient ParticleNet tagging, background processes still dominate after applying the selection described in the previous section. Among them, W+jets constitutes the largest contribution due to its high production cross section and the relatively large probability of jets being misidentified as hadronically decaying taus (τ_h). A comparable contribution arises from top-quark pair ($t\bar{t}$) production, which remains one of the most important backgrounds in this final state. The relative fractions of the main processes are shown in Figure 5.3.

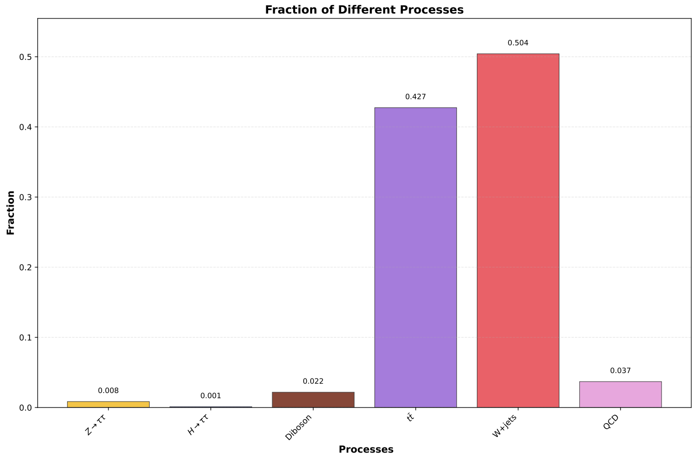


Figure 5.3.: Fractional contributions of the main background processes after the full event selection in $\mu\tau$ final state.

For the analysis, this dominance of backgrounds, especially W+jets, requires the use of dedicated control regions to constrain the background normalizations directly from data. This approach reduces the dependence on MC predictions and improves the accuracy of the background estimation in the signal region.

5.3.1. Monte Carlo Simulation Samples

The $H \longrightarrow \tau^+\tau^-$ signal is simulated using the POWHEG 2.0 [88] generator in the gluon-fusion production mode. To enhance sensitivity to boosted topologies, only events with $p_T^{fatjet} > 200$ GeVare selected.

The Drell–Yan background, $Z/\gamma^* \longrightarrow \tau^+\tau^-$, is simulated with the MadGraph5_aMC@NLO v2.6.5 [89] generator at next-to-leading order (NLO), using the FxFx jet matching and merging scheme [90]. This procedure provides a consistent description of different jet multiplicities and improves the modeling of additional hard QCD radiation. Similar to the signal selection, only events with $p_T^Z > 100$ GeVare considered to target the boosted regime relevant for this analysis.

The W+jets background is modeled at LO using MadGraph5_aMC@NLO v2.6.5 with the MLM jet matching and merging scheme [91]. HT-binned samples (`HT-100To200`, `HT-200To400`, etc.) are employed to increase the statistical precision in the high-p_{T} tails, which are critical for this boosted topology.

Top-quark pair production ($t\bar{t}$) and single top processes are generated at NLO using POWHEG 2.0, which provides an accurate modeling of the top-quark kinematics and spin correlations [92, 93].

Diboson processes (WW, WZ, and ZZ) are modeled using a mixture of NLO generators:

- POWHEG 2.0 for $ZZ \to 4\ell$,
- MadGraph5_aMC@NLO with the FxFx scheme for $ZZ \to 2q2\ell$, $WZ \to 2q2\ell$, and $WZ \to 3\ell\nu$.

QCD multijet samples are simulated at LO using Pythia 8.240 with the CP5 underlying-event tune. These samples are generated in p_{T}-binned intervals to improve statistical precision in the high-p_{T} regions relevant for this analysis.

The NNPDF 3.1 [94] Parton Distribution Functions (PDF) set is used for all simulated samples. Parton showering, hadronization, and underlying-event modeling are performed with Pythia 8.240 [90], using the CP5 tune [95], which is optimized for LHC Run 2 conditions. Additional pile-up is overlaid on all simulated events using Pythia, and the pile-up profile is reweighted to match that observed in data. The simulated samples are processed with a full simulation of the CMS detector based on Geant4 [96] and are normalized to the integrated luminosity of the analyzed dataset using the best available cross-section calculations.

In most $H \to \tau^+\tau^-$ analyses, the QCD background is estimated using data-driven techniques, such as the ABCD method or the Fake-Factor method [97]. These approaches typically exploit control regions enriched in QCD events, often defined by requiring two same-sign (SS) leptons. This strategy is motivated by charge conservation: processes involving neutral resonances decaying into two prompt SS leptons don't exist in the SM, making the SS region almost exclusively populated by QCD events with misidentified leptons.

However, in this analysis, the estimation of the QCD background is based entirely on simulated samples. The reason is that it is not possible to reliably define a SS control region in the boosted topology considered here. The

ParticleNet tagger, which is used to identify boosted $\mu\tau$ signatures inside fat jets, does not provide information about the individual charges of the reconstructed subobjects. It only classifies the jet as being consistent with a $\mu\tau$ decay, without distinguishing the charges of the muon or the hadronically decaying τ lepton. Consequently, the standard SS control-region strategy cannot be applied, and simulated QCD samples are used instead.

The data used in this analysis originate from the `SingleMuon` primary dataset. Specifically, events collected during the 2018 Run 2 data-taking period of the LHC are analyzed, corresponding to an integrated luminosity of 59.8 fb^{-1} [98]. The choice of 2018 data is motivated by its well-understood detector performance and stable operating conditions, which are crucial for studies involving advanced tagging algorithms.

Since the performance of the PARTICLENET tagger in boosted di-τ topologies has not yet been extensively validated, it is essential to rely on a dataset with well-characterized reconstruction and calibration. The 2018 data-taking period offers the most reliable reference for such studies, as it benefits from improved detector calibrations, mature reconstruction algorithms, and extensive cross-checks performed in previous CMS analyses.

5.4. Collinear mass as discriminating variable

In order to perform the measurement, one needs to define a suitable discriminating variable. At first glance, the softdrop mass (Figure 5.1) appears to be a reasonable candidate. This is motivated by the fact that the softdrop algorithm removes low-p_{T} constituents, which is crucial for reconstructing highly boosted resonances. However, as illustrated in Figure 5.1, this variable does not provide sufficient discriminating power, since the Higgs boson signal cannot be easily separated from Drell–Yan and other background processes.

A more promising approach is based on the *collinear mass approximation* [84]. The basic kinematics of this method are illustrated schematically in Figure 5.4.

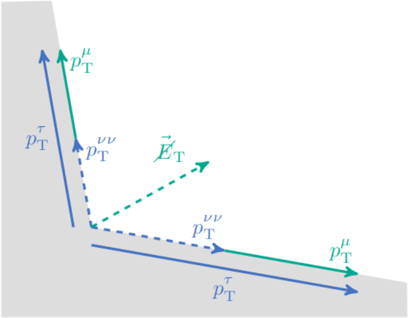


Figure 5.4.: Schematic illustration of the collinear approximation. The visible τ decay products and neutrinos are assumed to be collinear, and the full τ momentum is reconstructed from the visible momentum scaled by the neutrino momentum fraction [99].

The collinear approximation has been successfully used in $H \to \tau\tau$ analyses. Its application relies on the following assumptions:

- All MET originates from neutrinos produced in τ decays;
- The neutrinos are collinear with the visible τ decay products;
- The τ lepton momentum is much larger than its mass ($p_\tau \gg m_\tau$).

Under these assumptions, the τ lepton momentum can be written as

$$p(\tau_i) = \left(1 + x_{\nu\nu,i}\right)\ p(\mu_i), \qquad i = 1, 2, \tag{5.1}$$

where $p(\tau_i)$ and $p(\mu_i)$ denote the four-momenta of the τ lepton and the visible decay product (in this case, the muon), respectively, and $x_{\nu\nu,i}$ is the fraction of the τ momentum carried by the neutrinos. The missing transverse energy is then expressed as

$$\vec{E}_T^{\text{miss}} = x_{\nu\nu,1}\,\vec{p}_T(\mu_1) + x_{\nu\nu,2}\,\vec{p}_T(\mu_2). \tag{5.2}$$

Solving this system of equations yields the neutrino momentum fractions $x_{\nu\nu,1}$ and $x_{\nu\nu,2}$, which can then be used to reconstruct the collinear mass.

Since the individual kinematic properties of the τ lepton and muon cannot be disentangled inside a fatjet, subjets are used instead. A subjet is a smaller jet obtained by reversing the clustering sequence of a fatjet, effectively representing one of its hard prongs after grooming or declustering. For this analysis, one of the subjets is matched to the muon at generator level within $\Delta R < 0.8$.

The reconstructed collinear mass distribution is shown in Figure 5.5.

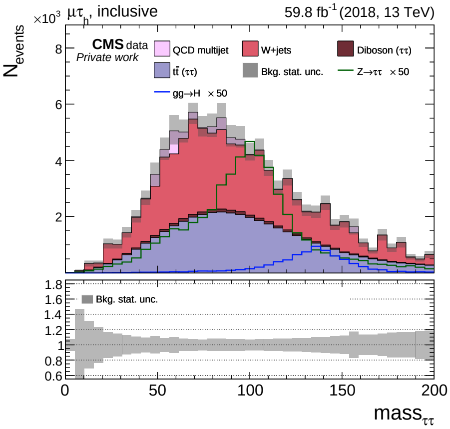


Figure 5.5.: Reconstructed collinear mass distribution using the subjet-based method. The Z and Higgs boson peaks are clearly visible but shifted to higher masses with respect to their true invariant masses. This shift originates from the use of subjets as proxies for τ leptons: subjets are groomed objects and thus lack low-p_T constituents.

A clear shift of the reconstructed mass with respect to the true Z and Higgs boson masses is therefore expected. The difference in p_T between the muon in the fatjet vicinity and the subjet matched to this muon is shown in Figure 5.6.

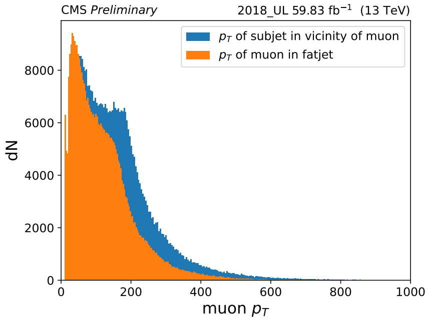


Figure 5.6.: Difference in transverse momentum between the muon reconstructed in the fatjet vicinity and the matched subjet. The subjet p_T is systematically larger due to grooming, which removes soft radiation and low-p_T constituents. This difference contributes to the observed mass shift in Figure 5.5.

Despite this systematic bias, the collinear mass distribution exhibits well-separated peaks for the Z and Higgs bosons, as well as a distinguishable background component. This demonstrates that the method provides sufficient discriminating power for the measurement.

5.5. Fitting strategy

In order to constrain the background yields, dedicated control regions (CRs) are defined, where the respective background processes dominate. The dominant backgrounds are W+jets and $t\bar{t}$ production (approximately 50% and 43% of the total background, respectively), while other processes such as diboson (2%), Z boson (0.8%), and QCD multijet events (3.7%) are negligible in comparison.

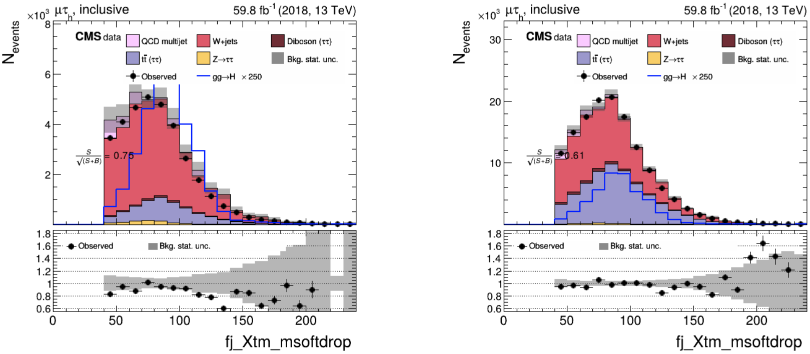
A W+jets enriched region is defined by $m_T > 60$ GeVand $N_{b\text{-jets}} = 0$, exploiting the fact that leptonic W decays typically produce large transverse masses, while vetoing b-jets suppresses top-quark contamination. A $t\bar{t}$ enriched region is defined by $m_T < 60$ GeVand $N_{b\text{-jets}} > 0$, where the presence of b-jets tags top-quark decays. The signal-enriched region, designed to select $H \to \tau\tau$ events, is defined by $m_T < 60$ GeVand $N_{b\text{-jets}} = 0$, which suppresses both W+jets and $t\bar{t}$ contributions.

The transverse mass is defined as

$$m_T = \sqrt{2\, p_T^{\ell}\, E_T^{\text{miss}} \;\left(1 - \cos\Delta\phi\left(\ell, E_T^{\text{miss}}\right)\right)}, \tag{5.3}$$

where p_T^{ℓ} is the transverse momentum of the visible lepton, E_T^{miss} is the missing transverse energy (MET), and $\Delta\phi(\ell, E_T^{\text{miss}})$ is the azimuthal angle between the lepton and MET.

The application of these optimized selection criteria significantly suppresses the dominant backgrounds, thereby improving the expected signal significance. Figure 5.7 shows the soft-drop mass distributions before and after applying the control region cuts.



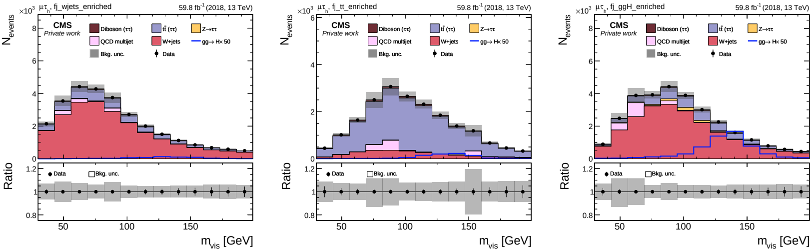
(a) Soft-drop mass distribution before applying control region cuts. Backgrounds dominate over the potential Higgs signal.

(b) Soft-drop mass distribution after applying control region cuts. Backgrounds are suppressed, improving the signal sensitivity.

Figure 5.7.: Comparison of soft-drop mass distributions before and after applying control region cuts. The background suppression leads to an enhanced expected signal significance.

The control regions are included in a simultaneous maximum-likelihood fit together with the signal-enriched region. The yields in the W+jets and $t\bar{t}$ control regions are used to constrain the normalization and shape of these dominant backgrounds, reducing systematic uncertainties and enabling a more precise extraction of the $H \to \tau\tau$ signal strength.

The pre-fit distributions of the control regions are shown in Figure 5.8. These regions correspond to the W+jets-enriched, $t\bar{t}$-enriched, and signal-enriched selections defined in the previous section. The pre-fit represents the expected background and signal contributions before performing the statistical fit to data. In this work, the pre-fit templates are derived from an Asimov dataset, which provides the nominal expectation of event yields without statistical fluctuations. This choice allows for a direct illustration of the relative background composition and normalization in the different control regions prior to the fit procedure.



(a) W+jets enriched region. **(b)** $t\bar{t}$ enriched region. **(c)** Signal-enriched region.

Figure 5.8.: Pre-fit distributions in the three control regions: (a) W+jets-enriched, (b) $t\bar{t}$-enriched, and (c) $H \to \tau\tau$ signal-enriched region. The pre-fit is derived from an Asimov dataset, representing the expected event yields without statistical fluctuations.

5.5.1. Statistical inference

The main goal of this thesis is to measure the cross section of the $H \to \tau\tau$ process in the high-p_T regime. To achieve this, a maximum-likelihood fit of the simulated MC predictions to the observed data is performed.

The total cross section times branching fraction is defined as

$$\sigma_{H\to\tau\tau} = \sigma(pp \to H) \times \mathcal{B}(H \to \tau\tau), \tag{5.4}$$

where $\sigma(pp \to H)$ denotes the Higgs boson production cross section and $\mathcal{B}(H \to \tau\tau)$ is the branching fraction of the Higgs boson decay to tau leptons.

Maximum-Likelihood Fit. The extraction of the cross section is based on a Maximum Likelihood Estimation (MLE) approach. The likelihood function is constructed as

$$\mathcal{L}(\mu, \theta) = \prod_i P(n_i \mid \mu\, s_i(\theta) + b_i(\theta)) \prod_k \pi(\theta_k), \tag{5.5}$$

where

- μ is the *signal strength modifier*, defined as the ratio of the measured cross-section to the SM prediction,
- n_i is the observed number of events in bin i,
- $s_i(\theta)$ and $b_i(\theta)$ are the expected signal and background yields, respectively for bin i,
- $P(n_i \mid \mu s_i + b_i)$ is the Poisson probability of observing n_i events, given the predictions based on μ, s_i and b_i,
- $\theta = \{\theta_k\}$ are nuisance parameters representing systematic uncertainties,
- $\pi(\theta_k)$ are prior probability density functions (typically Gaussian or log-normal) constraining the nuisance parameters.

The best-fit values of the parameters are obtained by maximizing the likelihood function with respect to μ and θ:

$$\hat{\mu},\ \hat{\theta} = \underset{\mu, \theta}{\operatorname{argmax}}\ \mathcal{L}(\mu, \theta). \tag{5.6}$$

Nuisance parameters. The nuisance parameters θ encode experimental and theoretical systematic uncertainties affecting both the normalization and the shape of the signal and background distributions. In the fit, these parameters are *profiled*, meaning that for each tested value of the parameter of interest μ, they are set to the conditional maximum-likelihood estimates that maximize the likelihood. This procedure allows the data to constrain the nuisance parameters directly, thereby reducing the impact of systematic uncertainties and properly propagating them to the uncertainty in μ.

Cross-Section Extraction. Once the best-fit value of μ is determined, the cross section times branching fraction is obtained as

$$\sigma(pp \to H)\times\mathcal{B}(H \to \tau^+\tau^-) \; = \; \mu\cdot\left[\sigma(pp \to H) \times \mathcal{B}(H \to \tau^+\tau^-)\right]_{\mathrm{SM}} . \quad (5.7)$$

The fit is performed simultaneously in three categories: a W+jets-enriched control region, a $t\bar{t}$-enriched control region, and a signal-enriched region. The inclusion of control regions constrains the background normalizations and reduces the impact of systematic uncertainties, thereby improving the precision of the signal extraction.

Asimov dataset. The fit is performed on an Asimov dataset, which represents the statistically expected event yields without statistical fluctuations. This approach provides the expected sensitivity and precision of the measurement in the absence of real data fluctuations.

5.6. Uncertainty model

There are two main categories of systematic uncertainties considered in this analysis: **yield (normalization) uncertainties**, which affect only the overall normalization of the distributions, and **shape uncertainties**, which can modify the shape of the distributions used in the maximum-likelihood fit.

The baseline uncertainty model follows the recommendations of the CMS analysis presented in Ref. [85].

Yield uncertainties. The following sources contribute to yield uncertainties:

- **Integrated luminosity:** The total integrated luminosity of the 2018 data-taking period is known with an uncertainty of 2.5% [100].
- **Muon trigger and identification efficiency:** The efficiency of the muon trigger and reconstruction is measured using the tag-and-probe method [101], resulting in a 2% uncertainty.
- **Tau identification efficiency:** The hadronic τ identification is assigned a conservative uncertainty of 10%, following the recommendations for boosted topologies.

- **Bin-by-bin statistical uncertainties:** Statistical uncertainties of the MC templates are incorporated using the so-called bin-by-bin method. Each bin of the template is assigned an independent nuisance parameter, allowing the fit to vary the predicted yield in that bin within its statistical uncertainty. This approach prevents the fit from being overly constrained by limited MC statistics and is particularly relevant for low-yield regions [102].

Shape uncertainties. Shape-modifying uncertainties are implemented as variations of the relevant physics objects in the simulation, propagated through the full analysis chain. The following sources are considered:

- **Jet energy scale (JES):** Uncertainties in the calibration of the jet energy scale are applied by shifting the jet four-momentum up and down by one standard deviation. The resulting changes in reconstructed kinematic distributions are propagated to all derived quantities, including τ_h reconstruction and the collinear mass [103].
- **Unclustered energy** Variations in the MET are obtained by coherently shifting the momenta of all reconstructed objects contributing to the MET calculation. This uncertainty has a direct impact on variables sensitive to missing transverse momentum, such as transverse and collinear masses [75].
- **b-tagging efficiency:** The efficiency of identifying b-jets is varied within its uncertainty by applying scale factor variations (up/down). This uncertainty mainly affects the normalization and composition of the $t\bar{t}$ background [104].

All systematic uncertainties are incorporated into the maximum-likelihood fit as nuisance parameters. Yield uncertainties are modeled using log-normal priors, while shape uncertainties are implemented through template morphing based on up and down variations. Correlations between different processes and categories are preserved whenever applicable, following CMS recommendations. The nuisance parameters are profiled in the fit, allowing the data to constrain their impact where possible. This procedure ensures a consistent treatment of both statistical and systematic uncertainties in the extraction of the signal strength and cross section.

5.7. Results

The best-fit signal strength, obtained from the maximum-likelihood fit to the Asimov dataset, is

$$\mu = 1^{+1.4}_{-1.5},$$

where μ denotes the ratio of the measured cross section to the SM expectation. The sensitivity achieved in this analysis is compatible with the results obtained in the $\mu\tau$ final state using the 2018 dataset from the previous CMS analysis [85]. The measurement is limited by statistical uncertainties, and the result quoted here corresponds exclusively to ggF production mode. This outcome is statistically consistent with the SM prediction and shows no significant deviation.

Part IV.

Conclusion and Outlook

6. Summary and suggestions for the future analyses

The primary motivation of this analysis is to probe the Yukawa structure of the Higgs sector and to test the predictions of the SM in the boosted regime. The $H \rightarrow \tau\tau$ decay channel provides an excellent opportunity for this purpose, as the τ lepton belongs to the third generation and possesses the strongest Yukawa coupling among all leptons. At the same time, the boosted topology introduces specific challenges, particularly in reconstruction. To address these, a novel approach based on machine-learning techniques was employed: ParticleNet. This jet-tagging algorithm treats the τ as a jet and assigns a probability for each candidate to be identified as a τ lepton.

A central difficulty arises in the reconstruction of the invariant mass, as neutrinos from τ decays remain undetected. To overcome this limitation, the collinear mass approximation was adopted. This method provides a practical mass observable that can serve as a discriminating variable in the statistical analysis to separate signal from background.

The results demonstrate that ParticleNet shows good agreement with collision data and provides a reliable performance for boosted τ identification. The collinear mass approximation also exhibits useful discriminating power, enabling efficient signal extraction.

Nevertheless, several limitations remain. For instance, the collinear mass approximation reconstructs the peaks of the Z and Higgs bosons with noticeable shifts relative to their true invariant masses. Furthermore, the kinematic properties of muons and τ leptons cannot be disentangled in this approach, which motivates the use of subjet kinematics instead. Systematic uncertainties also present a window for improvement. In particular, this work assumed a conservative 10% uncertainty on boosted τ identification efficiency. Employing more precise, data-driven estimates would enhance the robustness and sensitivity of the measurement.

The methodology developed here is not limited to the Higgs sector alone. Other analyses could benefit from these techniques, such as $Z' \rightarrow \tau\tau$ searches or investigations within extended Higgs sectors, including the MSSM and NMSSM . More broadly, improvements in boosted τ reconstruction and mass approximation methods strengthen the overall LHC program in searches for new physics and precision tests of the SM.

Looking ahead, this work can be extended using Run 3 data, where the LHC has already delivered a larger dataset than in Run 2. The increased luminosity will significantly improve the statistical power of boosted analyses. In addition, ongoing developments in detector technology (e.g. high-granularity calorimeters) and next-generation machine-learning methods (such as graph neural networks and domain adaptation) hold promise for further advances in τ identification and mass reconstruction. At the High-Luminosity LHC, such techniques will be crucial for precision Higgs coupling measurements and for exploring possible deviations that could point to physics beyond the SM.

In conclusion, this thesis contributes to the understanding of Higgs boson properties in the boosted $H \rightarrow \tau\tau$ channel by demonstrating the applicability of modern machine-learning methods and advanced mass reconstruction techniques. While challenges remain, the results presented here represent an important step toward unlocking the full potential of boosted analyses, paving the way for more sensitive searches and more precise measurements in the future.

Bibliography

[1] R. Oerter. *The theory of almost everything: The standard model, the unsung triumph of modern physics.* 2006.

[2] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: `10.1093/ptep/ptac097`.

[3] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13.16 (1964), pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`.

[4] David J. Griffiths. *Introduction to Elementary Particles.* New York: Wiley-VCH, 2008.

[5] T. Nakano and K. Nishijima. "Charge Independence for V-particles". In: *Prog. Theor. Phys.* 10 (1953), pp. 581–582. DOI: `10.1143/PTP.10.581`.

[6] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995. ISBN: 978-0201503975.

[7] Anthony Zee. *Quantum Field Theory in a Nutshell: Second Edition.* Princeton University Press, Feb. 2010. ISBN: 978-0-691-14034-6.

[8] Paul A. M. Dirac. "Quantum theory of emission and absorption of radiation". In: *Proc. Roy. Soc. Lond. A* 114 (1927), p. 243. DOI: `10.1098/rspa.1927.0039`.

[9] A. I. Miller. *Early quantum electrodynamics: A source book.* Cambridge University Press, Oct. 1995. ISBN: 978-0-521-56891-3.

[10] T. D. Lee and Chen-Ning Yang. "Question of Parity Conservation in Weak Interactions". In: *Phys. Rev.* 104 (1956), pp. 254–258. DOI: `10.1103/PhysRev.104.254`.

[11] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19.21 (1967), pp. 1264–1266. DOI: `10.1103/PhysRevLett.19.1264`.

[12] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Elementary Particle Theory: Relativistic Groups and Analyticity.* Ed. by N. Svartholm. Stockholm: Almqvist and Wiksell, 1968, pp. 367–377.

[13] D. J. Gross. “The discovery of asymptotic freedom and the emergence of QCD”. In: *Proc. Nat. Acad. Sci.* 102 (2005), pp. 9099–9108. DOI: `10.1073/pnas.0503831102`.

[14] Franz Gross et al. “50 Years of Quantum Chromodynamics”. In: *Eur. Phys. J. C* 83 (2023), p. 1125. DOI: `10.1140/epjc/s10052-023-11949-2`. arXiv: `2212.11107 [hep-ph]`.

[15] Chen-Ning Yang and Robert L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. In: *Phys. Rev.* 96 (1954). Ed. by Jong-Ping Hsu and D. Fine, pp. 191–195. DOI: `10.1103/PhysRev.96.191`.

[16] Paul A. M. Dirac. “The quantum theory of the electron”. In: *Proc. Roy. Soc. Lond. A* 117 (1928), pp. 610–624. DOI: `10.1098/rspa.1928.0023`.

[17] I. Hinchliffe. “The QCD Lagrangian”. In: (July 2004).

[18] S. Schael et al. “Precision electroweak measurements on the Z resonance”. In: *Phys. Rept.* 427 (2006), pp. 257–454. DOI: `10.1016/j.physrep.2005.12.006`. arXiv: `hep-ex/0509008`.

[19] S. Schael et al. “Measurement of the W boson mass and width in e^+e^- collisions at LEP”. In: *Eur. Phys. J. C* 47 (2006), pp. 309–335. DOI: `10.1140/epjc/s2006-02576-8`. arXiv: `hep-ex/0605011`.

[20] François Englert and Robert Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13.9 (1964), pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`.

[21] Gerald S. Guralnik, Carl R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13.20 (1964), pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`.

[22] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716.1 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`.

[23] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716.1 (2012), pp. 30–61. DOI: `10.1016/j.physletb.2012.08.021`.

[24] ALEPH et al. “Search for the Standard Model Higgs boson at LEP”. In: *Phys. Lett. B* 565 (2003), pp. 61–75. DOI: `10.1016/S0370-2693(03)00614-2`.

[25] CDF and D0 Collaborations. "Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data". In: *Phys. Rev. D* 88.5 (2013), p. 052014. DOI: `10.1103/PhysRevD.88.052014`.

[26] Benjamin W. Lee, Chris Quigg, and H. B. Thacker. "Weak interactions at very high energies: The role of the Higgs boson mass". In: *Phys. Rev. D* 16.5 (1977), pp. 1519–1531. DOI: `10.1103/PhysRevD.16.1519`.

[27] Aron J. Beekman, Louk Rademaker, and Jasper van Wezel. "An Introduction to Spontaneous Symmetry Breaking". In: *SciPost Phys. Lect. Notes* 11 (2019), p. 1. DOI: `10.21468/SciPostPhysLectNotes.11`. arXiv: `1909.01820 [hep-th]`.

[28] Janosh Riebesell. *Higgs Potential.* `https://tikz.net/higgs-potential/`.

[29] Chris Quigg. *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions.* 2nd. Princeton University Press, 2013. ISBN: 978-0691152674.

[30] Ta-Pei Cheng and Ling-Fong Li. *Gauge Theory of Elementary Particle Physics.* Oxford University Press, 1984.

[31] Particle Data Group. "Review of Particle Physics". In: *Prog. Theor. Exp. Phys.* 2024.8 (2024), p. 083C01. DOI: `10.1093/ptep/ptae073`.

[32] Csaba Csáki and Philip Tanedo. "Beyond the Standard Model". In: *2013 European School of High-Energy Physics.* 2015, pp. 169–268. DOI: `10.5170/CERN-2015-004.169`. arXiv: `1602.04228 [hep-ph]`.

[33] Carl D. Anderson. "The Positive Electron". In: *Phys. Rev.* 43 (6 Mar. 1933), pp. 491–494. DOI: `10.1103/PhysRev.43.491`. URL: `https://link.aps.org/doi/10.1103/PhysRev.43.491`.

[34] Stephen P. Martin. "A Supersymmetry Primer". In: *Adv. Ser. Direct. High Energy Phys.* 21 (2010). Updated version available at `https://arxiv.org/abs/hep-ph/9709356`, pp. 1–153. eprint: `hep-ph/9709356`.

[35] Howard Baer and Xerxes Tata. *Weak Scale Supersymmetry: From Superfields to Scattering Events.* Cambridge University Press, 2006.

[36] Howard E. Haber and Gordon L. Kane. "The Search for Supersymmetry: Probing Physics Beyond the Standard Model". In: *Phys. Rept.* 117 (1985), pp. 75–263. DOI: `10.1016/0370-1573(85)90051-1`.

[37] Ulrich Ellwanger, Cyril Hugonie, and Ana M. Teixeira. "The Next-to-Minimal Supersymmetric Standard Model". In: *Phys. Rept.* 496 (2010), pp. 1–77. DOI: `10.1016/j.physrep.2010.07.001`. eprint: `arXiv:0910.1785[hep-ph]`.

[38] Marcela Carena and Howard E. Haber. "Higgs boson theory and phenomenology". In: *Prog. Part. Nucl. Phys.* 50 (2003), pp. 63–152. DOI: `10.1016/S0146-6410(02)00177-1`. eprint: `hep-ph/0208209`.

[39] Manuel Maniatis. "The Next-to-Minimal Supersymmetric extension of the Standard Model reviewed". In: *Int. J. Mod. Phys. A* 25 (2010), pp. 3505–3602. DOI: `10.1142/S0217751X10049827`. eprint: `arXiv:0906.0777[hep-ph]`.

[40] Armen Tumasyan et al. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery." In: *Nature* 607.7917 (2022). [Erratum: Nature 623, (2023)], pp. 60–68. DOI: `10.1038/s41586-022-04892-x`. arXiv: `2207.00043 [hep-ex]`.

[41] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: 2/2017 (Oct. 2016). DOI: `10.23731/CYRM-2017-002`. arXiv: `1610.07922 [hep-ph]`.

[42] C. Grojean. "Higgs Physics". en. In: *CERN Yellow Reports* (2016), Vol 5 (2016): Proceedings of the 2015 CERN-. DOI: `10.5170/CERN-2016-005.143`. URL: `https://e-publishing.cern.ch/index.php/CYR/article/view/435`.

[43] Albert M Sirunyan et al. "Evidence for Higgs boson decay to a pair of muons". In: *JHEP* 01 (2021), p. 148. DOI: `10.1007/JHEP01(2021)148`. arXiv: `2009.04363 [hep-ex]`.

[44] Deniz Yilmaz, Mehmet Sahin, and Dogukan Hazar Yavuz. *Higgs production at next generation e^+e^- colliders*. 2023. arXiv: `2301.13463 [hep-ph]`. URL: `https://arxiv.org/abs/2301.13463`.

[45] Serguei Chatrchyan et al. "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons". In: *JHEP* 05 (2014), p. 104. DOI: `10.1007/JHEP05(2014)104`. arXiv: `1401.5041 [hep-ex]`.

[46] Georges Aad et al. "Observation and measurement of Higgs boson decays to WW* with the ATLAS detector". In: *Phys. Rev. D* 92.1 (2015), p. 012006. DOI: `10.1103/PhysRevD.92.012006`. arXiv: `1412.2641 [hep-ex]`.

[47] Benedetta Camaiani. "STXS and differential Higgs boson cross section measurements at CMS". In: *PoS* ICHEP2024 (2025), p. 047. DOI: `10.22323/1.476.0047`.

[48] Ilaria Brivio and Michael Trott. “The Standard Model as an Effective Field Theory”. In: *Phys. Rept.* 793 (2019), pp. 1–98. DOI: `10.1016/j.physrep.2018.11.002`. arXiv: `1706.08945 [hep-ph]`.

[49] C. P. Burgess. “Introduction to Effective Field Theory”. In: *Ann. Rev. Nucl. Part. Sci.* 57 (2007), pp. 329–362. DOI: `10.1146/annurev.nucl.56.080805.140508`. arXiv: `hep-th/0701053`.

[50] Raquel Gomez-Ambrosio. “Studies of dimension-six EFT effects in vector boson scattering”. In: *The European Physical Journal C* 79.5 (May 2019). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-019-6893-2`. URL: `http://dx.doi.org/10.1140/epjc/s10052-019-6893-2`.

[51] CMS Collaboration. *Combined effective field theory interpretation of Higgs boson, electroweak vector boson, top quark, and multi-jet measurements.* 2025. arXiv: `2504.02958 [hep-ex]`. URL: `https://arxiv.org/abs/2504.02958`.

[52] C. Grojean et al. “Very boosted Higgs in gluon fusion”. In: *Journal of High Energy Physics* 2014.5 (May 2014). ISSN: 1029-8479. DOI: `10.1007/jhep05(2014)022`. URL: `http://dx.doi.org/10.1007/JHEP05(2014)022`.

[53] Abdelhak Djouadi. “The anatomy of electroweak symmetry breaking”. In: *Physics Reports* 457.1–4 (Feb. 2008), pp. 1–216. ISSN: 0370-1573. DOI: `10.1016/j.physrep.2007.10.004`. URL: `http://dx.doi.org/10.1016/j.physrep.2007.10.004`.

[54] Malte Buschmann et al. “Mass effects in the Higgs-gluon coupling: boosted vs. off-shell production”. In: *Journal of High Energy Physics* 2015.2 (Feb. 2015). ISSN: 1029-8479. DOI: `10.1007/jhep02(2015)038`. URL: `http://dx.doi.org/10.1007/JHEP02(2015)038`.

[55] G. Acquistapace et al. “CMS, the magnet project: Technical design report”. In: (May 1997).

[56] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: *JINST* 12.10 (2017), P10003. DOI: `10.1088/1748-0221/12/10/P10003`. arXiv: `1706.04965 [physics.ins-det]`.

[57] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: `10.1088/1748-0221/3/08/S08004`.

[58] “The CMS muon project: Technical Design Report”. In: (1997).

[59] V. Karimäki. “The CMS tracker system project: Technical Design Report”. In: (1997). Ed. by M. Mannelli et al.

[60] "The CMS tracker: addendum to the Technical Design Report". In: (2000).

[61] "The CMS electromagnetic calorimeter project: Technical Design Report". In: (1997).

[62] P. Adzic et al. "Energy resolution of the barrel of the CMS electromagnetic calorimeter". In: *JINST* 2 (2007), P04004. DOI: `10.1088/1748-0221/2/04/P04004`.

[63] "The CMS hadron calorimeter project: Technical Design Report". In: (1997).

[64] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.06 (2018), P06015. DOI: `10.1088/1748-0221/13/06/P06015`. arXiv: `1804.04528 [physics.ins-det]`.

[65] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: `10.1088/1748-0221/12/01/P01020`. arXiv: `1609.02366 [physics.ins-det]`.

[66] T. Speer et al. "Track reconstruction in the CMS tracker". In: *Nucl. Instrum. Meth. A* 559 (2006). Ed. by J. Blumlein et al., pp. 143–147. DOI: `10.1016/j.nima.2005.11.207`.

[67] 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]`.

[68] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: `10.1088/1126-6708/2008/04/063`. arXiv: `0802.1189 [hep-ph]`.

[69] S. Catani et al. "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions". In: *Nucl. Phys. B* 406 (1993), pp. 187–224. DOI: `10.1016/0550-3213(93)90166-M`.

[70] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. DOI: `10.1103/PhysRevD.48.3160`. arXiv: `hep-ph/9305266`.

[71] Stan Bentvelsen and Irmtraud Meyer. "The Cambridge jet algorithm: Features and applications". In: *Eur. Phys. J. C* 4 (1998), pp. 623–629. DOI: `10.1007/s100707775200 50232`. arXiv: `hep-ph/9803322`.

[72] Yuri L. Dokshitzer et al. “Better jet clustering algorithms”. In: *JHEP* 08 (1997), p. 001. DOI: `10.1088/1126-6708/1997/08/001`. arXiv: `hep-ph/9707323`.

[73] “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018”. In: (2020). URL: `https://cds.cern.ch/record/2715872`.

[74] *Jet algorithms performance in 13 TeV data.* Tech. rep. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2256875`.

[75] Albert M Sirunyan et al. “Pileup mitigation at CMS in 13 TeV data”. In: *JINST* 15.09 (2020), P09018. DOI: `10.1088/1748-0221/15/09/P09018`. arXiv: `2003.00503 [hep-ex]`.

[76] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: `10.1103/PhysRevD.110.030001`.

[77] “Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2”. In: (2016).

[78] CMS collaboration. *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2.* Tech. rep. Geneva: CERN, 2016. URL: `https://cds.cern.ch/record/2196972`.

[79] A. M. Sirunyan et al. “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV”. In: *JINST* 13.10 (2018), P10005. DOI: `10.1088/1748-0221/13/10/P10005`. arXiv: `1809.02816 [hep-ex]`.

[80] Armen Tumasyan et al. “Identification of hadronic tau lepton decays using a deep neural network”. In: *JINST* 17 (2022), P07023. DOI: `10.1088/1748-0221/17/07/P07023`. arXiv: `2201.08458 [hep-ex]`.

[81] CMS Collaboration. *Identification of tau leptons using a convolutional neural network with domain adaptation in the CMS experiment.* `https://cds.cern.ch/record/2931189`. CMS Analysis Note CMS-TAU-24-001. 2024.

[82] Huilin Qu and Loukas Gouskos. “ParticleNet: Jet Tagging via Particle Clouds”. In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: `10.1103/PhysRevD.101.056019`. arXiv: `1902.08570 [hep-ph]`.

[83] Yue Wang et al. *Dynamic Graph CNN for Learning on Point Clouds.* 2019. arXiv: `1801.07829 [cs.CV]`. URL: `https://arxiv.org/abs/1801.07829`.

[84] A. Elagin et al. “A New Mass Reconstruction Technique for Resonances Decaying to di-tau”. In: *Nucl. Instrum. Meth. A* 654 (2011), pp. 481–489. DOI: `10.1016/j.nima.2011.07.009`. arXiv: `1012.4686 [hep-ex]`.

[85] Aram Hayrapetyan et al. “Measurement of the production cross section of a Higgs boson with large transverse momentum in its decays to a pair of τ leptons in proton-proton collisions at s=13TeV”. In: *Phys. Lett. B* 857 (2024), p. 138964. DOI: `10.1016/j.physletb.2024.138964`. arXiv: `2403.20201 [hep-ex]`.

[86] Aram Hayrapetyan et al. “Performance of CMS muon reconstruction from proton-proton to heavy ion collisions”. In: *JINST* 19.09 (2024), P09012. DOI: `10.1088/1748-0221/19/09/P09012`. arXiv: `2404.17377 [hep-ex]`.

[87] Serguei Chatrchyan et al. “Description and Performance of Track and Primary-Vertex Reconstruction with the CMS Tracker”. In: *JINST* 9.10 (2014), P10009. DOI: `10.1088/1748-0221/9/10/P10009`. arXiv: `1405.6569 [physics.ins-det]`.

[88] Simone Alioli et al. “NLO Higgs boson production via gluon fusion matched with shower in POWHEG”. In: *JHEP* 04 (2009), p. 002. DOI: `10.1088/1126-6708/2009/04/002`. arXiv: `0812.0578 [hep-ph]`.

[89] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[90] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: `10.1016/j.cpc.2015.01.024`. arXiv: `1410.3012 [hep-ph]`.

[91] Maximilian Reininghaus, Torbjörn Sjöstrand, and Marius Utheim. “Pythia 8 as hadronic interaction model in air shower simulations”. In: *EPJ Web Conf.* 283 (2023), p. 05010. DOI: `10.1051/epjconf/202328305010`. arXiv: `2303.02792 [astro-ph.HE]`.

[92] John M. Campbell et al. “Top-Pair Production and Decay at NLO Matched with Parton Showers”. In: *JHEP* 04 (2015), p. 114. DOI: `10.1007/JHEP04(2015)114`. arXiv: `1412.1828 [hep-ph]`.

[93] Simone Alioli et al. “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”. In: *JHEP* 09 (2009). [Erratum: JHEP 02, 011 (2010)], p. 111. DOI: `10.1088/1126-6708/2009/09/111`. arXiv: `0907.4076 [hep-ph]`.

[94] Richard D. Ball et al. “Parton distributions for the LHC Run II”. In: *JHEP* 04 (2015), p. 040. DOI: `10.1007/JHEP04(2015)040`. arXiv: `1410.8849 [hep-ph]`.

[95] 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]`.

[96] S. Agostinelli et al. “GEANT4 - A Simulation Toolkit”. In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: `10.1016/S0168-9002(03)01368-8`.

[97] Armen Tumasyan et al. “Measurements of Higgs boson production in the decay channel with a pair of τ leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 83.7 (2023), p. 562. DOI: `10.1140/epjc/s10052-023-11452-8`. arXiv: `2204.12957 [hep-ex]`.

[98] Angela Giraldi. “Precision luminosity measurement with proton-proton collisions at the CMS experiment in Run 2”. In: *PoS* ICHEP2022 (2022), p. 638. DOI: `10.22323/1.414.0638`. arXiv: `2208.08214 [hep-ex]`.

[99] Thomas Müller. “Analysis of Standard Model Higgs Boson Decays to Tau Pairs with the CMS Detector at the LHC”. PhD thesis. Karlsruhe Institute of Technology (KIT), 2015.

[100] *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ =* 13 TeV. Tech. rep. Geneva: CERN, 2019. URL: `https://cds.cern.ch/record/2676164`.

[101] Vardan Khachatryan et al. “Reconstruction and identification of τ lepton decays to hadrons and ν τ at CMS”. In: *JINST* 11.01 (2016), P01019. DOI: `10.1088/1748-0221/11/01/P01019`. arXiv: `1510.07488 [physics.ins-det]`.

[102] Roger J. Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: `10.1016/0010-4655(93)90005-W`.

[103] Garvita Agarwal. "Jet Energy Scale and Resolution Measurements in CMS". In: *PoS* ICHEP2022 (2022), p. 652. DOI: `10.22323/1.414.0652`. arXiv: `2301.02175 [hep-ex]`.

[104] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: `10.1088/1748-0221/13/05/P05011`. arXiv: `1712.07158 [physics.ins-det]`.

Part V.

Acknowledgment

Completing a PhD is a significant milestone, and I am grateful for the many colleagues who supported me along the way.

I would like to express my sincere thanks to my supervisors, Prof. Dr. Markus Klute and Priv. Doz. Roger Wolf, for their role in overseeing this thesis. The journey offered me the opportunity to develop through meaningful independent exploration, which ultimately became a defining and formative aspect of my doctoral experience.

My warm thanks go to the Tau POG group, and in particular to the conveners Anne-Catherine Le Bihan, Daniel Winterbottom, Andrea Cardini, and Michał Bluj. Their support was invaluable during my work with the group and throughout my time as an L3 convener, an experience that remains one of my favorite parts of the PhD.

I am especially grateful to my co-convener Irene Andreou, whose technical and organizational contributions were indispensable. Working together as a team was both productive and rewarding.

My appreciation extends to Lucas Russel, with whom I had the pleasure of collaborating on machine-learning projects, and from whom I learned many interesting and useful techniques.

I would also like to thank Izaak Neutelings for our collaboration as facilitators at the CMS Data Analysis School.

My colleague Xunwu Zuo deserves special acknowledgement for many insightful discussions about physics, analysis strategies, and the inner workings of CMS. His perspectives on career development within the collaboration were extremely helpful.

I am grateful to Nicolò Trevisani for his valuable input and for the many engaging conversations we shared.

My thanks go as well to the members of the KIT Tau group: Artur Gottmann, Ralf Schmieder, Christian Winter, Nikita Shadskiy, Moritz Molch, Artur Monsch, and Tim Voigtländer, for their contributions to my analysis and their assistance with the KIT infrastructure.

To my colleagues and friends Aritra Bal, Alejandro Quiroga Triviño, Maximilian Horzela, Alessandro Brusamolino, and Sofia Giappichini, thank you for the friendship, encouragement, and good moments we shared.

I would also like to thank Marco Link and Max Neukum for many enjoyable conversations and for their memorable contribution to my fluency in German.

Finally, I would like to thank Nils Faltermann, who was always exceptionally helpful with organizational matters and the teaching-assistance infrastructure.

A. Appendix: Samples

A.0.1. MC and data

The analysis is based on the 2018 UL (Ultra Legacy) dataset of proton–proton collisions recorded by the CMS experiment at the LHC, corresponding to an integrated luminosity of approximately 59.7 fb^{-1} at $\sqrt{s} = 13$ TeV.

DY+jets

- DYJetsToLL_LHEFilterPtZ-100To250 (NLO)
- DYJetsToLL_LHEFilterPtZ-250To400 (NLO)
- DYJetsToLL_LHEFilterPtZ-400To650 (NLO)
- DYJetsToLL_LHEFilterPtZ-650ToInf (NLO)

$\mathbf{t\bar{t}}$

- TTTo2L2Nu
- TTToHadronic
- TTToSemiLeptonic

Diboson

- WZTo3LNu
- ZZTo2Q2L
- ZZTo4L

Single top

- ST_t-channel_antitop_4f_InclusiveDecays
- ST_t-channel_top_4f_InclusiveDecays
- ST_tW_antitop_5f_inclusiveDecays

- ST_tW_top_5f_inclusiveDecays

W+jets

- WJetsToLNu (LO)

QCD

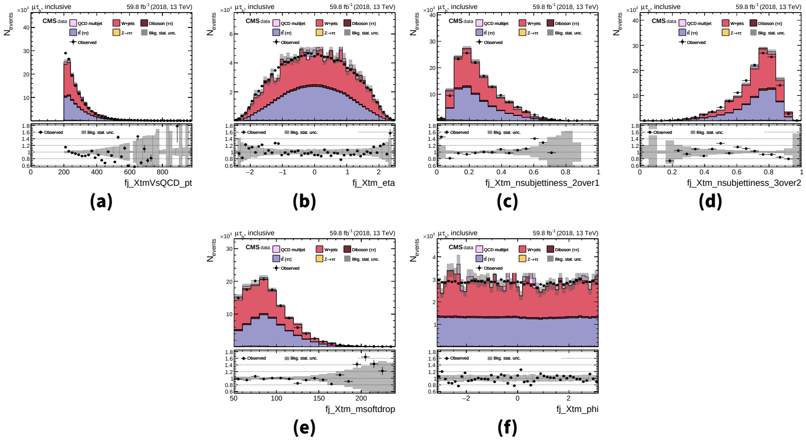
- QCD_Pt_170to300
- QCD_Pt_300to470
- QCD_Pt_470to600
- QCD_Pt_600to800
- QCD_Pt_1000to1400
- QCD_Pt_1400to1800
- QCD_Pt_1800to2400
- QCD_Pt_2400to3200
- QCD_Pt_3200toInf

Data

- SingleMuon_Run2018A-UL2018
- SingleMuon_Run2018B-UL2018
- SingleMuon_Run2018C-UL2018
- SingleMuon_Run2018D-UL2018

A.0.2. MC and Data agreement

In order to validate the analysis strategy, comparisons between data and simulated samples have been performed. The presented MC/data agreement plots are shown without including contributions from Higgs boson processes, and correspond to the $\mu\tau$ channel in events containing a fatjet tagged as a $\tau\tau$ candidate by the ParticleNet algorithm. This provides a detailed validation of the background modelling in the relevant signal-enriched phase space.



(a) **(b)** **(c)** **(d)**

(e) **(f)**

Figure A.1.: MC/data comparison plots.