

Master's thesis

Master's Programme in Particle Physics and

Astrophysical Sciences, Particle physics

Sensitivity to a heavy $H^{\pm}$ decaying to $\tau^{\pm}\nu_{\tau}$ in the fully hadronic final state using Run 2 CMS data

Jere Mäki-Ikola

February 27, 2025

Supervisor(s): Sami Lehti, Kenneth Österberg

University of Helsinki

Faculty of Science

PL 64 (Gustaf Hällströmin katu 2a)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Particle Physics and Astrophysical Sciences, Particle physics

Tekijä — Författare — Author

Jere Mäki-Ikola

Työn nimi — Arbetets titel — Title

Sensitivity to a heavy $H^{\pm}$ decaying to $\tau^{\pm}\nu_{\tau}$ in the fully hadronic final state using Run 2 CMS data

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Master's thesis	February 27, 2025	93

Tiivistelmä — Referat — Abstract

The Standard Model (SM) of particle physics is one of the most precise theories with the latest discovery being the Higgs boson in 2012. Yet, it has its shortcomings; for example, not being able to explain the nature of dark matter, the hierarchy problem, or the matter-antimatter-asymmetry in the Universe.

One way to solve these shortcomings is to expand the theory of SM. Some of these theories, such as the Two-Higgs-Doublet Model (2HDM), introduce an expanded Higgs sector which includes multiple new Higgs bosons. One of these new predicted Higgs bosons is the charged Higgs boson ($H^{\pm}$).

This thesis focuses on heavy charged Higgs bosons ($m_H^{\pm} > m_t$ where m_t is the mass of the top quark) in the fully hadronic $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ decay channel using the CMS data collected at CERN's Large Hadron Collider (LHC) during Run 2 (years 2016-2018). This study uses simplified uncertainties and blinded data.

The aim of this thesis is to calculate expected limits by using transverse mass (m_T) distributions as inputs. This is done for the combination of all Run 2 data together and separately by individual years. Additionally, this thesis deduces the most optimal selection from 15 different jet selection criteria using year 2017 data. Lastly, combined expected limits using all jet selection combinations and projections using Run 2 + Run 3 luminosity in calculations are also produced.

Avainsanat — Nyckelord — Keywords

particle physics, Higgs boson, charged, hadron, tau lepton, missing transverse energy, BSM

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

Acknowledgements **1**

1 Introduction **2**

2 Particle physics theory **3**

2.1 Standard Model . . . 3

2.1.1 Overview . . . 3

2.1.2 Electroweak theory . . . 6

2.1.3 Higgs mechanism . . . 7

2.1.4 Quantum chromodynamics . . . 10

2.1.5 Tau leptons . . . 12

2.2 Beyond the Standard Model . . . 14

2.2.1 Problems with the SM . . . 14

2.2.2 BSM Models with Extended Higgs Sectors . . . 16

2.2.3 Charged Higgs bosons . . . 17

3 Experimental setup **20**

3.1 Large Hadron Collider . . . 20

3.2 Compact Muon Solenoid . . . 23

3.2.1 Tracker . . . 26

3.2.2 Electromagnetic calorimeter . . . 29

3.2.3 Hadron calorimeter . . . 31

3.2.4 Superconducting magnet . . . 34

3.2.5 The muon system . . . 36

3.2.6 Trigger system . . . 38

3.3 Event reconstruction and particle flow . . . 39

3.3.1 Pileup . . . 41

4 Data analysis **46**

4.1 Analysis production 46
4.2 Selections and background 48
4.3 Statistics 54
4.3.1 CL_S method and limits 55
4.3.2 Uncertainties 57

5 Results **61**
5.1 Validation 62
5.2 Selection Optimization Study 67
5.3 Expected Limits for 2017, 2018, and Full Run 2 69
5.4 Combined Limit of All Selections and Projected Sensitivity for Run 2 + Run 3 75
5.5 Future studies 80

6 Summary **83**

Bibliography **85**

A The usage of AI in this thesis **92**

B Uncertainties in the analysis **93**

Acknowledgements

First and foremost, I wish to express my gratitude to my supervisors, Sami Lehti and Kenneth Österberg, for giving me the opportunity to work on this fascinating project. Working on this has broadened my horizons and given me so much experience.

I would also like to extend thanks to Aapo Kössi, whose assistance and collaboration have been truly invaluable. Without your help, I would still be stuck in step one.

I am deeply thankful and debtful to my partner, Luca, for their unwavering support, patience, and love, which have sustained me through this challenging yet rewarding process.

To Emilia, my close friend and colleague in this field, your encouragement, insights, and support have been invaluable to me.

To my family, Miia, Aki, and Aapo, thank you for your unconditional love and encouragement, which have given me strength and motivation.

To all my other close friends, your support and motivation has given me hope to get this through the finish line.

I am also grateful to my colleagues at the Helsinki Institute of Physics for providing an inspiring work environment. Your shared knowledge and discussions have enriched my experience in countless ways.

1 Introduction

The discovery of the Higgs boson in 2012 by the CMS and ATLAS collaborations at the Large Hadron Collider (LHC) marked a milestone in particle physics. However, unresolved questions such as the nature of dark matter, the matter-antimatter asymmetry, and the hierarchy problem highlight the need for new physics beyond the Standard Model (SM).

One approach to addressing these open questions is to extend the SM through models that introduce an expanded Higgs sector. The Two-Higgs-Doublet Model (2HDM), for instance, predicts the existence of additional Higgs bosons, including Higgs bosons with an electric charge ($H^{\pm}$), which is absent in the SM. The properties and decay modes of $H^{\pm}$ vary depending on the specific BSM (Beyond the Standard Model) scenario, making it a key target in collider searches.

This thesis focuses on the search for heavy charged Higgs bosons ($m_{H^{\pm}} > m_t$) decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in a fully hadronic final state, using proton-proton collision data collected by the CMS experiment during Run 2 (2016-2018). The τ lepton is particularly interesting due to its enhanced coupling to the charged Higgs boson in scenarios with large values of $\tan\beta$, the ratio of the vacuum expectation values of the two Higgs doublets. Additionally, the fully hadronic decay mode offers a distinct experimental signature, reducing background contamination from processes involving electrons or muons.

One of the main objectives of this thesis is to derive expected limits on the production of a heavy charged Higgs boson using statistical analysis methods Additionally, this thesis aims to optimize the event selection strategy to be used in this analysis. Advanced event selection techniques are employed to enhance signal sensitivity while minimizing dominant SM background contributions.

2 Particle physics theory

2.1 Standard Model

2.1.1 Overview

The Standard Model (SM) of particle physics is a comprehensive theoretical framework that describes the fundamental particles and their interactions, excluding gravity. Built on the principles of quantum field theory, it incorporates the electroweak [1–3] and strong forces [4–6], which are unified under the SU(3) $\times$ SU(2) $\times$ U(1) gauge symmetry group. These forces are mediated by specific gauge bosons, which govern the interactions between matter particles and define the behavior of the universe at the smallest scales. A detailed description of the SM can be found in Ref. [7–9]

The fundamental particles in the SM are divided into fermions, which are the constituents of matter, and bosons, which mediate forces. Fermions are classified into quarks and leptons. Quarks exist in six flavors—up, down, charm, strange, top, and bottom—and interact via the strong force, forming composite particles such as protons and neutrons. Leptons include the electron, muon, tau, and their associated neutrinos. These particles primarily experience the weak and electromagnetic interactions, with neutrinos interacting exclusively via the weak force.

The gauge bosons of the SM facilitate the fundamental interactions. The photon mediates the electromagnetic force, the W and Z bosons are responsible for weak interactions, and gluons mediate the strong force, which binds quarks together to form hadrons. The Higgs boson is a unique scalar particle associated with the Higgs field. This field plays a critical role in spontaneous symmetry breaking, giving mass to particles like the W and Z bosons while leaving the photon massless.

All of the particles, both fermions and gauge bosons, are listed with their masses, charges and spins in Figure 2.1.

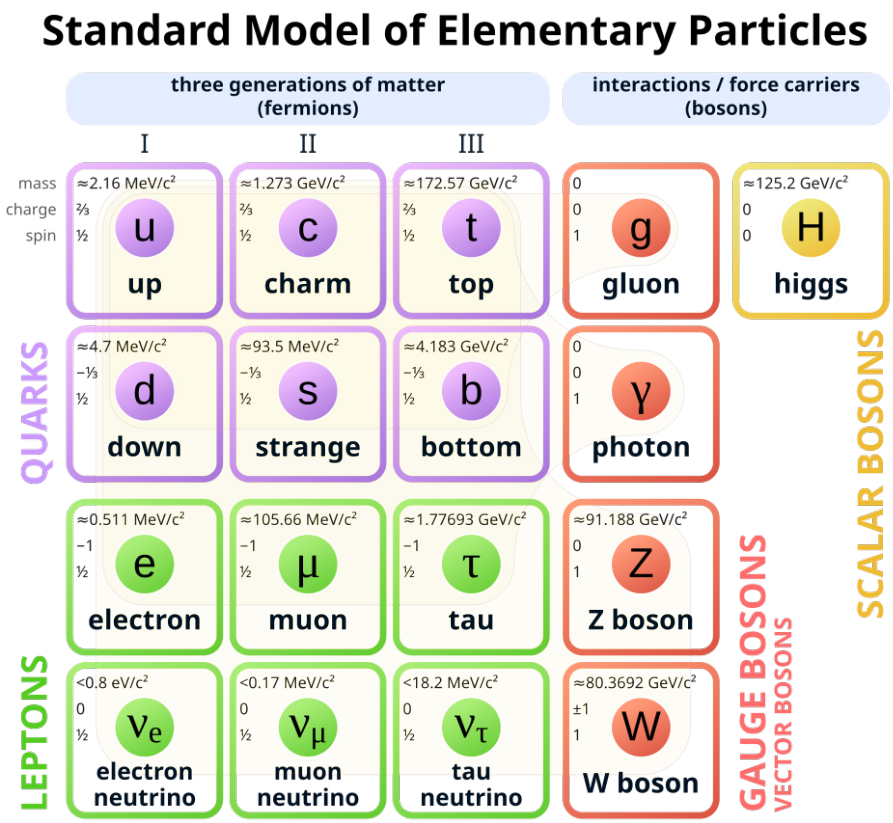


Figure 2.1: The particles of Standard Model divided into groups of fermions, bosons and other subgroups. The numbers on top-left corners represent the particle's mass, charge and spin. [10]

The theoretical structure of the SM is encoded in its Lagrangian, which combines contributions from the gauge fields, fermions, Higgs field, and Yukawa interactions. The complete Lagrangian is written as:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}.$$

The gauge term, $\mathcal{L}_{\text{gauge}}$, describes the dynamics of the gauge fields. For example, the SU(3) gauge term representing the strong force is given by:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu},$$

where $F^a_{\mu\nu}$ is the field strength tensor for gluons, defined as:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu.$$

Here, g_s is the strong coupling constant, and f^{abc} are the structure constants of the SU(3) symmetry group.

The fermion dynamics are encapsulated in $\mathcal{L}_{\text{fermion}}$, which describes the behavior of quarks and leptons. For a generic fermion field ψ, this term is written as:

$$\mathcal{L}_{\text{fermion}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi,$$

where D_μ is the covariant derivative, incorporating interactions with gauge fields.

The Higgs field introduces spontaneous symmetry breaking through $\mathcal{L}_{\text{Higgs}}$, given by:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi),$$

where ϕ is the Higgs field, D_μ is the covariant derivative, and $V(\phi)$ is the Higgs potential:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2.$$

This potential allows the Higgs field to acquire a nonzero vacuum expectation value, breaking the SU(2) $\times$ U(1) symmetry into U(1) electromagnetism.

Finally, $\mathcal{L}_{\text{Yukawa}}$ accounts for the interactions between the Higgs field and fermions, providing mass to particles after symmetry breaking:

$$\mathcal{L}_{\text{Yukawa}} = -y_f(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L),$$

where y_f are the Yukawa couplings.

The electromagnetic force arises from the U(1) symmetry, weak interactions from the SU(2) symmetry, and the strong force from the SU(3) symmetry. The unification of the electromagnetic and weak forces is achieved through the Higgs mechanism, which explains the distinct masses of the W and Z bosons compared to the massless photon.

2.1.2 Electroweak theory

Electroweak theory, a unification of electromagnetism and the weak nuclear force, is one of the cornerstones of the SM. This framework provides a unified description of how these two fundamental forces operate through the formalism of gauge theories. At its foundation, electroweak theory is based on the symmetry group $SU(2)_L \times U(1)_Y$. The $SU(2)_L$ component governs weak interactions, which affect left-handed particles, while $U(1)_Y$ represents the hypercharge symmetry underlying electromagnetic interactions. This structure is a non-Abelian gauge theory (a gauge theory with a non-commutative symmetry group), characterized by self-interactions of the gauge fields, giving rise to the complex dynamics of particle interactions.

The theory predicts that interactions are mediated by gauge bosons, with distinct roles for each particle. The photon (γ) governs electromagnetic interactions, which are long-range because the photon is massless. In contrast, the weak interactions are mediated by the massive $W^{\pm}$ and Z^0 bosons, limiting their range to subatomic distances. This distinction between massless and massive mediators is critical to the differing behaviors of the two forces. The interaction strengths are determined by two coupling constants, g for $SU(2)_L$ and g' for $U(1)_Y$. These constants relate to the Weinberg angle (θ_W), a fundamental parameter that describes how the photon and the Z^0 boson emerge as linear combinations of the original gauge fields W^3 and the $U(1)_Y$ field. The interplay of these constants and the symmetry structure underlies the unification of electromagnetic and weak forces within the electroweak framework.

Weak interactions are notable for their chiral nature, as they act exclusively on left-handed particles and right-handed antiparticles. This chiral asymmetry is intrinsic to the $SU(2)_L$ symmetry. Within this structure, fermions in the SM are organized into doublets under $SU(2)_L$, where only the left-handed components experience weak interactions, while right-handed components interact through $U(1)_Y$ alone. This arrangement explains the specific ways in which particles interact through the weak force. The dynamics of the weak force manifest in two forms of interactions. Charged current interactions, mediated by $W^{\pm}$, result in changes to the flavor of particles, such as in beta decay processes. Neutral current interactions, mediated by Z^0, do not alter the flavor of the particle but are responsible for

phenomena such as neutrino scattering. The theory's ability to account for both charged and neutral currents was a major breakthrough in unifying these forces under a common theoretical framework.

The interactions between quarks within the electroweak framework are further enriched by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, a cornerstone of the SM's description of flavor physics. The CKM matrix encodes the mixing between the weak interaction eigenstates (quantum states with definite values, or eigenvalues, for a specific observable like energy or angular momentum) and the mass eigenstates of the quarks. This misalignment means that the quark flavors as they appear in weak interactions are linear combinations of their mass eigenstates, a phenomenon that plays a crucial role in processes involving charged currents. The CKM matrix is a unitary 3×3 matrix that governs the transitions between the three generations of quarks: up, charm, and top in the up-type sector, and down, strange, and bottom in the down-type sector. Its elements, denoted V_{ij}, quantify the probability amplitude for a transition between a specific pair of up- and down-type quarks under the action of the $W^{\pm}$ bosons. For instance, the V_{ud} element represents the mixing between the up and down quarks, while V_{tb} corresponds to the coupling between the top and bottom quarks.

One of the most striking features of the CKM matrix is its role in CP violation, a phenomenon that introduces a fundamental asymmetry between matter and antimatter. The unitarity of the CKM matrix imposes constraints on its elements, leading to relationships known as unitarity triangles. The angles of these triangles encode the phases responsible for CP violation in weak interactions, a key aspect of explaining the observed baryon asymmetry of the universe. The structure of the CKM matrix reflects a hierarchy in the strengths of quark mixing, with transitions between quarks of the same generation (such as up to down or charm to strange) being dominant, while those between different generations (such as up to strange or charm to bottom) are progressively suppressed.

2.1.3 Higgs mechanism

The unification of electromagnetic and weak interactions within the framework of the SM relies on the $SU(2)_L \times U(1)_Y$ gauge symmetry as mentioned before. This symmetry,

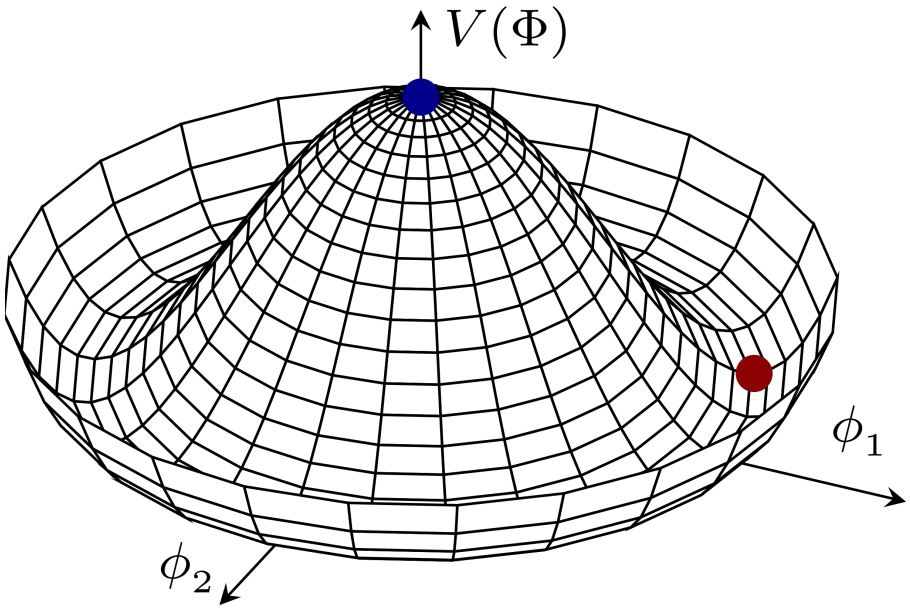


Figure 2.2: Illustration of the "Mexican hat" shaped Higgs potential. The unstable minimum is appointed as a blue ball and one of the possible stable minima is represented with a red ball.

however, is not manifest at low energies, where the two forces behave distinctly. Electromagnetism remains a long-range force mediated by a massless photon, while weak interactions are short-range due to the massive $W^{\pm}$ and Z^0 bosons. The theoretical explanation for this divergence lies in the phenomenon of electroweak symmetry breaking, achieved through the Higgs mechanism.

The Higgs mechanism, officially named Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism [11–16], introduces a scalar field, represented as a doublet under $SU(2)_L$. This field, denoted ϕ, interacts with the gauge bosons and is responsible for generating their masses. The scalar field is governed by a potential, expressed as $V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4$. For the choice of $\mu^2 < 0$, the potential takes on a characteristic shape resembling a "Mexican hat." The symmetry-breaking vacuum is not at the origin but lies on a circular trough of minima defined by the radial magnitude of ϕ. This setup spontaneously selects one of the minima, breaking the symmetry. The "Mexican hat" Higgs potential is illustrated in Figure 2.2. The spontaneous symmetry breaking is illustrated in Figure 2.3.

The field ϕ can be parameterized to reflect the symmetry-breaking configuration, with

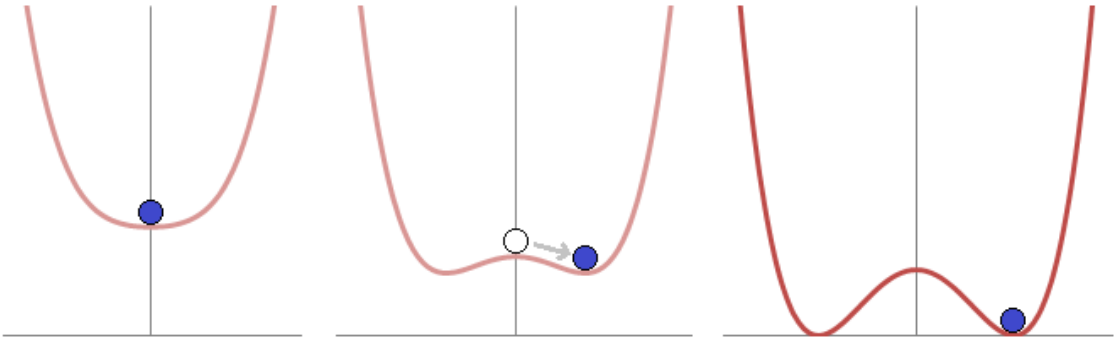


Figure 2.3: Illustration on how spontaneous symmetry breaking works. At low energy, the center becomes unstable and the stable minimum moves to a new position. The ball then spontaneously picks one of the minima choices and breaks the symmetry. $V(\Phi)$ is the Higgs potential where it is represented as a function of ϕ_1 and ϕ_2 real parameters. [17]

$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ where the neutral complex scalar field is defined usually as $\phi^0 = \frac{1}{\sqrt{2}}(v+h+i\chi)$. Here, $v = \sqrt{-\mu^2/\lambda}$ represents the vacuum expectation value (VEV), h is the physical scalar excitation corresponding to the Higgs boson, and χ is an unphysical degree of freedom eliminated through gauge fixing. The charged component ϕ^+ does not appear as a physical particle and it is not explicitly parameterized. This is because it is absorbed to provide the longitudinal polarization state for the $W^\pm$, giving them mass.

Gauge bosons acquire mass through interactions with the Higgs field's VEV. This occurs via the covariant derivative $D_\mu\phi$ in the Lagrangian $\mathcal{L}_{\text{Higgs}}$, where now D_μ encodes the gauge fields as $D_\mu = \partial_\mu - igW_\mu^a\tau^a - ig'B_\mu Y$. Here, W_μ^a are the $SU(2)_L$ gauge bosons, B_μ is the $U(1)_Y$ gauge boson, and τ^a and Y are their respective generators.

As the Higgs field acquires a VEV, the gauge bosons mix and form physical states. The charged bosons $W_\mu^\pm = (W_\mu^1 \mp iW_\mu^2)/\sqrt{2}$ emerge, while the neutral components undergo a rotation. The original fields W_μ^3 and B_μ mix to form the physical Z_μ and A_μ bosons:

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu, \quad A_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu.$$

The mixing angle θ_W, known as the weak mixing angle, is defined by the ratio of the gauge couplings: $\tan\theta_W = g'/g$. This angle encapsulates the relationship between the original gauge fields and their physical counterparts. Importantly, the photon A_μ remains

massless, ensuring the preservation of $U(1)_{\mathrm{EM}}$, while the Z_μ boson acquires a mass through its interaction with the VEV.

The masses of the gauge bosons are proportional to the VEV and the corresponding coupling constants. For the W and Z bosons, the masses are given by $M_W = gv/2$ and $M_Z = v/2\sqrt{g^2 + g'^2}$, respectively. The photon remains massless, reflecting the unbroken electromagnetic gauge symmetry.

In addition to providing masses for the gauge bosons, the Higgs mechanism also explains the origin of fermion masses. This is achieved through Yukawa interactions, which couple the scalar field to fermion fields. These interactions have the form $\mathcal{L}_{\mathrm{Yukawa}} = -y_f \bar{\psi}_L \phi \psi_R +$ h.c., where y_f is the Yukawa coupling. Upon symmetry breaking, the interaction yields fermion masses $m_f = y_f v/\sqrt{2}$.

The experimental confirmation of the Higgs mechanism came with the discovery of the Higgs boson in 2012 by the CMS and ATLAS collaborations [18–20]. This scalar particle, with a mass of approximately 125 GeV, has properties consistent with the SM predictions, including its couplings to gauge bosons and fermions [21–25]. Its discovery marked a monumental achievement, validating the theory of electroweak symmetry breaking and the unification of electromagnetic and weak forces.

2.1.4 Quantum chromodynamics

Quantum Chromodynamics (QCD) is the fundamental theory describing the strong interaction, one of the four known fundamental forces of nature. It governs the behavior of quarks and gluons, the elementary particles that form protons, neutrons, and other hadrons. QCD is a non-Abelian gauge theory based on the SU(3) symmetry group, which characterizes the "color charge" carried by quarks and gluons. This color charge is analogous to the electric charge in electromagnetism but comes in three types: red, green, and blue. These charges interact through the exchange of massless gauge bosons called gluons.

The dynamics of QCD are encoded in its Lagrangian, which describes the interactions between quarks and gluons. The QCD Lagrangian is given by:

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermion}} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi,$$

where ψ_q represents the quark fields of flavor q, and m_q denotes their respective masses. The covariant derivative, $D_\mu = \partial_\mu - ig_s T^a A^a_\mu$, ensures the gauge invariance of the theory under the SU(3) group. Here, A^a_μ are the gluon fields, g_s is the strong coupling constant, T^a are the SU(3) generators, and $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu$ is the gluon field strength tensor. The structure constants f^{abc} encapsulate the non-Abelian nature of the interactions.

QCD exhibits two defining properties: asymptotic freedom and confinement. Asymptotic freedom, a phenomenon discovered by Gross, Wilczek, and Politzer [26, 27], refers to the weakening of the strong interaction at high energies or short distances. This property enables the use of perturbative techniques in calculations involving large momentum transfers. Conversely, confinement ensures that quarks and gluons cannot exist as free particles but are permanently bound into color-neutral combinations called hadrons. This phenomenon becomes dominant at low energies and is inherently non-perturbative.

The behavior of the strong coupling constant, $\alpha_s = g_s^2/4\pi$, is scale-dependent and is governed by the QCD renormalization group equations. The evolution of α_s with the energy scale μ is described by the β-function:

$$\mu \frac{\partial \alpha_s}{\partial \mu} = -\beta(\alpha_s) = -\beta_0 \alpha_s^2 - \beta_1 \alpha_s^3 - \cdots,$$

where $\beta_0 = \frac{11 - 2n_f/3}{4\pi}$ is the leading-order coefficient, and n_f is the number of active quark flavors. The negative sign of β_0 underlies asymptotic freedom, while the divergence of α_s at low scales is indicative of confinement.

Confinement leads to hadronization, the process by which quarks and gluons transition into observable particles such as mesons and baryons. This process precludes the direct observation of free quarks or gluons in experiments. To study QCD in this regime, non-perturbative approaches such as lattice QCD are employed, where spacetime is discretized, and computations are performed numerically.

In high-energy collisions, QCD manifests in the production of jets, collimated sprays of particles arising from the fragmentation of quarks and gluons. These jets are crucial for probing QCD dynamics and for testing its predictions. Factorization theorems allow the separation of short-distance, perturbative processes from long-distance, non-perturbative effects, enabling precise calculations of cross-sections for scattering processes.

2.1.5 Tau leptons

The tau lepton (τ) is the heaviest of the three charged leptons in the SM, with a mass of approximately 1.777 GeV. As a third-generation lepton, τ is unique in its relatively long lifetime (2.9×10^{-13} seconds) and its capability to decay into a variety of final states.

As a spin-$\frac{1}{2}$ particle with a negative electric charge ($-e$), τ is accompanied by its corresponding neutrino (ν_τ). It participates in both electromagnetic and weak interactions, making it an important probe for investigating lepton universality and CP violation. Decays from τ can happen both leptonically and hadronically. In the SM, τ couplings to the Z and W bosons are identical to those of the electron and muon, barring effects of mass differences. Precision measurements of these couplings have consistently affirmed lepton universality.

The tau's relatively high mass opens up a range of decay channels, classified into leptonic and hadronic modes. Leptonic decays, such as $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$, proceed through W-boson-mediated interactions, similar to all other tau decays. τ hadronic decays constitute approximately 65% of all τ decays. The simplest hadronic mode is $\tau^- \to \pi^- \nu_\tau$, but more complex decays involving kaons (mesons composed of a strange quark and an up or down antiquark, or their antiparticles) and resonances, such as $\tau^- \to \rho^- \nu_\tau$ or $\tau^- \to a_1^- \nu_\tau$, also occur. These decays are mediated by the coupling of the τ to the hadronic current, offering a window into QCD at low energies. The branching fractions of the τ decays are listed in Table 2.1.

Table 2.1: Branching fractions for different τ decay channels based on [7,44]. $h^{\pm}$ represents a charged hadron which can be $\pi^{\pm}$ or $K^{\pm}$.

Decay Channel	**Resonance**	**Branching Fraction (%)**
Leptonic Decay Modes		35.6
$\tau^- \to e^- \nu_e \nu_\tau$		17.8
$\tau^- \to \mu^- \nu_\mu \nu_\tau$		17.4
Hadronic Decay Modes		
One-Prong Decays (excl. K^0)		48.2
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	ρ^-	25.9
$\tau^- \to h^- 2\pi^0 \nu_\tau$	a_1^-	9.6
$\tau^- \to h^- \nu_\tau + \geq 3\pi^0$		1.3
Three-Prong Decays (excl. K^0)		14.6
$\tau^- \to h^- h^- h^+ \nu_\tau$	a_1^-	9.5
$\tau^- \to h^- h^- h^+ \nu_\tau + \geq 1\pi^0$		5.1
Others (incl. decays with K^0)		≈ 2

Identifying τ in high-energy experiments is challenging due to the presence of neutrinos in their decay products. τ identification strategies rely on advanced algorithms to distinguish their decays from background processes. The identification of τ through their leptonic decays benefits from the well-understood properties of electrons and muons. In Higgs analyses, kinematic variables such as transverse momentum (p_T) and missing transverse energy (MET from hereon) are commonly used to suppress backgrounds from W- and Z-boson decays, while tau identification primarily relies on track multiplicity, decay vertex properties, and isolation criteria. For hadronic τ decays, jet substructure analysis is important. Discrimination against jets from quark or gluon fragmentation exploits variables like track multiplicity, decay vertex properties, and isolation criteria. Modern machine learning techniques are increasingly employed to enhance signal efficiency while

reducing background contamination.

2.2 Beyond the Standard Model

2.2.1 Problems with the SM

Before talking about possible extensions to the SM, the prominent problems are discussed. The SM is one of the most successful theoretical frameworks, describing fundamental particles and their interactions through electromagnetic, weak, and strong forces. Despite its predictive power and experimental success, the SM has significant limitations that highlight its incompleteness as a comprehensive theory of nature. This subchapter is based on and has a more thorough review in Ref. [7, 28]

One of the most prominent issues with the SM is its inability to incorporate gravity. While the SM is constructed within the framework of quantum field theory, gravity is described by general relativity, a classical field theory. The two frameworks are fundamentally incompatible, and attempts to unify them, such as string theory or loop quantum gravity, remain speculative and incomplete.

Another major challenge to the SM is the existence of neutrino masses, which contradict its original formulation. Neutrino oscillation experiments have conclusively shown that neutrinos have non-zero masses, yet the SM assumes neutrinos to be massless. The mechanisms responsible for these masses, such as the seesaw mechanism or new scalar fields, require physics beyond the Standard Model (BSM).

The SM also fails to explain dark matter, a critical component of the universe's energy density, inferred from astrophysical observations. Dark matter interacts gravitationally but not through electromagnetic, weak, or strong forces. The SM does not predict any particles that could serve as dark matter candidates, necessitating new physics in the form of Weakly Interacting Massive Particles (WIMPs), axions, or sterile neutrinos.

The asymmetry between matter and antimatter in the universe presents another significant problem. The SM predicts nearly equal production rates of matter and antimatter in the early universe. However, the observed universe is dominated by matter. While

CP violation in quark interactions is included in the SM, its magnitude is insufficient to account for the observed matter-antimatter asymmetry, indicating the need for additional sources of CP violation.

The hierarchy problem is another unresolved issue. This problem concerns the vast difference between the electroweak scale, on the order of hundreds of GeV, and the Planck scale, at approximately 10^{19} GeV, where gravitational effects become significant. The Higgs boson mass is highly sensitive to quantum corrections, and stabilizing it without fine-tuning appears unnatural. This disparity suggests that the SM might be incomplete or require extensions such as supersymmetry to resolve the hierarchy problem.

The SM also struggles with the strong CP problem. QCD, which describes the strong interaction, allows for CP violation through a term in its Lagrangian. Experimentally, such CP violation is extremely small, as indicated by the neutron's electric dipole moment. The absence of significant CP violation without fine-tuning remains unexplained.

Furthermore, while the SM successfully unifies the electromagnetic and weak forces, it does not unify the strong force or gravity. Theories such as Grand Unified Theories (GUTs) aim to merge all forces into a single framework, predicting phenomena like proton decay. Such predictions have not yet been observed experimentally, leaving unification an open challenge.

The SM incorporates the CKM matrix to describe quark flavor mixing and CP violation. However, it does not explain the origin of the specific values of CKM matrix elements or the observed hierarchies in particle masses and mixing angles.

Finally, the SM cannot account for the cosmological constant, a measure of dark energy driving the accelerated expansion of the universe. The observed value of the cosmological constant is many orders of magnitude smaller than what quantum field theory predicts, presenting a profound discrepancy.

These limitations underscore the need for a theoretical framework that extends BSM. Physicists continue to search for such a framework through experimental efforts and through the development of new theoretical models, including supersymmetry, extra dimensions, and additional gauge symmetries.

2.2.2 BSM Models with Extended Higgs Sectors

Several extensions to the SM propose modifications to the Higgs sector, which plays a critical role in electroweak symmetry breaking and particle mass generation. Among these, models with an extended Higgs sector, such as the Two Higgs Doublet Model (2HDM) [7, 28–35] and the Minimal Supersymmetric Standard Model (MSSM) [7, 28, 34, 36–39], offer compelling frameworks to address unresolved issues in the SM.

The 2HDM introduces an additional SU(2) scalar doublet to the SM, resulting in a Higgs sector with multiple scalar particles: two CP-even neutral Higgs bosons (h, H), one CP-odd neutral Higgs boson (A), and a pair of charged Higgs bosons ($H^{\pm}$). This extended Higgs sector leads to new physical phenomena and experimental signatures that are absent in the SM. [28, 32–34]

One motivation for the 2HDM is the natural incorporation of spontaneous CP violation [33]. The 2HDM addresses this issue by imposing discrete symmetries to control the couplings of Higgs fields to fermions. Depending on how the two doublets couple to up-type and down-type quarks, four distinct types of 2HDM are classified: Type I, Type II, Type X (lepton-specific), and Type Y (flipped) [35].

The scalar potential of the 2HDM is expressed as [7, 32]:

$$V(\Phi_1, \Phi_2) = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - \left[m_{12}^2 \Phi_1^\dagger \Phi_2 + \text{h.c.}\right] + \frac{\lambda_1}{2}(\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger \Phi_2)^2$$

$$+\lambda_3(\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) + \lambda_4(\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) + \left\{\frac{\lambda_5}{2}(\Phi_1^\dagger \Phi_2)^2 + \text{h.c.}\right\} + ...,$$

where Φ_1 and Φ_2 are the scalar doublets, m_{ij} are mass parameters, and λ_i are coupling constants. The spontaneous symmetry breaking of this potential leads to five physical Higgs bosons after absorbing three Goldstone bosons (massless scalar particles that arise from the spontaneous breaking of continuous symmetries).

The MSSM is an extension of the SM that incorporates supersymmetry, a symmetry between bosons and fermions. It predicts the existence of superpartners for all the SM particles and an extended Higgs sector with two Higgs doublets. In the MSSM, the Higgs potential is constrained by supersymmetry, leading to a distinct relationship among its parameters. The mass of the lightest Higgs boson h is bounded from above, typically

around 125 GeV in scenarios consistent with experimental observations. [36–39]

2.2.3 Charged Higgs bosons

Charged Higgs bosons ($H^{\pm}$) are distinctive features of models with an extended Higgs sector, such as the 2HDM. Unlike the neutral Higgs boson of the SM, $H^{\pm}$ carry an electric charge, leading to unique production and decay signatures. This subchapter looks closely at the predicted $H^{\pm}$ using Ref. [7, 28–39].

$H^{\pm}$ are produced in high-energy proton-proton collisions through various mechanisms, with the dominant process depending on their mass relative to the top quark mass (m_t). For $H^{\pm}$ lighter than the top quark, production could occur primarily through the decay of top quarks in top-antitop ($t\bar{t}$) pair events. In these scenarios, the top quark decays into an $H^{\pm}$ and a bottom quark ($t \to H^{+}b$), making this channel sensitive to the top-Higgs coupling and the parameter $\tan\beta$, which is the ratio of the vacuum expectation values of the two Higgs doublets in models like the 2HDM.

For heavier $H^{\pm}$, where $m_{H^{\pm}} > m_t$, direct production mechanisms should dominate. These include gluon-bottom fusion ($gb \to tH^{-}$), in which a gluon and a bottom quark collide to produce a top quark and an $H^{\pm}$. Additionally, $H^{\pm}$ can be produced in association with top quarks through processes such as gluon-gluon fusion ($gg \to t\bar{b}H^{-}$). These mechanisms involve intricate interplay between strong and electroweak interactions, making their cross sections sensitive to the underlying model parameters. Other, less significant processes, such as pair production of $H^{\pm}$ or their production in association with gauge bosons, provide complementary information about the extended Higgs sector but have smaller cross sections and are experimentally more challenging to observe. The gluon-gluon fusion process is illustrated with a Feynman diagram in Figure 2.4.

Once produced, $H^{\pm}$ decay through distinct channels, with the dominant modes depending on their mass and couplings. For lighter $H^{\pm}$, decays into lighter fermions such as charm and strange quarks ($H^{+} \to c\bar{s}$) are important at low values of $\tan\beta$, while decays into leptons, particularly a τ and its associated neutrino ($H^{+} \to \tau^{+}\nu_{\tau}$), dominate at high $\tan\beta$. For heavier $H^{\pm}$, decays into third-generation quarks, such as $H^{+} \to t\bar{b}$, become accessible and often dominate due to the large Yukawa couplings involved. The decay $H^{+} \to \tau^{+}\nu_{\tau}$

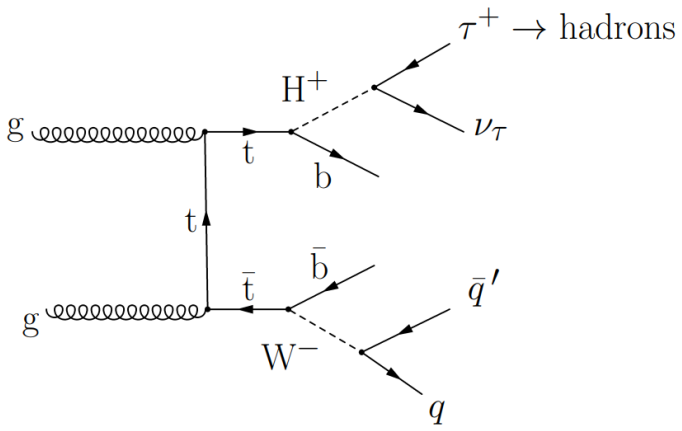


Figure 2.4: Feynman diagram of the gluon-gluon fusion which then produces H^+ and ultimately decays fully hadronically so that no distinct electrons or muons are present. [40]

remains an essential mode across a broad range of parameter space, particularly at high $\tan\beta$, where the coupling of $H^\pm$ to leptons is enhanced.

This thesis focuses specifically on the $H^+ \rightarrow \tau^+\nu_\tau$ decay channel, where the τ decays hadronically, producing a distinct experimental signature comprising a hadronic τ jet, MET due to the neutrino, and additional hadronic activity from associated production processes. The fully hadronic decay channel is of particular interest because it offers a cleaner experimental environment compared to other channels. In hadron-hadron collisions, fully hadronic τ decays benefit from more significant signal cross sections and reduced contamination from other processes involving neutrinos. If observed, the $H^+ \rightarrow \tau^+\nu_\tau$ channel provides direct sensitivity to $\tan\beta$ and $H^\pm$ mass. The calculated branching fractions of producing $H^\pm$ in Figure 2.5 are also in agreement with this. The branching fraction of the $H^+ \rightarrow \tau^+\nu_\tau$ channel stays very high across all possible masses of $H^\pm$. Although other channels have higher branching fractions with higher masses, the $H^+ \rightarrow \tau^+\nu_\tau$ channel stays as an important channel also due to how to discriminate the final state from the SM background processes.

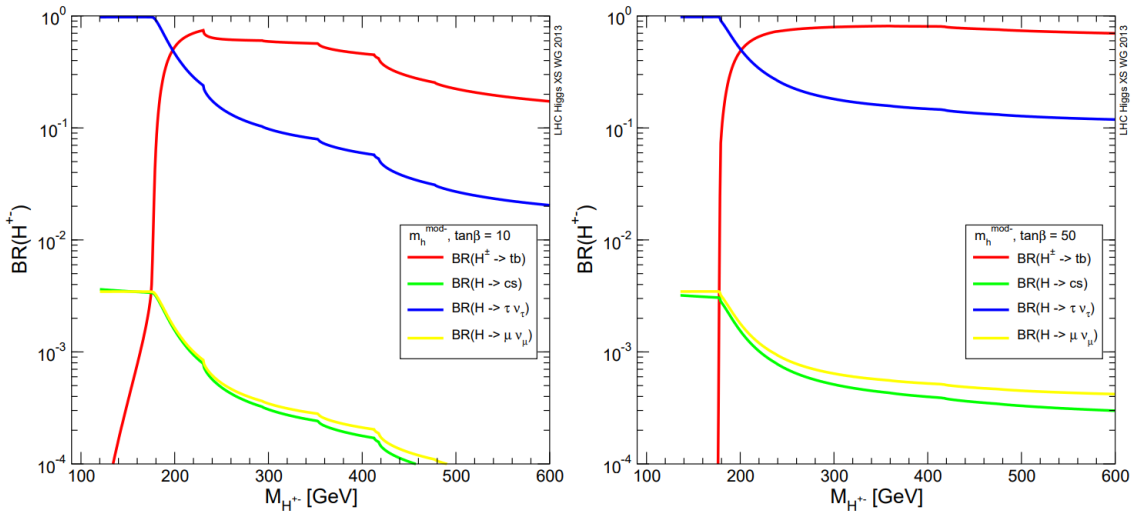


Figure 2.5: Branching fractions of the $H^{\pm}$ as a function of $M_{H^{\pm}}$. Different colours represent different decay channels where blue is the $H^{+} \to \tau^{+}\nu_{\tau}$ channel. On the left side, $\tan\beta = 10$ was assumed and on the right side, $\tan\beta = 50$ was assumed. [39]

3 Experimental setup

3.1 Large Hadron Collider

The Large Hadron Collider (LHC), located at CERN near Geneva, Switzerland, is the world's largest and most powerful particle accelerator. It was designed to explore fundamental questions in physics, including the nature of mass, the existence of dark matter, and the possibility of extra dimensions. An overview of LHC is provided in this chapter. A detailed description can be found in Ref. [41].

The LHC operates within a circular tunnel that spans 26.7 kilometers in circumference, lying between 45 and 170 meters underground. This tunnel was initially constructed for the Large Electron-Positron Collider (LEP) between 1984 and 1989. It is situated on the border between Switzerland and France, making the LHC an international project in both scope and collaboration. The tunnel itself cuts through regions of molasse rock and limestone, offering stable geological conditions necessary for the operation of such a large and sensitive machine.

At its core, the LHC is a hadron collider, meaning it accelerates hadron particles made of quarks, such as protons. The machine was built to accelerate two beams of protons in opposite directions around the circular track and bring them into head-on collisions at specific interaction points. By colliding these protons at extremely high energies, scientists can recreate conditions similar to those that existed shortly after the Big Bang. This allows them to study rare particle interactions and phenomena that cannot be observed under normal conditions. The LHC is designed to reach a center-of-mass collision energy of 14 TeV for proton-proton collisions, making it the highest energy particle collider ever built. Additionally, it can collide heavy ions, such as lead ions, with a center-of-mass energy of 2.8 TeV per nucleon, which is used to investigate states of matter like quark-gluon plasma - an extremely hot and dense form of matter believed to have existed in the early universe.

The LHC is made up of several highly advanced technologies, all working together to

accelerate and control particle beams with extraordinary precision. One of the most important components of the LHC is its system of superconducting magnets. These magnets are used to steer and focus the beams of protons around the circular accelerator. There are 1,232 dipole magnets in total, each approximately 15 meters long. These dipole magnets generate powerful magnetic fields of up to 8.33 Tesla, which are required to bend the paths of the particles as they circle the accelerator. Quadrupole magnets are also used to focus the beams, ensuring they stay tightly packed and concentrated.

To achieve superconductivity, these magnets must be kept at extremely low temperatures. The LHC's magnets are cooled to 1.9 K (-271.3°C) using superfluid helium, which is essential to maintain their superconducting state. At these ultra-low temperatures, the magnets can carry enormous electric currents without resistance, enabling them to generate the intense magnetic fields necessary for steering the high-energy beams. The LHC's cryogenic system is one of the largest in the world, capable of cooling large volumes of helium to temperatures close to absolute zero. This system is a critical part of the machine's ability to operate at such high energy levels.

Another vital component is the radiofrequency (RF) cavities, which are used to accelerate the protons. These cavities generate electromagnetic fields that push the protons forward, increasing their energy as they complete each circuit of the accelerator. The RF system operates at a frequency of 400 MHz, continuously imparting energy to the particle beams over several minutes until they reach nearly the speed of light. In addition, the vacuum system of the LHC ensures that the beams travel through an environment free of air and other particles. The beam pipe is maintained at an ultra-high vacuum (about 10^{-13} atmospheres) to prevent the protons from colliding with gas molecules, which could disrupt the beam and affect the experiments.

The LHC also houses four major detectors, each designed for specific types of particle physics research. CMS and ATLAS are two general-purpose detectors that focus on a wide range of physics phenomena. These detectors were instrumental in the discovery of the Higgs boson in 2012, one of the LHC's most significant achievements. Both CMS and ATLAS are used to search for new particles, study the properties of known particles, and explore the possibility of new physical theories BSM, such as supersymmetry and extra di-

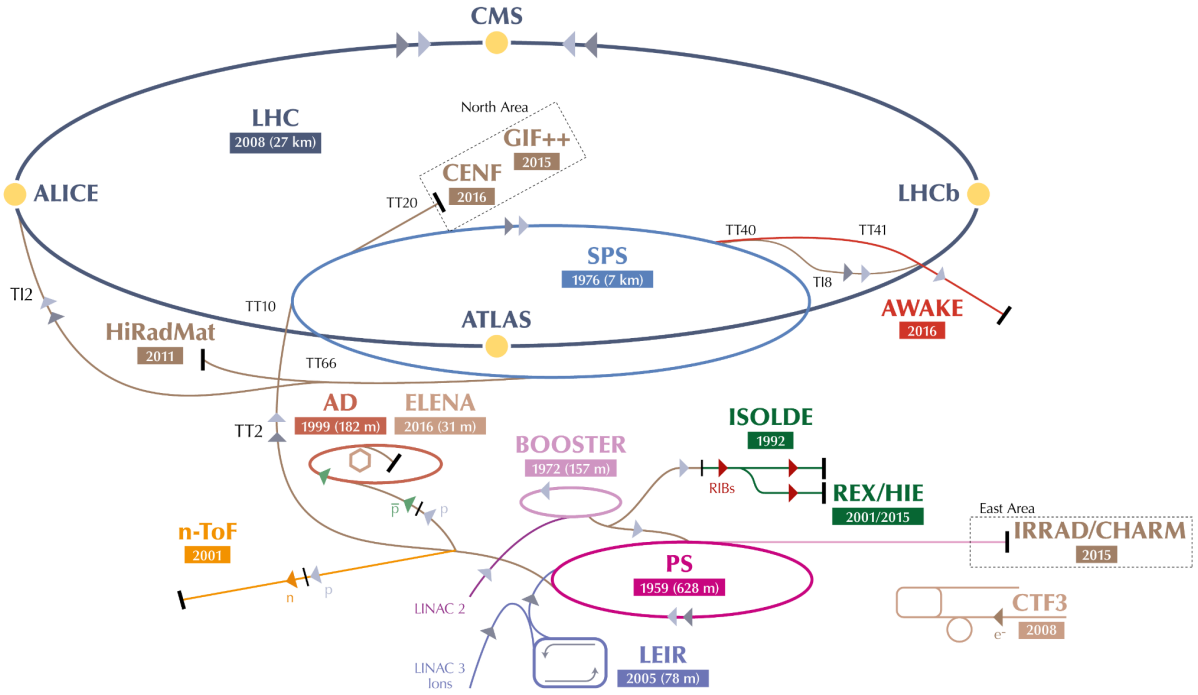


Figure 3.1: An illustration of the CERN accelerator complex with LHC, its predecessors, and other ongoing collaborations. [42]

mensions. ALICE, another of the LHC's major detectors, is specifically designed to study the quark-gluon plasma created in heavy-ion collisions. This state of matter, thought to have existed in the early universe, is critical for understanding how the fundamental forces of nature behaved just after the Big Bang. The LHCb detector focuses on studying particles containing bottom quarks, which are crucial for understanding the matter-antimatter asymmetry in the universe. The map of the LHC is illustrated in Figure 3.1 more closely featuring other collaborations than the four major detectors.

Looking to the future, the LHC is undergoing a series of upgrades aimed at increasing its performance and the amount of data it can collect. One of the most significant projects is the High-Luminosity LHC (HL-LHC), expected to start operating in the late 2020s. This upgrade will increase the number of particle collisions by a factor of 3-4, allowing scientists to gather even more data and observe rarer processes that occur in the collisions. The increased luminosity will enable the LHC to probe deeper into the structure of matter and test the predictions of new theoretical models with higher precision.

3.2 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of the primary experiments at the LHC at CERN located at point 5 in Cessy, France. The CMS detector is massive, measuring 21.6 meters in length and 14.6 meters in diameter, with a total weight of around 14000 tons. Built as a general-purpose detector, CMS is designed to investigate a wide range of physics phenomena, including the Higgs boson, dark matter, supersymmetry, and extra dimensions. Operating at the energy frontier, CMS studies high-energy proton-proton and lead-lead collisions, with a center-of-mass energy of 13 TeV for protons and 5.5 TeV for heavy ions, making it a key instrument for exploring the SM and potential new physics.

At the core of CMS lies a large superconducting solenoid magnet, which generates a magnetic field of 3.8 Tesla and is 6 m in diameter. This magnetic field bends the trajectories of charged particles, allowing the measurement of their momenta with high precision. Surrounding the magnet are several layers of particle detectors, each playing a crucial role in identifying and measuring the properties of the particles produced in collisions.

The innermost component is the tracking system, composed of a high-resolution silicon pixel and strip detector. These detectors measure the paths of charged particles as they traverse the magnetic field, providing precise information about their momentum and position. Beyond the tracking system are the calorimeters. First, the electromagnetic calorimeter (ECAL), made of lead-tungstate crystals, designed to measure the energy of photons and electrons. Surrounding this is the hadron calorimeter (HCAL), which measures the energy of hadrons like protons, neutrons, and pions. Both calorimeters work together to provide a complete view of the energy and momentum flow from particle collisions.

A defining feature of CMS is its advanced muon detection system. Muons, which are heavier counterparts of electrons, interact only minimally with the calorimeters and pass through most of the detector without being stopped. To detect them, CMS has four layers of muon detectors embedded within the return yoke of the solenoid magnet. The system allows precise measurement of the muons' paths, which is crucial for many important physics analyses, such as studies of the Higgs boson and searches for new particles. The

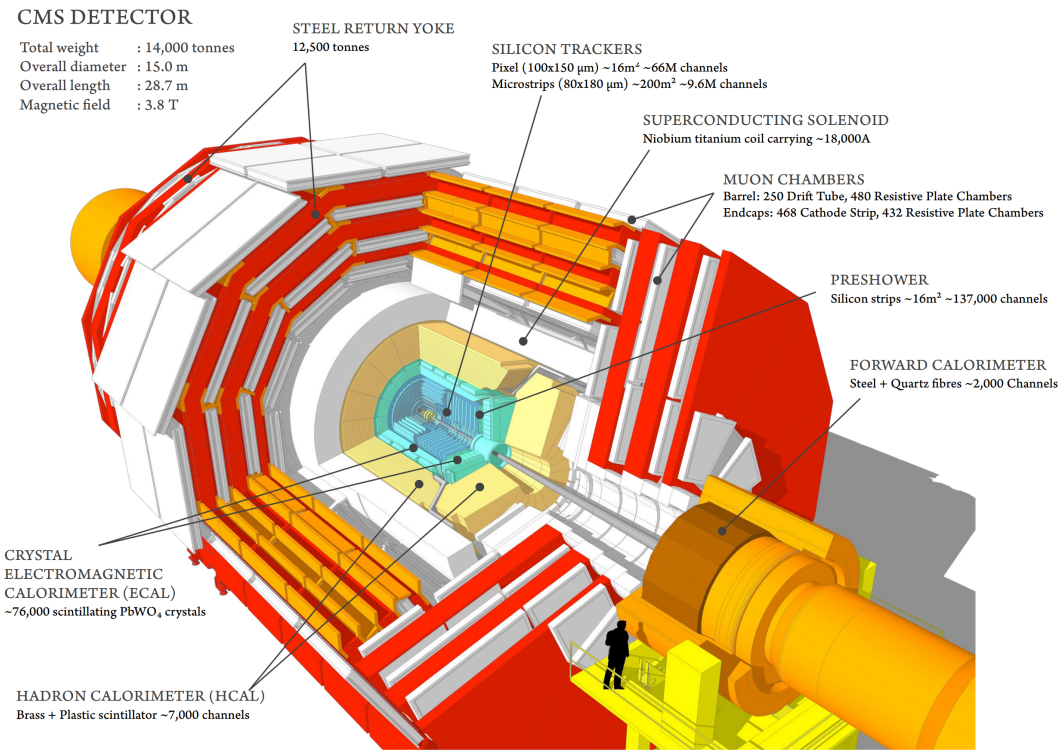


Figure 3.2: Illustration of the CMS detector and its subdetectors. [43]

whole structure of the CMS detector is illustrated in Figure 3.2.

In analyzing particle collisions, CMS physicists frequently use the concept of the transverse plane, which is perpendicular to the direction of the colliding beams. Since much of the interesting physics occurs in the plane transverse to the beamline, transverse momentum is often used as a key parameter to study the properties of the particles produced. Analyzing particle momenta in this plane helps focus on particles with significant momentum perpendicular to the beamline, filtering out background noise and low-energy particles that are less relevant.

Two important quantities used in CMS analyses to describe the angular distribution of particles are rapidity and pseudorapidity. Rapidity is a function of both the energy and the longitudinal momentum of a particle along the beam axis. It is particularly useful because rapidity differences remain invariant under Lorentz boosts along the beam direction, making it an ideal quantity for comparing particle distributions in high-energy collisions. For high-energy physics, rapidity is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right).$$

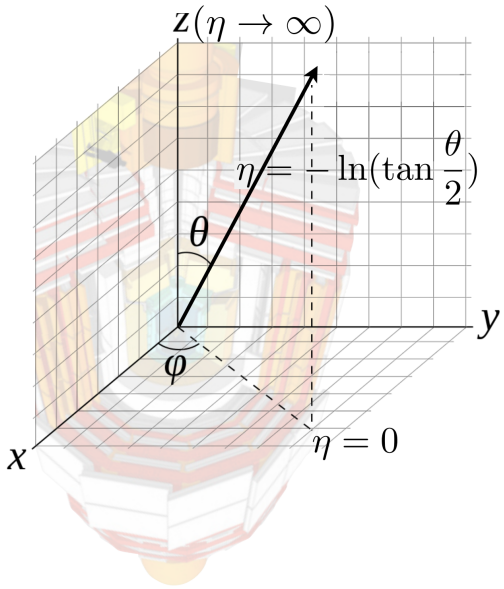


Figure 3.3: Illustration of the CMS coordinate system. [44]

When $m \ll p$, the mass can be ignored. It is then more convenient to use pseudorapidity because it only requires knowledge of the particle's polar angle relative to the beamline. Pseudorapidity is defined as:

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

where θ is the polar angle. Pseudorapidity is preferred over rapidity in CMS because the detector is optimized to measure the angular distribution of particles rather than their energy or longitudinal momentum directly. This simplifies event reconstruction and provides an efficient way to describe particle trajectories across the detector.

The use of pseudorapidity also aligns with the focus on transverse variables, as most interesting physics events occur in the transverse plane. Pseudorapidity enables the detector to describe the angular distribution of particles without the need to measure longitudinal momentum, which can be challenging in high-energy collisions. This geometric advantage ensures that measurements remain accurate and straightforward across the wide pseudorapidity range covered by the CMS detector. This can be seen illustrated in Figure 3.3.

In the following sections, an overview of CMS subdetectors is provided. A detailed description of the CMS detector components can be found in Ref. [45].

3.2.1 Tracker

The CMS Tracker is a highly complex and precise detector system at the core of the CMS experiment, designed to measure the trajectories of charged particles produced in high-energy collisions. It is composed of silicon-based detectors arranged in a cylindrical structure surrounding the interaction point, providing high-resolution spatial information that is crucial for reconstructing particle tracks and determining their momentum. The tracker is situated within a 3.8 T solenoidal magnetic field, which causes charged particles to curve depending on their momentum and charge. The tracker's primary function is to measure these curved paths with a high precision, enabling detailed reconstruction of the particles' origins, decay products, and momenta.

The CMS Tracker consists of two primary subsystems: the Silicon Pixel Detector and the Silicon Strip Tracker. While the entire tracker could ideally be composed of pixel detectors due to their superior resolution and ability to withstand high particle flux, the extensive use of pixel sensors is limited by cost and technological constraints. Consequently, the pixel detector is strategically placed in regions closest to the beamline, where the particle density and radiation levels are highest, ensuring optimal performance in these critical areas. The Silicon Strip Tracker, which is more cost-effective and covers a much larger area, complements the pixel detector by extending tracking coverage to regions further from the interaction point.

The innermost part of the tracker is the Silicon Pixel Detector, which is composed of three barrel layers and two end-cap disks on each side, closely surrounding the interaction point. This region experiences the highest particle flux and radiation levels due to its proximity to the beamline, making it essential to use finely segmented sensors. Each pixel in the detector is roughly 100×150 μm in size, providing a spatial resolution of about 10 μm in the transverse plane and 20 μm along the beam axis. This high resolution is achieved through charge-sharing effects, where the collected charge from a single particle interaction spreads across multiple neighboring pixels due to the Lorentz drift induced by the 4 T magnetic field of CMS. By reading out the analog pulse height and applying charge interpolation techniques, the detector can achieve a position resolution finer than the individual pixel size, enhancing the precision of vertex reconstruction and

track measurement. This fine granularity is essential for accurate vertex reconstruction, particularly for identifying the primary interaction vertex and for detecting displaced vertices from the decays of short-lived particles such as b-quarks or τ. The compact size and high precision of the Pixel Detector enable it to operate effectively even in the dense environment near the beamline, where multiple particles are produced in each bunch crossing.

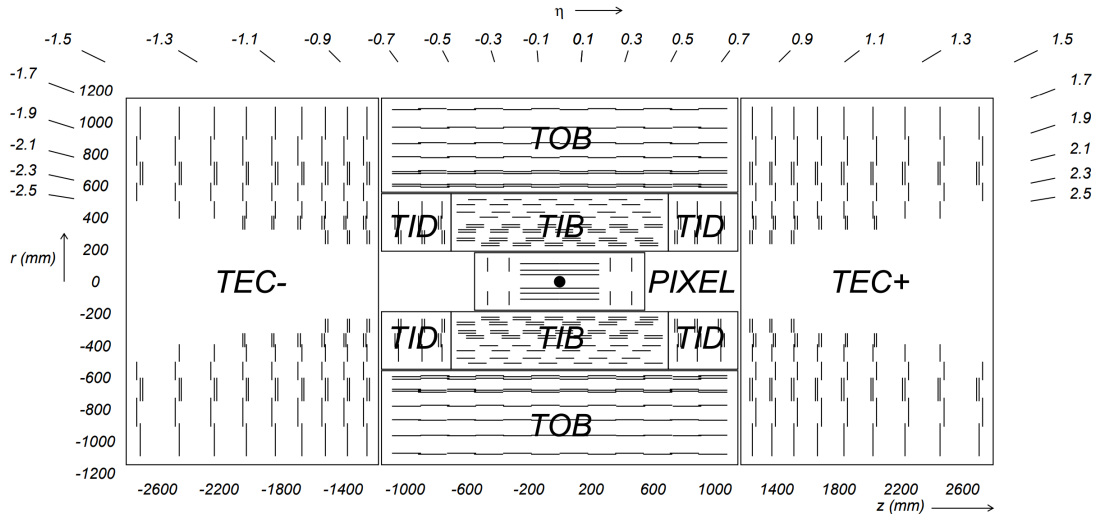


Figure 3.4: Illustration of the CMS Tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits. [45]

Surrounding the Pixel Detector is the Silicon Strip Tracker, which is divided into the Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker Inner Disks (TID), and Tracker End-Caps (TEC). This is illustrated in Figure 3.4. The strip sensors are designed to cover a much larger area than the pixel sensors, allowing for efficient tracking of particles across a wide range of pseudorapidities. The strip pitch ranges from 80 μm in the inner layers to 180 μm in the outer layers, striking a balance between spatial resolution and detector coverage. The Silicon Strip Tracker covers an area of over 200 m^2, making it the largest silicon detector ever constructed. This extensive coverage is essential for accurately reconstructing the tracks of particles with a wide range of transverse momenta, from low-energy particles that spiral tightly in the magnetic field to high-energy particles that follow near-straight paths.

The mechanical structure of the CMS Tracker is designed to ensure stability, alignment,

and minimal material budget, while supporting the complex arrangement of sensors. The tracker is housed within a cylindrical carbon fiber support structure, which provides rigidity and stability while maintaining a low mass to minimize multiple scattering and energy loss of particles passing through. The support structure incorporates carbon fiber rails, adjustable sliding pads, and precision-machined planes to maintain alignment during operation. The central support tube, which spans 5.30 meters in length with an inner diameter of 2.38 meters, is suspended on the hadronic calorimeter (HCAL) barrel and serves as the backbone of the entire system. This tube is a composite structure, featuring a carbon fiber skin with a lightweight Nomex honeycomb core, designed to handle the mechanical loads and thermal stresses that arise during operation.

In addition to supporting the silicon detectors, an independent structure holds the Pixel Detectors, central beam pipe, and radiation monitors. This modular design allows for independent access and maintenance of these components, reducing the risk of mechanical disturbances to the delicate strip detectors during interventions. The system's design also considers long-term stability, particularly in the high-radiation environment near the beamline, where radiation-induced damage can affect detector performance over time. Active cooling systems, made of lightweight materials such as carbon fiber and titanium, are integrated into the support structure to manage the heat generated by the densely packed electronics and maintain the operational temperature of the sensors within a stable range.

The combination of the highly granular Pixel Detector and the extensive coverage of the Silicon Strip Tracker enables the CMS Tracker to deliver precise measurements of particle trajectories, which are essential for reconstructing the complex final states that emerge from high-energy proton-proton collisions. The design allows for the efficient identification and reconstruction of charged particles in a wide range of momenta and pseudorapidities, making the CMS Tracker a crucial component for both precision measurements and the search for new physics phenomena.

3.2.2 Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) is a hermetic and homogeneous calorimeter that surrounds the CMS tracker. Its primary function is to measure the energy of photons and electrons by absorbing them and converting their energy into a detectable signal.

The components of ECAL are divided into three main sections: the barrel, the endcaps, and the preshower detector. The barrel and the endcaps consist of scintillating lead tungstate (PbWO4) crystals, while the preshower detector in front of the endcaps is a sampling calorimeter composed of lead radiators and silicon strip sensors. The ECAL Barrel (EB) consists of 61,200 crystals arranged in a cylindrical shape around the interaction point. The crystals are organized into 36 supermodules, each containing four modules of 1700 crystals. The use of supermodules allows for a modular construction, making the ECAL easier to assemble and maintain. Each supermodule is an independent unit that can be calibrated and tested separately, contributing to the overall reliability and efficiency of the ECAL. The supermodules also enhance the detector's ability to precisely measure the energy of particles by minimizing the gaps between crystals. The barrel covers the $|\eta| < 1.479$ region. The two ECAL endcaps (EE) close the ECAL at both ends of the barrel. Each endcap contains 7324 crystals arranged in a rectangular grid, optimized for detecting particles emitted at small angles relative to the beamline. The endcaps cover the $1.479 < |\eta| < 3.0$ region. The ECAL Preshower (ES) is positioned in front of the ECAL endcaps. The Preshower detector helps in distinguishing between single photons and closely spaced photon pairs, such as those produced in the decay of neutral pions. The positions and η ranges of the ECAL components are illustrated in Figures 3.5 and 3.6.

Within the endcap regions, the crystals are grouped into units known as supercrystals. Each supercrystal is a 5×5 matrix of crystals, forming a modular unit that simplifies assembly, calibration, and maintenance. These supercrystals are an essential structural element, providing the necessary spatial granularity to resolve complex electromagnetic showers, particularly in the high-radiation environment near the beamline. The endcap disks are split into two halves (Dees) which hold these supercrystals.

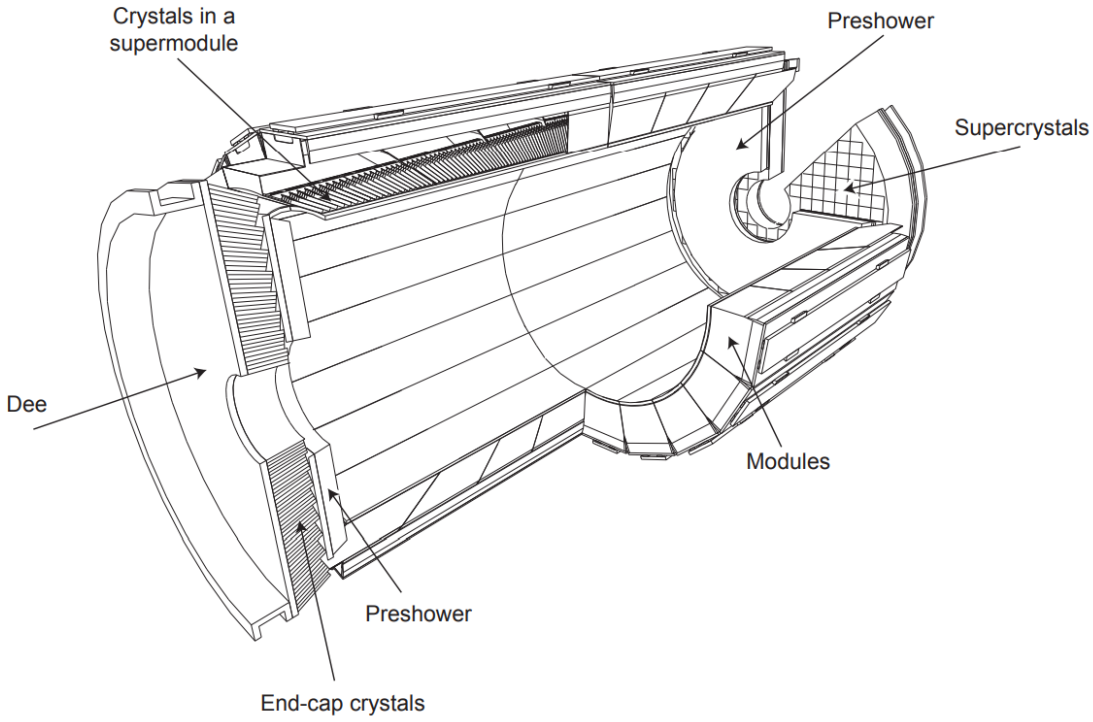


Figure 3.5: Illustration of ECAL detector's structure [45]

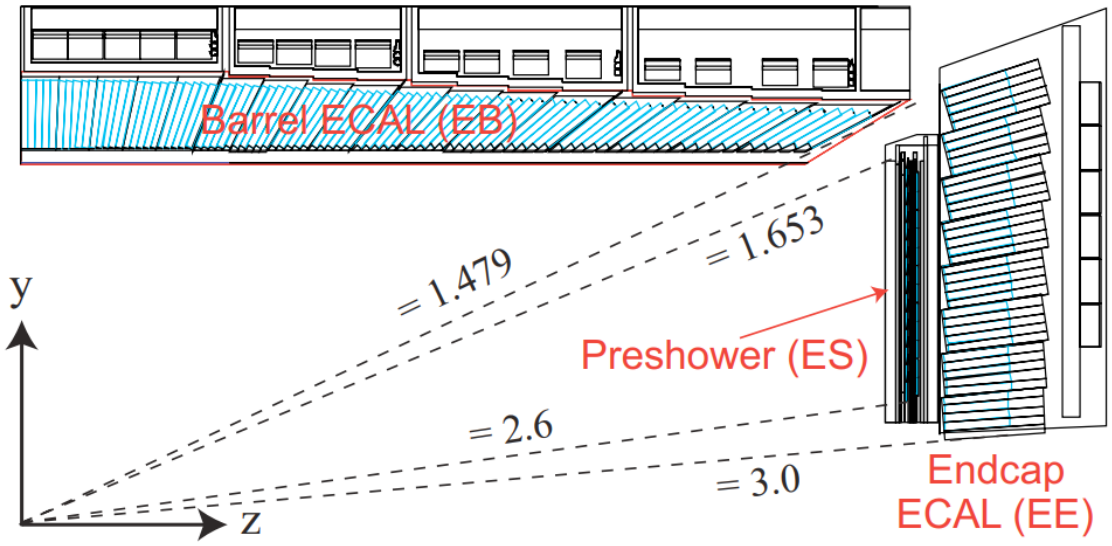


Figure 3.6: Illustration of the transverse section in ECAL. The numbers represent the pseudorapidities (η). [46]

The detection mechanism in the ECAL is based on scintillation, where the incoming electromagnetic particles interact with the dense lead tungstate crystals, producing light

(scintillation photons) proportional to the energy deposited. This light is then detected by photodetectors. In the barrel section, the scintillation light is detected by silicon Avalanche Photodiodes (APDs). APDs are chosen for their high quantum efficiency, ability to operate in a high magnetic field, and radiation hardness, which are crucial in the CMS environment. In the endcaps, the light is detected by Vacuum Phototriodes (VPTs). These are used due to their ability to withstand the harsher radiation environment closer to the beamline.

Achieving high precision in energy measurement is crucial for the ECAL, especially in the context of identifying electrons and photons in complex events. The energy resolution of the ECAL can be expressed as:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \tag{3.1}$$

where S is a stochastic term, N is a noise term, and C is a constant term. The ECAL has been designed to achieve an energy resolution of about 0.5% for electrons and photons in the central region.

Given the high radiation environment within the LHC, particularly in the CMS endcap regions, the ECAL components are designed to be radiation hard. The $PbWO_4$ crystals and photodetectors undergo radiation-induced changes over time, which affects light output and, consequently, the energy resolution. To mitigate this, the ECAL is equipped with a laser monitoring system that tracks the transparency of the crystals in real-time, allowing for corrections to be applied to calibration.

3.2.3 Hadron calorimeter

The next subsystem in the CMS detector is the Hadron Calorimeter (HCAL), which measures the energy of hadrons - particles composed of quarks like protons, neutrons, and pions (the lightest mesons, composed of up and down quarks and their antiquarks, mediating the strong nuclear force at low energies) - as well as jets and MET. The HCAL is a sampling calorimeter, composed of alternating layers of dense absorber materials and active scintillators. The absorbers, made primarily from brass in the barrel and endcaps due to its non-magnetic properties and higher density compared to iron, initiate hadronic showers, while the scintillators collect the light produced by the energy deposition in these

showers. In contrast, the Forward Hadron Calorimeter (HF) utilizes steel as the absorber material to handle the extreme particle flux and radiation levels experienced in the forward regions. This design allows the HCAL to measure both the position and energy of hadronic particles traversing it, contributing critically to event reconstruction in the CMS detector. The HCAL is subdivided mainly into four main components: the Hadron Calorimeter Barrel (HB), the Hadron Calorimeter Endcaps (HE), the Outer Hadron Calorimeter (HO), and the Forward Hadron Calorimeter (HF). Each of these subsystems is optimized for the specific geometric and energy requirements across a wide range of pseudorapidities (η). The HCAL and η ranges are illustrated in Figure 3.7.

The Hadron Calorimeter Barrel (HB) covers the central region of the CMS detector, spanning the pseudorapidity range of $|\eta| < 1.3$. The HB consists of two half-barrels, designated HB+ and HB-, which are wrapped around the inner electromagnetic calorimeter (ECAL). The structure of the HB includes 32 towers along each half-barrel, segmented radially into multiple layers of brass absorber plates and plastic scintillator tiles. The brass plates, chosen for their high density and relatively low cost, ensure effective containment of hadronic showers while maintaining a compact detector footprint. Each tower is approximately 0.087 in η and 0.087 in azimuthal angle (ϕ), providing a granularity suitable for jet reconstruction and transverse energy measurements. The scintillation light generated in each layer is collected by wavelength-shifting fibers and read out by hybrid photodiodes (HPDs), which convert the light into electrical signals. The HB's depth ranges from 5.8 to 10 interaction lengths (λ), ensuring nearly complete containment of hadronic showers originating from the interaction point.

In the forward region, the Hadron Calorimeter Endcaps (HE) extend the coverage of the HCAL to a pseudorapidity range of $|\eta| < 3$. Positioned at either end of the barrel, the HE is designed to cope with the higher particle flux and energy densities typical of the forward regions. The HE consists of 14 layers of brass absorbers, interleaved with scintillator tiles similar to those in the HB, but with greater depth to accommodate the more intense hadronic showers. The segmentation in η and ϕ is finer than in the HB, with tower sizes of approximately 0.087 in η and 0.174 in ϕ. This enhanced granularity is crucial for resolving closely spaced particles in jet-rich final states. The HE provides an energy resolution that is slightly worse than the HB due to the higher energies encountered

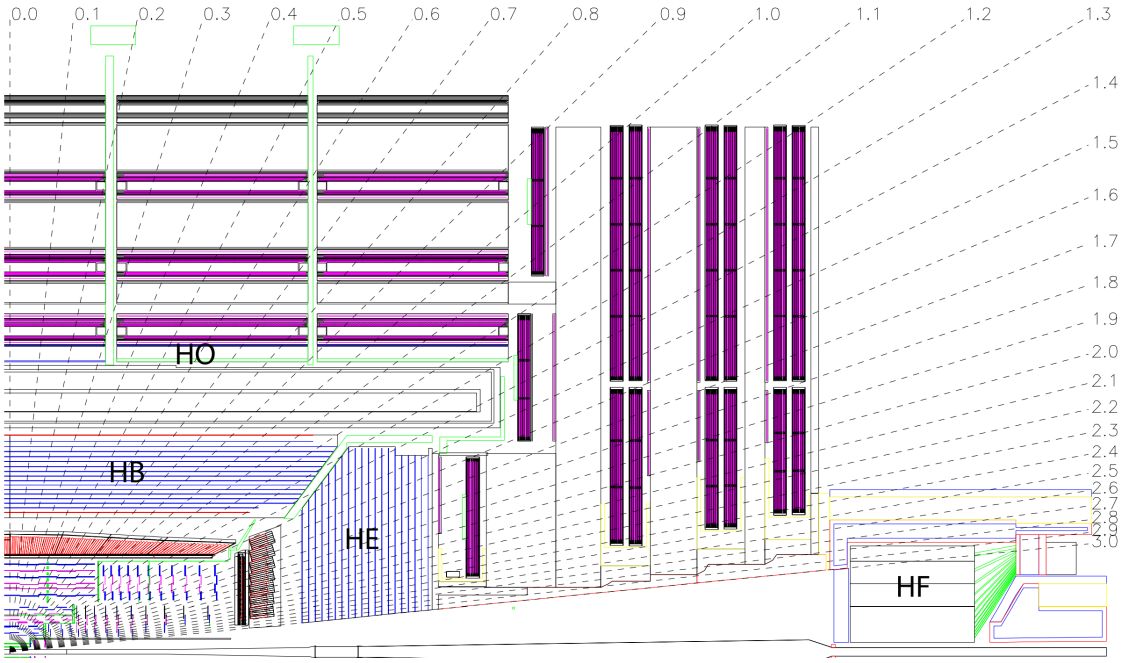


Figure 3.7: Illustration of longitudinal view of HCAL detector. Numbers represent η. [45]

but remains sufficient for precise measurements of forward jets and MET. The depth of the HE extends up to 10 λ, allowing it to fully contain hadronic showers, even at high energies.

To improve the energy resolution, particularly for particles with late-developing showers, the Outer Hadron Calorimeter (HO) is placed outside the solenoid magnet, surrounding the entire barrel region. The HO acts as a "tail catcher," providing additional sampling for showers that leak through the HB. The HO is composed of a single layer of scintillator tiles positioned behind the solenoid, adding an effective depth of 11 interaction lengths λ when combined with the HB. This ensures near-total containment of hadronic showers, reducing the likelihood of underestimating the energy of high-p_T jets. The HO plays a significant role in improving the overall energy resolution for jets and particles, especially in events with large transverse momentum, and aids in calculating accurate MET. The granularity of the HO matches that of the HB to facilitate seamless integration of data across the barrel region.

For extreme forward coverage, the Forward Hadron Calorimeter (HF) is placed at both ends of the detector, extending the pseudorapidity range up to $|\eta| = 5.2$. The HF is critical for detecting particles produced at small angles relative to the beamline, which

are essential for studying forward physics and minimizing pile-up effects. Unlike the other HCAL subsystems, the HF uses steel as the absorber material and quartz fibers as the active medium. Charged particles passing through the HF generate Cherenkov light (electromagnetic radiation emitted when a charged particle moves through a medium faster than the speed of light in that medium) in the quartz fibers, which is collected and transmitted to photomultiplier tubes (PMTs) for signal readout. The HF is divided into two parts: one optimized for electromagnetic showers and the other for hadronic showers, allowing for effective separation of the two. The granularity in the HF is coarser, with towers of 0.175 in η and 0.175 in ϕ, which is sufficient given the forward physics requirements. The energy resolution of the HF is lower compared to the HB and HE due to the intense radiation environment and the challenges of maintaining high resolution at such high pseudorapidities. Nevertheless, the HF is indispensable for reconstructing the total energy balance in high-energy events and plays a pivotal role in forward jet and MET analyses.

The comprehensive design of the CMS HCAL ensures precise energy measurements across a broad range of pseudorapidities. The information provided by the HCAL is critical for reconstructing hadronic jets, determining MET, and identifying complex event topologies. The HCAL enables accurate discrimination between signal and background by offering detailed information on jet energies, shapes, and distributions.

3.2.4 Superconducting magnet

The CMS superconducting magnet is designed to generate a powerful and uniform magnetic field that is crucial for the accurate measurement of particle momentum. This magnet is a solenoid with a large-bore superconducting coil that surrounds the inner detector systems. The solenoid's design is specifically tailored to achieve a magnetic field strength of 4 T within its 6 m diameter and 12.5-meter-long bore, which is essential for bending the paths of charged particles produced in high-energy collisions, thereby allowing precise momentum measurements.

The superconducting coil of the CMS magnet is constructed using a reinforced NbTi conductor, formed into a Rutherford-type cable. This cable is co-extruded with pure

aluminum and mechanically reinforced with an aluminum alloy to withstand the immense electromagnetic forces generated during operation. The solenoid itself is composed of four layers of this winding, carefully designed to balance the need for a strong magnetic field against the physical constraints of the detector's size and weight. With a cold mass of 220 tons, the CMS solenoid is notable for its high energy-to-mass ratio, storing up to 2.6 GJ of energy. This design, while effective, results in significant mechanical stresses during energization, necessitating robust structural support to maintain the integrity of the magnet.

The magnet's yoke, which provides the magnetic flux return, is a critical component of the system. It consists of 11 large elements, including six endcap disks and five barrel wheels. The central wheel is particularly important as it houses the solenoid coil and its cryostat. These yoke elements are not only crucial for returning the magnetic flux but also serve as structural components that support the assembly of various sub-detectors within the CMS. The assembly process is meticulously designed, with heavy-duty air and grease pads facilitating the smooth movement and precise alignment of these massive components, ensuring the entire system operates effectively.

The CMS magnet's electrical and protection systems are sophisticated, reflecting the complexity of its operation. The coil is coupled with an external mandrel, creating a "quench-back" effect that mitigates the risk of local overheating during a quench, where the superconducting state is lost, and the coil becomes resistive. This system is powered by a bipolar thyristor converter (a power electronic device used to control and switch high currents efficiently in both directions), which regulates the current through the magnet. In the event of a quench or power shutdown, a protection circuit, including a dump resistor, is activated to safely dissipate the stored energy, protecting the magnet from damage.

Maintaining the superconducting state of the CMS magnet requires precise thermal management, provided by a dedicated vacuum and cryogenic system. The vacuum system ensures effective thermal insulation within the coil cryostat, using a combination of pumps to maintain a high vacuum level. This is crucial for minimizing heat transfer and maintaining the low temperatures required for superconductivity. The cryogenic plant is capable of cooling the temperature to 4.45 K, which is vital for sustaining the superconducting

conditions in the NbTi coils.

3.2.5 The muon system

The CMS muon system is composed of three distinct types of gaseous detectors: Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). These detectors are strategically deployed across the detector to optimize coverage and performance across different pseudorapidity ranges, each designed to operate efficiently in the specific environmental conditions they encounter within the CMS experiment. The CMS muon system with pseudorapidities is illustrated in Figure 3.8.

Drift Tubes (DTs) are predominantly located in the barrel region of the detector, covering the pseudorapidity range of $|\eta| < 1.2$. This central region of the detector benefits from a relatively uniform magnetic field and lower radiation levels, making DTs particularly effective for providing high spatial resolution. The DTs are constructed from an array of rectangular drift cells, each containing a central anode wire. When a muon passes through the chamber, it ionizes the gas mixture inside the cells. The ionization electrons drift towards the anode wire, and the time taken for these electrons to reach the wire is used to calculate the muon's trajectory with high precision. The gas mixture used in the DTs typically consists of a blend of argon, carbon dioxide, and a small amount of isobutane. This mixture is chosen for its stable ionization properties, allowing for precise timing measurements and robust operation over extended periods.

In the endcap regions, which cover the pseudorapidity range of $0.9 < |\eta| < 2.4$, the CMS detector faces more intense and variable magnetic fields, along with higher levels of radiation. To meet these challenges, Cathode Strip Chambers (CSCs) are employed in these regions. CSCs are fast and highly resistant to radiation, making them ideal for the harsh conditions present in the endcaps. They consist of multi-wire proportional chambers, where the charge induced on segmented cathode strips allows for precise two-dimensional position measurement. The chemical mixture used in CSCs typically includes a combination of argon and carbon dioxide, with the possible addition of a small percentage of CF_4 (tetrafluoromethane) to enhance the chamber's performance in terms of stability and longevity under radiation. This specific gas mixture is optimized for fast response

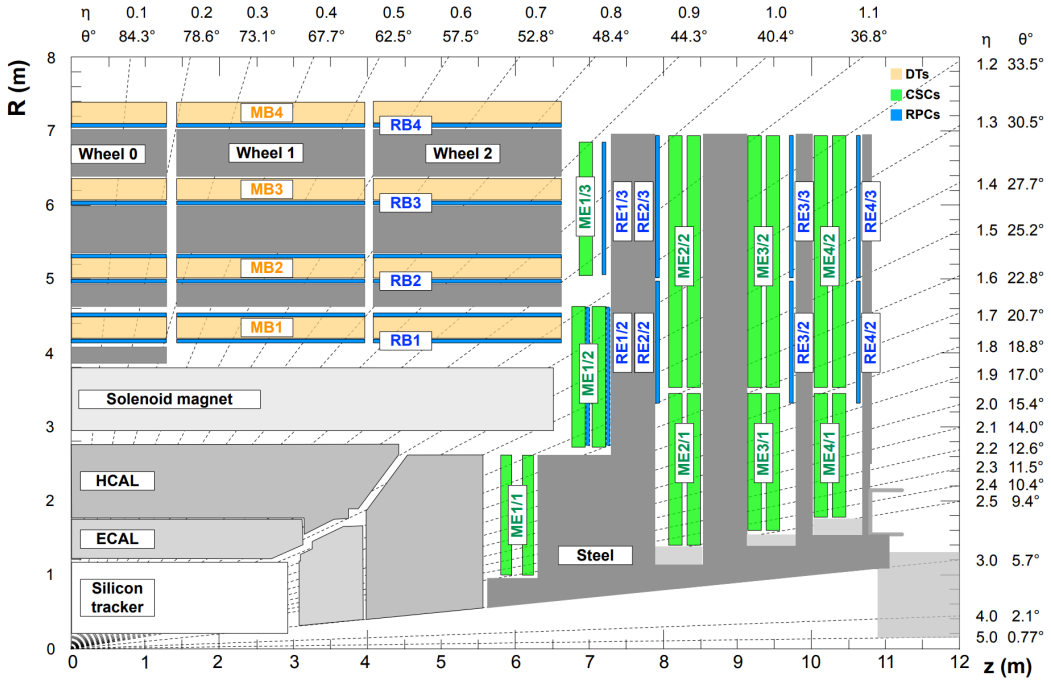


Figure 3.8: Illustration of one quadrant of the muon system in CMS. MB and RB are muon system in the barrel region and ME and RE are the muon system in the endcap region. [48]

times and minimal aging effects, ensuring reliable muon detection even in high-radiation environments.

The Resistive Plate Chambers, distributed across both the barrel and endcap regions, provide an additional layer of detection with their excellent timing resolution, crucial for the CMS's trigger system. RPCs are deployed to cover a broad pseudorapidity range, from $|\eta| < 1.6$ in the barrel to $|\eta| < 2.1$ in the endcaps. These chambers operate as fast gaseous detectors, capable of generating a trigger signal in less than a nanosecond, which is essential for distinguishing muon events from the background noise in real-time. The gas mixture used in RPCs typically comprises a combination of freon ($C_2H_2F_4$), isobutane (C_4H_{10}), and sulfur hexafluoride (SF_6). Freon acts as the main ionizing gas, isobutane as a quenching agent to prevent sparking, and SF6 is included to further stabilize the detector's performance, enhancing its capability to operate at high efficiency under the fast timing constraints required by the LHC's high collision rates.

The strategic combination of DTs, CSCs, and RPCs across the CMS detector ensures com-

prehensive coverage and redundancy in muon detection, allowing for robust and precise measurements. In the barrel region, the integration of DTs and RPCs provides both high spatial resolution and fast triggering capabilities, while in the endcaps, the combination of CSCs and RPCs ensures that detection efficiency is maintained even under challenging conditions. The muon system's design also addresses the varying pseudorapidity ranges across the detector, optimizing performance to ensure that muons are accurately identified and measured regardless of their point of origin within the collision events.

3.2.6 Trigger system

The Trigger system of the CMS detector enables the selection of relevant collision events from the enormous number of interactions produced by the LHC. Given the high collision rate at the LHC, reaching up to 40 million bunch crossings per second, it is impossible to record data from every event. The CMS Trigger system is designed to reduce this vast number of events to a manageable rate of approximately 1,000 events per second, focusing on those that are most likely to yield interesting physics results.

The CMS Trigger operates in two main stages: the Level-1 (L1) Trigger and the High-Level Trigger (HLT). These two stages work in tandem to filter out mostly irrelevant events in real-time while ensuring that potentially significant events are not missed.

The first stage of the CMS Trigger is the Level-1 (L1) Trigger, which is implemented using custom-designed hardware. The L1 Trigger is responsible for the initial reduction in event rate, processing the raw data from the detector at the full bunch-crossing rate of 40 MHz. This system is based on fast electronics and programmable logic devices which allow for rapid decision-making within a latency of a few microseconds.

The L1 Trigger analyzes data from the calorimeters and the muon detectors to identify high-energy particles and other signatures of interest, such as high transverse momentum (p_T) jets, isolated leptons, and large MET. The calorimeter trigger, for example, uses coarse-grained energy sums to identify regions of high energy deposition, which could indicate the presence of energetic jets or electrons. The muon trigger similarly looks for patterns in the muon chambers that correspond to high-p_T muons. The information from these subsystems is processed in parallel, and a global L1 decision is made based on a

predefined set of criteria, known as trigger menus.

To achieve the required reduction in data, the L1 Trigger rejects the vast majority of events, reducing the event rate from 40 MHz to about 100 kHz. This output rate was used during Run 2. This drastic reduction is necessary to ensure that the data can be transferred to the next stage of the Trigger system, the High-Level Trigger (HLT), without overwhelming the data acquisition system.

Following the L1 Trigger, HLT takes over, performing a more refined selection using the full granularity of the detector data. The HLT is implemented entirely in software, running on a large computing farm capable of processing the data at a rate of up to 100 kHz. The HLT algorithms are much more complex than those used in the L1 Trigger, as they have access to the full event information, including detailed data from the CMS Tracker, which provides high-resolution information on the trajectories of charged particles.

The HLT reconstructs the event using advanced pattern recognition algorithms, track fitting procedures, and particle identification techniques. For instance, it uses the precise measurements from the Silicon Pixel and Strip Detectors to reconstruct the tracks of charged particles and associate them with the energy deposits in the calorimeters and signals in the muon detectors. This comprehensive event reconstruction allows the HLT to apply more sophisticated selection criteria, such as identifying b-tagged jets, τ, and photons with high precision, as well as performing complex calculations like invariant mass reconstruction of particle pairs. The final output rate of the HLT is about 1 kHz, which corresponds to the number of events that are permanently recorded for offline analysis during Run 2.

3.3 Event reconstruction and particle flow

The Particle-Flow (PF) algorithm in CMS is designed to reconstruct all final-state particles—charged hadrons, neutral hadrons, photons, electrons, and muons—by integrating information from the tracker, ECAL, HCAL, and muon system. This holistic approach significantly improves the resolution of reconstructed physics objects such as jets and MET by providing a more granular and precise event description than calorimeter-based

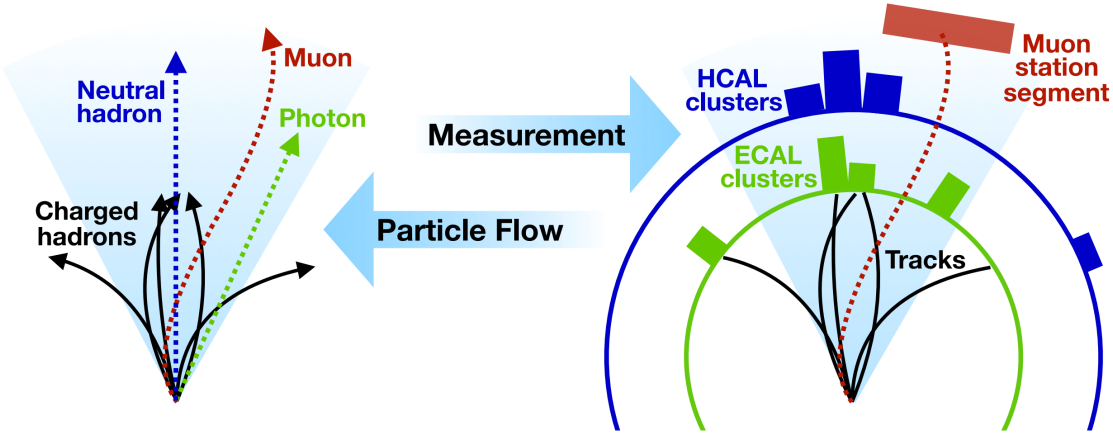


Figure 3.9: The PF algorithm illustration by [44], originally from [50]

reconstruction alone.

The reconstruction process begins with charged particles, whose trajectories are precisely measured using the silicon tracker. These charged particles leave tracks in the magnetic field of CMS, enabling their momenta to be determined with high accuracy. The PF algorithm associates these tracks with energy deposits in the calorimeters, refining their energy and directional measurements. For muons, additional information from the muon chambers is incorporated to improve identification and energy resolution.

Neutral particles, which do not leave tracks, are reconstructed entirely from calorimetric information. Photons deposit energy in the ECAL, where clustering algorithms identify and calibrate their signals. Neutral hadrons, such as neutrons, primarily interact in the HCAL, and their energy is measured through hadronic showers. Since the calorimeter response is nonlinear, corrections and calibrations are applied to neutral hadrons to improve energy resolution.

By linking charged-particle tracks to calorimeter energy deposits, the PF algorithm reduces ambiguities in energy measurement and ensures that energy deposits from different particles are correctly assigned. Once individual particles are reconstructed, higher-level objects such as jets and MET are built using algorithms like anti-k_T clustering. The inclusion of both charged and neutral particles in the jet reconstruction process significantly improves the energy resolution of jets. For MET, the PF algorithm calculates the negative vector sum of the transverse momenta of all reconstructed particles.

Muons are identified using a combination of the inner tracker and the muon system. Standalone muon tracks are first reconstructed within the muon system, using the drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC). These standalone muon tracks are then matched to tracks in the inner tracker to form "global muons." The global muon reconstruction benefits from the high momentum resolution of the tracker and the extended path length measured by the muon detectors, resulting in a more accurate momentum determination for the muons. The PF algorithm further refines the muon identification by linking the global muon track to any corresponding calorimeter deposits and ensuring that the muon energy profile is consistent with expectations for a minimally interacting particle. This reduces the chance of misidentifying hadrons as muons, particularly in cases where hadronic showers penetrate the muon system.

The key aspect of PF reconstruction is the ability to link tracks from the tracker with energy clusters in the calorimeters, creating a unified description of the particle's properties. For charged particles, the algorithm projects the track information into the calorimeter layers and searches for compatible energy clusters. This matching is based on spatial and momentum criteria, allowing the PF algorithm to link the correct calorimeter energy deposit to each track. For neutral particles, where no track is available, the PF algorithm uses the calorimeter clusters directly to reconstruct the energy and direction of the particles.

Once individual particles are reconstructed, the PF algorithm forms higher-level physics objects, such as jets and MET. Jets are reconstructed by clustering the identified particles using algorithms such as anti-k_T, which groups particles based on their proximity in (η, ϕ) space. Including both charged and neutral particles in the jet reconstruction leads to significant improvements in energy and directional resolution compared to traditional calorimeter-based methods. For MET, the PF algorithm calculates the negative vector sum of the transverse momenta of all reconstructed particles.

3.3.1 Pileup

Pileup is a phenomenon where multiple proton-proton interactions occur within the same or adjacent bunch crossings. This phenomenon arises because the intense beams used

in the LHC are composed of billions of protons, and while only a fraction participate in hard scattering processes that produce particles of interest, many others undergo soft interactions. These additional, less energetic interactions generate particles that traverse the detector simultaneously with those from the primary hard scatter, introducing significant challenges in event reconstruction and analysis. As luminosity increases, so does the average number of pileup interactions, making it one of the principal limiting factors for precision measurements and new physics searches at the LHC.

Pileup can be categorized into in-time and out-of-time pileup. In-time pileup refers to additional interactions occurring within the same bunch crossing as the primary event, while out-of-time pileup originates from interactions in previous or subsequent bunch crossings. The latter arises due to the finite response time of the detector, particularly in calorimetric measurements where energy deposits persist across multiple readout cycles. Out-of-time pileup can introduce systematic distortions in jet energy measurements and MET calculations, as signals from past interactions overlap with the current event.

The presence of pileup complicates several aspects of event reconstruction. One of the primary difficulties is the identification of the primary vertex – the collision point associated with the hard scatter that produces high transverse momentum (p_T) particles. In a typical LHC collision, the detector may register dozens of reconstructed vertices, each corresponding to a separate proton-proton interaction. Distinguishing the primary vertex from the others is crucial for correctly associating particle tracks and measuring the properties of jets, leptons, and MET. Failure to accurately identify the primary vertex can lead to incorrectly reconstructed events, reducing the sensitivity of analyses and potentially obscuring signals of new physics.

In addition to vertex confusion, pileup introduces significant contamination into calorimetric measurements. Particles from secondary interactions deposit energy in the electromagnetic and hadronic calorimeters, resulting in an overestimation of jet energies and distorting the calculation of MET. This contamination is particularly problematic in searches for processes involving invisible particles, such as neutrinos or potential dark matter candidates, where the imbalance in transverse momentum is a key observable. Overlapping energy deposits from pileup interactions can mask or mimic the signatures of such parti-

cles, leading to false signals or reduced sensitivity.

A further challenge arises from the overlap of particle tracks in the detector. In events with high pileup, the tracker registers numerous particle trajectories, many of which originate from secondary vertices rather than the primary hard scatter. This overlap complicates the reconstruction of individual particle momenta, particularly in the dense regions of jets, where tracks from multiple interactions may merge, leading to misidentified or incorrectly measured jets. The resulting degradation in track reconstruction can diminish the resolution of key observables, affecting the overall performance of the detector.

To address these challenges, the CMS experiment employs sophisticated mitigation techniques that leverage the fine spatial granularity and advanced reconstruction capabilities of the detector. A cornerstone of pileup mitigation is the PF algorithm discussed earlier which enables the identification of particles associated with pileup vertices, allowing them to be excluded or down-weighted in the reconstruction process.

One effective mitigation technique is charged hadron subtraction (CHS), which involves discarding tracks that are traced back to pileup vertices. By removing these tracks from the energy reconstruction of jets and MET, CHS significantly reduces the contamination introduced by pileup interactions. This method relies on the exceptional performance of the CMS tracker, which can distinguish tracks originating from different vertices with high precision. Another complementary technique is the pileup per particle identification (PUPPI) algorithm, which assigns a weight to each reconstructed particle based on the likelihood that it originated from the primary vertex. Unlike CHS, PUPPI does not discard pileup particles entirely but scales their contribution, preserving more information from the event while still mitigating the overall impact of pileup.

The combination of CHS and PUPPI has proven highly effective in maintaining the integrity of reconstructed jets and MET under high pileup conditions. This is particularly important for analyses sensitive to hadronic activity, such as studies of Higgs boson decays, searches for supersymmetry, and investigations of exotic particles. By ensuring that jet energies accurately reflect the particles produced in the hard scatter, these techniques enhance the precision of measurements and improve the discovery potential for new physics.

Beyond calorimetric measurements and track reconstruction, pileup mitigation also extends to vertex reconstruction. The CMS vertex reconstruction algorithm selects the primary vertex based on the number of associated high-p_T tracks, assuming that the hard scatter typically produces more high-energy particles than pileup interactions. This vertex is then used as the reference point for associating other particles, further refining the event reconstruction process. In addition, multivariate techniques employing machine learning have been developed to improve vertex identification by analyzing patterns of track distributions and energy deposits, offering a more robust solution to the vertex reconstruction challenge.

Pileup mitigation is essential not only for improving the resolution of individual events but also for ensuring the statistical accuracy of large datasets. As the LHC continues to operate at higher luminosities, with future runs expected to feature average pileup levels exceeding 50 interactions per bunch crossing, the importance of pileup mitigation will only increase. During Run 2 of the LHC, the average number of pileup interactions reached 27-37 per bunch crossing as seen in Figure 3.10.

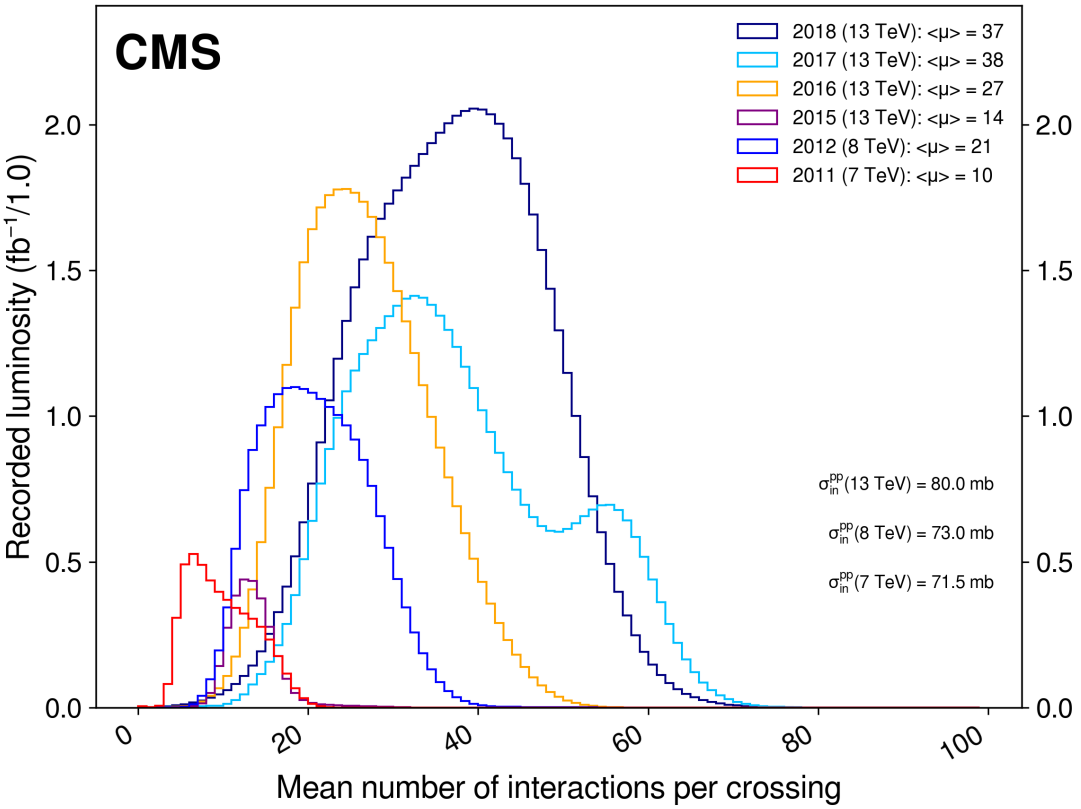


Figure 3.10: Distribution of pileup for different data-taking periods. The figure includes data from both Run 1 and Run 2, where Run 1 corresponds to collisions at $\sqrt{s} = 7$ and 8 TeV (2010-2012), and Run 2 corresponds to collisions at $\sqrt{s} = 13$ TeV (2015-2018). The overall mean values and minimum bias cross sections are also shown. [51]

4 Data analysis

4.1 Analysis production

The datasets used in this analysis were recorded by the CMS experiment during proton-proton collisions at a center-of-mass energy of 13 TeV. These datasets, corresponding to an integrated luminosity of 146.48 fb^{-1}, were recorded during years 2016-2018 collectively known as Run 2. In 2015, CMS also recorded 4.3 fb^{-1} of data. However, the data is not included in this analysis due to limited statistical impact and optimization difficulties. The combination of L1-trigger and HLT in CMS was used to record the most relevant events for analysis. [51]

To interpret the recorded collision data, detailed simulations are generated, replicating both the physics of the interactions and the response of the CMS detector. Signal modeling is performed for charged Higgs bosons decaying into a τ and a neutrino ($H^{\pm} \to \tau_h \nu_\tau$). These simulations span a wide range of hypothesized $H^{\pm}$ masses, from 80 GeV to 3 TeV, emphasizing τ decays into fully hadronic final states to enhance the visibility of the signal against the background and leptonic final states. These are generated using MadGraph [52, 53].

The analysis also requires understanding of the background processes. Simulated datasets are generated for dominant background events, including top-quark pair production ($t\bar{t}$), W+jets processes. These are modeled using Monte Carlo event generators (software tools that simulate particle collisions using random sampling techniques based on theoretical models) such as MadGraph [52,53], POWHEG [54–56], and Pythia [57], with cross-section inputs derived from next-to-leading-order (NLO) calculations. The simulation of detector effects, including resolution and efficiency, is incorporated using tools like Geant4 for full detector simulation [58, 59].

Once data are recorded by the CMS experiment, they are stored in a RAW format, which consists of the direct digital output from various subdetectors. This format encapsulates the raw electronic signals and detector responses, requiring extensive processing to extract

meaningful physical observables. To complement the recorded data, simulated events are generated to model the underlying physics processes and detector effects. These simulations begin with the generation of the hard scattering process and subsequent hadronization using Monte Carlo event generators such as MadGraph, POWHEG, and Pythia. The simulated particles are then passed through a full Geant4-based detector simulation, which accurately reproduces the energy deposition and detector response, ultimately producing an output in the same RAW format as real data. This allows for a direct comparison between experimental data and simulated events, ensuring a consistent event reconstruction procedure. The next step in the analysis involves event reconstruction, which was covered in chapter 3.3.

The statistical analysis is designed to distinguish a potential $H^{\pm}$ signal from the dominant SM background. Events are categorized based on τ helicity and jet multiplicity to enhance the analysis sensitivity.

These events are analyzed using columnar analysis (illustrated in 4.1). In high-energy physics, data analysis traditionally employs an event loop paradigm, where the analysis processes each event (row) individually. Each event comprises multiple fields detailing properties of particles from a collision. The analysis code manipulates these structures sequentially to produce derived quantities or summary statistics, such as histograms. Columnar analysis operates on entire columns of data across multiple events simultaneously while also taking into account irregular data structures. [60]

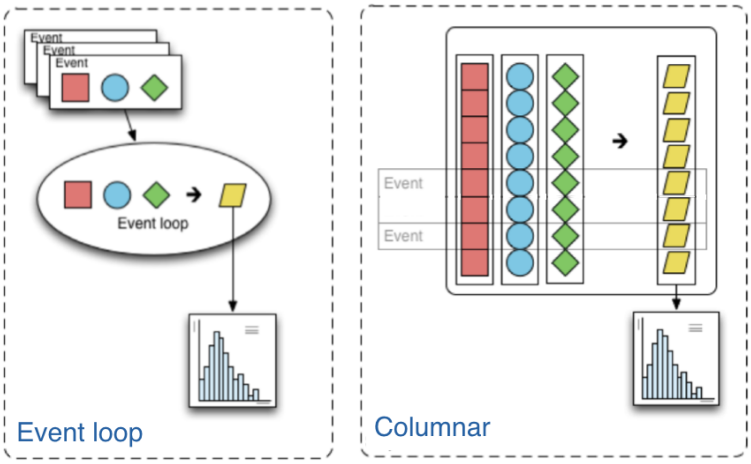


Figure 4.1: Illustration how columnar analysis (right) works compared to a normal event loop (left). [60]

4.2 Selections and background

The fully hadronic final state in searches for charged Higgs bosons ($H^{\pm}$) involves the identification and categorization of events characterized by hadronic tau decays (τ_h), MET, and additional jets. The event selection strategies, background suppression techniques, and categorization methodologies are critical for optimizing signal sensitivity and reducing background contamination. A detailed description can be found in Ref. [61].

The selection process begins with a trigger algorithm designed to capture events of interest. For the fully hadronic final state, the trigger requires at least one τ_h candidate with transverse momentum (p_T) > 50 GeV and pseudorapidity ($|\eta|$) < 2.1. It also demands MET estimated from calorimeter information, with a threshold of MET > 90 GeV. Trigger efficiencies are measured using the tag-and-probe technique with $Z/\gamma^* \to \tau^+\tau^-$ events. In order to get enough statistics, the tau pt and MET thresholds are on the rising part of the trigger turn-on.

Offline selection criteria ensure a high signal-to-background ratio while maintaining sensitivity. Candidates for τ_h must have $p_T > 50$ GeV, $p_T^{\text{track}} > 30$ GeV, and $|\eta| < 2.1$. Events with isolated electrons or muons are vetoed. Originally, at least three jets with $p_T > 30$ GeV and $|\eta| < 4.7$ are required, with at least one jet tagged as originating from a b quark. However, for this thesis, the number of jets and the number of b-tagged jets are

varied to find the most optimal selection and combining all the possible events to a single expected limit. Events must satisfy MET > 90 GeV. To suppress QCD background from misidentified τ_h, selections are applied based on angular variables. One of them is $R_{\text{min}}^{\text{bb}}$ which is defined as:

$$R_{\text{min}}^{\text{bb}} = \min_{n} \sqrt{(180^\circ - \Delta\phi(\tau_h, \vec{p}_T^{\text{miss}}))^2 + (\Delta\phi(\text{jet}_n, \vec{p}_T^{\text{miss}}))^2}, \tag{4.1}$$

where $\Delta\phi$ is the azimuthal angular difference and $\vec{p}_T^{\text{miss}}$ is MET. This variable is designed to distinguish different event topologies, particularly separating cases where τ_h and MET are back-to-back (τ, MET in opposite directions) from those where they are collinear. Events are required to satisfy $R_{\text{min}}^{\text{bb}} > 40^\circ$ to reduce contamination from QCD multijet background. Initially, a stricter selection was applied to the collinear region (upper-left corner of Figure 4.2), but this requirement was later removed to retain more signal events while still maintaining sufficient background suppression. The final selection ensures that mismeasured hadronic activity does not lead to artificially high MET values.

Events are categorized to enhance sensitivity to different $H^\pm$ mass hypotheses and polarization effects. Events are split based on the polarization variable $R_\tau = p_T^{\text{track}}/p_T^{\tau_h}$. High R_τ values indicate tau leptons from $H^\pm$ decays, while low values are dominated by background events, primarily from $W^\pm$ decays. This difference arises from the spin nature of the originating bosons; the charged Higgs boson is a spin-0 scalar, which results in right-handed tau leptons, whereas the $W^\pm$ boson is a spin-1 vector particle, leading to left-handed tau leptons. The polarization of the tau affects the energy distribution among its decay products, causing tau leptons from $H^\pm$ to exhibit higher R_τ values, reflecting a larger fraction of energy carried by the leading charged hadron compared to the more evenly distributed energy in $W^\pm$ decays. The threshold chosen for this categorization is $R_\tau = 0.75$ (In chapter 5, events with $R_\tau > 0.75$ are referred to as $R_{\tau,more}$ and events with $R_\tau \leq 0.75$ are referred to as $R_{\tau,less}$). Events are further classified by the number of jets and b-tagged jets, enhancing discrimination between signal and background, particularly for higher $H^\pm$ masses.

The primary backgrounds in the fully hadronic final state include QCD multijet events, top quark backgrounds ($t\bar{t}$ and single top), and electroweak processes. QCD multijet

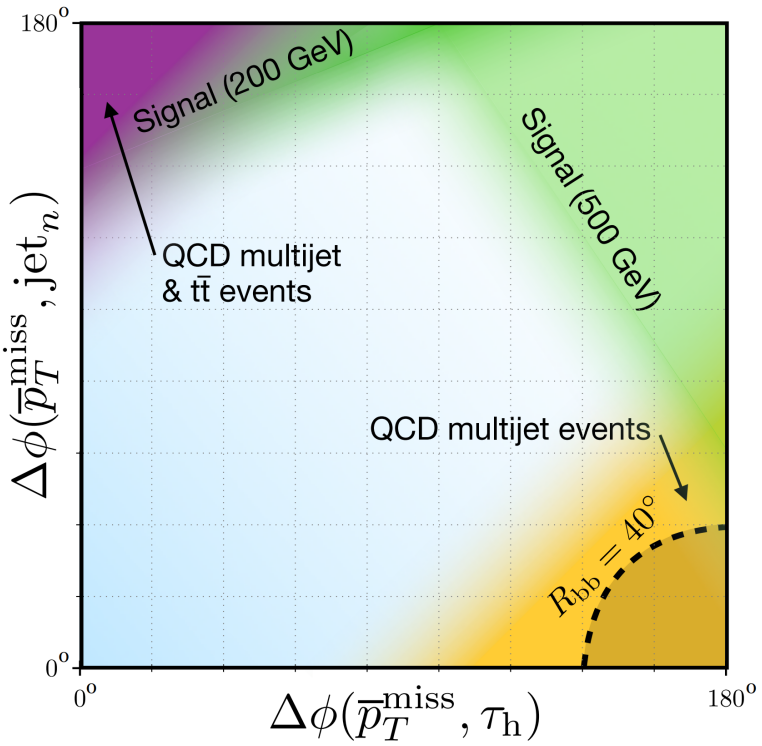


Figure 4.2: Distribution of R^{bb}_{min} (R_{bb} in the picture) for selected events. The different colors represent contributions from various background processes: red corresponds to $t\bar{t}$ events, blue represents QCD multijet background (so purple is the combination of these two), green indicates W+jets events, and yellow corresponds to EWK backgrounds. Events with $R^{bb}_{min} > 40°$ are selected to suppress QCD background, but as shown in the figure, $t\bar{t}$ events tend to populate the upper left corner. This indicates that the selection primarily targets other background processes rather than completely rejecting $t\bar{t}$ events.

events are suppressed using angular selections, $R_{\text{min}}^{\text{bb}}$, isolation criteria, and transverse mass constraints to reduce contributions from misidentified hadronic jets and fake MET sources. Top quark backgrounds, which inherently contain b-jets, are reduced by applying vetoes on events with multiple b-tagged jets and imposing high MET requirements to suppress possible semileptonic decays. Contributions from W + jets, Z/γ^*, and diboson events are minimized using isolation criteria and transverse mass selections to distinguish real missing energy from misreconstructed hadronic activity. The jet-to-τ_h misidentification rate is estimated from data using control regions where τ_h isolation is inverted. The inverted region is normalized to the expected level of events in the signal region.

To evaluate the effectiveness of the selection strategy, transverse mass (m_T) distributions are constructed from the visible $H^\pm$ decay products, τ and MET:

$$m_T = \sqrt{2p_T^{\tau_h} p_T^{\text{miss}} (1 - \cos\Delta\phi(\tau_h, \vec{p}_T^{\text{miss}}))}.$$

Unlike the invariant mass (m), which requires full knowledge of all decay products' momenta, m_T is specifically useful when undetected neutrinos are present in the final state. Since neutrinos are not directly observed in the CMS detector, their longitudinal momentum (p_z) is unknown, making it impossible to reconstruct the full invariant mass of the system. However, transverse momenta are well-measured, and MET serves as an estimator for the neutrino's transverse momentum.

Moreover, the m_T distribution has a well-defined kinematic endpoint at the mass of the decaying particle for processes such as $W \to \ell\nu$ or $H^\pm \to \tau\nu$. This feature makes m_T particularly useful for distinguishing signal from background, as QCD and fake MET backgrounds typically exhibit different distributions. Thus, applying a selection based on m_T helps suppress backgrounds while preserving signal sensitivity.

These m_T distributions are used in a binned maximum likelihood fit to extract potential $H^\pm$ signals. Signal sensitivity is maximized by combining categories with varying signal-to-background ratios.

The selection strategy for the fully hadronic final state emphasizes efficiency in capturing $H^\pm \to \tau^\pm\nu_\tau$ events while suppressing backgrounds. Categorization and angular selections

Table 4.1: Summary of selection criteria for the fully hadronic final state based on [61]

Selection Criteria	**Requirements**
Trigger τ_h candidate	$p_T > 50$ GeV, $\|\eta\| < 2.1$
Trigger MET	MET > 90 GeV
Offline τ_h selection	$p_T > 50$ GeV, $p_T^{\text{track}} > 30$ GeV, $\|\eta\| < 2.1$
Electron/muon veto	Events with isolated electrons or muons are vetoed
Jet selection	default: ≥ 3 with $p_T > 30$ GeV, $\|\eta\| < 4.7$; the amount of jets is varied
b-jet tagging	default: > 0; the amount is varied
p_T^{miss}	$p_T^{\text{miss}} > 90$ GeV
Angular variable	$R_{\text{min}}^{\text{bb}} > 40°$
Helicity-based categorization	$R_\tau = p_T^{\text{track}}/p_T^{\tau_h}$; threshold $R_\tau = 0.75$

play pivotal roles in enhancing sensitivity across a wide range of $H^\pm$ masses, enabling comprehensive exploration of the parameter space in models predicting charged Higgs bosons. All of the selections are listed in Table 4.1.

The background estimation in searches for charged Higgs bosons in the fully hadronic final state is a critical component of the analysis. In this context, the dominant background processes include QCD multijet production, top-quark pair (tt) production, single top-quark production, and electroweak processes such as W+jets, Z/γ^*, and diboson production. These backgrounds significantly impact the sensitivity of the analysis and require robust methodologies for their estimation and mitigation.

The QCD multijet background predominantly involves jets that mimic the signal signature, arising from the misidentification of hadronic jets as τ_h. The ttbar ($t\bar{t}$) and single top-quark processes contribute to backgrounds through their decay chains that can mimic the signal. For example, $t\bar{t}$ events frequently include jets and τ_h-like signatures. Electroweak processes such as W+jets and Z/γ^* are significant contributors due to their production of leptons and jets, while diboson events (WW, WZ, ZZ) add minor but non-negligible backgrounds. A crucial challenge in background estimation is differentiating between genuine τ_h and jets misidentified as τ_h, especially in QCD multijet events.

The estimation of backgrounds from jets misidentified as τ_h (denoted as jet$\rightarrow \tau_h$) is conducted using control samples enriched with such misidentifications. Events are selected by inverting the τ_h isolation requirement to enhance the presence of jet$\rightarrow \tau_h$ misidentifications. Genuine τ_h contributions are subtracted using simulation. The ratio of yields in signal-like and control regions, known as fake factors, is determined. Fake factors are measured in bins of p_T and η of the τ_h to account for dependencies on kinematic variables and detector response. The jet$\rightarrow \tau_h$ background consists of QCD multijet events and electroweak/top processes. Their contributions are separated by modeling p_T and MET distributions in control regions.

The electroweak and top backgrounds with genuine taus are estimated from simulation. The simulated event samples are normalized to their respective theoretical cross sections. Corrections for τ_h, electron, and muon efficiencies ensure the simulation reflects the detector performance. Variations in jet energy scale (JES), τ_h energy scale, and MET resolution are considered to account for modeling discrepancies.

The QCD multijet background predominantly arises from jets misidentified as τ_h candidates, leading to contamination in the signal region. To suppress these events, a combination of tau-ID criteria, MET selections, and angular variables is applied. Tau-ID plays a role in reducing fake tau contributions by distinguishing genuine τ_h decays from quark- and gluon-initiated jets using isolation, decay mode reconstruction, and advanced machine-learning-based discriminants. Additionally, since QCD multijet events rarely contain genuine MET, requiring a minimum MET threshold helps further suppress these backgrounds by reducing events where MET arises from jet mismeasurements rather than real neutrinos. Furthermore, angular selections such as R_{bb}^{min} and m_T constraints are employed to reject events where fake MET is produced by jet misreconstruction, ensuring that background contamination is minimized while maintaining signal efficiency. These combined strategies effectively enhance the purity of the selected event sample.

To validate the background estimation methods, predictions from the estimation techniques are compared against observed yields in signal-depleted validation regions. These validation regions are designed to be kinematically similar to the signal region but dominated by background processes, allowing for an assessment of the estimation accuracy.

Discrepancies between the predicted and observed event yields can indicate potential biases in the background modeling and help refine the applied corrections.

Additionally, closure tests were previously performed by [44] to verify the reliability of the background estimation methods. These tests ensured that the techniques accurately reproduced expected background yields, and their results are trusted for the analysis presented in this thesis.

Maximum likelihood fits to MET distributions are performed to determine the normalization of QCD multijet and electroweak/top components. The impact of systematic uncertainties on background yields and shapes is evaluated as different uncertainties can affect either the overall normalization or the kinematic distributions of events.

4.3 Statistics

Statistical analysis is essential to particle physics, providing the tools to test hypotheses, interpret data, and determine the presence or absence of signals. When searching for new particles such as the charged Higgs bosons, experimental results are characterized by low signal rates on top of significant background noise. This makes statistical methods essential for extracting meaningful conclusions. The frequentist framework is widely adopted for such studies due to its objectivity and its emphasis on compatibility between observed data and theoretical predictions. A detailed description can be found in Ref. [62].

The search typically involves two hypotheses. The null hypothesis, denoted as H_b, assumes that the observed data arise purely from background processes. The alternative hypothesis, H_{s+b}, incorporates contributions from both signal and background processes. The goal of the analysis is either to reject H_b and establish evidence for the signal or to constrain the strength of the signal within a predefined confidence level.

To achieve these objectives, likelihood-based statistical tests are employed. The likelihood function quantifies the probability of observing the data given specific parameters of the model, including those describing the signal and background. These likelihoods are the foundation of test statistics used to assess the consistency of the data with each hypothesis.

In the context of particle physics, the likelihood function, $L(\mu, \theta)$, is central to statistical

inference. Here, μ is the signal strength parameter, which scales the signal contribution relative to a nominal prediction, and θ represents nuisance parameters that account for uncertainties in background and signal modeling. The ratio of likelihoods forms the basis of the test statistic, which evaluates the relative support for competing hypotheses.

The profile likelihood ratio, $\lambda(\mu)$, is defined as the ratio of the likelihood maximized over nuisance parameters for a given signal strength to the globally maximized likelihood. Mathematically, this is expressed as:

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}_\mu)}{L(\hat{\mu}, \hat{\theta})},$$

where $\hat{\theta}_\mu$ represents the maximum likelihood estimators of the nuisance parameters for a fixed μ, and $\hat{\mu}$ and $\hat{\theta}$ are the maximum likelihood estimates over all parameters.

From the profile likelihood ratio, a test statistic q_μ is constructed to test the compatibility of the data with a given signal strength:

$$q_\mu = -2 \ln \lambda(\mu).$$

This test statistic quantifies how much the data deviate from the hypothesized value of μ. For discovery analyses, where the goal is to reject the null hypothesis H_b of no signal, a different test statistic, q_0, is used:

$$q_0 = -2 \ln \lambda(0).$$

These test statistics follow specific asymptotic distributions under the null and alternative hypotheses, which are leveraged for computational efficiency.

4.3.1 $\mathrm{CL_S}$ method and limits

The CL_S method [63,64] is a modified frequentist approach specifically designed for setting expected limits. Its modification addresses an issue inherent in traditional frequentist confidence intervals: overly stringent limits can result from statistical fluctuations in the data, particularly when background contributions are low.

In the CL_S framework, two confidence levels are defined. The first, CL_{s+b}, represents the probability of observing data at least as incompatible with the signal+background hypothesis as the observed data, assuming that both signal and background processes contribute to the observations. It is given by:

$$CL_{s+b} = P(q_\mu \geq q_\mu^{\mathrm{obs}} | H_{s+b}),$$

where q_μ^{obs} is the observed value of the test statistic. The second confidence level, CL_b, quantifies the probability of observing such deviations under the background-only hypothesis:

$$CL_b = P(q_\mu \geq q_\mu^{\mathrm{obs}} | H_b).$$

The modified confidence level, CL_S, is then defined as the ratio:

$$CL_S = \frac{CL_{s+b}}{CL_b}.$$

This modification ensures that the exclusion of a signal hypothesis is not overly influenced by downward fluctuations in the background. A signal hypothesis is excluded at the $1-\alpha$ confidence level if $CL_S < \alpha$, where α is typically chosen as 0.05 for a 95% confidence level.

The CL_S method is not only used to exclude hypotheses but also to quantify the significance of an observed signal. The discovery of a new phenomenon, such as the charged Higgs boson, requires rejecting the null hypothesis H_b with high confidence. In particle physics, a discovery is conventionally defined as achieving a significance level of at least 5σ, corresponding to a p-value of approximately 2.87×10^{-7}. This stringent criterion ensures robustness against statistical fluctuations.

The significance Z is related to the p-value p by the equation:

$$Z = \Phi^{-1}(1-p),$$

where Φ^{-1} is the inverse cumulative Gaussian distribution function of the standard normal distribution. A 5σ significance means that the observed data is as noticeable as, what would be expected in the background-only hypothesis in fewer than one in a million experiments.

To facilitate computationally efficient limit setting and sensitivity estimation, Asimov data set is employed. The Asimov data set is an artificial, representative dataset constructed to match the median expectations of the experiment under a given hypothesis. Using this dataset, one can calculate expected confidence levels and significances without relying on extensive Monte Carlo simulations. [65]

Asymptotic approximations derived from results provide further simplifications. For large datasets, the test statistic q_μ approximately follows a chi-squared distribution under the null hypothesis, reducing the computational overhead associated with repeated pseudo-experiments. [65]

By combining results from multiple final states and experiments, the sensitivity of the analysis is maximized. For this thesis, only the fully hadronic final state is taken into account.

4.3.2 Uncertainties

In this chapter, we explore the detailed framework for addressing uncertainties in the analysis of charged Higgs bosons in fully hadronic final states. The study of uncertainties is a critical aspect of high-energy physics analyses, as it defines the reliability and precision of the conclusions drawn from the data. These uncertainties arise from multiple sources, broadly categorized into experimental, theoretical, and statistical origins, each of which requires careful quantification and propagation to ensure robust results. Among these, systematic uncertainties often represent the largest and most challenging component, as they stem from biases and imperfections in the experimental methodology, reconstruction techniques, and theoretical modeling.

Experimental uncertainties are among the most immediate challenges in any high-energy physics experiment. These include inaccuracies in detector calibration, reconstruction of particle energies, and identification efficiencies. In this analysis, uncertainties in the

jet energy scale (JES) and jet energy resolution (JER) play a significant role, as jets constitute the primary experimental signature of the fully hadronic decay channels. The JES uncertainties are derived by calibrating simulated jets against data across a range of kinematic conditions, accounting for factors such as pileup contributions, variations in detector response as a function of pseudorapidity η, and differences in jet transverse momentum (p_T). These uncertainties are generally small in the central detector regions, often below 1% for $p_T > 30\,\text{GeV}$, but can rise in forward regions due to limited coverage and resolution. Similarly, the JER, which quantifies the resolution of jet momentum reconstruction, introduces uncertainties that depend on the p_T and flavor of the jets, ranging from 15–20% for low-p_T jets to about 5% for jets at the TeV scale. These effects are corrected using a combination of measurements and simulated calibration techniques. [66–69]

Another key experimental uncertainty arises in the identification and reconstruction of τ, which are central to the charged Higgs decay signature in this analysis. The identification algorithm, known as the hadron-plus-strips (HPS) algorithm, is designed to distinguish true taus from jets, electrons, and muons misidentified as τ candidates. Systematic uncertainties in τ identification are driven by pileup effects. Misidentification probabilities, particularly for jets falsely reconstructed as taus, are carefully estimated using data-driven methods. The energy scale of τ also contributes an additional layer of uncertainty, as discrepancies between simulated and measured τ momentum distributions can impact the reconstructed mass of the τ-neutrino system, which is a crucial discriminating variable. [66, 68]

Pileup effects, arising from multiple proton-proton interactions within the same bunch crossing, constitute another experimental source of uncertainty. Pileup impacts the reconstruction of jets, τ candidates, and the MET, which is an essential observable in this analysis. Correction methods, such as the hybrid jet area method, are used to mitigate these effects by subtracting the contribution of additional interactions from the measured energies. However, residual uncertainties remain, especially in high-luminosity conditions, where pileup multiplicity becomes significant. [66, 70]

Theoretical uncertainties are equally important, as they influence the modeling of both

signal and background processes. These uncertainties arise from several sources, including parton distribution functions, renormalization and factorization scales, and modeling of hadronization and parton showering. Variations in PDFs, which describe the momentum distribution of quarks and gluons within protons, can affect the predicted production cross-sections and kinematics of the signal and background events. The impact of PDF uncertainties is quantified by employing multiple PDF sets and evaluating the spread in predictions. Similarly, the choice of renormalization and factorization scales introduces uncertainties in the calculations of signal and background cross-sections. These scales are varied up and down by a factor of two to estimate their contribution to the overall theoretical uncertainty. [66, 71]

Monte Carlo simulations, which are essential for estimating signal and background yields, are subject to statistical uncertainties due to their finite sample size. These uncertainties are especially relevant in analyses that rely on bin-by-bin comparisons between data and simulations. The Barlow-Beeston method is a well-established technique for incorporating these uncertainties into the likelihood fitting procedure, ensuring that both data and simulation errors are accounted for. [72]

Systematic uncertainties are among the most challenging aspects of the analysis, as they represent potential biases in the experimental and theoretical methodologies. These uncertainties can arise from misestimated backgrounds, mismodeled detector effects, and interpolation between discrete simulation points. In this analysis, background estimation relies heavily on data-driven methods to reduce dependence on simulation. For instance, the background from jets misidentified as taus is estimated directly from control regions in the data, where the τ candidates are predominantly fake. However, systematic uncertainties associated with these methods, such as the choice of control region and the extrapolation to the signal region, must be carefully evaluated and propagated. [66, 68]

Finally, the correlations between different sources of uncertainty must be addressed to avoid over- or underestimating their combined impact. For example, JES and pileup corrections are often correlated, as both depend on the detector's response to hadronic activity. These correlations are incorporated into the statistical analysis through covariance matrices or nuisance parameter marginalization, ensuring a consistent treatment of

uncertainties across all observables. [66]

5 Results

This study focuses on deriving expected limits for discovering the heavy charged Higgs boson ($H^{\pm}$) in the decay channel $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ using the CMS Run 2 dataset (2016-2018). By leveraging the increased dataset size and incorporating methodological advancements, the analysis aims to enhance sensitivity to the $H^{\pm}$ signal, particularly in the fully hadronic final state.

Compared to the previous analysis conducted on data recorded in 2016, this work introduces three main differences. First, a simplified set of systematic uncertainties was included, focusing on the most impactful sources of uncertainty while excluding others deemed negligible. The list of the uncertainties used in this analysis can be found in B.1.

Second, this analysis implements DeepTau for hadronic tau lepton identification [73, 74]. DeepTau is a deep neural network-based algorithm that significantly enhances the discrimination between genuine hadronic tau decays and background contributions from quark and gluon jets, electrons, and muons. Compared to previous multivariate-based tau identification methods, DeepTau utilizes a more comprehensive input set, incorporating lower-level information from reconstructed particles in the vicinity of the tau candidate. This results in a substantial improvement in both efficiency and background rejection, which is particularly relevant for analyses involving hadronic tau final states, such as the one presented here.

Third, the 2016 reference limits were derived using the DeepCSV b-tagging algorithm [75], whereas this analysis employs the DeepJet b-tagging algorithm [76]. DeepJet represents an improvement over DeepCSV by incorporating more detailed information from jet constituents, including charged and neutral particles, secondary vertices, and advanced deep-learning techniques [77]. This change affects the selection efficiency of b-tagged jets and, consequently, the expected event yields. Despite this, the overall expected sensitivity should become slightly better, as DeepJet provides better background rejection while maintaining signal efficiency. The impact of this switch is accounted for through scale factors derived from CMS b-tagging performance studies [77].

The statistical analysis was performed with using Combine [78] and CombineHarvester [79] tools.

5.1 Validation

To validate the analysis and ensure consistency with previous results, we first reproduce the transverse mass (m_T) distributions and expected limits using the 2016 data. Since the analysis remains blinded, only the expected limits are considered, derived from the background estimation and statistical framework.

The transverse mass (m_T) serves as the primary discriminating variable for the search, distinguishing potential $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ signal events from Standard Model background processes. The key background contributions arise from top-quark pair production ($t\bar{t}$), single-top processes, diboson production (VV), Drell-Yan (DY), and W+jets.

Figure 5.1 presents the blinded m_T distributions for the first era of 2016, while Figure 5.2 shows the distributions for the second era of 2016. These are comparing the two different selections of the R_{τ} variable. Different selections of R_{τ} modify the event composition, impacting the background-to-signal ratio and thus affecting the expected sensitivity. R_{τ} selection was discussed in 4.2.

Using the reconstructed m_T distributions as input (illustrated in Figures 5.1 and 5.2), we derive the expected limits following the CL_S method. Figures 5.3 and 5.4 show the expected limits on $\sigma \times BR(H^{\pm} \to \tau^{\pm}\nu_{\tau})$ as a function of $m_{H^{\pm}}$ for the 2016 dataset. To ensure consistency in methodology and background modeling, the new 2016 limits are compared to the reference limits.

The differences between Figures 5.3 and 5.4 can be seen in the overlap picture in Figure 5.5. The main difference between the 2016 reference and this analysis is the simplified systematic uncertainty treatment. With full systematics the new limit is expected to get slightly worse and showing similar results to the reference. The other main differences are in using the DeepJet b-tagging algorithm compared to DeepCSV and DeepTau for distinguishing genuine hadronic tau decays as mentioned before. One of the possible differences between this analysis and the 2016 reference results is the change in binning

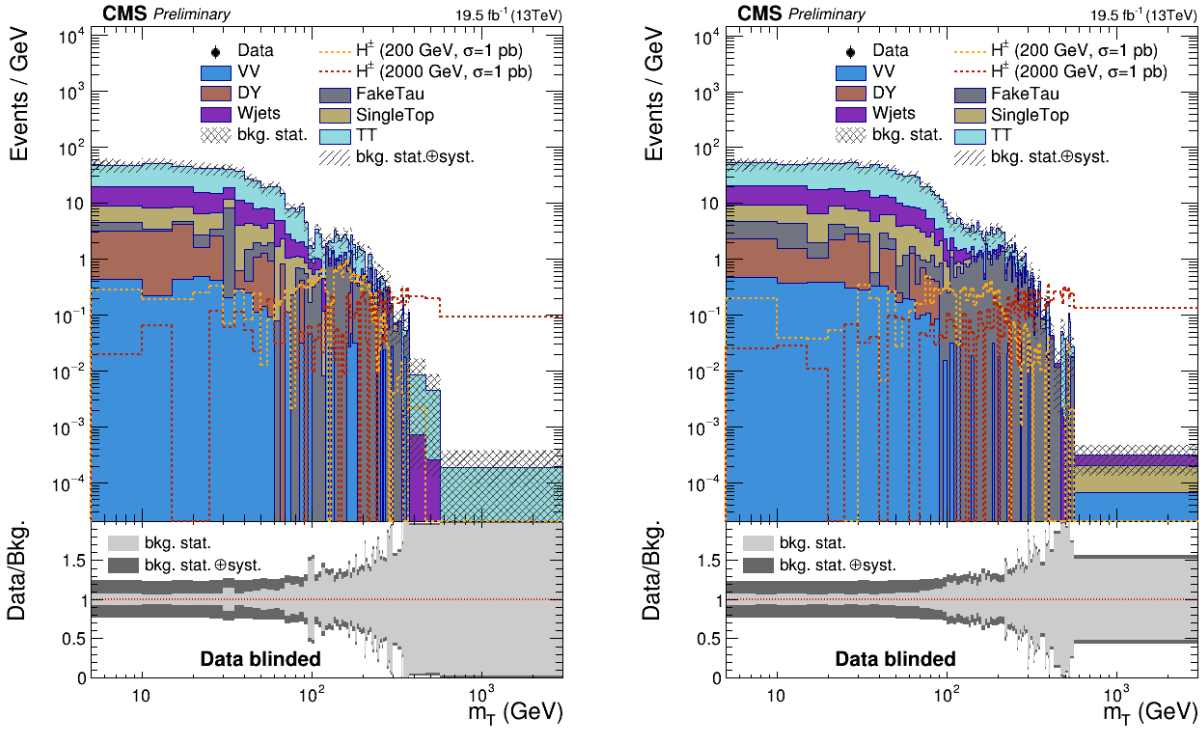


Figure 5.1: Transverse mass distributions in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ fully hadronic channel for the first era of year 2016. Year 2016 is divided into two eras because of hardware-related change in the tracker [80]. The left uses $R_{\tau} \leq 0.75$ categorization while the right uses $R_{\tau} > 0.75$.

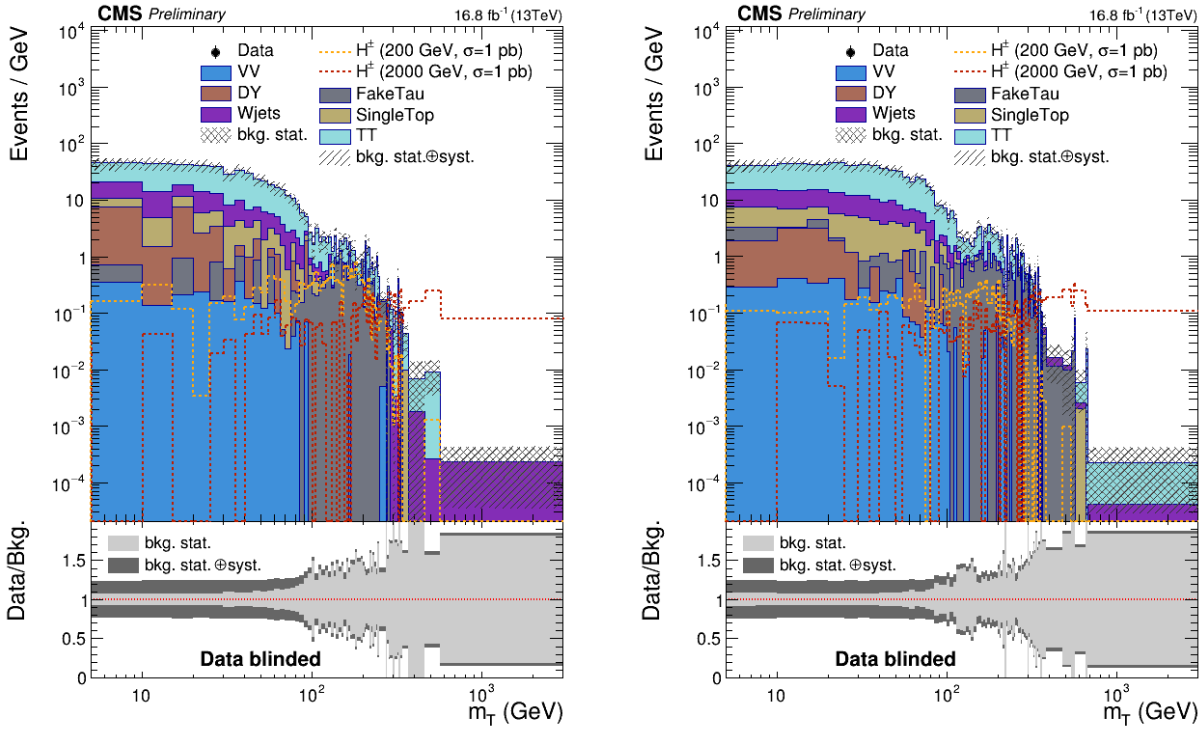


Figure 5.2: Transverse mass distributions in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ fully hadronic channel for the second era of year 2016. Year 2016 is divided into two eras because of hardware-related change in the tracker [80]. The left uses $R_{\tau} \leq 0.75$ categorization while the right uses $R_{\tau} > 0.75$.

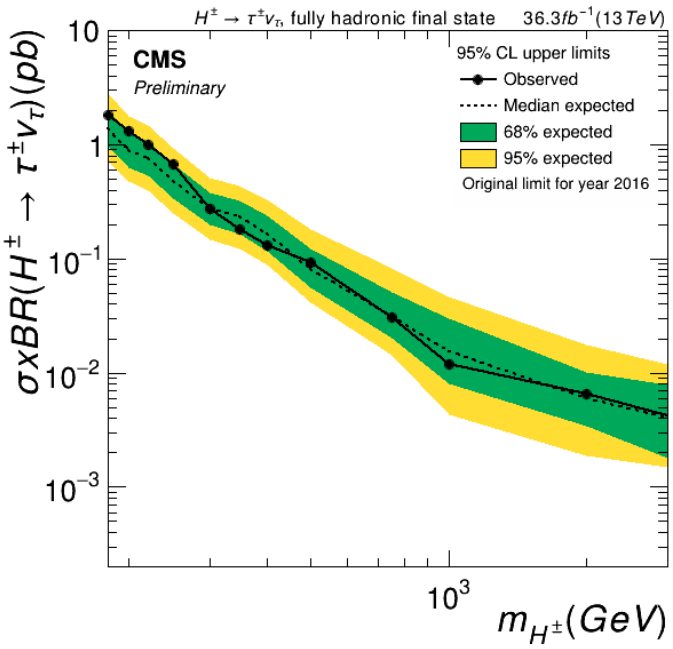


Figure 5.3: The published 2016 expected limit used as a reference [81]. The limit calculations has used the default selection of $N_{\text{jets}} \geq 3$ and $N_b > 0$.

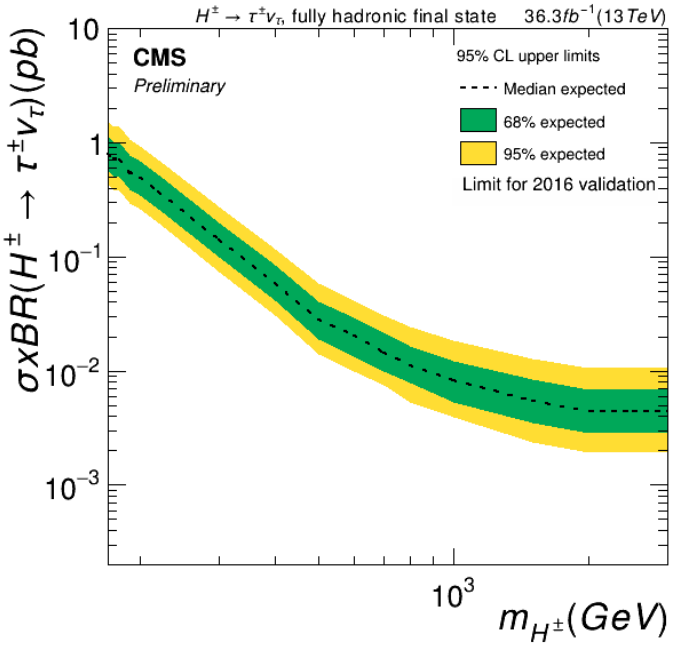


Figure 5.4: Year 2016 expected limit produced in this study. The limit calculation uses the default selection of $N_{\text{jets}} \geq 3$ and $N_b > 0$.

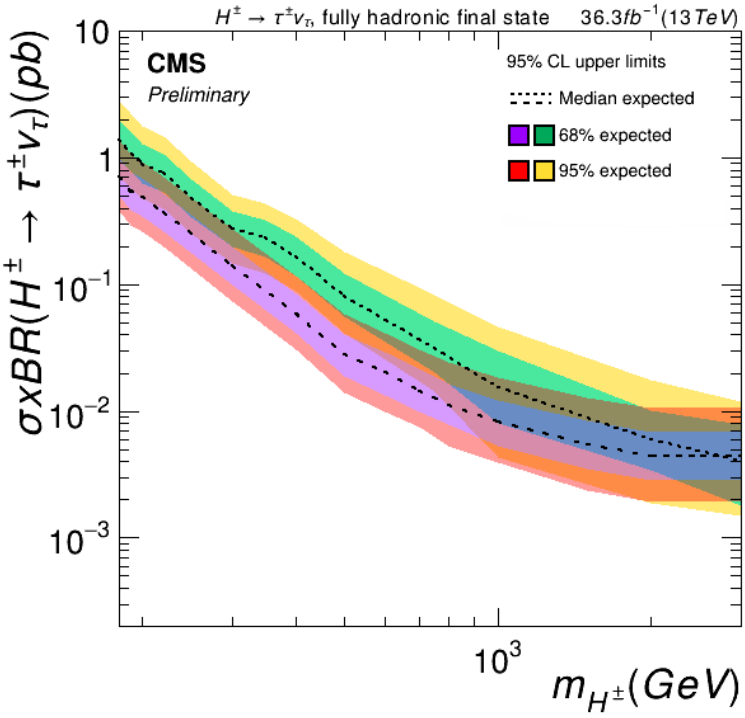


Figure 5.5: Overlap of the 2016 reference expected limit plot with the current 2016 expected limit plot. The 2016 reference expected limit plot has the original green and yellow colours while the current expected limit plot has purple and red. The 2016 reference median expected limit has short-dashed lines and the current 2016 limit has long spaces in between lines.

scheme. Due to dataset formatting constraints, it is no longer possible to maintain the same bin structure as the observed 2016 analysis. Some of the bins and datasets used in the reference 2016 limit are not available in the new 2016 limit and vice versa. So some of the neighboring bins in both analyses might contain events that are included in another. Because of this, fully comparing the two plots is actually impossible. Despite all of this, Figure 5.5 shows that the new 2016 limit is lower than the 2016 reference across all masses. As the simplified treatment of systematic uncertainties is expected to give overoptimistic results, it is also expected that the DeepTau and DeepJet algorithms are working better than the older algorithms used in the reference study.

Therefore, this validation provides confidence that the methodology can be extended to the 2017, 2018, and full Run 2 datasets while maintaining sensitivity to a charged Higgs boson signal.

5.2 Selection Optimization Study

The selection criteria for identifying $H^{\pm}$ events play a role in maximizing the sensitivity of this analysis. To determine the most optimal event selection strategy, a study was conducted using only the 2017 dataset, systematically varying the number of jets (N_{jets}) and the number of b-tagged jets (N_b). The goal of this study was to identify the selection criteria that yield the lowest expected limits, thereby improving the search sensitivity for $H^{\pm} \to \tau^{\pm}\nu_{\tau}$.

The selection categories analyzed in this study included different requirements on N_{jets}, ranging from low-multiplicity final states to higher-multiplicity events, where the presence of additional jets can arise from associated production or background processes such as $t\bar{t}$. Simultaneously, different thresholds for b-tagged jet multiplicity N_b were tested, leveraging CMS b-tagging algorithms to suppress non-top backgrounds. Since b-quark-associated jets are often soft, they may not be reconstructed as separate jets or may complicate b-tagging identification. Therefore, we systematically test different jet and b-jet multiplicities in the final state and statistically combine different categories. The summary of all the combinations of these two selections can be found in 5.1.

This revision incorporates the reasoning suggested in the comment while keeping your original structure intact. Let me know if you need further refinements!

Table 5.1: Jet categorization combinations used in the optimal selection analysis. The case of ($N_{\text{jets}} = 2$) and $N_b > 1$ is ignored because it contains only events where all jets are b-jets. To combine all the jet categories for a single expected limit plot, only the bottom 2×2 (highlighted in blue) cases are taken into account since they contain mutually exclusive events.

$N_{jets} \backslash N_b$	$== 1$	> 1	$== 0$	> 0
$== 3$	$N_{jets} = 3, N_b = 1$	$N_{jets} = 3, N_b > 1$	$N_{jets} = 3, N_b = 0$	$N_{jets} = 3, N_b > 0$
> 3	$N_{jets} > 3, N_b = 1$	$N_{jets} > 3, N_b > 1$	$N_{jets} > 3, N_b = 0$	$N_{jets} > 3, N_b > 0$
$== 2$	$N_{jets} = 2, N_b = 1$	$N_{jets} = 2, N_b > 1$	$N_{jets} = 2, N_b = 0$	$N_{jets} = 2, N_b > 0$
≥ 3	$N_{jets} \geq 3, N_b = 1$	$N_{jets} \geq 3, N_b > 1$	$N_{jets} \geq 3, N_b = 0$	$N_{jets} \geq 3, N_b > 0$

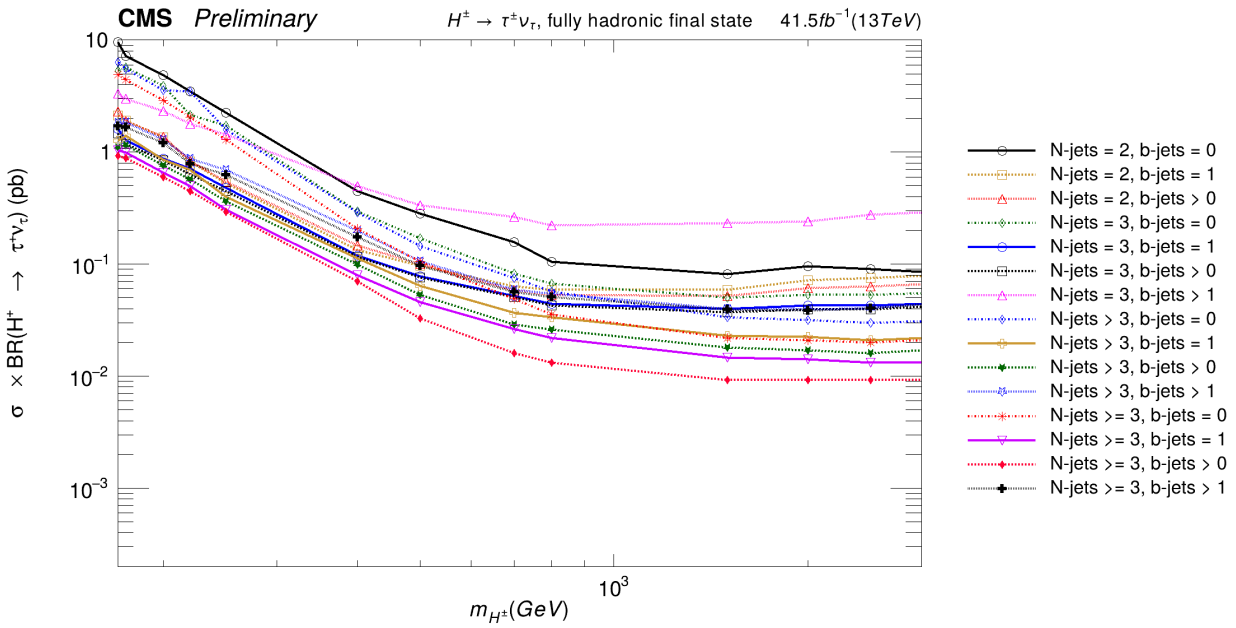


Figure 5.6: Plot of the expected median points for all jet categorizations separately.

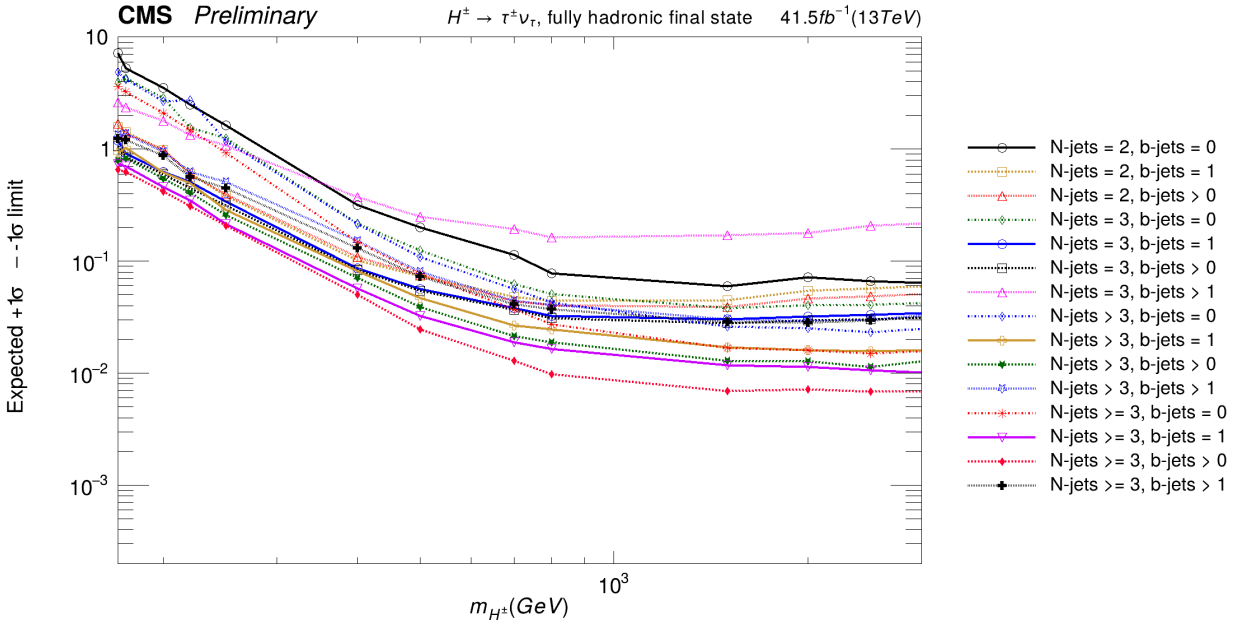


Figure 5.7: Plot of the expected 1σ band width for all jet categorizations separately.

Figure 5.6 presents the expected median limits for the different combinations of N_{jets} and N_b jet categorizations. Figure 5.7 presents the 1σ band width of all different combinations of N_{jets} and N_b jet categorizations. The study shows that the optimal selection, defined as the configuration yielding the lowest expected limit across all $H^{\pm}$ masses, corresponds to $N_{\text{jets}} \geq 3$ and $N_b > 0$. One crucial thing to note is that this jet categorization was also used as a default in this analysis and the 2016 reference. This means that the sensitivity to discover $H^{\pm}$ is highest with default selection for all of Run 2 as previously calculated by [44].

In conclusion, the selection optimization study for 2017 provides a well-motivated choice of N_{jets} and N_b for other years and full Run 2.

5.3 Expected Limits for 2017, 2018, and Full Run 2

With the validation of the analysis framework using the 2016 dataset, the expected limits are now derived for the 2017, 2018, and full Run 2 datasets. These limits are obtained following the same CL_S method, utilizing m_T distributions (Figures 5.8 and 5.9) as inputs, as with 2016 data.

The expected limits for the 2017 dataset are shown in Figure 5.10. The integrated luminosity collected in 2017 is approximately 41.5 fb^{-1}. Figure 5.11 presents the expected limits for 2018, which benefits from the largest individual dataset in this study, with an integrated luminosity of 59.8 fb^{-1}.

To achieve the highest possible sensitivity for Run 2, the datasets from 2016, 2017, and 2018 are combined to produce the full Run 2 expected limits, as shown in Figure 5.12. The total integrated luminosity of 138.6 fb^{-1} allows for a significant enhancement in sensitivity compared to other Run 2 individual years.

To compare all of the Run 2 plots with each other, all of the expected medians and 1σ band width for multiple masses across all and combined Run 2 years is shown in Figures 5.13 and 5.14 respectively.

One of the interesting results from Figure 5.13 and 5.14 is that the expected limits for year 2017 seem to be worse than the 2016 results. A similar trend is observed in some

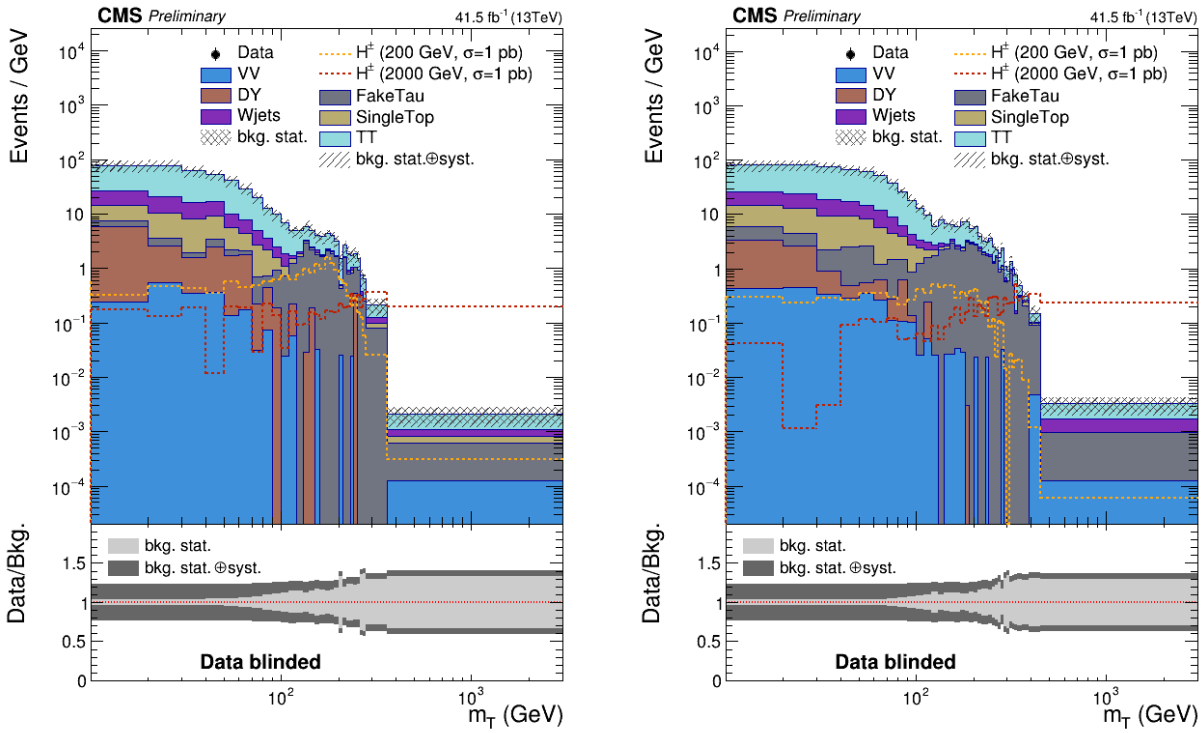


Figure 5.8: Transverse mass distributions in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ fully hadronic channel for year 2017. The left uses $R_{\tau} \leq 0.75$ categorization while the right uses $R_{\tau} > 0.75$.

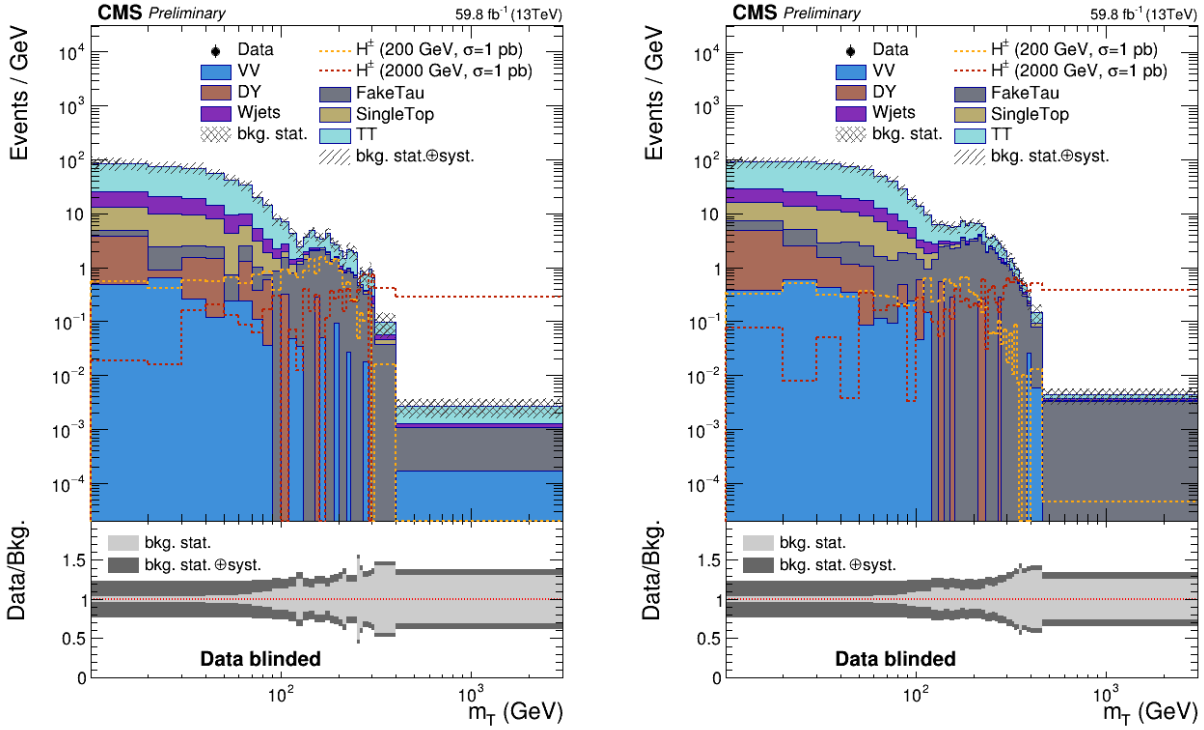


Figure 5.9: Transverse mass distributions in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ fully hadronic channel for year 2018. The left uses $R_{\tau} \leq 0.75$ categorization while the right uses $R_{\tau} > 0.75$.

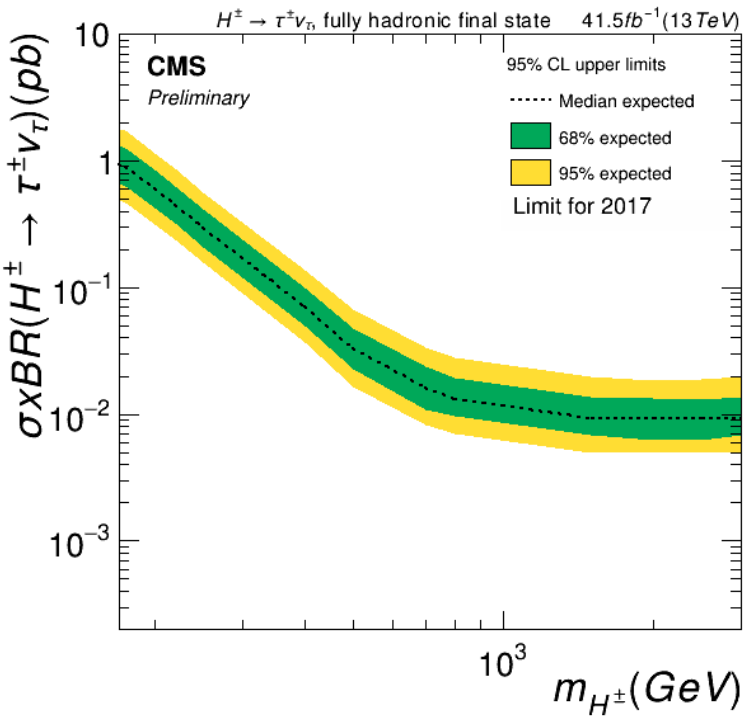


Figure 5.10: Year 2017 expected limit. The limit calculation uses the default selection of $N_{\text{jets}} \geq 3$ and $N_b > 0$.

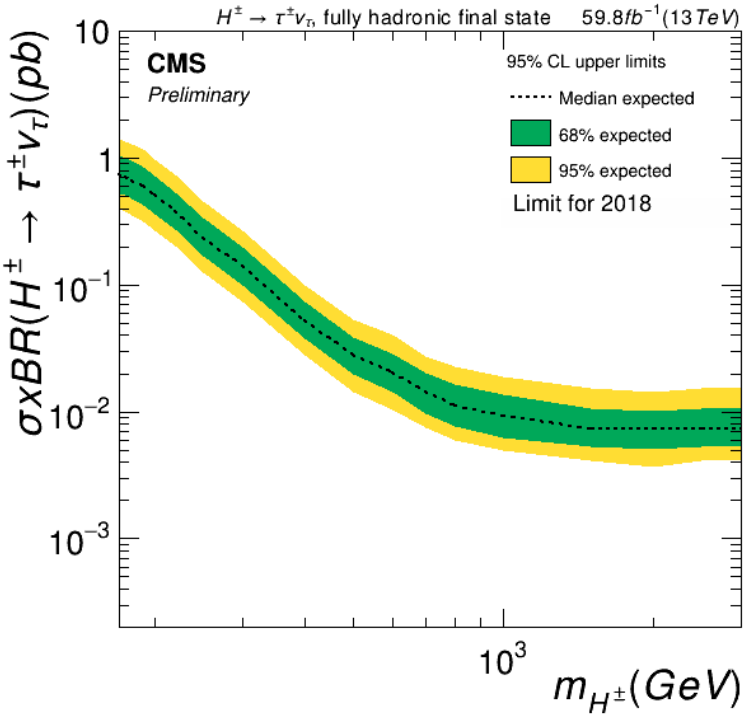


Figure 5.11: Year 2018 expected limit. The limit calculation uses the default selection of $N_{\text{jets}} \geq 3$ and $N_b > 0$.

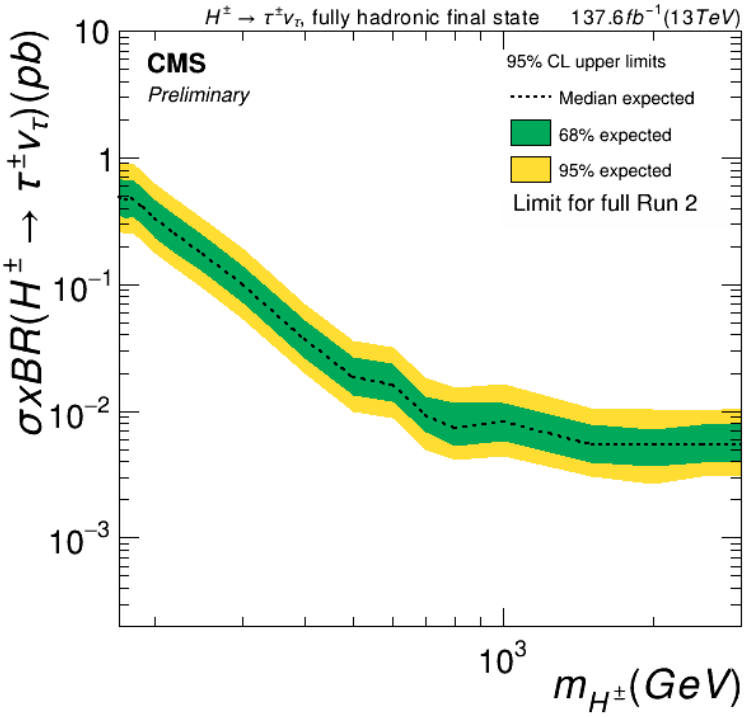


Figure 5.12: Expected limit for combining all Run 2 years. The limit calculation uses the default selection of $N_{\text{jets}} \geq 3$ and $N_b > 0$.

mass values of 2018 data. This suggests that factors beyond integrated luminosity are influencing the sensitivity. The limit of 2016 reference also at highest masses gives better values than combined Run 2 results. This could be most likely due to differences in systematic uncertainty treatment. This can also be explained by the very high 1σ band width which is still one of the largest across all mass values.

It is worth noting that the binning appears slightly inconsistent for 2017 and the 2016 reference compared to other years, introducing minor variations in the results, as discussed in Chapter 5.1. One possible explanation is the increased pileup in 2017 and 2018, which may have affected event reconstruction and background rejection efficiency (see chapter 3.3.1). Additionally, the observed differences in expected limits between 2016, 2017, and 2018 can likely be attributed to differences in trigger selections across these datasets. In particular, the trigger tags used in 2017 and 2018 differ from those in 2016, potentially affecting the efficiency of selecting signal-like events. These variations in trigger paths across years may have led to differences in event selection, resulting in systematic shifts in background composition and overall efficiency.

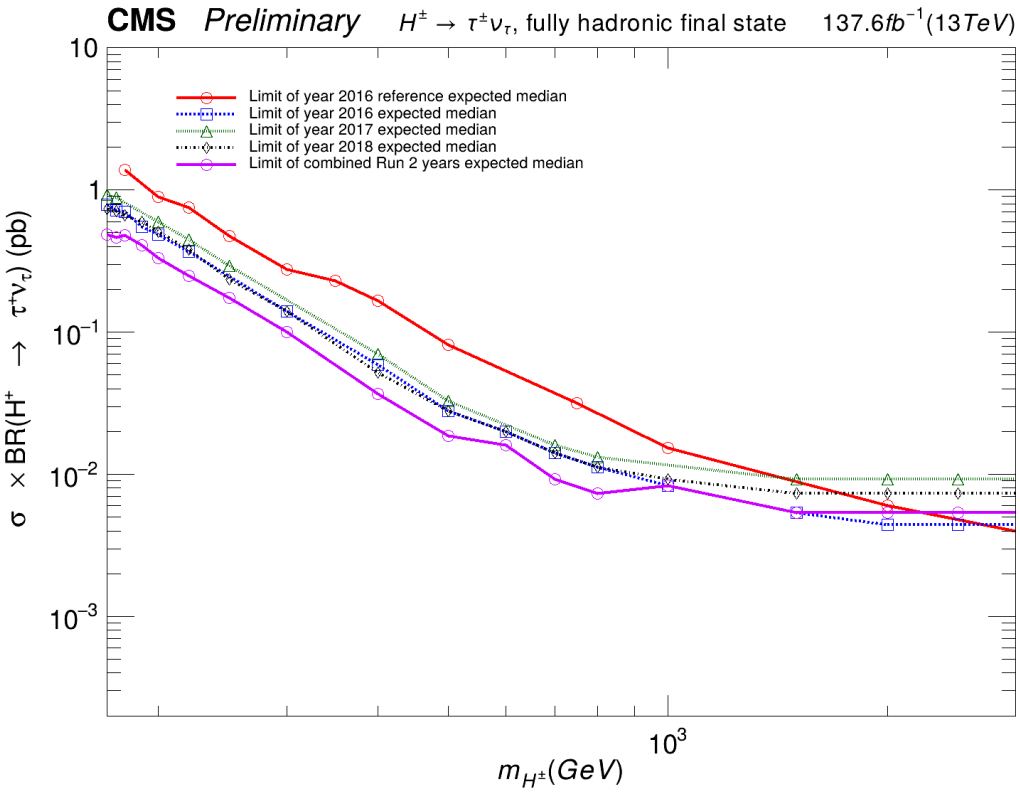


Figure 5.13: Plot of the expected median points for all Run 2 years separately and combined. The result of 2016 reference is also added for comparison.

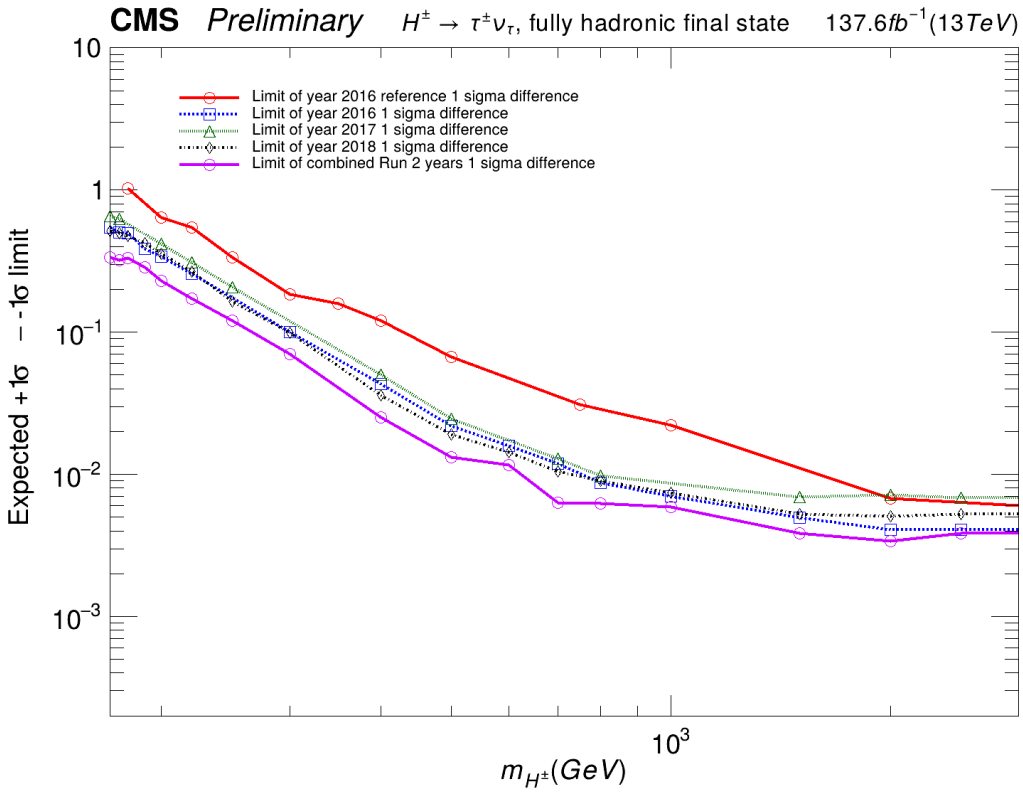


Figure 5.14: Plot of the expected 1σ band width for all Run 2 years separately and combined. The result of 2016 reference is also added for comparison.

5.4 Combined Limit of All Selections and Projected Sensitivity for Run 2 + Run 3

A combined limit calculation using the Table 5.1 combinations was also performed, incorporating all studied jet categories (highlighted in blue) instead of restricting the analysis to only the default ($N_{\text{jets}} \geq 3$ and $N_b > 0$) category. This is presented in Figure 5.15. By including multiple selection categories, the analysis benefits from improved statistical power, allowing different event topologies to contribute to the overall expected sensitivity.

The area highlighted in blue in Table 5.1 is one of the possible orthogonal combination of the jet categories. Others could be used but this orthogonal combination is used for its simplicity and it includes the default jet categorization used in the 2016 reference limits and this thesis. This default category was also calculated to give the lowest expected limit of all jet categories in Chapter 5.2.

In addition to the full Run 2 combined limit for default and all selections, a projected study was performed for Run 2 + Run 3 total integrated luminosity, 334.0 fb^{-1} (see [51] for more current values), which more than doubles the total integrated luminosity of Run 2 alone. This projection is based on the Run 2 results and projecting to the expected Run 2 + Run 3 integrated luminosity. (such as the same uncertainties as in Table B.1). These results are presented in Figures 5.16 and 5.17.

The median expected limits and 1σ band width derived from these combinations are presented in Figures 5.18 and 5.19. The median expected limits and 1σ band width are cut off at $m_{H^\pm} = 1500$ GeV and $m_{H^\pm} = 2500$ GeV for the Run 2 + Run 3 projections. This happens because of technical limitations in the Combine program. Despite this, the expected limits for higher mass values are predicted to be lower than Run 2 results. Compared to the limits obtained using a single selection criterion, the combined approach gives an improved expected limit for the $H^\pm \to \tau^\pm \nu_\tau$ channel.

Overall, the combined limit calculation for Run 2 demonstrates the advantage of including multiple selection scheme, leading to better expected limits than single selection scheme. The projected sensitivity for Run 2 + Run 3 highlights the significant potential of future

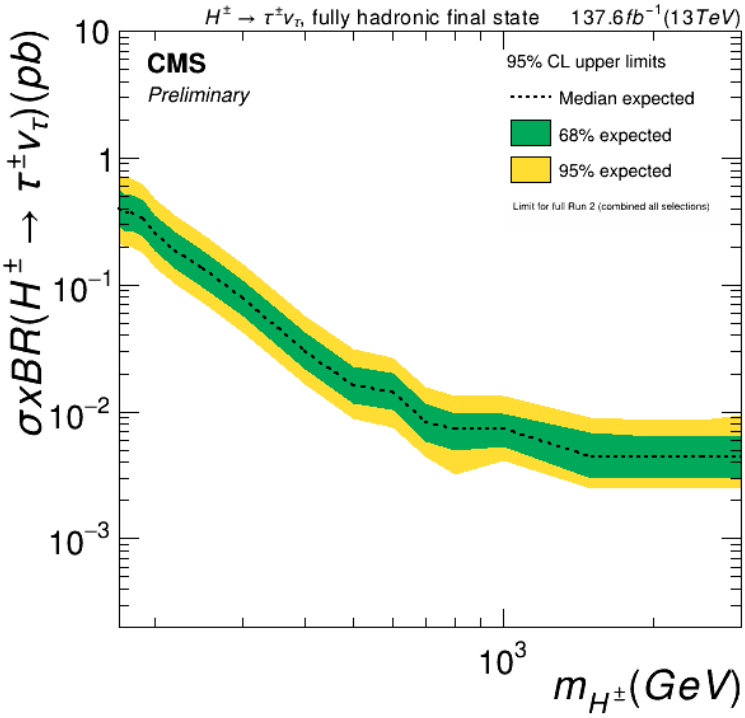


Figure 5.15: Expected limit for combining all Run 2 years and all of the jet categorizations.

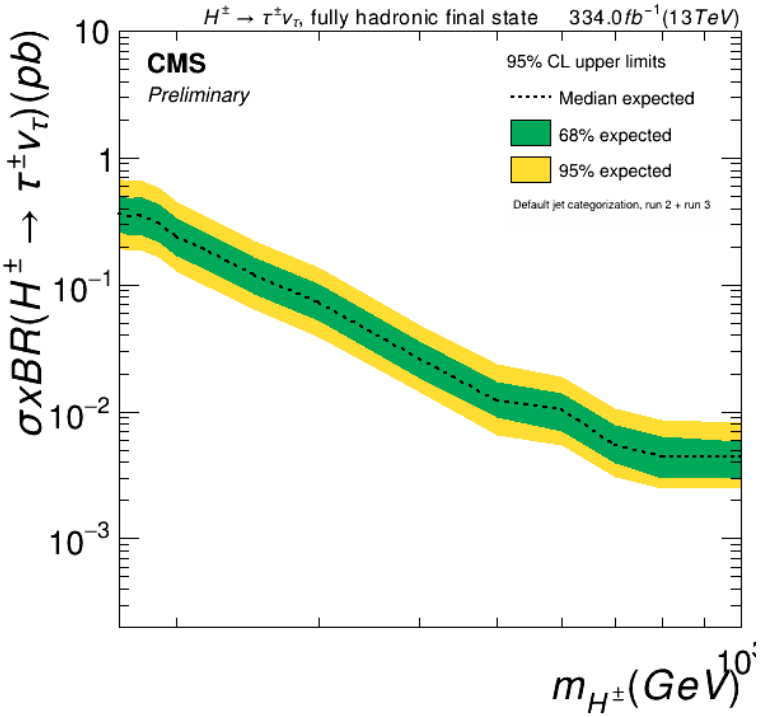


Figure 5.16: Expected limit for combining all Run 2 years using only the default selection ($N_{\text{jets}} \geq 3$ and $N_b > 0$), and while using Run 2 + Run 3 luminosity in calculations.

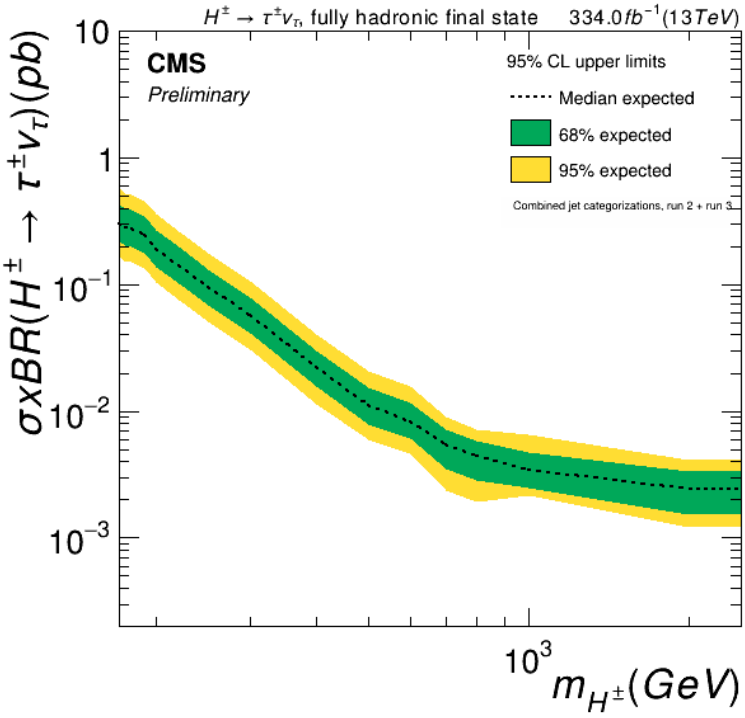


Figure 5.17: Expected limit for combining all Run 2 years and all of the jet categorizations while using Run 2 + Run 3 luminosity in calculations.

LHC data in extending the charged Higgs boson search reach, reinforcing the importance of continued optimization efforts in event selection and background suppression. Additionally, the projected limits for Run 2 + Run 3 give that the increased luminosity will continue to improve the search for $H^{\pm}$.

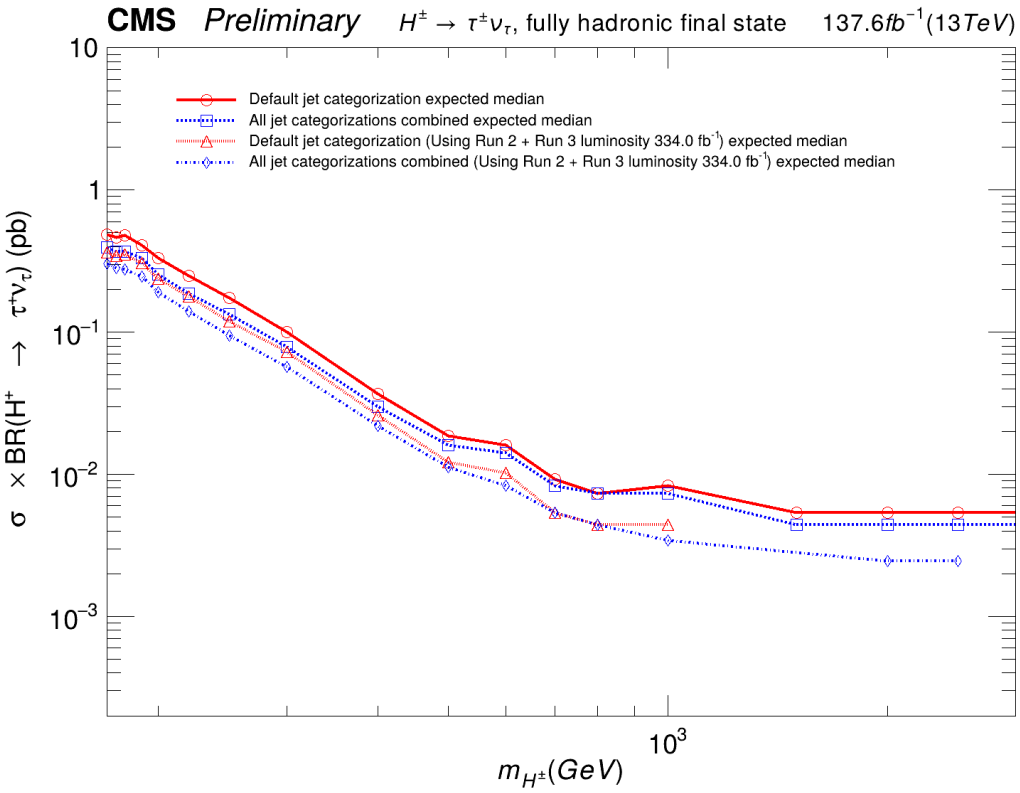


Figure 5.18: Plot of the expected median points for all Run 2 years combined using either default jet categorization of $N_{\text{jets}} \geq 3$ and $N_b > 0$ or combined all of the varied jet categorizations. The plot also has the projected expected median points using Run 2 + Run 3 luminosity for both default and combined jet categorization cases.

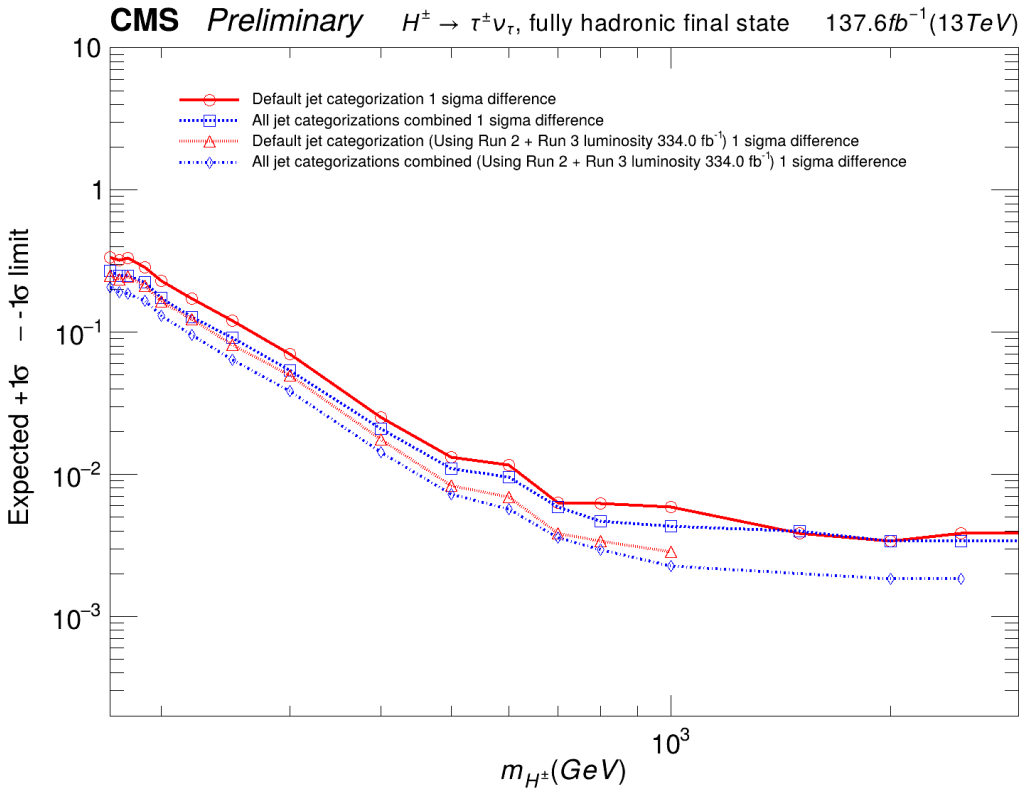


Figure 5.19: Plot of the expected 1σ band width for all Run 2 years combined using either default jet categorization of $N_{\text{jets}} \geq 3$ and $N_b > 0$ or combined all of the varied jet categorizations. The plot also has the projected expected median points using Run 2 + Run 3 luminosity for both default and combined jet categorization cases.

5.5 Future studies

Further studies with full shape systematics are crucial for future iterations of this study. While several key uncertainties were incorporated into the current analysis, some sources of systematic variation could be further refined or added. A comprehensive uncertainty treatment will improve the reliability of the expected limits, leading to a more precise interpretation of the results. Additionally, the inclusion of impact plots to better understand the sensitivity of various systematic uncertainties on the final expected limits is important. Impact plots provide a quantitative visualization of how different sources of uncertainty, such as jet energy scale corrections, tau identification efficiencies, and background modeling, influence the statistical results.

The evolution of the expected limit median points and 1σ band width across multiple analyses is illustrated in 5.20 and 5.21 respectively. Expanding the datasets by including Run 3 data from the LHC will significantly improve the statistical power of this analysis. With increased luminosity, imporoved algorithms (such as DeepJet and DeepTau), and improved detector performance, future studies are expected to improve the sensitivity to discover the charged Higgs boson.

Finally, an exploration of machine learning techniques for event selection and background suppression offer additional benefits as in Ref. [82]. Advanced classification methods, such as deep neural networks or boosted decision trees, are already shown to improve the signal efficiency while reducing systematic biases. Implementing such techniques alongside traditional methods leads to improved sensitivity to $H^{\pm}$ production.

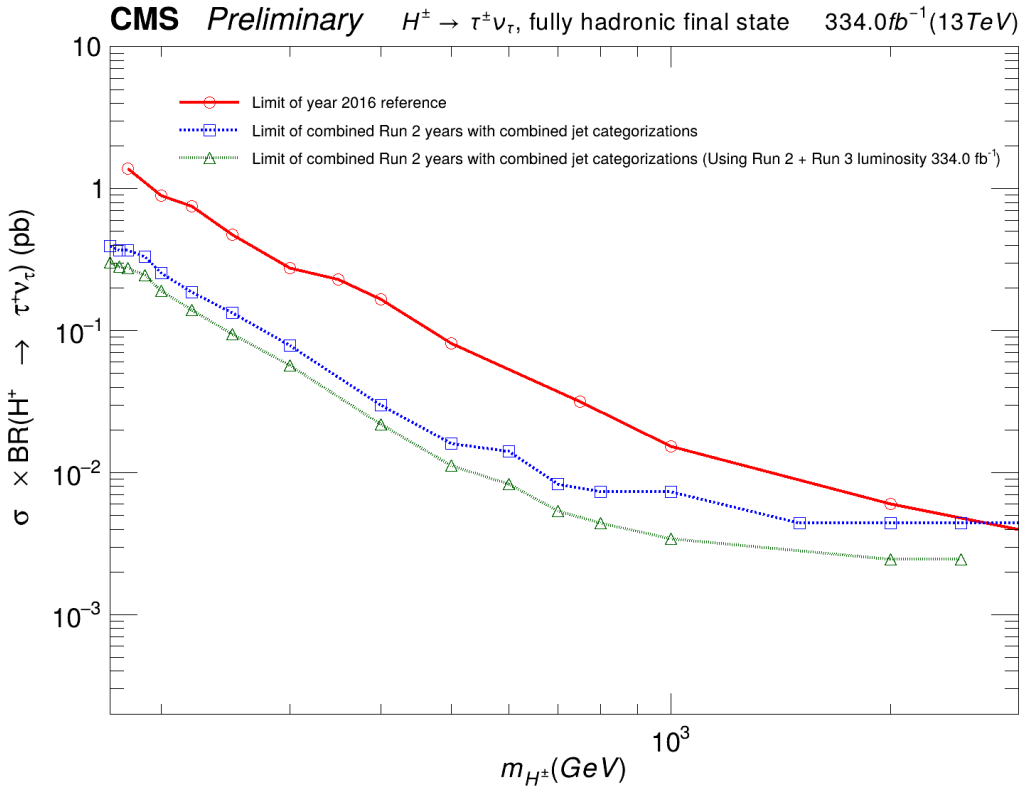


Figure 5.20: Plot of how the expected exclusion limits have evolved. This plot includes results presented in year 2016 reference and the two lowest yielding expected limits presented in this study; they are the combined Run 2 years with combined jet categorizations and the projection of the combined Run 2 years with combined jet categorizations using Run 2 + Run 3 luminosity of 334.0 fb^{-1}.

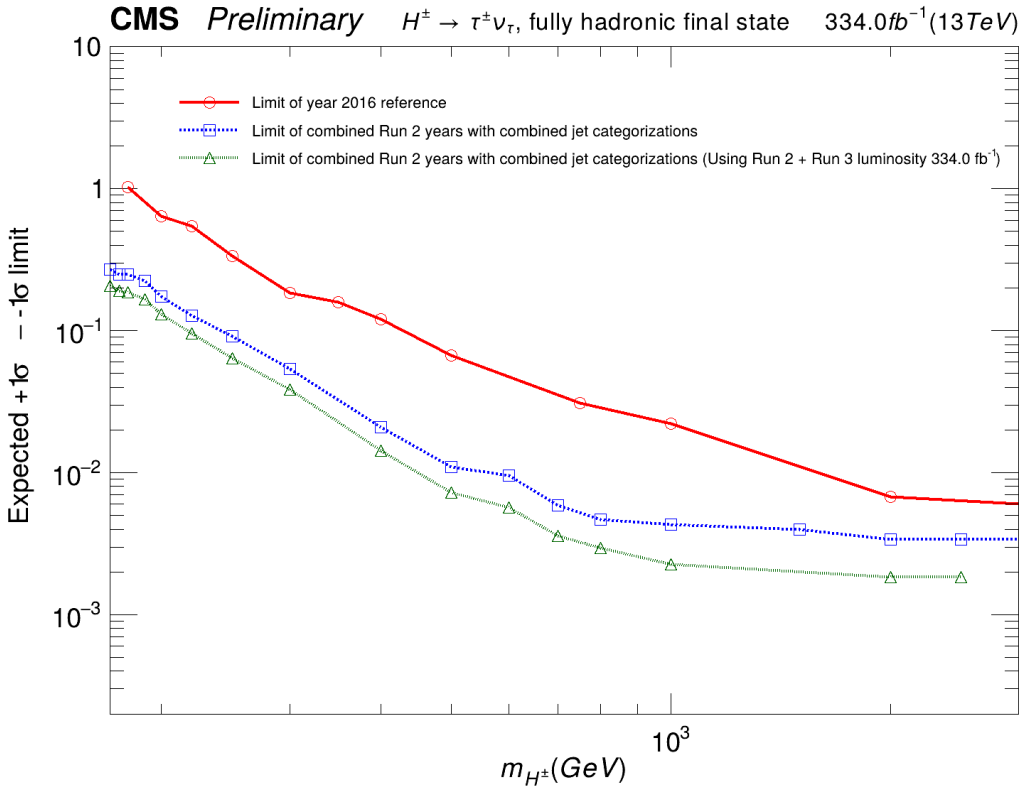


Figure 5.21: Plot of how the uncertainty of the expected limit has evolved. This plot includes results presented in year 2016 reference and the two lowest yielding expected limits presented in this study; they are the combined Run 2 years with combined jet categorizations and the projection of the combined Run 2 years with combined jet categorizations using Run 2 + Run 3 luminosity of 334.0 fb^{-1}.

6 Summary

The search for physics beyond the SM remains a key objective in particle physics, particularly in understanding the Higgs sector. This thesis investigates the sensitivity to discover a heavy charged Higgs boson ($H^{\pm}$) in the fully hadronic decay channel $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ using CMS Run 2 data collected at the LHC. The expected limits were derived separately for data recorded in 2016, 2017, and 2018, as well as for the combined Run 2 data, utilizing the $\mathrm{CL_S}$ statistical method. The combination of all data provided improved sensitivity over previous individual dataset analyses. Compared to earlier searches, such as [44, 81], this analysis demonstrates notable improvements in expected limits due to both the increased dataset and methodological refinements. Observed limits are not derived as the data is still blinded.

One key difference on this analysis is the treatment of systematic uncertainties, which were simplified to focus on the most dominant sources. The primary sources of uncertainty are identified as jet energy scale variations, tau identification efficiency, and differences in trigger selection across different data-taking years. The increased pileup in 2017 and 2018 is also found to impact expected limits. While some sources of uncertainty are mitigated through improved selection criteria, others remained a limiting factor in further sensitivity gains. Additionally, the usage of DeepJet and DeepTau algorithms yield better expected limits compared to the published 2016 results.

To maximize the analysis sensitivity, an optimization study is performed on the 2017 dataset, testing various event selection criteria, including jet multiplicity and b-tagging requirements. The optimal selection strategy, determined to be $N_{\mathrm{jets}} \geq 3$ and $N_b > 0$, is then applied to the full Run 2 dataset. This selection strategy leads to the lowest expected limits across all tested mass hypotheses.

A comparison with earlier studies highlights additional improvements in analysis techniques, such as the more comprehensive treatment of systematic uncertainties. Nonetheless, the overall performance of this study surpassed previous results, reaffirming the benefits of optimized selection criteria and improved tagging algorithms. Finally, a pro-

jection was made for CMS Run 2 + Run 3, assuming an integrated luminosity of 334.0 fb^{-1}. These projections indicate that future data collection and improved selection strategies could significantly enhance the sensitivity to the charged Higgs boson, reinforcing the importance of continued experimental searches in this channel.

In conclusion, this study has contributed to improving the expected exclusion limits for the heavy charged Higgs boson compared to previous results. The results demonstrate the feasibility of searching for $H^{\pm}$ in the fully hadronic τ-decay final state and provide a solid foundation for further studies extending the analysis to LHC Run 3.

Bibliography

[1] Glashow, S. L. (1961). Partial-symmetries of weak interactions. Nuclear Physics, 22(4), 579–588. https://doi.org/10.1016/0029-5582(61)90469-2

[2] Weinberg, S. (1967). A Model of Leptons. Physical Review Letters, 19(21), 1264-1266. https://doi.org/10.1103/physrevlett.19.1264

[3] Salam, A. (1994). Weak and electromagnetic interactions. Selected Papers of Abdus Salam, 244–254. https://doi.org/10.1142/9789812795915_0034

[4] Gell-Mann, M. (1961). THE EIGHTFOLD WAY: A THEORY OF STRONG INTERACTION SYMMETRY. https://doi.org/10.2172/4008239

[5] Zweig, G. (1964). An SU(3) model for strong interaction symmetry and its breaking. Version 2. 22–101.

[6] Fritzsch, H., Gell-Mann, M., & Leutwyler, H. (1973). Advantages of the color octet gluon picture. Physics Letters B, 47(4), 365–368. https://doi.org/10.1016/0370-2693(73)90625-4

[7] S. Navas et al. (Particle Data Group) (2024), Phys. Rev. D 110, 030001, doi: 10.1103/PhysRevD.110.030001

[8] Aitchison, I. J., & Hey, A. J. (2012). Gauge Theories in Particle Physics: A Practical Introduction, Fourth Edition - 2 Volume set. CRC Press.

[9] Schwartz, M. D. (2014). Quantum field Theory and the standard model. Cambridge University Press.

[10] File:Standard Model of Elementary Particles.svg. (2024). Wikimedia Commons. Retrieved 00:38, December 11, 2024 from https://commons.wikimedia.org/w/index.php?title=File:Standard_Model_of_Elementary_Particles.svg&oldid=917147923.

[11] Higgs, P. W. (1964). Broken symmetries, massless particles and gauge fields. Physics Letters, 12(2), 132–133. https://doi.org/10.1016/0031-9163(64)91136-9

[12] Higgs, P. W. (1964). Broken Symmetries and the Masses of Gauge Bosons. Physical Review Letters, 13(16), 508–509. https://doi.org/10.1103/physrevlett.13.508

[13] Guralnik, G. S., Hagen, C. R., & Kibble, T. W. B. (1964). Global Conservation Laws and Massless Particles. Physical Review Letters, 13(20), 585–587. https://doi.org/10.1103/physrevlett.13.585

[14] Higgs, P. W. (1966). Spontaneous Symmetry Breakdown without Massless Bosons. Physical Review, 145(4), 1156–1163. https://doi.org/10.1103/physrev.145.1156

[15] Kibble, T. W. B. (1967). Symmetry Breaking in Non-Abelian Gauge Theories. Physical Review, 155(5), 1554–1561. https://doi.org/10.1103/physrev.155.1554

[16] Englert, F., & Brout, R. (1964). Broken Symmetry and the Mass of Gauge Vector Mesons. Physical Review Letters, 13(9), 321–323. https://doi.org/10.1103/physrevlett.13.321

[17] File:Spontaneous symmetry breaking (explanatory diagram).png. (2020). Wikimedia Commons. Retrieved 00:33, December 11, 2024 from https://commons.wikimedia.org/w/index.php?title=File:Spontaneous_symmetry_breaking_(explanatory_diagram).png&oldid=480725564.

[18] ATLAS Collaboration (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Physics Letters B, 716(1), 1–29. https://doi.org/10.1016/j.physletb.2012.08.020

[19] CMS Collaboration (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Physics Letters B, 716(1), 30–61. https://doi.org/10.1016/j.physletb.2012.08.021

[20] CMS Collaboration (2013). Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. Journal of High Energy Physics, 2013(6). https://doi.org/10.1007/jhep06(2013)081

[21] CMS Collaboration (2015). Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. Physical Review D, 92(1). https://doi.org/10.1103/physrevd.92.012004

[22] CMS and ATLAS Collaborations (2015). Combined Measurement of the Higgs Boson Mass inppCollisions at s=7 and 8 TeV with the ATLAS and CMS Experiments. Physical Review Letters, 114(19). https://doi.org/10.1103/physrevlett.114.191803

[23] ATLAS Collaboration (2015). Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. The European Physical Journal C, 75(10). https://doi.org/10.1140/epjc/s10052-015-3685-1

[24] CMS and ATLAS Collaborations (2016). Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. Journal of High Energy Physics, 2016(8). https://doi.org/10.1007/jhep08(2016)045

[25] CMS Collaboration (2017). Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV. Journal of High Energy Physics, 2017(11). https://doi.org/10.1007/jhep11(2017)047

[26] Gross, D. J., & Wilczek, F. (1973). Ultraviolet Behavior of Non-Abelian Gauge Theories. Physical Review Letters, 30(26), 1343–1346. https://doi.org/10.1103/physrevlett.30.1343

[27] Politzer, H. D. (1973). Reliable Perturbative Results for Strong Interactions? Physical Review Letters, 30(26), 1346–1349. https://doi.org/10.1103/physrevlett.30.1346

[28] Nagashima, Y. (2014). Beyond the standard model of elementary particle physics. Wiley-Vch.

[29] Lee, T. D. (1973). A Theory of SpontaneousTViolation. Physical Review. D. Particles, Fields, Gravitation, and Cosmology/Physical Review. D. Particles and Fields, 8(4), 1226–1239. https://doi.org/10.1103/physrevd.8.1226

[30] Dorsch, G. C., Huber, S. J., & No, J. M. (2013). A strong electroweak phase transition in the 2HDM after LHC8. Journal of High Energy Physics, 2013(10). https://doi.org/10.1007/jhep10(2013)029

[31] Andersen, J. O., Gorda, T., Helset, A., Niemi, L., Tenkanen, T. V. I., Tranberg, A., Vuorinen, A., & Weir, D. J. (2018). Nonperturbative analysis of the electroweak phase transition in the two Higgs doublet model. Physical Review Letters, 121(19). https://doi.org/10.1103/physrevlett.121.191802

[32] Branco, G. C., Ferreira, P. M., Lavoura, L., Rebelo, M. N., Sher, M., & Silva, J. P. (2012). Theory and phenomenology of two-Higgs-doublet models. Physics Reports, 516(1-2), 1–102. https://doi.org/10.1016/j.physrep.2012.02.002

[33] Ginzburg, I. F., Krawczyk, M., & Osland, P. (2024). Two-Higgs-Doublet Models with CP violation. ArXiv.org. https://www.arxiv.org/abs/hep-ph/0211371

[34] Gunion, J. F., Haber, H., Kane, G., & Dawson, S. (2000). The Higgs Hunter's Guide. Westview Press.

[35] Barger, V., Hewett, J. L., & Phillips, R. J. (1990). New constraints on the charged Higgs sector in two-Higgs-doublet models. 41(11), 3421–3441. https://doi.org/10.1103/physrevd.41.3421

[36] Weinberg, S. (2013). The Quantum Theory of Fields: Volume 3, Supersymmetry. Cambridge University Press.

[37] Djouadi, A. (2008). The anatomy of electroweak symmetry breaking Tome II: The Higgs bosons in the Minimal Supersymmetric Model. Physics Reports, 459(1-6), 1–241. https://doi.org/10.1016/j.physrep.2007.10.005

[38] Carena, M., Heinemeyer, S., Stål, O., Wagner, C. E. M., & Weiglein, G. (2013). MSSM Higgs boson searches at the LHC: benchmark scenarios after the discovery of a Higgs-like particle. The European Physical Journal C, 73(9). https://doi.org/10.1140/epjc/s10052-013-2552-1

[39] LHC Higgs Cross Section Working Group Collaboration (2013). ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH Handbook of LHC Higgs cross sections: 3. Higgs Properties Report of the LHC Higgs Cross Section Working Group. https://doi.org/10.5170/CERN-2013-004

[40] CMS Collaboration (2012). Search for a light charged Higgs boson in top quark decays in pp collisions at $\sqrt{s} = 7\ TeV$. Journal of High Energy Physics, 2012(7). https://doi.org/10.1007/jhep07(2012)143

[41] Evans, L., & Bryant, P. (2008). LHC Machine. Journal of Instrumentation, 3(08), S08001. https://doi.org/10.1088/1748-0221/3/08/s08001

[42] Mobs, E. (2016). The CERN accelerator complex. Complexe des accélérateurs du CERN. https://cds.cern.ch/record/2197559.

[43] Sakuma, T., & McCauley, T. (2014). Detector and Event Visualization with SketchUp at the CMS Experiment. Journal of Physics Conference Series, 513(2), 022032. https://doi.org/10.1088/1742-6596/513/2/022032

[44] Laurila, S. H. (2019). Search for Charged Higgs Bosons Decaying to a Tau Lepton and a Neutrino with the CMS Experiment. CERN Document Server. https://cds.cern.ch/record/2696066.

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

[46] CMS Collaboration (2006). "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software", CMS Technical Design Report CERN-LHCC-2006-001, https://cds.cern.ch/record/922757.

[47] Corrin, E. P. (2002). Development of Digital Readout Electronics for the CMS Tracker.

[48] CMS Collaboration (2018). "The performance of the CMS muon system with data at $\sqrt{\mathrm{s}} = 13$ TeV", Technical Report CMS-CR-2018-167,

[49] CMS Collaboration (2017). "Particle-flow reconstruction and event description with the CMS detector", *JINST* **12** P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[50] CMS Collaboration (2015). "Mini-AOD: A new analysis data format for CMS", *J. Phys. Conf. Ser.* **664** 072052, doi:10.1088/1742-6596/664/7/072052, arXiv:1702.04685.

[51] CMS Collaboration (2024), CMS Luminosity - Public Results, Cern.ch. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, Accessed 3 December 2024

[52] Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., Shao, H.-S. ., Stelzer, T., Torrielli, P., & Zaro, M. (2014). The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. Journal of High Energy Physics, 2014(7). https://doi.org/10.1007/jhep07(2014)079

[53] Alwall, J., Höche, S., Krauss, F., Lavesson, N., Lönnblad, L., Maltoni, F., Mangano, M., Moretti, M., Papadopoulos, C., Piccinini, F., Schumann, S., Treccani, M., Winter, J., & Worek, M. (2007). Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic

collisions. The European Physical Journal C, 53(3), 473–500. https://doi.org/10.1140/epjc/s10052-007-0490-5

[54] Nason, P. (2004). A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. Journal of High Energy Physics, 2004(11), 040–040. https://doi.org/10.1088/1126-6708/2004/11/040

[55] Frixione, S., Nason, P., & Oleari, C. (2007). Matching NLO QCD computations with parton shower simulations: the POWHEG method. Journal of High Energy Physics, 2007(11), 070–070. https://doi.org/10.1088/1126-6708/2007/11/070

[56] Alioli, S., Nason, P., Oleari, C., & Re, E. (2010). A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. 2010(6). https://doi.org/10.1007/jhep06(2010)043

[57] Sjöstrand, T., Ask, S., Christiansen, J. R., Corke, R., Desai, N., Ilten, P., Mrenna, S., Prestel, S., Rasmussen, C. O., & Skands, P. Z. (2015). An introduction to PYTHIA 8.2. Computer Physics Communications, 191, 159–177. https://doi.org/10.1016/j.cpc.2015.01.024

[58] GEANT4 Collaboration (2003). Geant4—a simulation toolkit. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 506(3), 250–303. https://doi.org/10.1016/s0168-9002(03)01368-8

[59] Allison, J., Amako, K., Apostolakis, J., Araujo, H., Dubois, P. A., Asai, M., Barrand, G., Capra, R., Chauvie, S., Chytracek, R., Cirrone, G. A. P., Cooperman, G., Cosmo, G., Cuttone, G., Daquino, G. G., Donszelmann, M., Dressel, M., Folger, G., Foppiano, F., & Generowicz, J. (2006). Geant4 developments and applications. IEEE Transactions on Nuclear Science, 53(1), 270–278. https://doi.org/10.1109/TNS.2006.869826

[60] Gray, L., Smith, N., Tovar, B., Chen, Y., Krommydas, I., Novak, A., Chakraborty, J., Krug, N., Fackeldey, P., Valsecchi, D., Chopra, S., Simeon, R., Watts, G., Thain, D., Stark, G., Rübenach, J., Galewsky, B., Fischer, B., Taylor, D., Woodard, A. (2024). CoffeaTeam/coffea: v0.7.23 (v0.7.23). Zenodo. https://doi.org/10.5281/zenodo.14206140

[61] CMS Collaboration (2019). Search for charged Higgs bosons in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. Journal of High Energy Physics, 2019(7). https://doi.org/10.1007/jhep07(2019)142

[62] CMS and ATLAS Collaborations (2011). The LHC Higgs Combination Group, Procedure for the LHC Higgs boson search combination in Summer 2011. CERN Document Server. https://cds.cern.ch/record/1379837

[63] Junk, T. (1999). Confidence Level Computation for Combining Searches with Small Statistics. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 434(2-3), 435–443. https://doi.org/10.1016/S0168-9002(99)00498-2

[64] Read, A. L. (2002). Presentation of search results: theCLstechnique. Journal of Physics G Nuclear and Particle Physics, 28(10), 2693–2704. https://doi.org/10.1088/0954-3899/28/10/313

[65] Cowan, G., Cranmer, K., Gross, E., & Vitells, O. (2011). Asymptotic formulae for likelihood-based tests of new physics. The European Physical Journal C, 71(2), 1554. https://doi.org/10.1140/epjc/s10052-011-1554-0

[66] CMS Collaboration (2015). Search for a charged Higgs boson in pp collisions at $\sqrt{s} = 8$ TeV. Journal of High Energy Physics, 2015(11). https://doi.org/10.1007/jhep11(2015)018

[67] CMS Collaboration (2017). Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. Journal of Instrumentation, 12(02), P02014–P02014. https://doi.org/10.1088/1748-0221/12/02/p02014

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

[69] CMS Collaboration (2018). Performance of missing transverse momentum in pp collisions at sqrt(s)=13 TeV using the CMS detector. CERN Document Server. https://cds.cern.ch/record/2628600

[70] CMS Collaboration (2018). Measurement of the inelastic proton-proton cross section at $\sqrt{s}$=13 TeV. Journal of High Energy Physics, 2018(7). https://doi.org/10.1007/jhep07(2018)161

[71] Butterworth, J., Carrazza, S., Cooper-Sarkar, A., De Roeck, A., Feltesse, J., Forte, S., Gao, J., Glazov, S., Huston, J., Kassabov, Z., McNulty, R., Morsch, A., Nadolsky, P., Radescu, V., Rojo, J., & Thorne, R. (2016). PDF4LHC recommendations for LHC Run II. Journal of Physics G Nuclear and Particle Physics, 43(2), 023001. https://doi.org/10.1088/0954-3899/43/2/023001

[72] Barlow, R., & Beeston, C. (1993). Fitting using finite Monte Carlo samples. Computer Physics Communications, 77(2), 219–228. https://doi.org/10.1016/0010-4655(93)90005-W

[73] CMS Collaboration (2022). Identification of hadronic tau lepton decays using a deep neural network. Journal of Instrumentation, 17(07), P07023. https://doi.org/10.1088/1748-0221/17/07/p07023

[74] CMS Collaboration (2019). Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons. CERN Document Server. https://cds.cern.ch/record/2694158

[75] CMS Collaboration (2018). Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. Journal of Instrumentation, 13(05), P05011–P05011. https://doi.org/10.1088/1748-0221/13/05/p05011

[76] E. Bols, J. Kieseler, M. Verzetti, Stoye, M., & A. Stakia. (2020). Jet flavour classification using DeepJet. Journal of Instrumentation, 15(12), P12012–P12012. https://doi.org/10.1088/1748-0221/15/12/p12012

[77] CMS Collaboration (2018) Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector. CERN Document Server. https://cds.cern.ch/record/2646773

[78] CMS Collaboration (2024). The CMS Statistical Analysis and Combination Tool: Combine. Computing and Software for Big Science, 8(1). https://doi.org/10.1007/s41781-024-00121-4

[79] CombineHarvester: Introduction. (2025). Github.io. https://cms-analysis.github.io/CombineHarvester/

[80] CMS Collaboration (2020) Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-probe technique. CERN Document Server. https://cds.cern.ch/record/2724492/

[81] CMS Collaboration (2016). Search for charged Higgs bosons with the $H^{\pm} \rightarrow \tau^{\pm}\nu_{\tau}$ decay channel in the fully hadronic final state at $\sqrt{s} = 13$ TeV. CERN Document Server. https://cds.cern.ch/record/2223865

[82] Kössi, A. (2023). Implementing a machine learning classifier for the proton-proton to H+ and X, H+ to a tau lepton and a neutrino channel in CMS in the COFFEA framework. https://urn.fi/URN:NBN:fi:aalto-202301311856

A The usage of AI in this thesis

The large language model ChatGPT-4o (developed by OpenAI) was used in this thesis as writing help. It was used to generate initial drafts for various sections based on specific input, such as outlined subsections and key points to address. However, all generated drafts were thoroughly reviewed and revised to ensure accuracy and relevance. Additionally, ChatGPT-4o was employed for spelling and grammar checks throughout the text. ChatGPT-4o was not used in the maturity test/abstract section of this thesis. More info about using AI in the University of Helsinki can be found in https://studies.helsinki.fi/instructions/article/using-ai-support-learning.

B Uncertainties in the analysis

Table B.1: Systematic uncertainties used in the analysis.

Description
Renormalization and factorization scale variation for H+ signal
Renormalization and factorization scale variation for electroweak background
Renormalization and factorization scale variation for top background
PDF uncertainties for H+ signal
PDF uncertainties for electroweak background
PDF uncertainties for top background
b-tagging efficiency uncertainty
Electron veto efficiency uncertainty
Muon veto efficiency uncertainty
MET trigger efficiency uncertainty (MC)
MET trigger efficiency uncertainty (data)
Tau identification efficiency uncertainty
Tau trigger efficiency uncertainty (MC)
Tau trigger efficiency uncertainty (data)
b-jet mistag rate uncertainty
Electron-to-tau fake rate uncertainty
Muon-to-tau fake rate uncertainty
Pileup reweighting uncertainty
Jet energy resolution uncertainty
Jet energy scale uncertainty
MET scale uncertainty
Tau energy scale uncertainty
QCD scale uncertainty for Drell-Yan background
QCD scale uncertainty for single top background
QCD scale uncertainty for top pair background
QCD scale uncertainty for diboson background
QCD scale uncertainty for W+jets background
Luminosity uncertainty (13 TeV)
Top quark mass uncertainty
PDF uncertainty for Drell-Yan background
PDF uncertainty for single top background
PDF uncertainty for top pair background
PDF uncertainty for diboson background
PDF uncertainty for W+jets background