UNIVERSITÀ
DI SIENA
1240

Department of
Physical Sciences, Earth and Environment

Ph.D. in Experimental Physics
XXXVI Cycle

Search for resonant double Higgs production in the bb$\tau\tau$ final state at the CMS experiment.

Supervisor:
Prof. Maria Agnese Ciocci

Candidate:
Valeria D'Amante
Mat. 106371

Valeria D'Amante

Academic Year 2023-2024

Alla mia mamma e al mio papà,
Isabella e Cosmo.
Ogni giorno mi donate la vita.

Contents

List of Figures

List of Tables

Abstract

This thesis presents a search for resonant Higgs bosons pair (HH) production via the Gluon-gluon Fusion (ggF) mechanism, predicted in different Beyond Standard Model (BSM) scenarios to be observable with the current Large Hadron Collider (LHC) dataset. The search focuses on the $HH \rightarrow b\bar{b}\tau^+\tau^-$ final state, which offers a good balance between branching fraction and experimental feasibility but presents challenges due to neutrinos in τ decays and background discrimination. Advanced techniques are employed for τ reconstruction and identification, along with sophisticated modeling of the relevant backgrounds.

The analysis, performed on proton-proton (pp) collision data collected by the Compact Muon Solenoid (CMS) experiment during Run 2 (2016–2018) with an integrated luminosity of 137.1 fb^{-1}, employs advanced techniques for τ reconstruction and background modeling. The $\tau^+\tau^-$ pair is studied through three decay channels: $\tau_h\tau_h$ (fully hadronic) and $\tau_h\tau_\ell$, where τ_h represents a tau lepton decaying into hadrons plus a ν_τ , and τ_ℓ corresponds to the leptonic decay of the tau into either an electron ($\ell = e$) or a muon ($\ell = \mu$). A key aspect of the analysis is the use of machine-learning-based identification algorithms, including *DeepTau* for τ identification. The primary results are obtained using *DeepTau2017v2p1*, the stable version recommended for Run 2 analyses. Additionally, a study with *DeepTau2018v2p5*, an improved and re-trained version optimized for future runs, demonstrates better data/MC agreement, confirming its suitability for upcoming analyses. As part of the CMS internal review process, the analysis remains blinded to avoid bias. The 95% Confidence Level upper limits on the HH production cross section are set as a function of resonance mass and spin, interpreted within the Warped Extra Dimensions theory (WED) model for masses from 250 to 3000 GeV under spin-0 and spin-2 hypotheses.

Introduction

The discovery of a scalar boson with a mass of approximately 125 GeV[1] by the ATLAS [1] and CMS [2] collaborations in 2012 at the LHC [3] at CERN was a landmark event in particle physics. This particle, consistent with the Higgs boson predicted by the Standard Model (SM), confirmed the mechanism of electroweak symmetry breaking and marked a new era of precision measurements and searches for new phenomena. The Higgs boson discovery not only validated the theoretical framework of the Standard Model but also opened up new avenues to explore physics beyond it. Despite its successes, the Standard Model is not a complete theory of nature. It does not explain phenomena such as neutrino masses, dark matter, dark energy, or the baryon asymmetry of the universe. Moreover, it does not incorporate gravity in a quantum framework. These limitations motivate the search for theories Beyond Standard Model (BSM), which aim to address these open questions and resolve theoretical challenges such as the naturalness and hierarchy problems.

This thesis focuses on resonant Higgs bosons pair (HH) production in the $b\bar{b}\tau^-\tau^+$[2] final state, using the framework of Warped Extra Dimensions theory (WED) theory, which predicts the existence of heavy resonances that couple to pairs of Higgs bosons. The study explores resonance masses ranging from 250 GeV to 3000 GeV, considering two spin hypotheses: spin-0 (scalar) and spin-2 (graviton-like).

Data from proton-proton (pp) collisions at the CMS experiment at the LHC during Run 2 (2016–2018) is analyzed, with a total integrated luminosity of 137.1 fb^{-1}. The analysis fouses on Higgs boson pair production via gluon-gluon fusion in the $bb\tau^-\tau^+$ final state, chosen due to the high branching fraction of the decay $H \to b\bar{b}$ (57.5%) and the clean signal from τ-lepton decays. This decay channel offers a good balance between statistical sensitivity and experimental challenges, though it requires precise identification of τ-leptons and b-jets, along with effective background suppression.

The analysis includes three primary decay modes of the $\tau\tau$ system: leptonic decays

[1] In this thesis the system of natural units is used with $\hbar = c = 1$, where $\hbar = h/2\pi = 6.582119569(26) \times 10^{-22}$MeV s and $c = 299\,792\,458$ m s^{-1}

[2] For simplicity throughout this thesis the leptons charge and the particle-antiparticle nature of the quarks will not be specified, thus $\tau\tau$ corresponds to $\tau^+\tau^-$, bb corresponds to $b\bar{b}$ and in general qq stands for $q\bar{q}$ unless differently specified.

($\tau_e\tau_h$, $\tau_\mu\tau_h$) and hadronic decays ($\tau_h\tau_h$)[3]. The research builds on previous CMS searches by utilizing the full Run 2 dataset, employing advanced particle identification methods, and optimizing reconstruction algorithms tailored for this final state.

A crucial aspect of this analysis is the identification of hadronic τ decays, where the *DeepTau* algorithm plays a key role. *DeepTau2017v2p1*, used in Run 2, was trained solely on Monte Carlo (MC) simulations, leading to discrepancies between data and MC that required correction through scale factors. This thesis introduces the implementation and evaluation of *DeepTau2018v2p5*, an updated version incorporating adversarial machine learning techniques to mitigate these mismodeling effects at the training level. By training on both MC and real collision data, *DeepTau2018v2p5* aims to reduce systematic uncertainties and improve τ identification, ultimately enhancing the sensitivity of searches involving hadronic τ decays. Testing *DeepTau2018v2p5* on Run 2 data is crucial to validate its performance in well-understood datasets and ensure compatibility with previous analyses.

The 95% Confidence Level (CL) upper limit on the HH production cross section are set as a function of resonance mass, under the two spin hypotheses is presented. At the time of writing, the analysis is still under review, and the statistical analysis has been performed with a blind approach, meaning only expected results are presented, and the actual observed data is excluded. The goal is to document the methodology, results, and implications of this search, contributing to the ongoing exploration of the Higgs sector beyond the Standard Model.

The theoretical framework and motivations for this search are introduced in Chapter 1. Chapter 2 describes the LHC and CMS detector used to acquire the analyzed data. The sophisticated algorithms used to reconstruct and identify the physical objects (charged lepton, neutrino and jets) from the collected data are summarized Chapter 3. Chapter 4 describes the final state and the analysis strategy, from the trigger requirements to the physics objects selection and the events categorization. The data composition and modelling are discussed in Chapter 5. The estimation of the systematic uncertainties is presented in Chapter 6. Finally, the statistical analysis techniques adopted and the results are reported in Chapter 7.

[3]Throughout the thesis, the $\tau_e\tau_h$, $\tau_\mu\tau_h$ and $\tau_h\tau_h$ final states are labeled as $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ respectively.

Chapter 1

Theoretical framework and experimental state of the art

The Standard Model (SM) of particle physics is a renormalizable Quantum Field Theory (QFT) describing all known fundamental particles and three of the four fundamental forces: strong, weak, and electromagnetic. It represents the most rigorous and experimentally validated framework of particle physics to date.
The SM completeness was confirmed with the 2012 discovery of the Higgs boson by the ATLAS and the CMS experiments at the LHC, marking the observation of its final predicted component. However, the SM is not without limitations.
Discrepancies with experimental observations and astrophysical phenomena, as well as unresolved theoretical questions, suggest the existence of physics Beyond Standard Model (BSM). Extensions to the SM scalar sector provide promising avenues for exploring such new physics, with the Higgs boson playing a central role. In particular, double Higgs production is pivotal for probing the scalar sector structure, as it directly relates to Higgs self-interactions and offers a means to uncover BSM effects.

This chapter provides an overview of the theoretical framework, highlighting the importance of studying Higgs pair (HH) production in both the SM and BSM physics and of the current status of the experimental results for Higgs and double Higgs production. In Section 1.1, the SM framework is introduced; Section 1.2 provides the description of the Spontaneous symmetry breaking (SSB) with a specific focus on the Brout-Englert-Higgs (BEH) mechanism, the Higgs boson and its properties and couplings. Then, Section 1.4 discusses the phenomenology of single Higgs boson production at the LHC. The production of Higgs boson pairs is reviewed in Section 1.5, covering both the SM predictions and the context of BSM scenarios, with a detailed focus on the Warped Extra Dimensions theory (WED), as it predicts the resonances searched in this thesis. Finally, Section 1.6 summarizes the experimental status of H and HH searches at the LHC.

1.1. The Standard Model of particle physics

The Standard Model (SM) of particle physics [4,5] describes the microscopic world in terms of its fundamental constituents and their interactions. It provides an elegant theoretical framework for three of the four fundamental forces - the electromagnetic, the weak and the strong interactions — and the elementary particles that constitute matter. However, the SM does not incorporate the gravitational interaction, as gravity is negligible at subnuclear scales.
Fundamental particles in the SM are represented by fermion fields of spin-1/2, which are grouped into two families: leptons and quarks. Each family consists of three generations, with each generation containing a doublet of two particles. This structure applies equally to particles and their antiparticles[1].
Fermions across generations share identical properties, differing only in their masses. These mass differences arise from their varying coupling strengths to the Higgs scalar field (H), as discussed in Section 1.2.1.
The lepton doublet in each generation comprises a charged massive lepton and a neutral particle known as a neutrino. The electric charge of leptons is expressed as a fraction of the elementary electromagnetic charge e[2]. In the SM neutrinos are considered massless[3].
Similarly, quark doublets consist of two types of quarks (flavors) with fractional electric charges of +2/3 and -1/3, respectively. Quarks are massive and interact via both the strong and electroweak forces. Unlike leptons, quarks possess an additional property called color charge, which comes in three types: "red" (R), "green" (G) and "blue" (B). This property governs their interaction through the strong nuclear force. Quarks do not exist as free particles in nature; instead, they combine to form color-neutral states known as hadrons. Hadrons are classified into mesons, which consist of a quark-antiquark pair, and baryons, which are composed of three quarks.
Experimental evidence confirms the existence of six quark fields and six lepton fields, amounting to 12 fermion fields and fixing the number of SM generations at three. Stable visible matter in the Universe is composed of first-generation fermions: up and down quarks, electrons, and neutrinos. Particles from the second and third generations, except neutrinos, decay rapidly into first-generation particles. Thus, the known Universe is predominantly made of first-generation particles. The remaining particles described by the SM are typically produced in particle accelerators, energetic cosmic-ray collisions, or existed in the early Universe. Due to their short lifetimes, these particles play no significant role in the present-day Universe.
Fermions interact via the exchange of spin-1 bosons, which mediate the SM fundamental forces. The photon (γ) mediates electromagnetic interactions among all charged fermions, excluding neutrinos, which are electrically neutral. Eight massless and elec-

[1]Antiparticles have the same mass, spin, and lifetime as their corresponding particles but opposite charges.

[2]The elementary charge is approximately $e \simeq 1.602 \times 10^{-19}$C

[3]Experimental evidence from neutrino flavor oscillations demonstrates that neutrinos have tiny but non-zero masses. These modifications to the SM are discussed in [6,7].

trically neutral gluons mediate the strong nuclear interaction among quarks, carrying combinations of two color charges. The $W^{\pm}$ and Z bosons mediate the weak nuclear interaction, which involves all fermions.
The final component of the SM is the Higgs boson, a neutral spin-0 particle associated with a scalar field. Through the mechanism of SSB, the Higgs field imparts mass to both fermions and gauge bosons (discussed in Sections 1.2.1 and 1.2.1).
A summary of the experimental properties of the quark and lepton families, force carriers, and the Higgs boson is presented in Fig. 1.1.

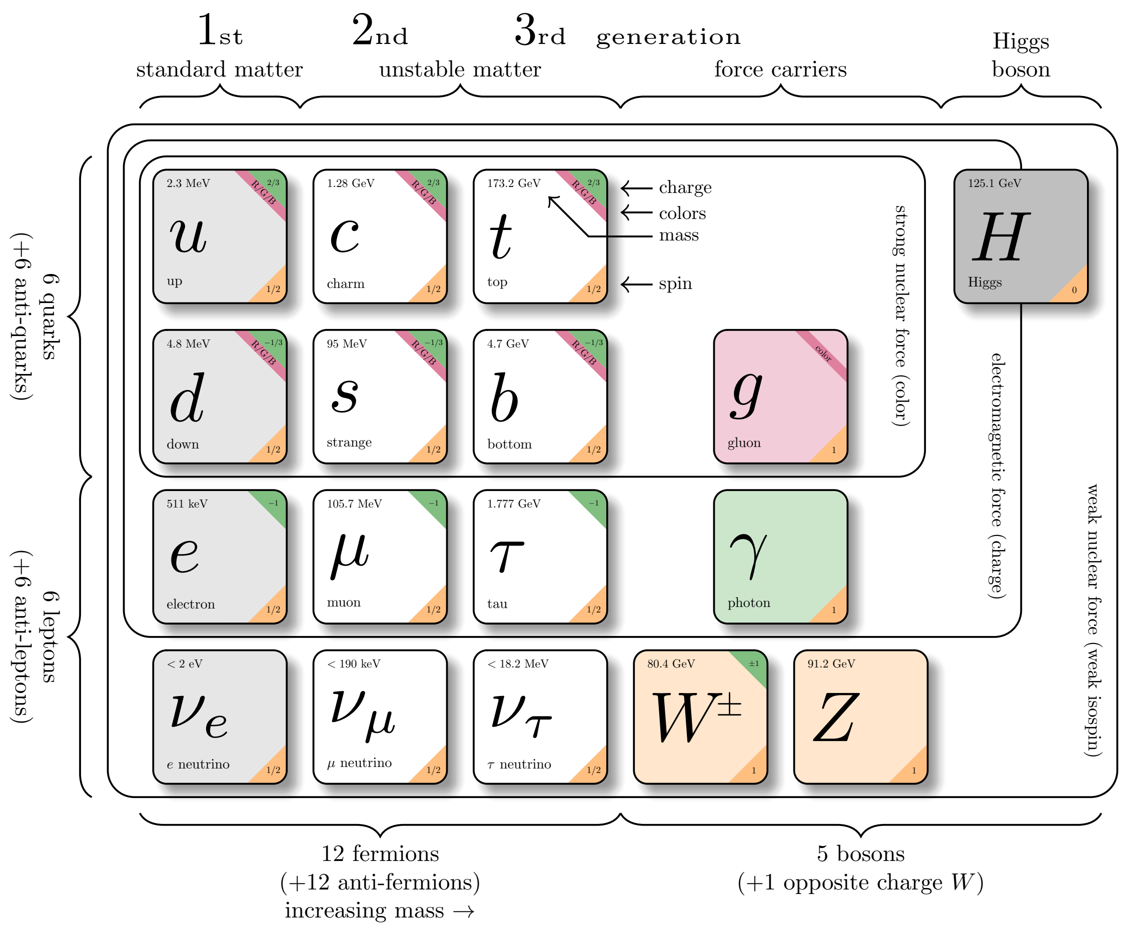


Figure 1.1: Summary diagram of all the SM particles: the fermions (on the left), which are the matter particles, are divided into quarks (in violet) and leptons (in green), while the bosons (on the right), are classified into gauge bosons, i.e. the carriers of the interactions (in red) and the scalar Higgs boson (in yellow). For each particle, the mass, electric charge and spin values are reported.

The gauge theory (in Section 1.1.1) and the SSB in gauge theories (Section 1.2) form the backbone of the SM, each grounded in distinct but interrelated theoretical principles.

1.1.1 Standard Model as a gauge theory

The SM is a renormalizable QFT that accurately describes all observed phenomena at subnuclear scales. It unifies the strong and electroweak forces and includes a scalar sector responsible for the BEH mechanism (Section 1.2.1), which breaks the electroweak symmetry. This mechanism generates the masses of quarks, leptons, the $W^{\pm}$ and Z bosons, and introduces a new spin-0 boson: the Higgs boson.
The foundation of the SM lies in its lagrangian density[4], a concise formulation of the theory that governs the dynamics of particles and interactions. The SM's interactions arise from the principle of local gauge invariance under the non-Abelian symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$, representing the strong and electroweak interactions. To satisfy gauge and Lorentz invariance, vector bosons are introduced as the force carriers.
Throughout this discussion, natural units are used, with $\hbar = c = 1$. Energy, mass, and momentum are measured in eV and multiples.

The strong interaction is described by Quantum Chromodynamics (QCD), a quantum field theory based on the gauge symmetry group $SU(3)_C$. This group features eight generators, represented mathematically by the Gell-Mann matrices, which correspond to eight physical gauge fields known as gluons. The quark field of the flavour f is a color triplet of $SU(3)_C$. Gluons mediate the strong force by coupling to the color charge of quarks. Unlike photons, gluons themselves carry color charge. This unique feature arises from the non-Abelian nature of the $SU(3)_C$ group and allows gluons to interact among each other. To preserve the gauge invariance, gluons and quarks are massless in the QCD formulation.
The strength of the strong interaction is determined by the strong coupling constant, $\alpha_s(\mathcal{Q})$, which depends on the energy scale $\mathcal{Q}$. This scale dependence gives rise to two hallmark features of QCD: *color confinement* and *asymptotic freedom* [8,9].

At low energy scales ($\mathcal{Q} = \Lambda_{\text{QCD}} \sim 200\,\text{MeV}$), the strong coupling constant α_s becomes large, leading to the phenomenon of *color confinement*. In this regime, the force between quarks increases with distance. As a result, quarks cannot exist as free particles and are instead bound together into color-neutral composite particles known as *hadrons*: *baryons*, composed of three quarks and *mesons*, consisting of a quark-antiquark pair. The color confinement explains why free quarks have never been observed in nature and why visible matter is composed entirely of confined states [10].

At high energy scales ($\mathcal{Q} \gg \Lambda_{\text{QCD}}$), the strong coupling constant α_s decreases due to the renormalization group behavior of the $SU(3)_C$ gauge symmetry group and the effects of virtual particles in quantum loops. This phenomenon, known as *asymptotic freedom*, implies that quarks and gluons interact weakly at short distances or high energies, behaving almost as free particles. Asymptotic freedom is crucial for enabling perturbative calculations in high-energy processes, such as those occurring in hadronic collisions at the Large Hadron Collider (LHC). Unlike the electroweak interactions, the

[4] For simplicity, the term "lagrangian density" will be referred to as the "lagrangian" throughout this discussion.

$SU(3)_C$ symmetry remains unbroken at all experimentally accessible energy scales, even after the introduction of the BEH mechanism (Section 1.2.1), leaving gluons massless. Details of the construction and properties of the QCD lagragian can be found in [11].

The electroweak theory, developed by Glashow, Salam, and Weinberg in the 1960s, unifies the electromagnetic and weak interactions.

The electromagnetic interaction is described by Quantum Electrodynamics (QED), and is governed by the $U(1)_{em}$ symmetry, with the electric charge as its generator and the photon (A_μ) as its gauge boson. For further details, see [12]. The photon is massless as a consequence of the gauge invariance, and it is not self–interacting because of the abelian property of the group.

The weak interaction arises from the $SU(2)_L$ symmetry and is mediated by the charged $W^\pm$ bosons and the neutral Z boson, affecting particles with a non-zero isospin charge. Experimental observations of parity violation in weak interactions are included into the theoretical framework. Left-handed fermions form $SU(2)_L$ doublets, while right-handed ones are singlets [5]. This distinction implies that left-handed and right-handed fermion field components transform differently under local gauge transformations.

The electroweak theory is based on the requirement of local gauge invariance of the lagrangian under the $SU(2)_L \times U(1)_Y$ symmetry group [13,14].
The $SU(2)_L$ gauge group is a non-Abelian symmetry associated with weak isospin (I_3). Gauge invariance under this group results in the presence of three gauge bosons fields, W^i_μ ($i = 1, 2, 3$). Fermion fields with left-handed chirality are represented as $SU(2)_L$ doublets, while right-handed fermions are $SU(2)_L$ singlets and do not interact with the W^i_μ fields.
The $U(1)_Y$ gauge group is Abelian and corresponds to weak hypercharge (Y). A single gauge boson, B_μ, arises from $U(1)_Y$ gauge invariance and interacts with both left- and right-handed fermion fields.
The generators of the $SU(2)_L$ and $U(1)_Y$ groups are, respectively, the weak isospin operators, $I \equiv (I_1, I_2, I_3)$, and the weak hypercharge Y. Gauge invariance associates the W^i_μ bosons with the weak isospin generators, while the B_μ boson is linked to the hypercharge generator.

Both quarks and leptons are structured into three generations (or families), each consisting of $SU(2)_L$ left-handed doublets, while right-handed chiral fermions remain singlets under the $SU(2)_L \times U(1)_Y$ group, as summarized in Tab. 1.1. This distinction ensures that weak interactions violate parity (P-symmetry), as they act exclusively on left-handed particles. Experimentally, the mass eigenstates of down-type quarks do not coincide with their weak eigenstates. The mixing between these two bases is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [5], which quantifies the probabilities of flavor-changing transitions between quark generations.

The electroweak charge, known as the hypercharge, unifies weak and electromag-

Type	1st gen.	2st gen.	3st gen.	I_3	Y	Q
Quarks	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$	$\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$
	u_R	c_R	t_R	0	$\frac{4}{3}$	$\frac{2}{3}$
	d_R	s_R	b_R	0	$-\frac{2}{3}$	$-\frac{1}{3}$
Leptons	$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} -1 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$
	e_R	μ_R	τ_R	0	-2	-1
	$\nu e, R$	$\nu\mu, R$	$\nu\tau, R$	0	0	0

Table 1.1: The SM fermion fields according to the $SU(2)_L$ representation, with left-handed doublets and right handed singlets, and their $SU(2)_L \times U(1)_Y$ quantum numbers: the third component of the weak isospin, I_3, the electroweak hypercharge Y and the electric charge Q.

netic charges through the Gell-Mann–Nishijima relation [15]:

$$Q = \frac{Y}{2} + I_3$$

where Q is the electric charge, I_3 is the weak isospin component, and Y is the weak hypercharge.
At this stage, by comparing experimental currents with those required in the Lagrangian to describe electroweak interactions [], it is evident that the gauge fields W^i_μ and B_μ do not directly correspond to the physical mediators of the electroweak interactions. Instead, the physical bosons are given by linear combinations of these gauge fields:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu), \qquad Z_\mu = \frac{gW^3_\mu - g'B_\mu}{\sqrt{g^2 + g'^2}}, \qquad A_\mu = \frac{gW^3_\mu + g'B_\mu}{\sqrt{g^2 + g'^2}} \tag{1.1}$$

where g and g' are the electroweak coupling constants for $SU(2)_L$ and $U(1)_Y$, respectively. To reproduce the observed electromagnetic current, these couplings relate to the electromagnetic coupling e via the Weinberg angle θ_W (experimentally measured [**?**]), as follows:

$$e = g\sin(\theta_W) = g'\cos(\theta_W). \tag{1.2}$$

Notably, in the original formulation of the SM, right-handed neutrinos were excluded, as neutrinos were assumed to be massless. However, the observation of neutrino oscillations in 2016 [6] demonstrated that neutrinos do have mass. Extensions of the SM incorporating neutrino masses often introduce right-handed neutrinos. Since neutrinos oscillate, the mixing between weak and mass eigenstates is described by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [16]. This mixing enables neutrino flavors to change as they propagate, further highlighting the parity-violating nature of weak interactions, which act exclusively on left-handed particles.

The Glashow-Salam-Weinberg model unifies electromagnetic and weak interactions within the $SU(2)_L \times U(1)_Y$ symmetry group. However, in the SM, the electroweak and strong interactions initially treat the W and Z bosons, as well as all fermions, as massless. This assumption arises because directly including mass terms for gauge bosons and fermions fields in the Lagrangian would violate gauge invariance. However, experimental evidence contradicts this, as the W and Z bosons, along with nearly all fermions, are observed to be massive [17]. Many theoretical approaches were proposed to address this issue. The breakthrough came in 2012, when the ATLAS and CMS collaborations at CERN announced the discovery of the Higgs boson [1,2], confirming the validity of the BEH mechanism as the source of mass for observed fermions and gauge bosons. The BEH mechanism introduces a scalar field, ϕ, governed by a potential $V(\phi)$. Through the Spontaneous symmetry breaking (SSB) (Section 1.2), this scalar field breaks electroweak symmetry and generates mass for the $W^{\pm}$ and Z bosons as well as fermions [18–23], providing a consistent theoretical explanation for how these particles acquire mass while preserving gauge invariance.

A detailed discussion of the BEH mechanism and its role in generating particle masses is presented in Section 1.2.1. This framework underpins the understanding of the Higgs boson as a fundamental component of the Standard Model and a cornerstone in explaining the masses of elementary particles.

1.2. Spontaneous Symmetry Breaking

The BEH mechanism, also known as Higgs mechanism, is based on the spontaneous symmetry breaking for a complex scalar field. Particles acquire mass through the interaction with this field after the SSB. Assuming a complex scalar field, $\phi(x)$, with the form:

$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x)), \tag{1.3}$$

undergoing to the potential:

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2, \tag{1.4}$$

where the first term is "mass like", and the second one drives the self interactions of the scalar field, the lagrangian is expressed as:

$$\begin{aligned}\mathcal{L} &= T - V(\phi) = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - \mu^2(\phi^\dagger\phi) - \lambda(\phi^\dagger\phi)^2 \\ &= \frac{1}{2}(\partial_\mu\phi_1)(\partial^\mu\phi_1) + \frac{1}{2}(\partial_\mu\phi_2)(\partial^\mu\phi_2) - \frac{1}{2}\mu^2(\phi_1^2+\phi_2^2) + \frac{1}{4}\lambda(\phi_1^2+\phi_2^2)^2\end{aligned} \tag{1.5}$$

For the potential to have a finite minimum $\lambda > 0$, whereas on μ^2 sign there is no constraint. If μ^2 is chosen to be positive, the potential has a paraboloidal shape and the minimum value is for $\phi = 0$, as shown in the left panel of Fig. 1.2. On the other hand, if $\mu^2 < 0$ the minimum lays on the circumference

$$\phi_1^2 + \phi_2^2 = -\mu^2/\lambda \equiv v^2, \tag{1.6}$$

implying an infinite set of possible values, which is shown in the right panel of Fig. 1.2 [15]. In the hypothesis of $\mu^2 < 0$, the spontaneous symmetry breaking occurs

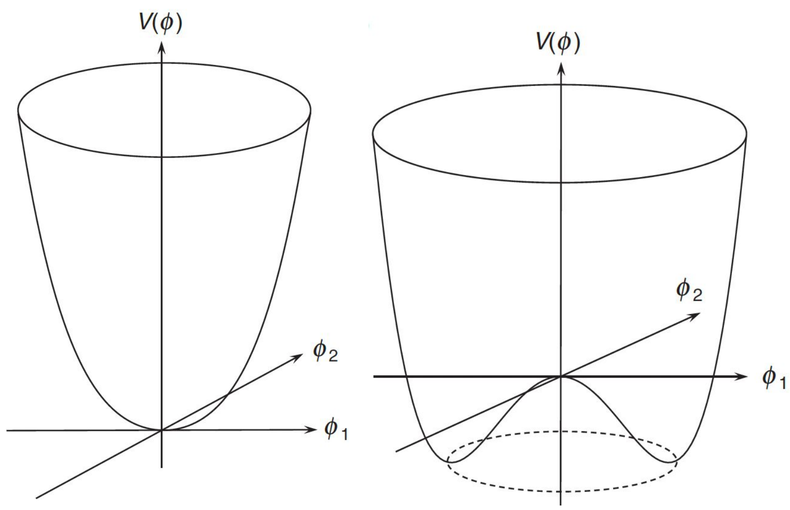


Figure 1.2: The $V(\phi)$ potential for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). [15]

by choosing any of the points in the ϕ_1, ϕ_2 plane verifying Eq. 1.6. The v quantity is called Vacuum Expectation Value (VEV). Since all points are physically equivalent in principle, for calculation simplicity it is convenient to use the point $(\phi_1, \phi_2)_{\text{min}} = (v, 0)$:

$$\phi_{\text{min}} = \frac{1}{\sqrt{2}}(\phi_{1,\text{min}} + i\phi_{2,\text{min}}) = \frac{v}{\sqrt{2}} \tag{1.7}$$

The excitations of the field, which describe physical particles, are obtained by considering perturbations of the field ϕ around the vacuum state:

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)) \tag{1.8}$$

Recomputing $\mathcal{L}$ using Eq. 1.8, substituting $\mu^2 = -\lambda v^2$ (obtained from Eq. 1.6), and simplifying, the largangian will have the form:

$$\mathcal{L} = \frac{1}{2}\left[(\partial_\mu \eta)(\partial^\mu \eta) + (\partial_\mu \xi)(\partial^\mu \xi)\right] - \lambda v^2 \eta^2 + \frac{\lambda}{4} v^4 - V_{int}(\phi) \quad (1.9)$$

The term $\lambda v^2 \eta^2$ corresponds to a massive boson raised from the vacuum state excitations, with $m_\eta = \sqrt{2\lambda v^2}$. On the other hand, the ξ corresponds to a massless boson. The term $\frac{\lambda}{4} v^4$ is a constant and it's not physically relevant. Finally, the $V_{int}(\phi)$ contains the interaction terms with triple and quartic vertices between η and ξ and self couplings, as shown in Fig. 1.3:

$$V_{int}(\phi) = \lambda v \eta^3 + \frac{\lambda}{4}\eta^4 + \frac{\lambda}{4}\xi^4 + \lambda v \eta \xi^2 + \frac{\lambda}{2}\eta^2 \xi^2 \quad (1.10)$$

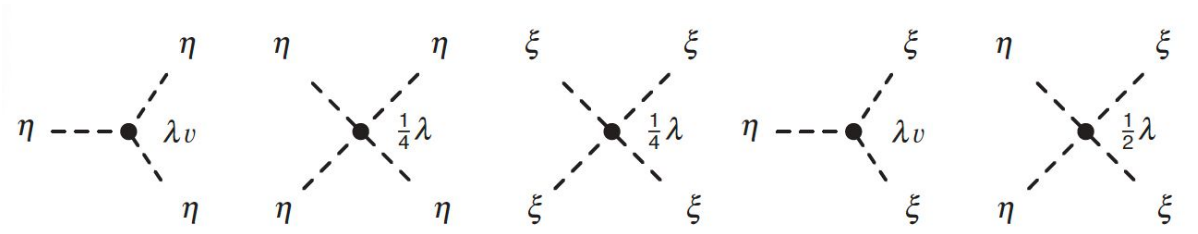


Figure 1.3: The scalar interactions obtained by breaking the symmetry for a complex scalar field [15].

When the symmetry is broken by the choice of a ground state, a massive boson and a massless one are introduced. The η excitations are in the direction where the potential is (to first order) quadratic and describe the scalar massive boson. In contrast, the particle described by the massless scalar field ξ corresponds to excitations in the direction where the potential does not change, as a consequence of the Goldstone theorem [24]: this massles scalar particle is known as Goldstone boson.

1.2.1 The Brout-Englert-Higgs Mechanism

The BEH mechanism consists in the SSB applied to a scalar field where also local invariance under the $SU(3)_L \times SU(2)_L \times U(1)_Y$ group is required [18–23] and in the finel stap an unitary gauge is applied to reabsorb the three unwanted massless Goldstone bosons arising from the symmetry breaking, giving mass to the $W^\pm$ and Z vector bosons, while the photon remains massless, in agreement with the unbroken $U(1)_{em}$ symmetry, and originating a new scalar massive particle, the Higgs boson. Considering then the interaction of the Higgs field with the fundamental fermions also the mass of the experimentally massive constituents are generated.
In the minimal Higgs model the field ϕ is a self interacting $SU(2)_L$ complex doublet with two components (with four degrees of freedom): one must be neutral, ϕ^0, and the other, ϕ^+, must be charged such that $\phi^+\phi^- = \mathbb{1}$, with $\phi^- = (\phi^+)\dagger$, to give masses to

both neutral and charged bosons.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.11}$$

With

$$\mathcal{L}_{BEH} = (\partial^\mu \phi)\dagger(\partial_\mu \phi) - V(\phi) \tag{1.12}$$

To preserve the invariance under local $SU(2)_L \times U(1)_Y$ gauge transformation, the partial derivative ∂_μ will be redefined:

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu + ig\frac{\sigma^a}{2} \cdot W^a_\mu + ig'\frac{Y}{2}B_\mu \tag{1.13}$$

Where: σ^a $(a = 1, 2, 3)$ are the three Pauli matrices, the generators of the $SU(2)$ group, g and g' are the weak couplings associated to the W_μ and B_μ gauge bosons. Finally, Y is the hypercharge defined as $Y = 2(Q - T_3)$. The hypercharge of the doublet is therefore 1.

Assuming the same expression of the potential in Eq. 1.4, being the most general $SU(3)_C \times SU(2)_L \times U(1)_Y$, and, as in the previous section, imposing $\lambda > 0$ to have a finite minimum and $\mu^2 < 0$, the minima lie on the circumference of Fig. 1.4:

$$\phi^\dagger_{min}\phi_{min} = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{-\mu^2}{2\lambda} \equiv \frac{v}{2} \tag{1.14}$$

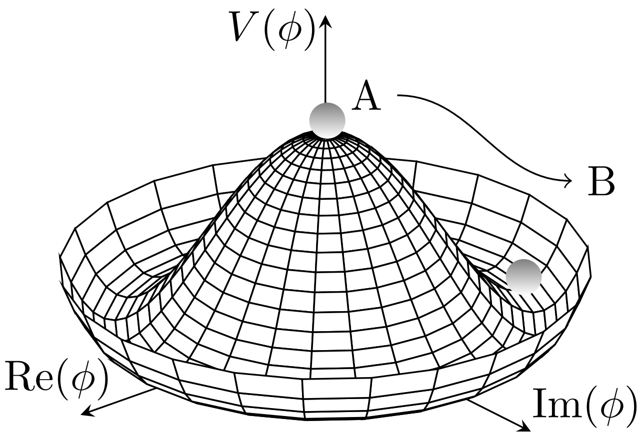


Figure 1.4: A schematic illustration of SSB via the BEH mechanism. The surface represents the Higgs potential, and the ball indicates the transition from an unstable local maximum to the continuum of ground states. [25]

When a particular vacuum is chosen for ϕ in Eq. 1.14, corresponding to $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$, the Higgs field acquires a VEV, leading to SSB. This is expressed as:

$$\phi_{min} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \qquad \text{with} \qquad v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{1.15}$$

where the parameter v represents the VEV of the Higgs field and determines the scale of electroweak symmetry breaking. This choice of vacuum state breaks the original $SU(2)_L \times U(1)_Y$ symmetry down to the electromagnetic $U(1)_{em}$ subgroup, ensuring the conservation of electric charge, while the Goldstone theorem predicts the emergence of massless bosons. Since three out of the four generators of $SU(2)_L \times U(1)_Y$ are broken, three massless Goldstone bosons arise as a consequence. The expansion of the Higgs field around the ground state of the $SU(2)_L$ doublet (Eq. 1.17) can be parametrized as:

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left(\frac{i\sigma_i}{2}\theta^i(x)\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} = U^\dagger \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.16}$$

where $H(x)$ and $\theta^i(x)$ are four real scalar fields. The $\theta^i(x)$ fields are three massless Goldstone bosons, whose existence is predicted by the Goldstone theorem. These massless fields, not observed in nature, can be removed with the following unitary gauge transformation ($U^\dagger U = \mathbb{1}$):

$$\varphi \rightarrow \phi'(x) = U\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.17}$$

After this transformation, only the real scalar field $H(x)$ remains associated to the physical particle Higgs boson and ensures that the fields appearing in the Lagrangian correspond to the physical particles.
Substituting the covariant derivative from Eq. 1.13, ϕ from Eq. 1.17, as well as the physical $W^{\pm}_{\mu}$, Z_μ fields introduced in Eq. 1.1, into Eq. 1.12, the BEH Lagrangian can be rewritten as:

$$\begin{aligned} \mathcal{L}_{BEH} &= \partial^\mu H(\partial_\mu H) - \frac{1}{2}(2\lambda v^2)H^2 \\ &+ \left[\left(\frac{gv}{2}\right)^2 W^{+\mu}W^-_\mu + \frac{1}{2}\frac{(g^2+g'^2)v^2}{4} Z^\mu Z_\mu\right]\left(1+\frac{H}{v}\right)^2 \\ &- \lambda v H^3 - \frac{\lambda}{4}H^4 + \frac{\lambda}{4}v^4 \end{aligned} \tag{1.18}$$

Higgs and vector bosons masses

The first line is the Klein-Gordon equation of the Higgs scalar field, that has a mass:

$$m_H^2 = 2\lambda v^2 = 2\mu^2 \tag{1.19}$$

Higgs mass is is a free parameter of the model, directly related to the scalar potential parameter μ, and there is no theoretical prediction for it within the SM.
The second line of Eq. 1.18 describes the interactions of the weak bosons with the

Higgs field. There are a HWW and a HZZ interactions from the 2H/v term and a HHWW and a HHZZ interaction from the H^2/v^2 term.

From the spontaneous symmetry breaking in the Higgs mechanism, the masses of the physical electroweak bosons are:

$$m_{W^\pm} = \frac{gv}{2} \qquad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2} \tag{1.20}$$

It can be noted that the A_μ field (the photon) remains massless, and this is due to the unbroken $U(1)_{em}$ symmetry. The three Goldstone bosons, eliminated by the gauge transformation (Eq. 1.16), became the longitudinal polarization components of the $W^\pm$ and Z vector bosons. Initially massless, with only transverse polarizations, these bosons acquired mass through the BEH mechanism.

Higgs self-interactions

The third line of Eq. 1.18 shows that cubic and quartic self-interactions of the Higgs boson are predicted. The BEH potential can be rewritten in terms of a trilinear and a quadrilinear coupling as:

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda_{HHH}H^3 + \lambda_{HHHH}H^4 - \frac{1}{4}\lambda v^4, \tag{1.21}$$

with

$$\lambda_{HHH} = \lambda v = \frac{m_H^2}{2v} \tag{1.22}$$

and

$$\lambda_{HHHH} = \frac{\lambda}{4} = \frac{m_H^2}{8v^2} \tag{1.23}$$

And from Eq. 1.19 the expression of λ is:

$$\lambda = \frac{m_H^2}{2v^2} \tag{1.24}$$

Higgs boson self-interactions are directly linked to the scalar potential parameters and fully determined by the Higgs mass and VEV, making their precise measurement a key test of the Standard Model. Unlike weak boson self-interactions, which arise from gauge symmetry, Higgs self-interactions originate purely from the scalar sector and are responsible for the Higgs boson mass. The BEH potential includes a constant term that, while irrelevant in the SM, contributes to vacuum energy and is linked to the cosmological constant governing the Universe curvature. However, the SM prediction for this constant contradicts astronomical observations, suggesting the need for a quantum theory of gravity or a mechanism to reduce the Higgs vacuum energy.

Fermions masses

Fermions have been assumed to massless until now. The spontaneous symmetry breaking allows also fermions to acquire mass through interactions with the Higgs field: the Yukawa interactions.
As for bosons, also fermionic mass terms are not invariant under $SU(2)_L \times U(1)_Y$, as the mass term would have the form:

$$-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \tag{1.25}$$

Due to the chiral nature of the $SU(2)_L$ symmetry group, the transformation properties of left- and right-handed chiral states are different, and hence Eq. 1.25 is not gauge invariant.
The fermions obtain mass via the interactions with the Higgs field, known as Yukawa interactions:

$$\mathcal{L}_{Yukawa} = -\hat{h}_{d_{i,j}}\bar{q}_{L_i}\phi d_{R_j} - \hat{h}_{u_{i,j}}\bar{q}_{L_i}\phi_C u_{R_j} - \hat{h}_{\ell_{i,j}}\bar{\ell}_{L_i}\phi e_{R_j} + h.c. \tag{1.26}$$

where: $\phi_C = i\sigma_2\phi^\dagger$, q_L, ℓ_L are the quark and lepton $SU(2)_L$ doublets and u_R, d_R, e_R are the $SU(2)_L$ right handed singlets of up and down type quarks and charged leptons, and h.c. stands for hermitian conjugate. Finally, the $\hat{h}_{d_{i,j}}$, $\hat{h}_{u_{i,j}}$ and $\hat{h}_{\ell_{i,j}}$ are 3×3 matrices in the quarks and leptons family phase space[5]. It is interesting to notice that the leptonic term equivalent to $\hat{h}_{u_{i,j}}\bar{q}_{L_i}\phi_C u_{R_j}$ is absent. This is because the ν_R transforms as singlet in all the known interactions, so with the current knowledge and energies reached at colliders, there is no proof of its existance.

After the spontaneous symmetry the scalar doublet and its conjugate have the form:

$$\phi(x) = \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \qquad \phi_C(x) = \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix}, \tag{1.27}$$

then the Eq. 1.26 will contain mass terms and interactions between fermions and the Higgs boson. The mass terms have the form:

$$M_{d_{ij}}\bar{d}_{L_i}d_{R_j} + M_{u_{ij}}\bar{u}_{L_i}u_{R_j} + M_{\ell_{ij}}\bar{\ell}_{L_i}\ell_{R_j} \tag{1.28}$$

Where $M_{f_{ij}}$, with $f = u, d, \ell$, are non-diagonal mass matrices: $M_{f_{ij}} = \hat{h}_{f_{ij}} \cdot v/\sqrt{2}$. To have the physical terms corresponding to fermion masses, a rotation is applied to the $M_{(u,d,\ell)_{ij}}$ matrices that diagonalizes also the Higgs-fermions interactions. In this way, $M_{f_{ij}} \to y_i\delta_{ij}$ and the fermions will acquire mass defined as:

$$m_{f_i} = \frac{y_i v}{\sqrt{2}} \tag{1.29}$$

Where y_i are defined as *Yukawa couplings* which also quantifies the strength of the

[5] The indices $i, j = 1, 2, 3$ refer to the three families of the up/down type quarks or charged leptons.

fermions interactions with the Higgs boson.
The experimental measurements evidence a large gap between the the fermion masses, hence the range of couplings values with the Higgs boson is remarkably large, spanning from $y_t \sim (1)$ for the top quark to $y_{\nu_e} < 10^{-11}$, without any specific reason. One of the most relevant limitations of the SM is that the couplings are not deterministic but come directly from the experimental measurements and limits on the particle masses, as discussed in Section 1.3.

Higgs boson interactions

Expanding Eq. 1.18, the terms containing Higgs couplings with fermions, bosons and self coupling:

$$\mathcal{L} = -g_{Hff}\bar{f}fH + \lambda_{HHH}vH^3 + \frac{\lambda_{HHHH}}{4}H^4 + \delta_V V_\mu V^\mu \left(g_{HVV}H + \frac{g_{HHVV}}{2}H^2\right) \tag{1.30}$$

with λ_{HHH} and λ_{HHHH} defined respectively in Eq. 1.22 and Eq. 1.23, and:

$$g_{Hff} = \frac{m_f}{v}, \quad g_{HVV} = \frac{2m_V^2}{v}, \quad g_{HHVV} = \frac{2m_V^2}{v^2}$$

where $V = W^{\pm}, Z$ and $\delta_V = 1$ for $W^{\pm}$and 1/2 for Z.
As a result, the dominant mechanisms for Higgs boson production and decay involve the coupling of H to W, Z and/or the third generation quarks and charged leptons. For Higgs bosons production at hadronic colliders, one of the main production process is the gluon fusion (Section 1.4). The Higgs interactions with gluons at Leading Order (LO) are induced by a one-loop process where the H couples to a virtual $t\bar{t}$ pair (with minor contribution from lighter quarks). Likewise, the Higgs boson couplings to photons is also generated via loops, although in this case the one-loop diagram involves a virtual W^+W^- pair.

Stability of the Higgs boson potential

The BEH mechanism has two free parameters: the VEV v, which sets the electroweak symmetry-breaking scale and the Higgs boson mass m_H. The VEV can be determined using the experimental value of the Fermi constant and the Eq. 1.20 for the mass of the charged vector bosons. From the expression of the Fermi constant it is possible to get the value of v:

$$\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8M_W^2} \implies v^2 = \frac{1}{\sqrt{2}G_F} \tag{1.31}$$

The G_F value is determined with a precision of 0.6 ppm from muon decay measurements [26], and substituting it in Eq. 1.31 one gets $v \simeq 246\,\text{GeV}$. If the VEV is obtained by the G_F measurement, on the self-coupling parameter, λ (defined in Eq. 1.24), the only constraint is the requirement of $\lambda > 0$ to have a finite minimum, but there is no prediction for it, and therefore also for the Higgs mass value. The discovery of the Higgs boson with $m_H \simeq 125\,\text{GeV}$ (Section 1.4.1) has crucial implications for the SM, especially for understanding the stability of the Higgs potential at high energy scales. From the experimentally determined values of the VEV and m_H, one gets $\lambda \simeq 0.13$ and $|\mu| \simeq 88.4\,\text{GeV}$.

At energies above the electroweak scale ($E > 246\,\text{GeV}$), the electroweak symmetry remains unbroken, and the gauge bosons W^i_μ, and B_μ are massless, indicating a unified electroweak interaction. Below this scale, the BEH mechanism spontaneously breaks this symmetry, distinguishing the weak and electromagnetic interactions. The weak bosons ($W^\pm, Z$) acquire mass, while the photon (A_μ) remains massless, reflecting the unbroken $U(1)_{em}$ symmetry. The physical $W^\pm, Z$, and γ bosons are linear combinations of the gauge fields W^i_μ and B_μ: At low energies, the weak and electromagnetic interactions manifest differently, consistent with experimental observations [27]. The unification of these interactions within the electroweak framework remains one of the greatest successes of the SM.

1.3. Limits of the Standard Model

The SM is often assumed to be valid up to the Planck scale ($M_{Planck} \sim 10^{19}$GeV) due to its remarkable consistency with experimental observations and its ability to describe particle interactions with high precision across a wide range of energy scales. However, the SM has several limitations:

- It does not account for gravitational interactions, which are described by general relativity. The quantization of gravity has proven to be non-renormalizable in a perturbative sense, making it incompatible with the SM framework.
- The SM lacks an explanation for baryogenesis, leaving the observed matter-antimatter asymmetry in the Universe unexplained.
- The description of particles and interactions provided by the SM accounts of only $\sim 5\%$ of the total matter and energy in the Universe, addressing only the particles and their interactions directly observed. The remaining portion, composed of dark matter and dark energy, falls outside the scope of the SM.
- There is no theoretical justification for the existence of three fermion families and for the wide range of coupling constants and masses observed among these families. In fact these SM parameters (masses, couplings, charges) are not predicted by the theory but must be determined experimentally, with no insight into the implications of their values.
- The observation of neutrino oscillations in 2016 [6] demonstrated that neutrinos do have mass [7]. Like quarks, neutrinos exhibit mixing between their weak interaction states and mass eigenstates, governed by the PMNS matrix [16]. This leads to the oscillation phenomenon, which changes the flavor of neutrinos. So to account for neutrino oscillations, the PMNS matrix is included, but the theoretical basis for this inclusion remains unclear.
- In the SM the Higgs boson mass, m_H, is another free parameter. There is no theoretical prediction for its value. Additionally, m_H is highly sensitive to

quantum fluctuations involving extremely large energies, leading to the hierarchy problem, discussed in Section 1.3.1.

These unsolved problems of the SM suggest that it may not be the definitive theory of particle physics. BSM theories could address these limitations, preserving the SM predictive successes while explaining new phenomena at higher energies.

1.3.1 The hierarchy problem and naturalness

The cutoff scale is an energy threshold beyond which a physical theory is no longer considered valid or complete. It marks the energy where new physics, corrections, or unknown phenomena are expected to emerge, requiring a more fundamental description. The SM cutoff scale is set to the Planck scale ($\Lambda \sim 10^{19}$ GeV). This is a pragmatic choice based on its theoretical robustness and experimental success, but it also serves as a framework to probe the limits of the current state-of-art physics understanding and search for new physics.

The *hierarchy problem* highlights a profound inconsistency between the small observed m_H value and the large quantum corrections it should receive due to contributions from virtual particles at high energy scales. Specifically, these corrections scale quadratically with the cutoff energy Λ, leading to contributions as large as Λ^2. This creates a severe fine-tuning problem, as the bare Higgs mass must cancel these large corrections almost exactly to yield the physical mass.

So either nature involves an extremely precise fine-tuning of parameters to cancel these corrections — a scenario that is considered unnatural - or there exists new physics at or near the TeV scale to regulate these corrections.

The concept of *naturalness* requires that a theory should not involve extreme fine-tuning of parameters. This implies that the quantum corrections to m_H should not vastly exceed its observed value.

Several extensions to the SM aim to address the hierarchy problem while restore naturalness. The Super symmetry (SUSY) [28] introduces superpartners for all SM particles, which contribute opposite-sign quantum corrections to the Higgs mass. This can cancel the quadratic dependence on Λ^2, leaving only logarithmic corrections, and the Higgs mass is stabilized naturally if superpartners exist at the TeV scale, making them potentially observable at the LHC. In some models, the problem is addressed by introducing a WED [29]. The large scale difference between the Planck scale and the TeV scale is explained geometrically via the warp factor. Composite Higgs models assume that the Higgs is not an elementary particle but a bound state of more fundamental constituents. This can alleviate the sensitivity to high-energy scales.

If no new physics exists near the TeV scale, it raises the possibility that fine-tuning is a fundamental feature of nature, or that alternative frameworks, like anthropic reasoning in the multiverse, might explain the small Higgs mass [17].

These considerations motivate the searches at LHC of new physics at the TeV scale. In this thesis, the proposed search for new resonances decaying in Higgs boson pairs explore the WED scenario.

1.4. Higgs boson phenomenology

1.4.1 Higgs boson discovery

The BEH mechanism received experimental confirmation in 2012 with the discovery of the Higgs boson by the ATLAS and CMS collaborations at the LHC. This discovery, based on data collected in proton-proton (pp) collisions at center-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV during LHC Run 1 (2010-2013), relied on analyses targeting different H decays, particularly the two-photon ($H \to \gamma\gamma$) and four-lepton ($H \to ZZ \to 4\ell$) final states. The experimental measurements by the ATLAS and CMS experiments of the invariant masses of the decay products in these channels provide clean signatures of the presence of the Higgs boson as a clear excess over background in correspondance of $m_{\gamma\gamma/4\ell} \sim 125\,\text{GeV}$, as shown in Fig. 1.5. The significance level of these measurements are 5.9σ and 5σ respectively for ATLAS and CMS experiments, leading to the announcement of the Higgs boson discovery on July, 4th 2012[6] [1,2].

Following its discovery, precise measurements of the Higgs boson mass (m_H) and properties were performed. A combined measurement at the end of Run 1 yielded $m_H = 125.09 \pm 0.21\,(\text{stat}) \pm 0.11\,(\text{syst})\,\text{GeV}$ [30]. Additional studies confirmed its spin-parity as $J^P = 0^+$ [31,32], and showed that its production cross-section, predominant decay modes and couplings to fermions and bosons align minimal scalar sector expected in the SM. A comprehensive summary of Run 1 measurements can be found in [31,33,34]. ATLAS and CMS performed several other measurements using Run 2 data (2015-2018), but an overall combination of all available measurements with these data across the two experiments has not yet been performed.

1.4.2 Single Higgs production at LHC

The left panel of Fig. 1.7 shows the SM Higgs boson production cross sections for the dominant processes as a function of $\sqrt{s}$ at the LHC. These proceses, in decreasing order of cross section, are:

- The **Gluon-Gluon Fusion (ggF)**, $gg \to H + X$, the blue line of Fig. 1.7 (left). The LO diagram of this process is shown in Fig. 1.6(a). It occurs via the exchange of a virtual top quark, with contributions from other quarks being suppressed. It has the largest cross section at LHC energies, with $\sigma^{\text{ggF}}_{\text{N3LO}} = 48.6^{+2.22}_{-3.27}\,\text{pb}$ at $\sqrt{s} = 13\,\text{TeV}$ for $m_H = 125\,\text{GeV}$ [35].
- **Vector Boson Fusion (VBF)**, $qq \to qq + H$, where qq can represent both quarks and antiquarks (red line of Fig. 1.7 left). The LO diagrams are shown in Fig. 1.6(b). Although its cross-section is smaller than that of ggF, VBF has a distinctive signature, as the scattered quarks appear as two high-energy hadronic forward jets (Section 3.3.4) with a large spatial separation and invariant mass. This

[6] The term "discovery" is a technical definition implying the provided measurement has a significance level $> 5\sigma$, corresponding to a $p-$value $< 3 \times 10^{-7}$.

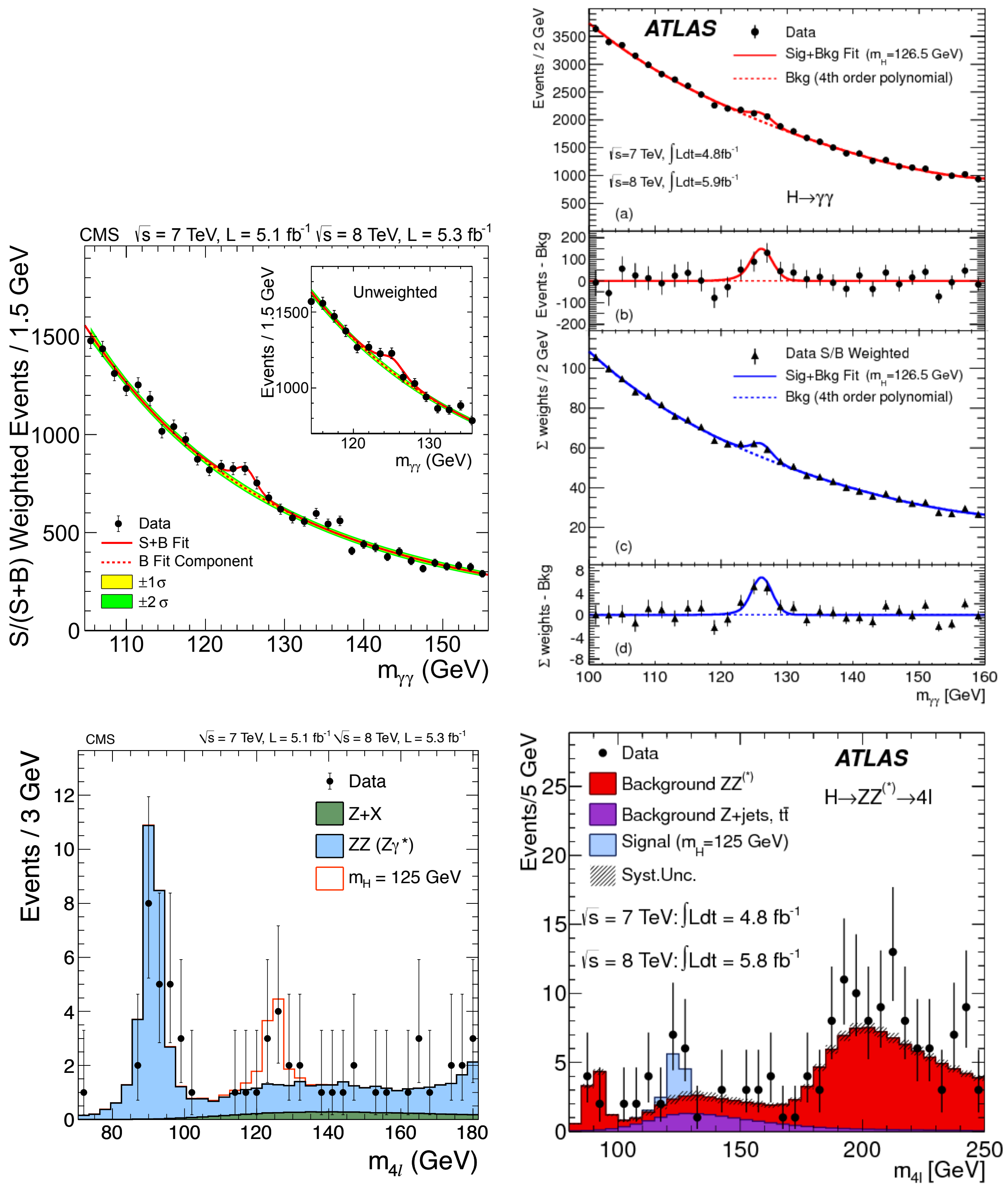


Figure 1.5: The diphoton invariant mass distribution with weighted by $S/(S+B)$ and unweighted events (upper panels), and the four-lepton invariant mass distribution for the $ZZ \to 4\ell$ analysis compared to background expectations (lower panels) obtained by CMS (left) and ATLAS (right). [1,2]

peculiar pattern aids in separating the VBF H production from background processes when performing experimental data selection and analysis.

- **Associated Production with a Vector Boson (VH)**, green and brown lines in Fig. 1.7 (left): in this mode, the Higgs boson is produced in association with a weak gauge boson, $V = W$ or Z. The LO Feynman diagram of this process is shown in Fig. 1.6(c).

- **Top-Quark Associated Production ($t\bar{t}H$)**, dark purple line in Fig. 1.7 (left). The LO diagram of this rare process is shown in Fig. 1.6(e): it gives access to direct measurement of top-Higgs Yukawa coupling (y_t).
- **Bottom-quark associated production ($b\bar{b}H$)**, pink line of Fig. 1.7 (left). The LO diagram of this rare process in shown in Fig. 1.6(e): it is suppressed by the b quark mass, but at high energies, logarithmic corrections increase the cross section. Moreover, in some BSM models where y_b is enhanced, it could become the main production mode.
- **Single Top-Quark Associated Production (tHq)**, liliac line in Fig. 1.7 (left). As for the $t\bar{t}H$ process, also the tHq production gives access to y_t, providing valuable insights on the y_t sign. This is due to the nearly complete destructive interference between two significant contributions: one where the Higgs boson couples to a space-like W boson (Fig. 1.6(f)) and another where it couples to the top quark (Fig. 1.6(g)).

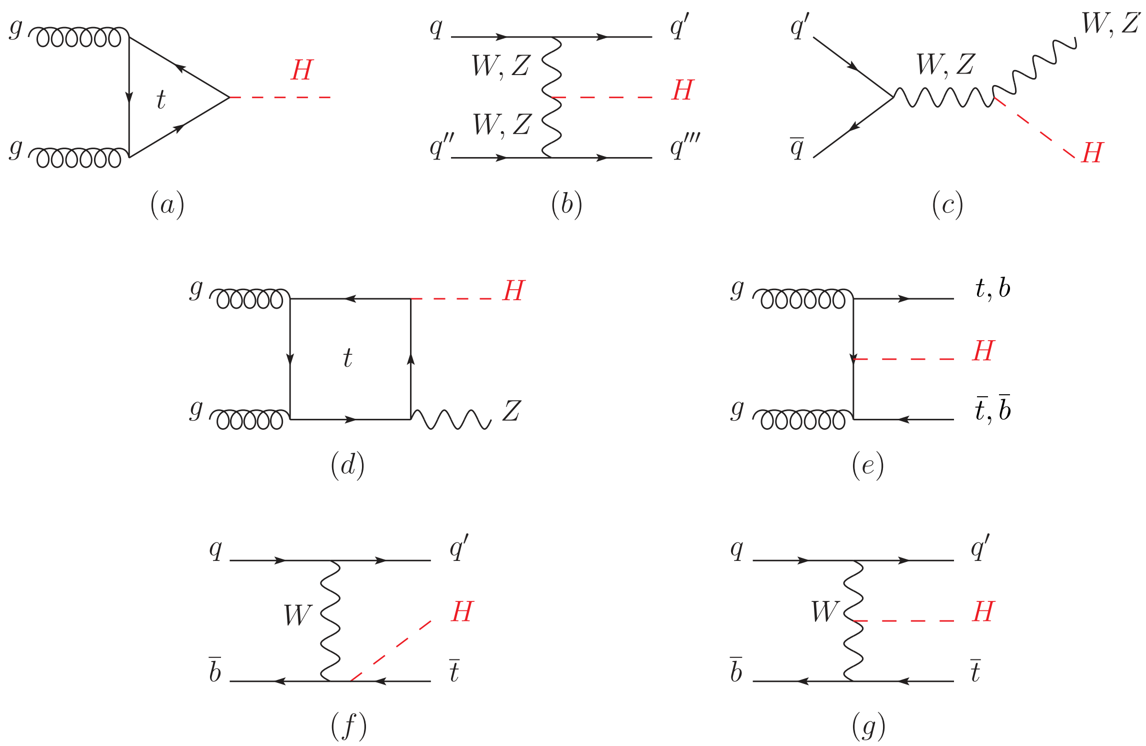


Figure 1.6: Main leading order Feynman diagrams contributing to the single Higgs boson production: (a) ggF, (b) VBF, (c) VH from $q\bar{q}$, (d) loop-induced VH from gluons, (e) $t\bar{t}H/b\bar{b}H$ and (f,g) tHq. [17]

1.4.3 Higgs decays

The Higgs boson couplings with SM particles can be investigated by considering the different Higgs boson decay modes. The Higgs boson directly couples to all massive

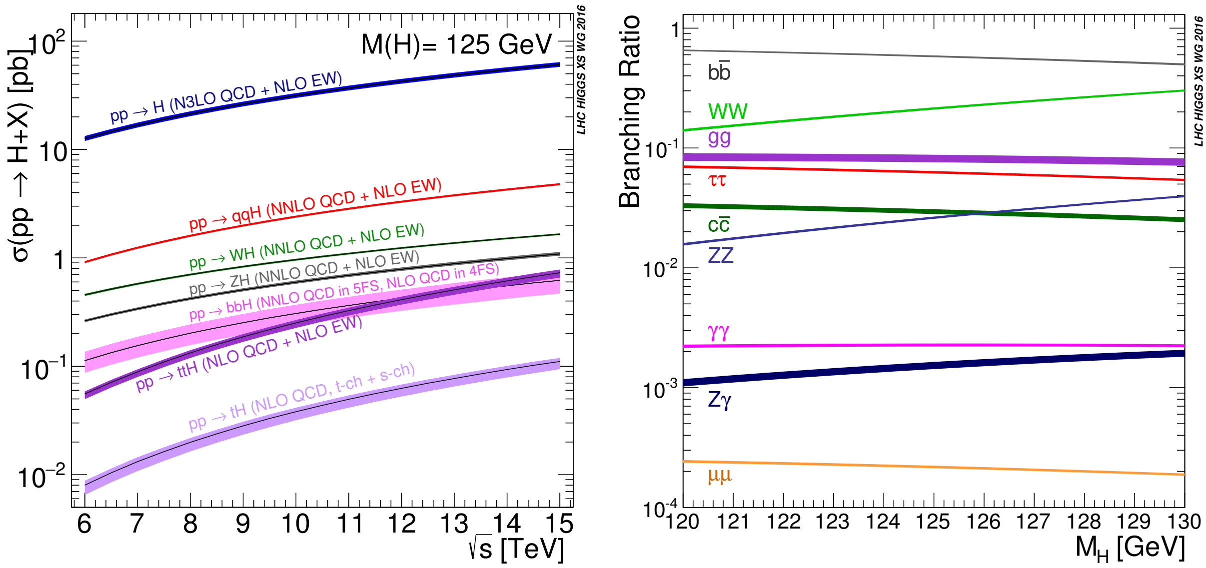


Figure 1.7: Left: SM Higgs boson production cross section as a function of the pp centre of mass energy, $\sqrt{s}$, for different production mechanisms. In each curve the order of perturbation theory is indicated. Right: branching ratios for the main decays of the Higgs boson as a function of its mass, near the experimentally measured value of $m_H = 125\,\text{GeV}$ [36].

SM particles and to massless particles (gluons and photons) via intermediate loops. The right panel of Fig. 1.7 shows the Higgs branching fraction as a function of the Higgs boson mass, m_H. The main decay modes of the Higgs boson are the H → b$\bar{\text{b}}$, with the largest Branching Ratio (BR) (as the direct $H \to t\bar{t}$ decay is not kinematically allowed for $m_H = 125\,\text{GeV}$), $H \to VV^*$ (with V=W$^{\pm}$,Z), with one of the two bosons produced off-shell, $H \to gg$, H $\to \tau^+\tau^-$, $H \to c\bar{c}$. Decays with smaller BRs involve $H \to \gamma\gamma$, $H \to \mu^+\mu^-$, $H \to Z\gamma$. The decays involving gluons and photons in the final state are loop induced at LO, thus they provide indirect information on the Higgs boson couplings to VV and t$\bar{\text{t}}$ in different combinations. The most relevant decay modes and their corresponding BR with uncertainty are summarized in Tab. 1.2. Fig. 1.8 illustrates the Higgs boson couplings to fermions and bosons as measured by the CMS collaboration, as a function of the mass of each particle [37]. All currently measured values are consistent with SM predictions.

1.5. Higgs bosons pair production

In Section 1.2.1, the Higgs boson self-coupling and its implications for potential stability are discussed, in the hypothesis that the Higgs boson potential that induces the SSB is the SM one. Any modification of the potential could impact the potential stability. A direct measurement of the Higgs boson self-coupling is essential to determine the exact form of the BEH potential, as well as to explore potential new physics beyond the SM and to understand the Higgs field role in high-energy phenomena [38].

Decay mode	Branching Ratio (%)	Rel. uncertainty
$H \to \bar{b}b$	5.82×10^{-1}	$^{+1.2\%}_{-1.3\%}$
$H \to WW^*$	2.14×10^{-1}	$\pm 1.5\%$
$H \to \tau^+\tau^-$	6.27×10^{-2}	$\pm 1.6\%$
$H \to ZZ^*$	2.62×10^{-2}	$\pm 1.5\%$
$H \to \bar{c}c$	2.89×10^{-2}	$^{+5.5\%}_{-2.0\%}$
$H \to \gamma\gamma$	2.27×10^{-3}	$\pm 2.1\%$
$H \to Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \to \mu^+\mu^-$	2.18×10^{-4}	$\pm 1.7\%$

Table 1.2: Higgs decays branching ratios'' and the relative uncertainty for a SM Higgs boson with $m_H = 125\,\text{GeV}$ [17].

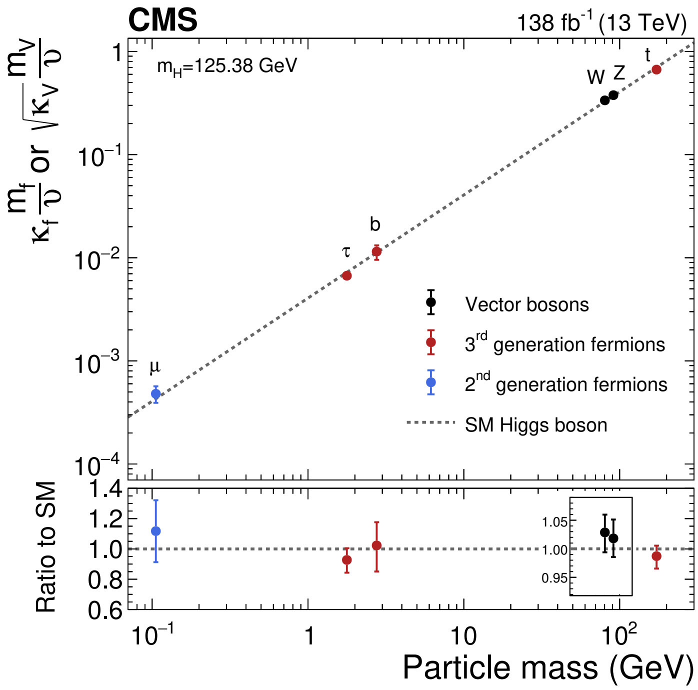


Figure 1.8: A summary of the CMS measurements of the Higgs boson couplings to the other fundamental particles, with the predictions by SM indicated by the dashed black line. In the lower panel, the ratio between the measured coupling and the standard model prediction is shown [37].

1.5.1 Standard model HH production at LHC

At LHC the rate of Higgs pair production is approximately 1000 times smaller than single Higgs one, making a direct measurement extremely challenging. The cross sections for the main double Higgs production mechanisms as a function of the pp center-of-mass energy ($\sqrt{s}$) are shown in Fig. 1.10, and the values of the cross sections at LHC for $\sqrt{s} = 13\,\text{TeV}$ are shown in Tab. 1.3.

The dominant Higgs pair production process in the SM is the **Gluon-gluon Fusion (ggF)**, the red line in Fig. 1.10, which at LO is induced through a quark loop, primarily involving top quarks with smaller contributions from bottom quarks. This process includes box and triangle diagrams, respectively left and right diagrams in Fig. 1.9(a), with the latter involving the Higgs trilinear coupling λ_{HHH}. The two LO diagrams undergo to destructive interference, impacting the invariant mass distribution of the Higgs pair, as shown in Fig. 1.11.
The double Higgs production via **Vector Boson Fusion (VBF)**, the green line in Fig. 1.10, has a smaller cross section than the ggF one. Nevertheless, it is easier to experimentally identify due to its experimental distinctive signature, including two high-transverse momentum component (p_T) boosted quarks, which in the detectors correspond to two forward and well separated hadronic jets (see Section 3.3.4). As shown in Fig. 1.9(b), this mechanism includes at LO contributions where two Higgs bosons radiate off virtual $W^{\pm}$ or Z bosons, and diagrams where an off-shell Higgs boson decays into a Higgs pair.
The **Double Higgs bremsstrahlung off top quarks** ($t\bar{t}HH$), whose LO diagrams are reported in in Fig. 1.9(d), has a very low cross section at LHC energies but increases with energy, as shown in Fig. 1.10 (pink line).
The **Double Higgs-strahlung** (VHH) refers to the associated production of Higgs pairs with a $W^{\pm}$ or Z boson. This process has a significantly smaller cross section than ggF and VBF (purple lines in Fig. 1.10). The LO Feynman diagrams are shown in Fig. 1.9(c).

1.5.2 Double Higgs production beyond the SM

In Section 1.3 the limitations of the SM have been discussed, as well as the motivations to search for new physics. This thesis investigates the double Higgs production, as in SM context it allows direct access to the Higgs self coupling, λ, and at the same time it serves as a promising probe for many BSM scenarios.
From an experimental point of view, the deviations in the self coupling can manifest as anomalies in production rates and/or in the production cross section. The experimental searches focus on two different types of BSM signatures in Higgs pair production, expressed as **resonant** and **non-resonant** production.

In the resonant HH production, a new heavy particle, X, decays into a Higgs bosons pair, creating a distinctive peak in the HH system invariant mass distribution if $m_X > 2m_H$. This process, for some parameters values, could lead to an enhancement in the Higgs pair production cross section and rate in correspondance of the resonance, compared to SM predictions, making the anomaly potentially detectable even at LHC energies and with data collected up to now. Many BSM models can be tested, from the Minimal Supersymmetric Standard Model (MSSM) to the WED, which is the model this thesis focuses on.

On the other hand, the non-resonant double Higgs production tests BSM models predicting modifications to Higgs couplings arising from extended symmetries or new processes that induce changes in quantum loops in the HH production. Experimentally, this reflects in changes of the kinematics of the process and in the enhancement of the

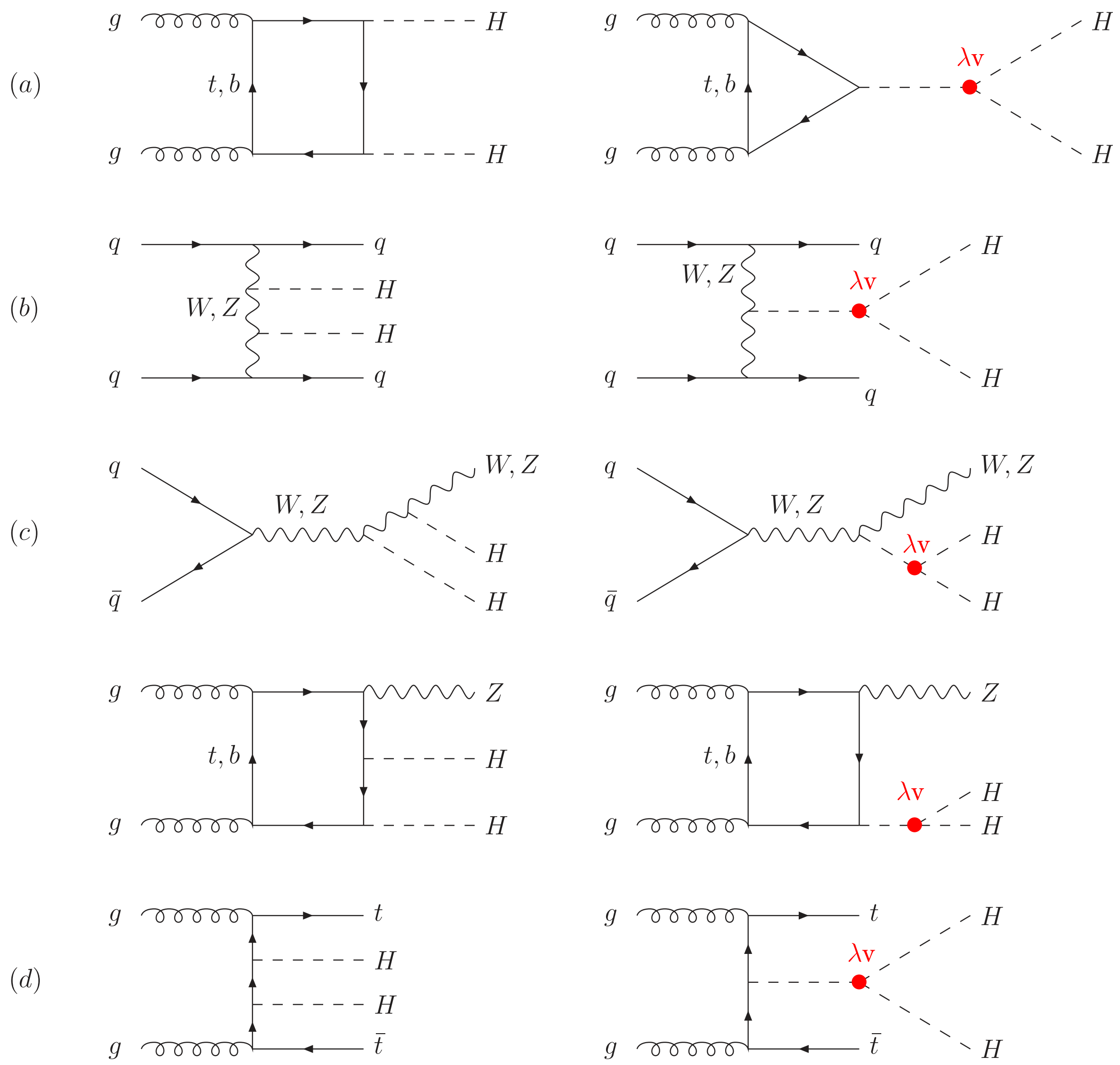


Figure 1.9: Main leading order diagrams contributing to Higgs pair production: (a) ggF, (b) VBF, (c) VHH from $q\bar{q}$ (d) loop-inducd VHH from gluons via top quarks. The trilinear Higgs coupling contribution is marked in red [39].

HH production rate.

1.5.3 Non resonant pair production

Effective Field Theory (EFT) approaches describe the non-resonant scenario, modifying or adding new Higgs interactions without introducing new particles [39]. These interactions may include contact terms affecting the production signatures observed at experiments. The central concept of the EFT approach is that interactions of any complexity, occurring at short distances, can be systematically approximated by a lagrangian with a finite set of parameters. This lagrangian offers an “effective” description of any underlying physical model. Specifically for HH production, the lagrangian is based on the SM one, supplemented with corrections represented by the

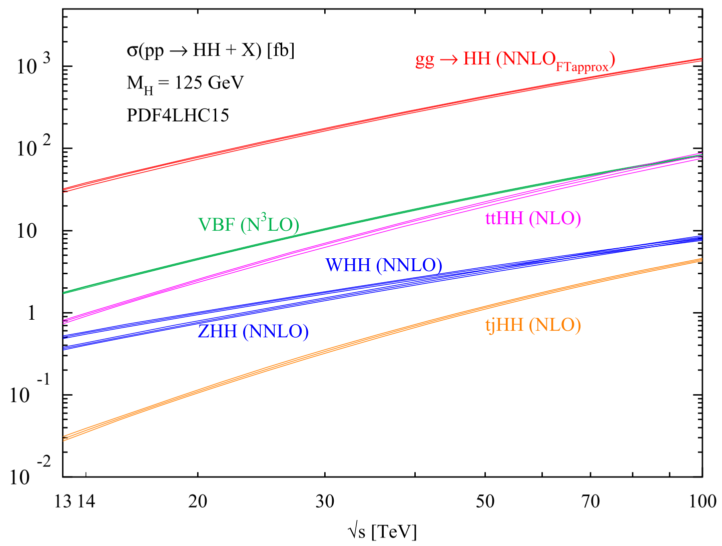


Figure 1.10

Process	σ (pb) ($\sqrt{s} = 13$ TeV)
ggF HH	$31.05^{+2.2\%}_{-5.0\%} \pm 3.0\%$
VBF HH	$1.73^{+0.03\%}_{-0.04\%} \pm 2.1\%$
ZH H	$0.363^{+3.4\%}_{-2.7\%} \pm 1.9\%$
$W^{+}HH$	$0.329^{+0.32\%}_{-0.41\%} \pm 2.2\%$
$W^{-}HH$	$0.173^{+1.2\%}_{-1.3\%} \pm 2.8\%$
$ttHH$	$0.775^{+1.5\%}_{-4.3\%} \pm 3.2\%$
$tjHH$	$0.0289^{+5.5\%}_{-3.6\%} \pm 4.7\%$

Table 1.3

Fig. 1.10: Production cross sections for Higgs pairs within the SM via gluon fusion, vector-boson fusion, double Higgs-strahlung and double Higgs bremsstrahlung off top quarks. PDF4LHC15 parton densities have been used with the scale choices according to Tab. 1.3. The size of the bands shows the total uncertainties originating from the scale dependence and the PDF+α_s uncertainties. Tab. 1.3: Signal cross sections (in fb) for Higgs bosons pair (HH) production including the available QCD corrections according to the recommendations of the LHC Higgs Cross Section Working Group [39].
Tab. 1.3: Signal cross sections (in fb) for HH production with uncertainties. The λ is defined in Eq. 1.24.

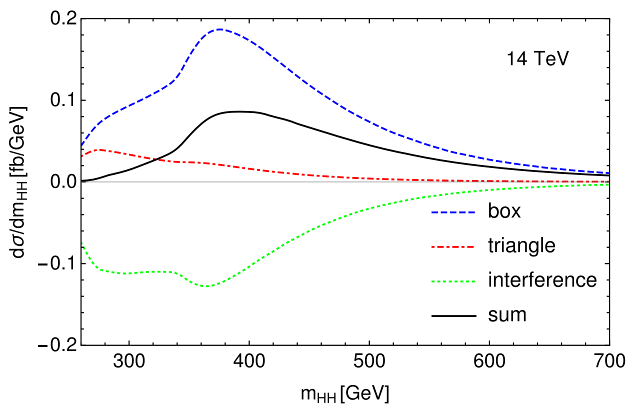


Figure 1.11: Higgs pair invariant mass distribution at leading order for the different contributions to the gluon fusion production mechanism and their interference. [39]

addition of local operators.

In a renormalizable lagrangian, all terms are operators with dimension $d \leq 4$. Hence, the effects of BSM physics can be locally approximated by extending the SM lagrangian to include higher-dimensional operators ($d > 4$). These additional operators are suppressed by powers of the BSM interaction scale M to lower their dimensions and make them renormalizable. Given that dimension-8 (and higher) operators contribute minimally due to power suppression, and dimension-5 and dimension-7 operators involve lepton number violation and neutrino mass terms (and can therefore be disregarded for LHC processes), only dimension-6 operators $\mathcal{O}_i^6$ are relevant [40]. Thus, the EFT lagrangian takes the form:

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{M^2} \mathcal{O}_i^6 \tag{1.32}$$

where c_i are the Wilson coefficients, which parameterize the contributions of BSM physics. This yields a finite set of parameters describing the most general modifications to the SM at short distances.
The EFT lagrangian can be re-expressed in terms of Higgs boson couplings, and the relevant terms for ggF HH production are:

$$\begin{aligned} \mathcal{L}_{HH} = & \frac{1}{2}(\partial_\mu H \partial^\mu H) - \frac{1}{2} m_H^2 H^2 - \kappa_\lambda \lambda v H^3 \\ & - \frac{m_t}{v}\left(v + \kappa_t H + \frac{c_2}{v} H^2\right)(\bar{t}_L t_R + \text{h.c.}) \\ & + \frac{\alpha_s}{12\pi v}\left(c_g H - \frac{c_{2g}}{2v} H^2\right) G^a_{\mu\nu} G^{a,\mu\nu}, \end{aligned} \tag{1.33}$$

Bringing the introduction of three new BSM contact interactions at the gHH (c_g), $ggHH$ (c_{2g}), and $ttHH$ (c_2) vertices, while modifying the Higgs trilinear coupling (κ_λ) and the top Yukawa coupling (κ_t). Deviations from the SM can be expressed as the ratio of the modified coupling to the SM expectation: $\kappa_\lambda = \lambda/\lambda^{\text{SM}}$ for the trilinear coupling and $\kappa_t = y_t/y_t^{\text{SM}}$ for the top Yukawa coupling. These five new parameters (κ_λ, κ_t, c_2, c_g, and c_{2g}) describe the lower-level interactions shown in Fig. 1.12, where modified couplings are highlighted in red [40]. Similarly, other new parameters arise when expanding the lagrangian for other HH production mechanisms (as VBF, ttHH ...).

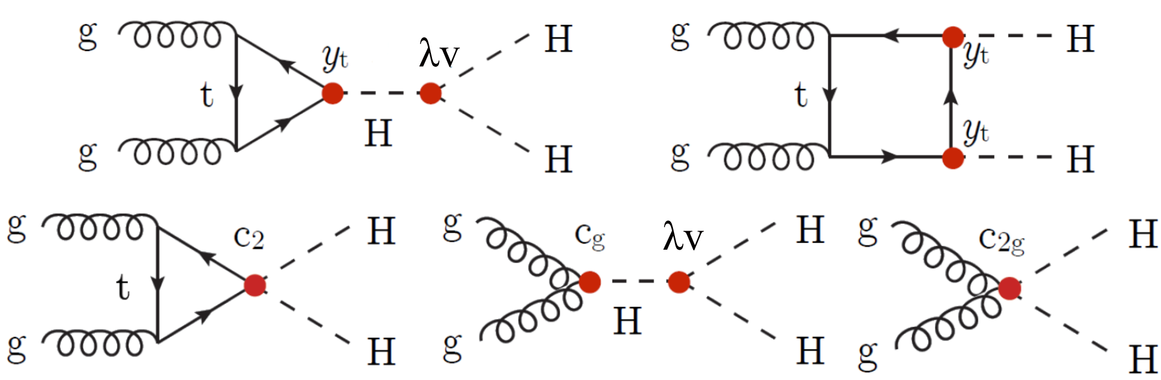


Figure 1.12: Feynman diagrams that contribute to non resonant Higgs boson pair production via ggF at LO. The two upper diagrams correspond to SM-like processes, while in the bottom row the diagrams correspond to pure BSM effects: (c) exploits the contact interaction of two Higgs bosons with top quarks, (d) and (e) describe contact interactions between the Higgs boson and gluons. [41]

To simplify the challenge of exploring all possible combinations of the EFT parameters in experimental searches, a "shape benchmarks" strategy is utilized by the ATLAS and CMS experiments. This approach classifies regions of the five-dimensional parameter space with similar kinematic features by analyzing approximately 1500 sample points on a grid. From this analysis, 12 representative benchmark shapes are defined, each corresponding to distinct kinematic topologies. This method enables a model-independent investigation of non-resonant Higgs boson pair (HH) production, accommodating various BSM scenarios, such as those involving extended scalar sectors, vector-like quarks, or composite Higgs models. Further details are provided in Ref. [41].

Even when focusing on a single EFT parameter, such as the trilinear Higgs coupling (λ_{HHH}), valuable insights can be gained. For example, variations in κ_λ, the normalized trilinear coupling, affect both the HH production cross section and its kinematics. These variations lead to a minimum cross section at $\kappa_\lambda = 2.45$, as illustrated in Fig. 1.13, assuming all other EFT parameters are fixed to their SM values. Furthermore, κ_λ influences the invariant mass and transverse momentum distributions of the Higgs boson pairs, which are critical observables for experimental analyses.

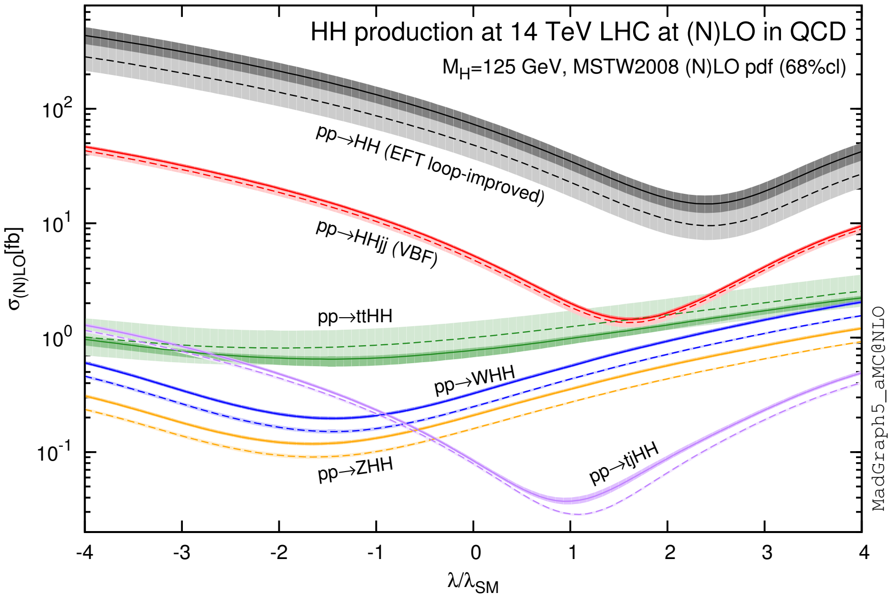


Figure 1.13: Total cross sections at the LO and NLO in glsxtrshortQCD for HH production channels, at $\sqrt{s} = 14\,\text{TeV}$ LHC as a function of the self-interaction coupling λ. The dashed (solid) lines and light- (dark-)colour bands correspond to the LO (NLO) results and to the scale and PDF uncertainties added linearly. The SM values of the cross sections are obtained at $\lambda/\lambda^{SM} = 1$. [42]

1.5.4 Resonant pair production

This thesis presents an experimental search for resonant double Higgs production through s-channel resonances that decay to SM Higgs bosons. The study focuses the WED framework, specifically in the search of two predicted heavy resonances: the

spin-2 graviton and the spin-0 radion. WED models are well-motivated extensions of the SM, providing a natural mechanism to address several key problems, including the hierarchy problem (see Section 1.3.1).
By stabilizing the extra dimension, the WED framework explains the separation of physical scales with the geometry of the fifth dimension, where the localization of fields within the extra dimension determines their respective physical scales in the 4D effective theory. The 5D WED fields can be expressed in terms of Kaluza Klein (KK) modes, arising as massive excitations in the compactified extra dimension. Their masses are determined by the geometry and compactification of the 5D space. The first modes have masses naturally around the TeV-scale, resolving the Planck-weak hierarchy problem. This ensures the KK modes play a role in stabilizing Electroweak symmetry breaking (EWSB) and the possibility of being in the kinematic range allowed for searches at LHC with a centre-of-mass energy of 13 TeV.
Moreover, the TeV-scale masses regulate quantum corrections to the Higgs boson mass, preventing it from being destabilized by large radiative corrections from high-scale physics.

Basics of the WED framework

The five-dimensional (5D) spacetime in WED has coordinates $x^M = (x^\mu, y)$, where $\mu = 0, 1, 2, 3$, $M = 0, \ldots, 5$, and y is the 5th dimension coordinate [7]. To avoid experimental detection of the fifth dimension, y must be finite and it is compactified on an orbifold $S^1/\mathbb{Z}_2$, connecting two 4-dimensional subspaces within the $(4+n)$-dimensional space time, also known as "3-brane". Therefore the finite 5th dimension has a 3-brane at each endpoint. The two 3-branes are defined as the UV brane (or Planck brane) at $y = 0$, and the IR brane (or TeV brane) at $y = \pi R$, which is the "visible brane". The 5D bulk consists of a warped anti-de Sitter (AdS_5) space [43]. In Fig. 1.14 an illustration of the WED geometry is illustrated.

The AdS_5 geometry is driven by the presence of a negative bulk cosmological constant, $\Lambda_5 < 0$ in the solution of Einstein's equations.This configuration allows gravity fluctuations (gravitons) to be expanded in KK modes on the TeV brane, further refining the relationship between Planck and electroweak scales.

The most general solution of the the 5D Einstein field equations for this configuration yields a warped metric:

$$ds^2 = g_{MN} dx^M dx^N = e^{-2\sigma} \eta_{\mu\nu} dx^\mu dx^\nu + dy^2 \tag{1.34}$$

where $\eta_{\mu\nu}$ is the 4D Poincarè invariant Minkowski metric $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$, $\sigma = k|y|$, k is the AdS_5 space curvature, $k \sim \mathcal{O}(M_{Planck})$, and y is the 5th dimension coordinate, spanning from 0 to $L = \pi r_C$. The 5D metric is not factorizable, but rather the four-dimensional metric is multiplied by a "warp" factor which is a rapidly changing function of an additional dimension, e^{-2ky}. This warp factor effectively

[7] In this section, the capital Latin letters will be used to indicate the 5D coordinate indices, whereas the Greek letters are used for the 4D coordinate indices.

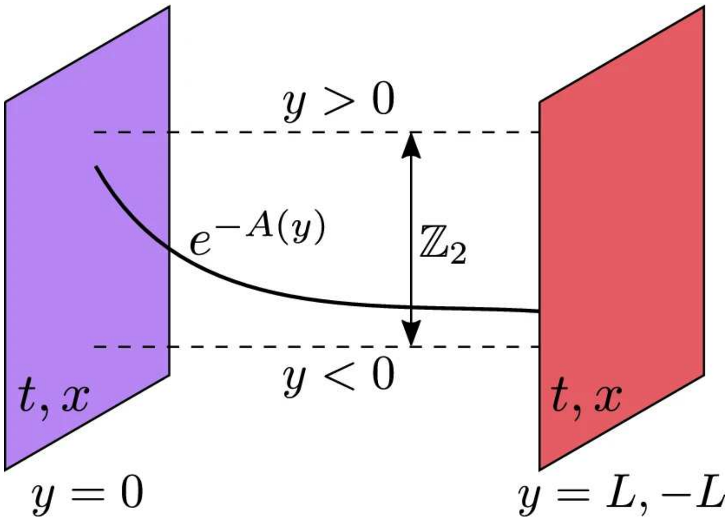


Figure 1.14: Schematic representation of the extra dimension in the Randall-Sundrum model. The left plane is the UV "hidden" 3-brane, and the red one is the IR "visible" 3-brane. [44]

generates two distinct physical scales: the Planck scale M_{P} on the UV brane and a lower scale $M_{\mathrm{P}}e^{-kL}$ on the IR brane (where k is on the order of M_{P}). The compactification scale r_C is chosen such that the warp factor e^{-kL} brings the Planck scale on the UV brane down to the TeV scale on the IR brane [29], thus providing a natural solution to the hierarchy problem by creating a separation of physical scales across the extra dimension. A summary of the branes definition is provided in Tab. 1.4.
For example, the choice of $kR \sim 12$ bring the the electroweak scale to emerge on the IR brane at approximately the TeV scale.

Brane	UV	IR
y position	$y = 0$	$y = L \equiv \pi r_C$
Scale	Planck	TeV
4D spacetime	hidden	visible

Table 1.4: Summary of the 3-branes in the WED model.

The action can be written as:

$$\begin{aligned} \mathcal{S} = -\int d^4x \int_{\pi r_C}^{+\pi r_C} dy \sqrt{-g} \Big[\Lambda_5 + \frac{1}{2} M_5^3 \mathcal{R} + \delta(y)(\Lambda_{(0)} - \mathcal{L}_{(0)}) \\ + \delta(y - \pi r_C)(\Lambda_{(\pi r_C)} - \mathcal{L}_{(\pi r_C)}) \Big] \end{aligned} \tag{1.35}$$

Where M_5 is the 5D Planck scale and $\mathcal{R}$ is the five-dimensional Ricci scalar, constructed from the five-dimensional metric g_{MN}. Denoting the trace of the 5D metric as $g = -g^X_X$, the volume unity is $\int d^5x\sqrt{-g}$. Finally, Λ_5, $\Lambda_{(0)}$ and $\Lambda_{(\pi r_C)}$ are the cosmological constants in the bulk and on the branes, with $\Lambda_{(0)} = e^{-k\pi r_C}\Lambda_{(\pi r_C)}$ [29].

The metric solution in Eq. 1.34, is valid if the bulk and boundary cosmological constants are related by:

$$\Lambda_{(0)} = -\Lambda_{(\pi r_C)} = -\frac{\Lambda_5}{k} \tag{1.36}$$

Obtaining:

$$k = \sqrt{\frac{-\Lambda_5}{24M_5^2}} \tag{1.37}$$

Here, M_5 is the 5D Planck mass [29].

The quantum fluctuations around the classical metric solution give rise to 4D tensor (spin-2 particle) $\otimes$ two 4D vectors (spin-1 particles) $\otimes$ one scalar component, corresponding to a spin-0 particle:

$$\delta g_{MN}(x,\phi) = \begin{pmatrix} h_{\mu\nu}(x,\phi) & h_{\mu,5}(x,\phi) \\ h_{\mu,5}(x,\phi) & h_{55}(x,\phi) \end{pmatrix} \tag{1.38}$$

For the purposes of this thesis, the axial gauge is chosen, defined by fix $h_{\mu 5} = 0$, decoupling the dynamics of tensor and scalar perturbations. Moreover, the contribution from spin-1 particles is not relevant for double Higgs production for spin conservation and for theoretical arguments related to boundary conditions on the 3-branes [45]. The perturbed space-time metric on on this gauge has the form:

$$ds^2 = e^{-2(\epsilon_r h_{55} + \sigma(\phi, r_C))}\left(\eta_{\mu\nu} + \epsilon_g h_{\mu\nu}(x,\phi)\right)dx^\mu dx^\nu + r_C^2\left(2\epsilon_r h_{55}(x,\phi)+1\right)^2 d\phi^2. \tag{1.39}$$

The tensor fluctuations $h_{\mu\nu}$ correspond to Graviton modes and the scalar fluctuation (h_{55}) correspond to Radion mode. By expanding the gravitation fluctuations, $h_{\mu\nu}(x,y)$, in terms of KK modes:

$$h_{\mu\nu}(x,y) = \sum_{n=0}^{\infty} \chi_{\mu\nu}^{(n)}(x)\xi_n(y), \tag{1.40}$$

each 4D expansion mode is known as the n-th KK mode, corresponding to the wave function of a particle in the effective four dimensional theory. The 5th dimensional part, ξ_n is defined as *profile* of the n-th mode. The profiles for all the known fields are described by a combination of exponential and Bessel functions [29].

The zero mode of the Graviton field correspond to the mediator of gravitational force, and its profile is $\xi^0(y) \propto e^{-ky}$, localizing the 4D graviton on the UV brane.

Unlike the zero mode there is no freedom to delocalize the KK (nonzero) modes away from the IR brane, thus the KK modes for all types of bulk fields are always localized near the IR brane. This means that the coupling of an individual KK excitation to matter or to other gravitational modes is set by the weak scale [29].

The KK gravitos have masses given approximately by $m_n \sim nke^{-\pi k r_C}$ TeV for $n = 1, 2, \ldots$. The first KK mode ($n = 1$) typically has a mass around the TeV scale, reflecting the IR scale.

The h_{55} mode describes small fluctuations in the separation between two branes and is effectively represented by a massless particle called the "radion." This mode is associated with the presence of the IR brane, persisting even if the UV brane is removed. To stabilize the inter-brane separation L at $r_C \approx 50/k$ without significant fine-tuning, a stabilization mechanism is required. The radion corresponds to variations in the size of the extra dimension and acquires mass via the inter-brane stabilization mechanism.

Two different scenarios

Depending on the specific scenario within the WED framework, SM particles are either confined to the IR brane (Randall-Sundrum scenario [46,47]) or allowed to propagate across the fifth dimension ("Bulk" scenario [48,49]). The propagation probability of the fermions, bosons and KK-gravitons are represented in Fig. 1.15, respectively in the RS1 scenario (left panel) and in the Bulk scenario (right panel). The bulk scenario is particularly compelling as it addresses both the Planck-weak hierarchy problem and the fermion masses hierarchy.

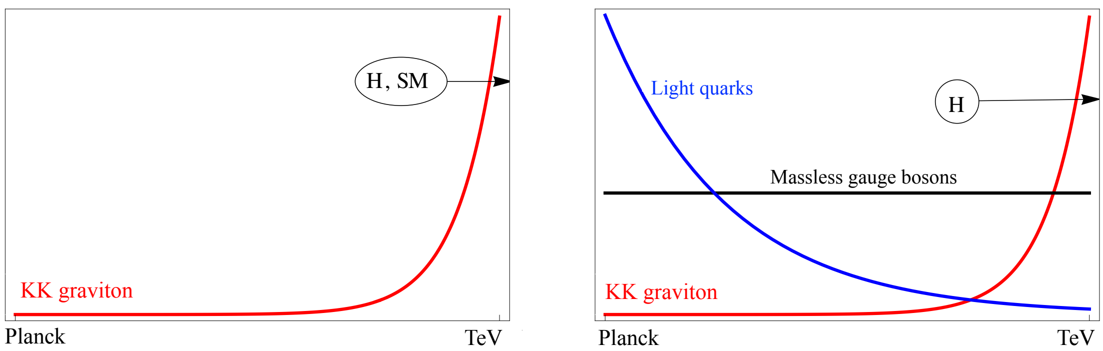


Figure 1.15: Localization of fields on the branes, in different scenarios: RS1 (left) and bulk (right). The x-axis represents the 5th dimension with the Planck brane on the left and the TeV brane on the right. The y-axis is the probability density. [29,50].

The SM is now fundamentally 5-dimensional [45,51], but its lightest modes appear as the familiar 4D SM particles, with phenomenological properties deriving from their extra-dimensional wavefunctions.
Fig. 1.16 illustrates this setup, highlighting the warped geometry, UV and IR branes, bulk fields, and hierarchical localization of SM particles.

The Higgs boson resides near the IR brane, where the energy scale is warped down to the TeV range as $m_H \propto M_5 e^{-kL}$. This configuration naturally resolves the Planck-weak hierarchy problem through the exponential warp factor, which explains the large disparity between the Planck and weak scales.

The SM fermions exhibit hierarchical localization in the extra dimension, with heavier fermions such as the top quark positioned closer to the IR brane. This localization reflects their larger Yukawa couplings, determined by the overlap of their wavefunctions with the Higgs boson near the IR brane. Similarly, massive KK modes of fermions, gauge bosons, and gravitons are localized near the IR brane, with their

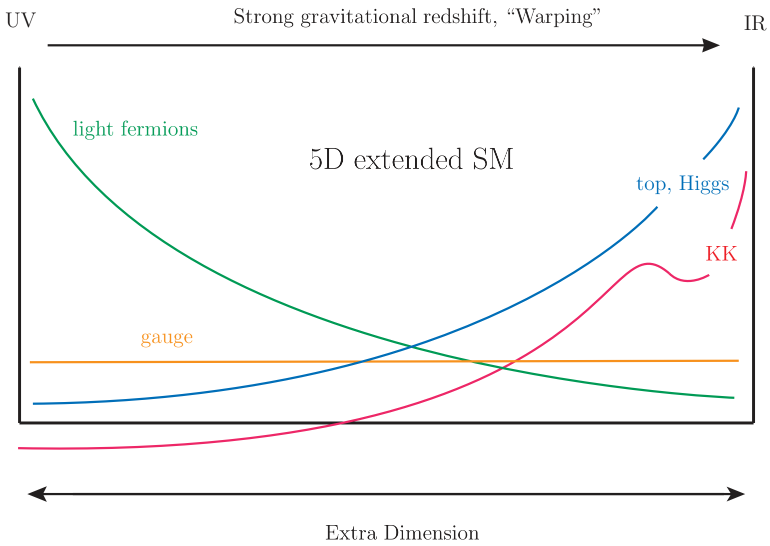


Figure 1.16: Warped extra dimensional model with SM fields in bulk (standard framework). Schematic shapes of extra-dimensional wavefunctions for various particles (zero modes and a generic KK mode) are shown [52].

masses set by the TeV-scale energy of the IR brane. From a 4D perspective, the apparent weakness of gravity arises from the small overlap of the graviton wavefunction with the IR brane.
This setup leads to unique phenomenological signatures, such as KK gravitons appearing as spin-2 resonances with TeV-scale masses, detectable at LHC through their decay products. The radion, a scalar field associated with fluctuations in the size of the extra dimension, also has a TeV-scale mass determined by the stabilization mechanism, making it accessible in collider experiments [29].

This framework provides a robust mechanism for the SM flavor structure, with fermion wavefunctions naturally generating hierarchical Yukawa couplings. The AdS/CFT correspondence offers a dual interpretation, linking this structure to the mechanism of partial compositeness, further enhancing the theoretical appeal of the bulk scenario.

The free parameters for this theory include the graviton and radion masses (m_{G^*}, m_R) and the warping parameters (r_C or k).

WED Phenomenology at LHC

In the following subsections, predictions for the production cross sections and decay branching ratios of the bulk graviton (or simply graviton) (G^*) and radion (R) are given, with benchmark values of $kL = 35$ and $\Lambda_R = 3$ TeV. These values have been chosen since the results can be rescaled for other parameter choices using analytic formulas [29]. The mass range considered for these predictions spans between 200 GeV and 6.5 TeV, which is $E_{CM}/2$ for the LHC centre-of-mass energy considered in this thesis (13 TeV).

Graviton production cross section and decays

In Fig. 1.17, the cross sections for the graviton at LHC with a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ are plotted as functions of the graviton mass in both the RS1 and bulk scenarios. The QCD-induced production mechanisms considered include gluon fusion (GF) and quark fusion (QF). Electroweak (EW) processes include photon fusion (PF) and production associated with two prompt jets (Xjj). Among these, gluon fusion (GF) is the dominant production channel for the graviton (G^*), and it is the production mechanism explored in this thesis.

The uncertainty from the parton distribution functions (δ_{PDF}) becomes significantly large for $m_{G^*} > 3\,\text{TeV}$.

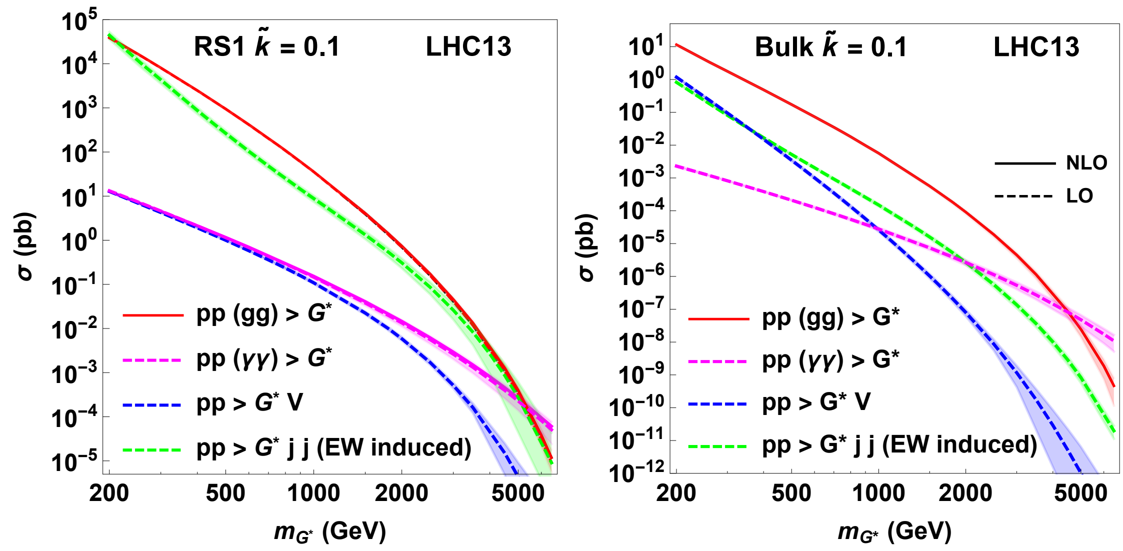


Figure 1.17: Production cross sections in pb for the KK-graviton with $k/M_{Pl} = 0.1$ in the RS1 (left) and Bulk (right) scenarios. The red curves correspond to the inclusive production, the green ones to the associated production with two jets, blue for associated production with a vector bosons and the magenta ones to photonfusion production. The NLO calculations are shown in continuous lines while the LO in dashed. When the NLO QCD results is available the uncertainty bands correspond to that calculation. [29]

The branching fractions for G^* decays to SM particles are displayed in Fig. 1.18 for both the RS1 (left) and bulk (right) scenarios. The thick curves correspond to the case of a fully elementary top quark[8], while the thin dashed curves assume no coupling of G^* to the top quark [29].
The localization of the SM matter fields for the graviton resonance impacts the kinematics of the signal and drastically modifies its production and decay properties [53], so that a distinction of the RS1 and bulk models is necessary for G.
In RS1, with all the SM fields localized on the TeV brane, a heavy graviton would decay to a wide range of final states with significant branching fractions as shown in the left panel of Fig. 1.18, and constraints on the RS1 model are mainly obtained from

[8]The elementary top quark hypothesis assumes right-handed top quark localized very near the TeV brane.

fermionic channels [54].

In the bulk scenario, G^* predominantly decays into pairs of massive particles, such as gauge bosons and top quarks, particularly when $m_{G^*} > 1\,\text{TeV}$. This behavior motivates searches for bulk gravitons that couple to pairs of Higgs bosons. The maximum branching fraction to a pair of Higgs bosons is below 10% under the hypothesis of an elementary H boson, as shown inthe right panel of Fig. 1.18. Accordingly, the HH final state is usually not the most important one for placing constraints on the bulk model, where the largest sensitivity arises from searches in WW, ZZ, or tt signatures. However, the branching fraction to HH can reach 25% if the top quark coupling becomes small, such that investigations of HH signatures are necessary in the context of bulk models, because the branching fractions are very model dependent.

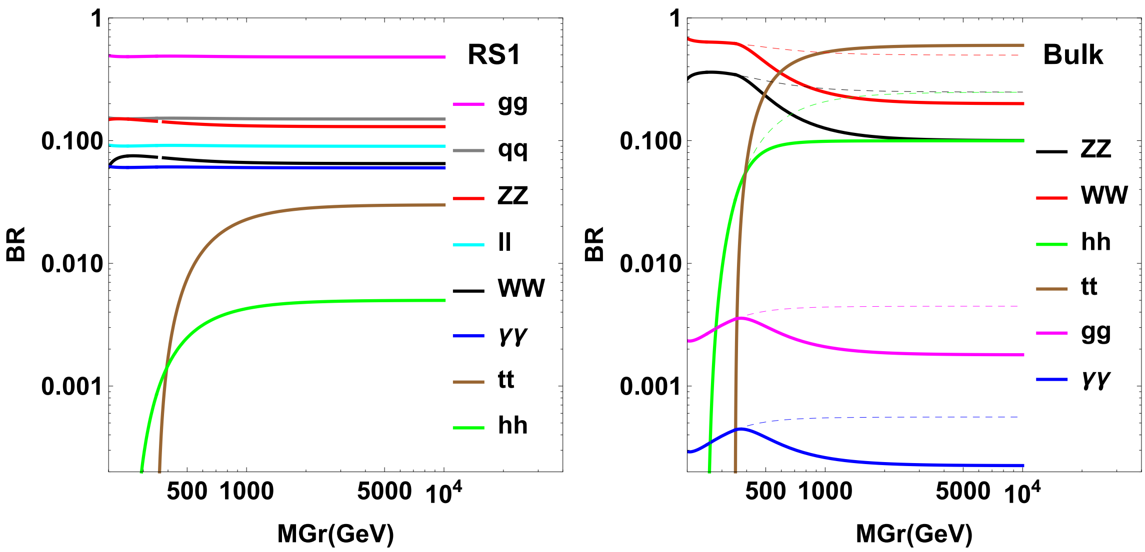


Figure 1.18: Branching fractions of G*. Left: RS1 scenario. The symbol q stands for the sum of light quarks (u; d; s; c; b), while l represents the sum of the three flavors of leptons (e; μ; τ) or neutrinos. Right: Bulk scenario comparing two hypothesis of fermion embedding. The branching ratios are independent of k parameter. [29]

Radion production cross section and decays

The radion is predicted to be the lightest particle in certain scenarios, exhibiting similarities to a heavy Higgs boson under the narrow width approximation. Thus the production and decay behavior of the radion can often be studied using techniques developed for heavy Higgs bosons. However, specific production mechanisms—such as VBF and vector boson exchange (VX)—reveal distinct differences in its behavior.

Differently from the gravitons, the physics of the radion depends only very weakly on the choice of the model [55], which obviates the need to distinguish the RS1 and bulk possibilities for the radion. More details on WED models can be found in [29].

Fig. 1.19 presents the production cross section for the radion, assuming $kL = 35$ and $\Lambda_R = 3\,\text{TeV}$, in both the bulk (solid lines) and RS1 (dashed lines) scenarios. With

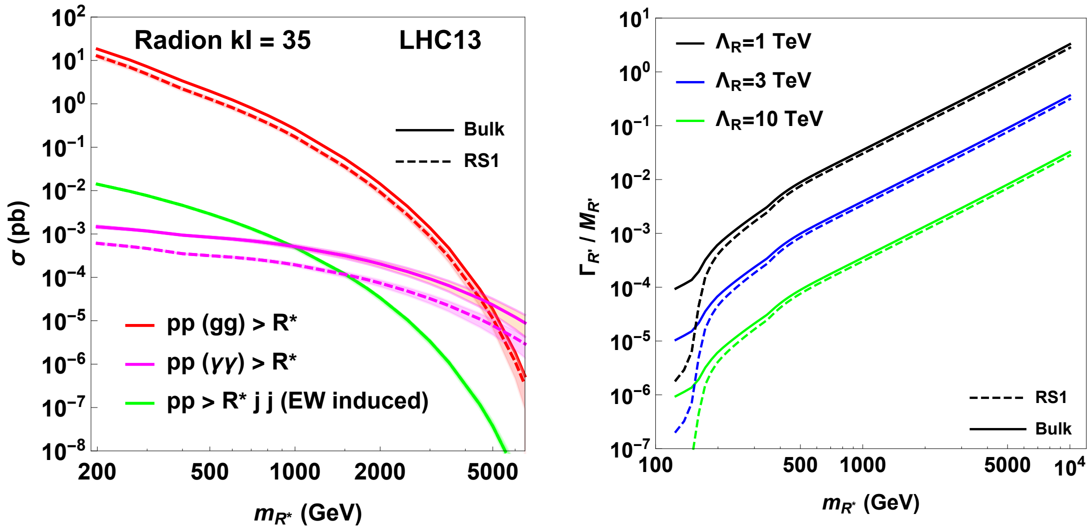


Figure 1.19: Left: Production cross sections in picobarns for the Radion, with $kL = 35$ and $\Lambda_R = 3\,\text{TeV}$. The GF (red) and VBF (Green) production modes for both to bulk (continuous lines) and RS1 (dashed lines) scenarios, for LHC at $\sqrt{s} = 13\,\text{TeV}$. Right: Radion total width and branching fractions. The continuous lines stand for the RS1 scenario while the dashed ones for the bulk scenario [29]

these fixed parameters, bulk terms significantly impact radion production: they double the gluon fusion (GF) production rate across the entire mass spectrum, while leaving VBF production unaffected. This suggests only a minor influence on the differential distributions. In this thesis the production mode considered for Radion is the GF.

Regarding decay characteristics, the radion has been extensively studied [55–58], particularly in terms of its total width and branching ratios. As shown in Fig. 1.19, the total width Γ_R is inversely proportional to Λ_R^2 but the individual branching ratios are not sensitive to it.

In Fig. 1.20, the individual branching fractions for the radion are shown, using $kL = 35$ as a benchmark. Generally, the radion exhibits a strong preference for decays into massive particles, with branching fractions to bosons increasing significantly with radion mass. In contrast, decays to fermions are relatively suppressed at higher radion masses, with decays to top quark pairs becoming particularly rare for $m_R > 1\,\text{TeV}$.

The branching ratios demonstrate a low sensitivity to variations in the bulk parameter kL, indicating stability in the decay patterns across different model setups. The dominant R decay modes are into pairs of massive gauge bosons, H bosons, and top quarks, as shown in Fig. 8 (right). Since the couplings are determined by the masses of the final-state particles, and these masses arise from the H boson localized on the TeV brane, the RS1 and bulk couplings are identical at LO. For large resonance mass m_X, the corresponding widths are

$$\Gamma(R \to HH) = \Gamma(R \to ZZ) = \frac{\Gamma(R \to WW)}{2} = \frac{1}{32\pi} \frac{m_X^3}{\Lambda_R^2} \tag{1.41}$$

which can be compared to

$$\Gamma(R \to t\bar{t}) = \frac{3}{8\pi}\left(\frac{m_t}{m_\chi}\right)^2 \frac{m_\chi^3}{\Lambda_R^2} \tag{1.42}$$

For large radion masses, the branching fraction to HH is approximately 25%, independently on Λ_R because the contribution from decays to tt is suppressed by $(m_t/m_X)^2$. This makes the $R \to$ HH channel an important one in the search for a radion resonance [50].

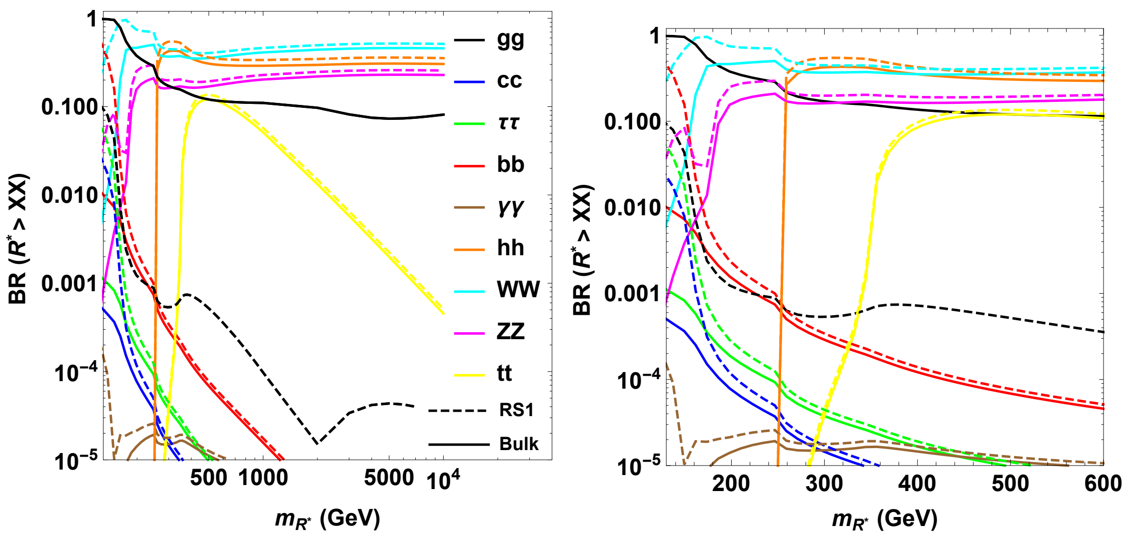


Figure 1.20: Radion branching fractions to bosons and fermions. The dashed lines stand for RS1 scenario, while the continuous ones for the bulk. The figure on the right is a zoom of the figure on the left, following the same legend. [29]

.

1.6. State of art of Single Higgs measurements

After 10 years of the Higgs boson discovery, the CMS and ATLAS collaborations accumulated a huge statistics during two LHC runs, Run 1 (2010-2013 at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$) and Run 2 (2015-2018, at $\sqrt{s} = 13\,\text{TeV}$).
The results with Run 2 data are reported in the 10 years ATLAS and CMS papers [37,59]. The relevant points of these updated results with higher energy and statistics are:

- The possibility to increase the precision in the already measured channels, as shown in Fig. 1.21, where the distribution of the decay products of the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels are shown respectively in the left and right panels [37].
- The inclusion of new unexplored channels, as the $H \to WW^*$ in Fig. 1.22a and the evidence of $H \to Z\gamma$ combining ATLAS and CMS measurements, in Fig. 1.22b [37,60].
- The first measurements or evidences for $H \to f\bar{f}$, where $f = \tau, b, \mu$ [37].

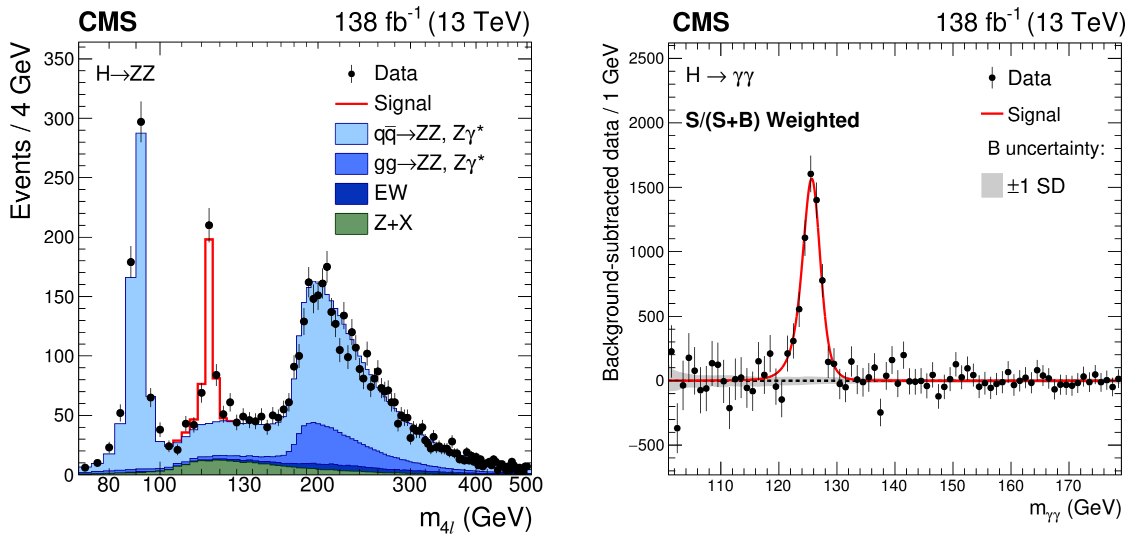
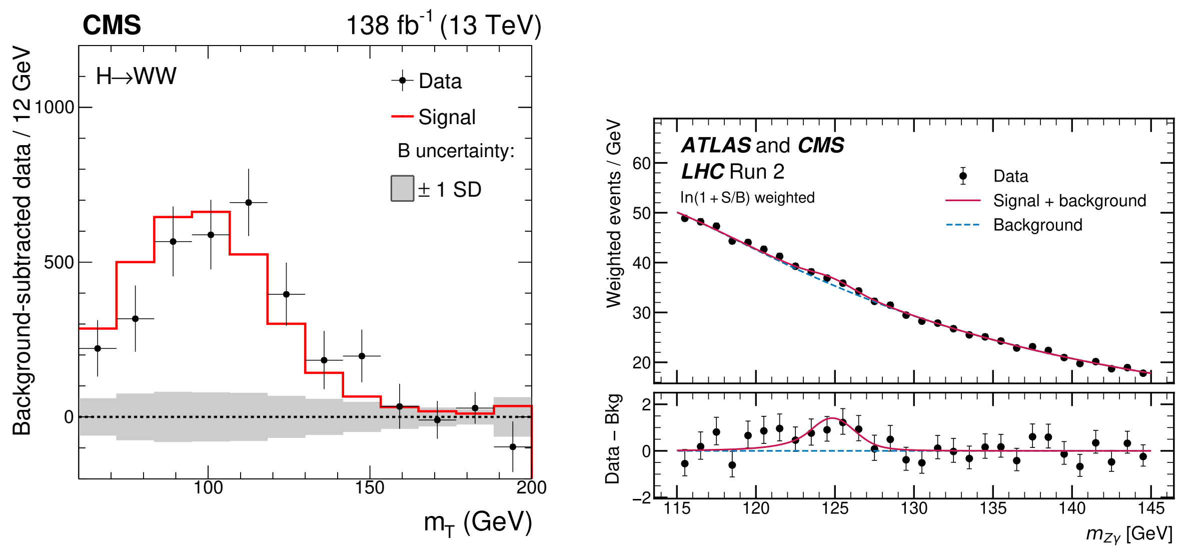


Figure 1.21: The background-subtracted diphoton invariant mass distribution targeting the study of the decay channel $H \to \gamma\gamma$ (left) and the invariant mass distribution of four charged leptons targeting the study of the decay channel $H \to ZZ^* \to 4\ell$ (right) [37].



(a) **(b)**

Figure 1.22: Left: The background-subtracted transverse mass m_T distribution targeting the study of the decay channel $H \to WW^*$. In red the SM predictions for $m_H = 125.38\,\text{GeV}$. The grey band is the 1 standard deviation uncertainty in the background subtraction [37]. Right: the $Z\gamma$ invariant mass distribution. Different $m_{Z\gamma}$ ranges are used in CMS and ATLAS analyses, only the common subrange is shown. The data distribution is weighted by $ln(1 + S/B)$, where S and B are the observed signal and background yields in the $120 \div 130\,\text{GeV}$ interval. The lower panelshows the background-subtracted results [37,60].

1.7. State of art of double Higgs measurements and searches

Double Higgs production is a very rare process at the LHC, so it is critical to study different final states of Higgs decays. The various Higgs decay channels provide a wide range of ongoing research using complementary techniques to maximize sensitivity and allow to explore a wide space of anomalous coupling parameters. Combining the results of these different final states is essential to achieve maximum sensitivity in double Higgs production.

1.7.1 Double Higgs decay channels

The choice of decay channels where to study HH production demands a trade-off between BR and background contamination, therefore there is no single "golden channel". Channels with the highest BRs, such as final states involving b-quark pairs, are significantly affected by substantial background noise, particularly from multijetQCD processes.

Fig. 1.23 displays the BRs for various HH final states, excluding rare combinations with negligible BR. At LHC, two experiments include the search of Higgs boson pair production as one of the main goals of their physics programme: ATLAS and CMS (Section 2.1.1).

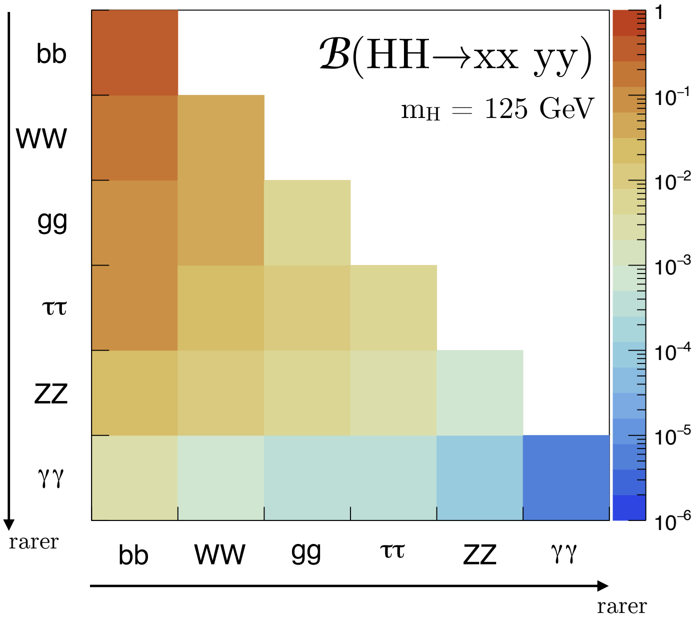


Figure 1.23: Summary of the branching fractions for the main HH final states, assuming the SM Higgs boson with $m_H = 125\,\text{GeV}$ [61].

Currently, several decay channels are under investigation at ATLAS and CMS:

- $HH \rightarrow b\bar{b}b\bar{b}$: despite this is the final state with the largest branching fraction ($BR \simeq 34\%$), it is not the most sensitive channel as it is affected by a large contamination which mimic the signal signature, mainly due to QCD process with multiple jets (QCD multi-jet) background. However it can profit from the larger signal yield to probe mass regions up to $m_{HH} \simeq 3\,\text{TeV}$.
- The HH $\rightarrow bb\tau\tau$ channel has a branching fraction ($BR \simeq 7.3\%$) lower than that of $HH \rightarrow b\bar{b}b\bar{b}$, but it provides a cleaner and more efficient signature thanks to the

$H \to \tau^+\tau^-$ decay. While this channel is affected by background contamination from QCD multijet and $t\bar{t}$ processes, it represents a compromise between two extreme cases: $HH \to b\bar{b}b\bar{b}$, which has the highest branching fraction but suffers from an overwhelming QCD multijet background, and $HH \to b\bar{b}\gamma\gamma$, which offers a very clean signature but has a significantly lower branching fraction due to the $H \to \gamma\gamma$ decay.
- The $HH \to b\bar{b}VV$ (with $VV = W^{\pm}W^{\mp}$ or ZZ) channel has a relatively high total branching fraction ($BR \simeq 28.1\%$) and experiences lower background contamination compared to the $HH \to b\bar{b}b\bar{b}$ final state. However, it is significantly affected by background from the to $t\bar{t}$ process. Typically searches focus on the very clean $VV \to 2\ell 2\nu$ decay mode, although the total branching fraction is approximately an order of magnitude lower ($BR \simeq 2.7\%$).
- The $HH \to b\bar{b}\gamma\gamma$ channel has a very small branching ratio ($BR \simeq 0.26\%$), but the clean signature of the photon pairs and of their kinematic results in a high signal selection efficiency and purity.

The decay channels listed above represent those expected at LHC to be the most sensitive, and their combination offers the broadest coverage of possible HH topologies. In particular, the first three channels provide the largest sensitivity to λ. Nevertheless, the collaborations have explored several other decay modes that, with the current dataset, yield significantly weaker sensitivity. During LHC Run 2, the increased center-of-mass energy and consequently the higher predicted HH production cross-section has opened new possibilities for HH searches. However, there is still potential to be unlocked in some of these other channels as more data are collected in the ongoing Run 3 and in the future LHC runs, when those decay channels become less statistically limited.

The HH $\to bb\tau\tau$ channel

As mentioned before, HH $\to bb\tau\tau$ channel is a very promising channel thanks to the relatively high branching ratio and reasonably clean signature compared to other decay channels, thanks to the H $\to \tau^+\tau^-$ decays.
Since τ leptons can decay into either leptons or hadrons, searches must target various final states. The main challenge in reconstructing and identifying these final states is the presence of neutrinos, which prevents a complete reconstruction of the event, and the significant background noise, both requiring to improve current reconstruction and selection techniques. However, $HH \to bb\tau\tau$, as highlighted in several studies, provides optimal sensitivity for BSM scenarios in both resonant and non-resonant HH signal hypotheses [38, 62].

In this thesis, the HH $\to bb\tau\tau$ channel is investigated across three distinct final states: the "fully hadronic" channel $HH \to bb\tau_h\tau_h$, where τ_h represents a τ lepton decaying into hadrons plus a ν_τ (referred as *hadronic tau*[9]) and the "semileptonic channel" $HH \to bb\tau_\ell\tau_h$, where one τ lepton decays into an e or μ (denoted as τ_ℓ,

[9] or, equivalently *hadronically decaying tau*.

and referred as *leptonic tau*) plus their associated neutrinos, while the other decays hadronically (τ_h). The fully hadronic channel accounts for 42% of the Higgs di-tau decays, while the semileptonic channel occurs in 46% of cases.

The characteristics of the $bb\tau\tau$ decay channel, along with the primary background contributions, are detailed in Section 4.1 and throughout the remainder of this thesis.

1.7.2 HH production experimental searches

The cross section for HH production predicted by the SM is extremely small, making it undetectable at the LHC with current data. As a result, experimental searches set upper limits on the production cross section.

Several HH searches for the non-resonant and the resonant production have been performed by both the CMS and ATLAS collaborations at the LHC.
The latest results on BSM HH searches are presented in the following.

Non-resonant HH production

Both ATLAS and CMS are deeply involved in these searches, exploiting many final states and focusing on ggF and VBF HH production, setting 95% Confidence Level (CL) limits on the production cross section and on the coupling modifiers involved in the Feynman diagrams for each production mode.
In the following, the latest results on HH production cross section and on κ_λ are shown. A more complete set of limits set to the coupling modifiers and EFT parameters are available in the ATLAS paper [63] and CMS Physics Analysis Summary [64].
Fig. 1.24 shows the 95% CL observed (expected) upper limit on the inclusive production cross section relative to the SM expectation, in searches using different final states and their combination, obtained from CMS (left) and ATLAS (right). These results are obtained analysing proton-proton collisions data produced between 2016 and 2018 at the LHC a centre-of-mass energy of 13 TeV and collected by the two experiments. Assuming all the coupling modifiers to be 1 (SM couplings), the values found by CMS for observed (expected) limits are 3.5 (2.5), while ATLAS ones are 2.9 (2.4).

In Fig. 1.25, the combined expected and observed 95% CL upper limits on the HH production cross section from CMS (left) and ATLAS (right), expressed as a function of κ_λ which is left free to vary, while the other coupling modifiers are fixed to the unity (SM couplings). CMS set 95% CL exclusion limit to HH production cross section for values of κ_λ outside the range between -1.39 and 7.02 (expected range is $-1.02 < \kappa_\lambda < 7.19$). The same results are obtained from ATLAS, constraining κ_λ to be in the interval $-1.2 < \kappa_\lambda < 7.2$ (expecting $-1.6 < \kappa_\lambda < 7.2$).
The theoretical cross section predictions (red curves) display a characteristic dip near SM values (when all coupling modifiers are set to 1) due to destructive interference between contributing amplitudes in the two Feynman diagrams in Fig. 1.9(a). Experimental limits (black lines) strongly depend on κ_λ, reflecting the impact of interference effects on the total cross section and kinematic properties of HH events. These kinematic changes influence the efficiency of signal detection, emphasizing the interplay

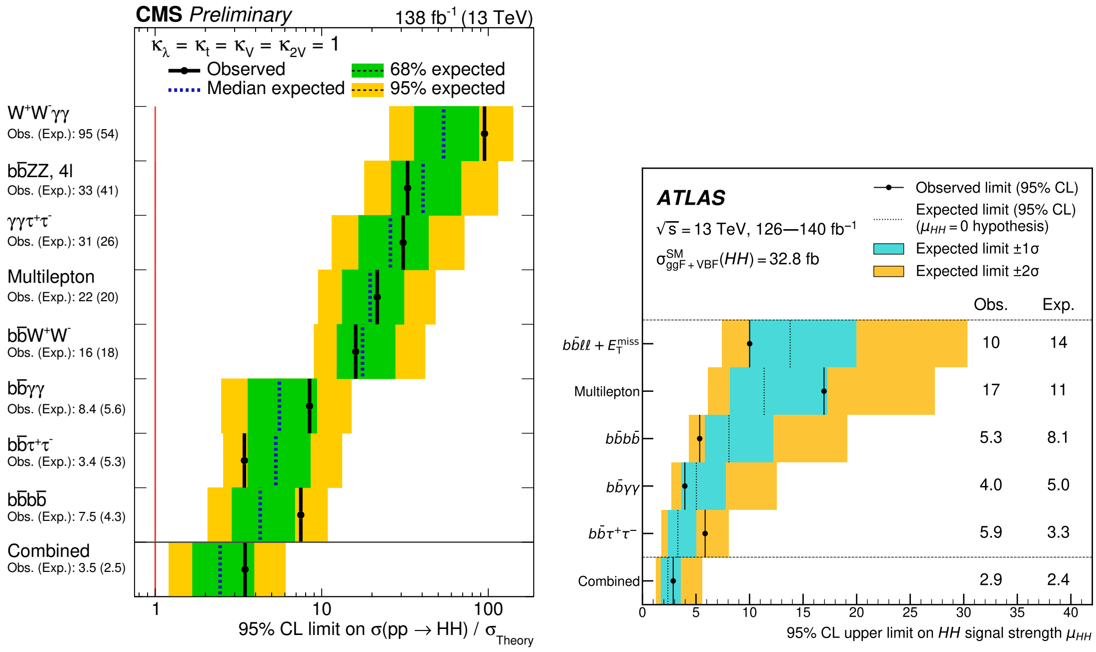


Figure 1.24: Observed and expected 95% CL confidence level limits on the ratio of experimentally estimated production cross section (CMS, left panel) and signal strength (ATLAS, right panel) and the expectation from the SM in searches using different final states and their combination. The HH production mechanisms considered are the ggF and VBF ones.The overall combination of all searches are shown by the lowest entry in both figures. [63,64].

between theoretical predictions and experimental sensitivity.

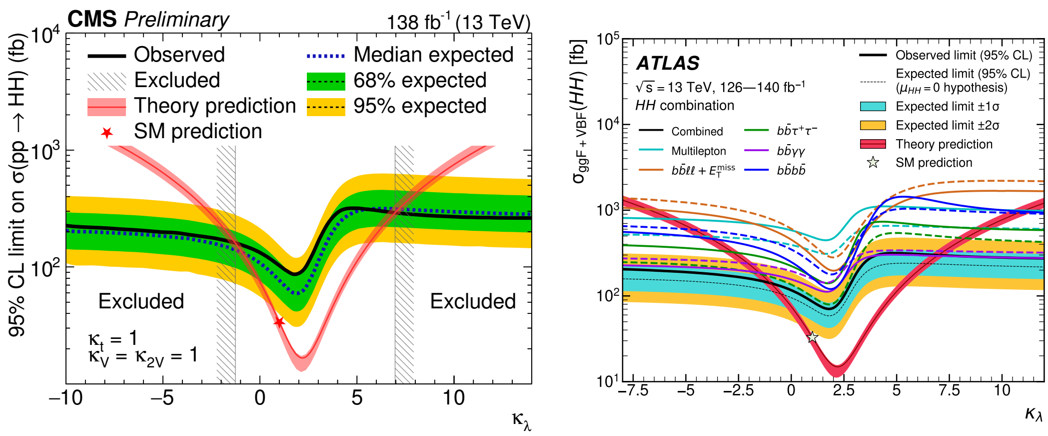


Figure 1.25: Combined expected and observed 95% CL upper limits on the HH production cross section obtained at the CMS (left) and ATLAS (right) experiments, for different values of κ_λ. The blue/green and yellow bands represent, respectively, the 1 and 2 standard deviations (σ) extensions beyond the expected limit; the red solid line (band) shows the theoretical prediction for the HH production cross section (for 1σ uncertainty).

H + HH results

ATLAS and CMS combined single-Higgs (H) and double-Higgs (HH) analyses with minimal statistical overlap, enabling a global fit framework to simultaneously constrain κ_λ and other coupling modifiers. This combined approach enhances constraints on κ_λ compared to HH-only analyses, assuming all other Higgs couplings match their SM values. The combination of single-Higgs and double-Higgs data from the ATLAS experiment yields observed (expected) constraints on κ_λ of $-1.4 < \kappa_\lambda < 6.1$ ($-2.2 < \kappa_\lambda < 7.7$) at 95% CL in a less model-dependent fit, as shown in left and right upper panels of Fig. 1.26 for observed and expected results. All other coupling modifiers remain consistent with SM predictions within uncertainties. This combined analysis allows for a more general interpretation, simultaneously constraining κ_λ and other Higgs couplings based on the data [65].
The combination of event categories enriched in single-H and HH events at CMS [66] leads to the exclusion of values of κ_λ outside the interval $-1.2 < \kappa_\lambda < 7.5$ at 2σ CL. This result is compatible with the expected range of -2.0 < κ_λ < 7.7 under the assumption that all other Higgs boson couplings are equal to their standard model predicted values. Relaxing the assumption on the Higgs couplings to fermions and vector bosons the observed (expected) κ_λ interval is constrained to be within $-1.4 < \kappa_\lambda < 7.8$ (-2.3 < κ_λ < 7.8) at 2σ CL.

HH resonant production

Resonant searches at CMS and ATLAS target a heavy, spin-0 X or spin-2 G resonance, which has a narrow-width compared to the experimental mass resolution. The spin-0 X resonance corresponds to the radion and the spin-2 resonance is denoted as G, having the graviton in mind.

Fig. 1.27 presents the upper limits on the resonance production cross section ($\sigma(pp \to X/G \to HH)$) as functions of the resonance mass for both spin-0 (left) and spin-2 (right) hypotheses. The exclusions in terms of the cross section reach values as low as 0.2 fb for both spin scenarios. The highest sensitivity at low masses is achieved through the diphoton search, while at high masses the searches involving boosted topologies provide the best sensitivity. The results of the statistical combination is depicted by the red lines. These combined results are shown separately in Fig. 1.28, along with the ±1 and ±2 standard deviation intervals for the expected limits in green and yellow respectively. No deviation exceeding 2 standard deviations from the expected limits is observed. Significant improvements in sensitivity are achieved relative to the best individual channels in the radion mass range $m_X \sim 0.5 \div 1$ TeV, where multiple channels contribute approximately equally to the combination. Below 0.32 TeV and above 0.8 TeV, the combination sets the strongest observed limits to date on resonant HH production. These results are obtained combining the CMS searches in different channels.

The results of the individual searches and their combinations are interpreted within specific models. In Fig. 1.29 the HH searches are used to extract limits to the WED

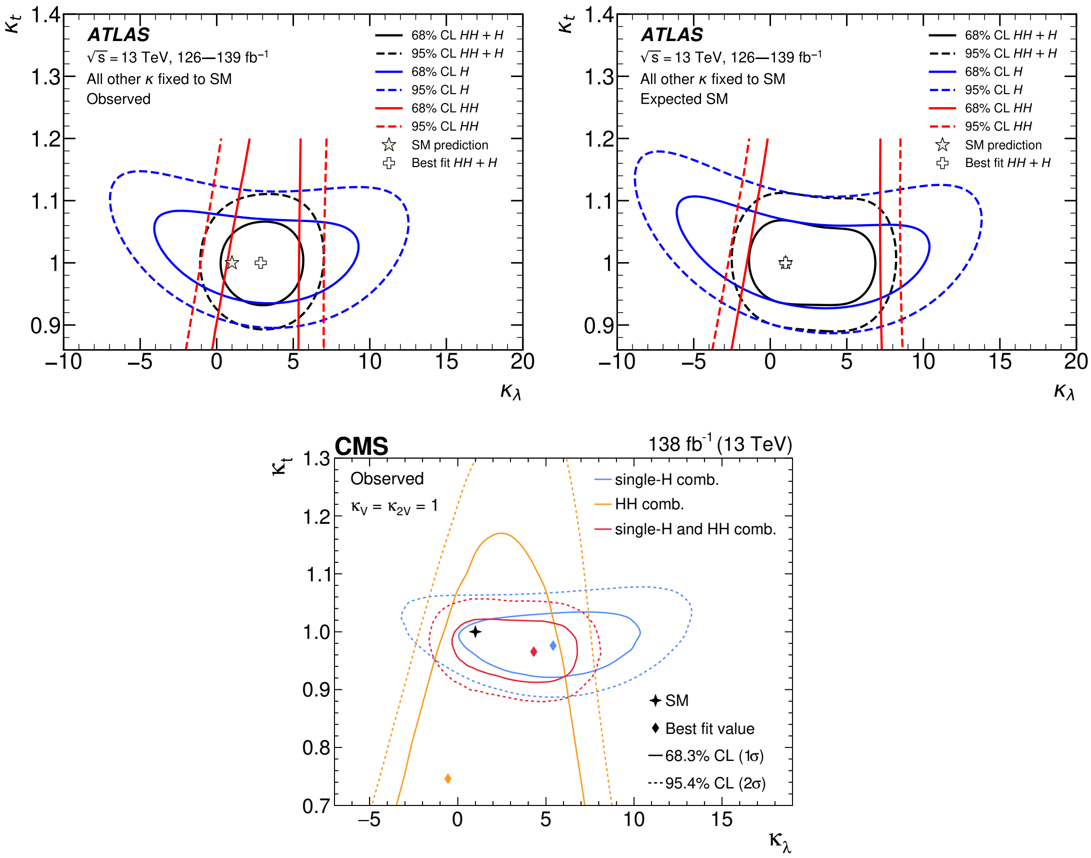


Figure 1.26: Top: ATLAS observed (left) and expected in the SM scenario (right) constraints in the κ_λ VS κ_t plane from single-Higgs (blue) and double-Higgs(red) analyses, and their combination (black). The solid (dashed) lines show the 68% (95%) CL contours. The double-Higgs contours are shown for values of κ_t smaller than 1.2. The observed constraint for the single- and double-Higgs combination for κ_t values below unity is slightly less stringent than that for the single-Higgs fit alone due to the slightly higher best-fit value for this coupling modifier. Bottom: CMS observed two-dimensional likelihood scans of (κ_λ,κ_t) comparing the full combination of single-H and HH (red) to the combinations of only single-H (blue) or only HH (yellow) channels [66].

parameters. In the left panel, the lower limits on the bulk radion ultraviolet cutoff parameter Λ_R, as a function of the radion mass m_R, and upper limits (right) on the parameter $\tilde{k} = k/M_{Pl}$ [46] of the spin-2 bulk graviton G, as a function of m_G. Excluded areas are indicated by the direction of the hatching along the exclusion contours

The resonant ATLAS HH searches set 95% CL limits on resonant HH production cross-section, $\sigma(X \to H)$, and shown in Fig. 1.30 in the $bb\gamma\gamma$, $bb\tau\tau$ and $bbbb$ final states, and their statistical combinations.

The $b\bar{b}\gamma\gamma$ search demonstrates the highest sensitivity at low masses, while the $b\bar{b}\tau^+\tau^-$ search dominates in the $350 \div 800\,\mathrm{GeV}$ range, and the $b\bar{b}b\bar{b}$ search is the most sensitive at higher masses, highlighting the complementarity of these three search channels.

The combined results are consistent with the predictions from the SM backgrounds across the tested mass range of m_X from 250 GeV to 5 TeV. The right panel of Fig. 1.30

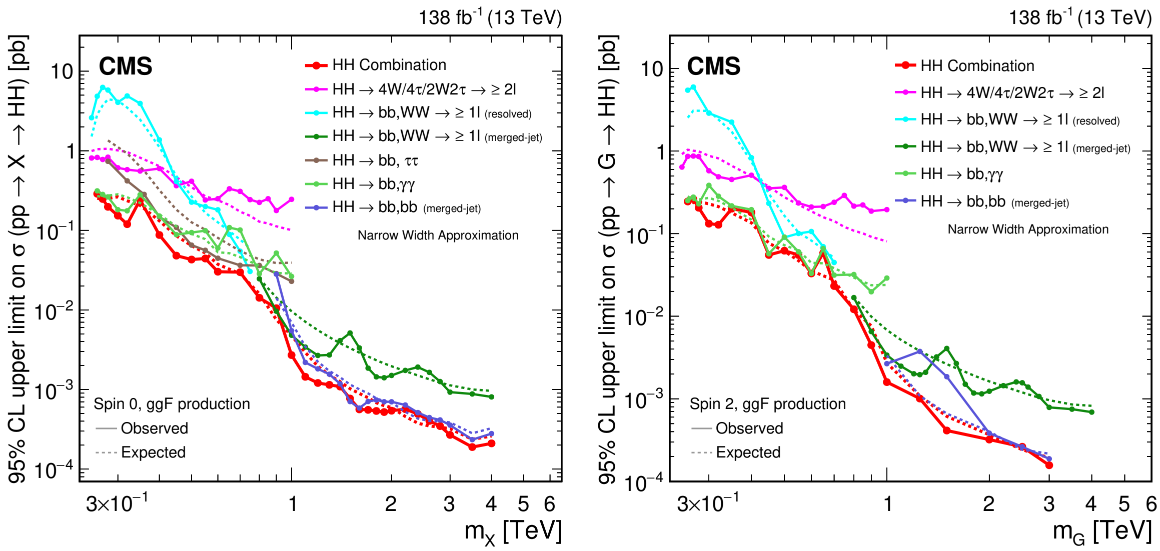


Figure 1.27: Observed and expected 95% CL upper limits on the cross section σ for the production of a (left) spin-0 resonance X and (right) a spin-2 resonance G at CMS. The results of the individual analyses presented in this report and the result of their combined likelihood analysis are shown. The observed limits are indicated by markers connected with solid lines and the expected limits by dashed lines. [50]

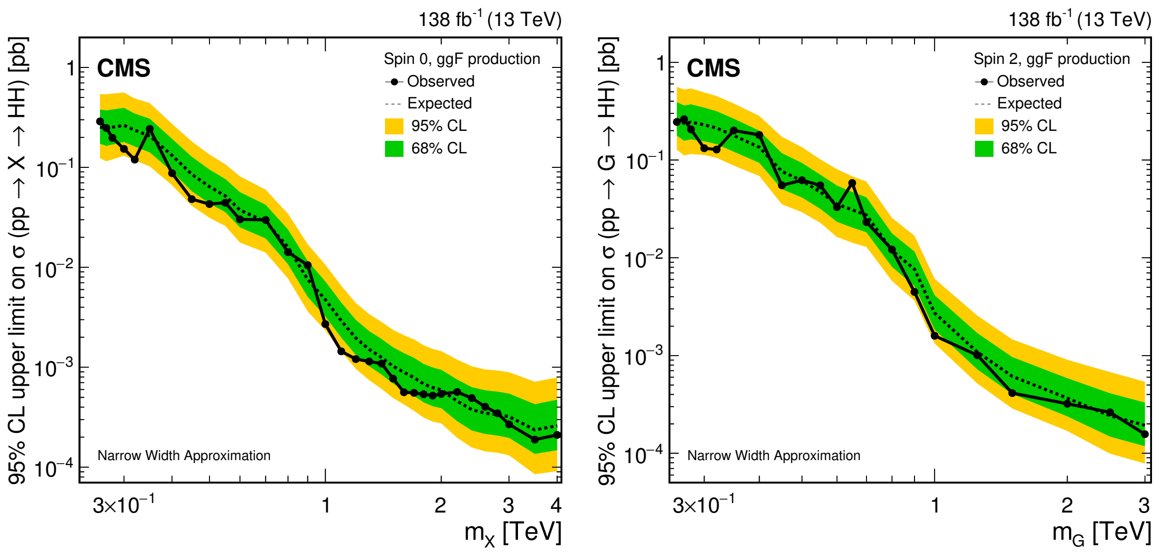


Figure 1.28: HH-Combination observed and expected 95% CL upper limits on the cross section σ for the production of a spin-0 resonance X (left) and a spin-2 resonance G (right), via gluon-gluon fusion at CMS, as obtained from the combined likelihood analysis of all contributing individual analyses presented in this report and shown in Fig. 1.27. In addition to the limit from the combined likelihood analysis the 68 and 95% central intervals for the expected upper limits in the absence of a signal are shown as coloured bands. [50]

shows the local p-value as a function of m_X for a narrow resonance decaying into a pair of SM Higgs bosons. The largest deviation is observed at $m_X = 1.1$ TeV, corresponding to a local significance of 3.3σ, driven primarily by the $b\bar{b}\tau^+\tau^-$ channel. The global significance of this excess is estimated to be 2.1σ, using the method outlined in Ref [68].

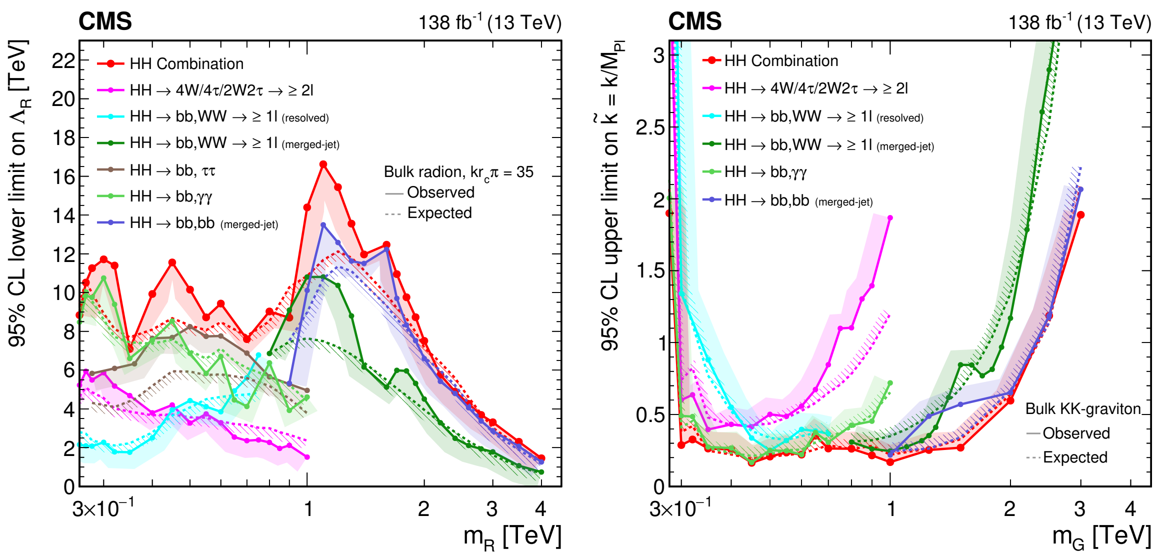


Figure 1.29: Observed and expected limits, at 95% CL, for WED parameters as obtained from the $X \to HH$ analyses used to compute the limits in Fig. 1.27 and their combined likelihood analysis at CMS. [50]

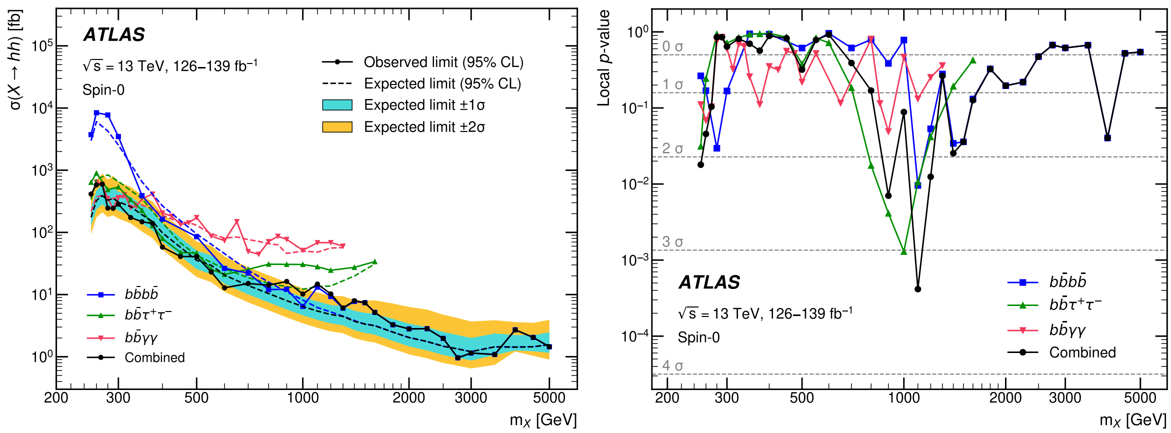


Figure 1.30: Observed and expected upper limits (left) at the 95% CL on the resonant Higgs boson pair production cross section and local p-value (right) as a function of the resonance mass m_X at the ATLAS experiment. The symbol h denotes a SM Higgs boson with $m_H = 125\,\text{GeV}$ [67].

The observed (expected) upper limits range from 0.96 fb to 600 fb (1.2 fb to 390 fb), depending on m_X.

So far there is no observation of any deviation from the SM prediction. However, searches for resonant Higgs boson pair production are crucial as they provide a direct probe of new physics, including scenarios with heavy resonances such as extended Higgs sectors, WED, or new gauge interactions.

A breakthrough in this area would not only reveal new fundamental particles but also provide critical insights into the structure of EWSB and the potential enhancement of Higgs self-coupling, which is key to understanding the stability of the electroweak vacuum.

Chapter 2

The experimental apparatus

The search presented in this thesis uses proton-proton (pp) collisions data collected by the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) between 2016 and 2018 at a centre of mass energy of $\sqrt{s} = 13\,\text{TeV}$. This chapter provides an overview of the LHC structure and operation (Section 2.1) and of the CMS detector (Section 2.2), focusing on its subdetectors and features that are relevant for the analysis presented in this thesis.

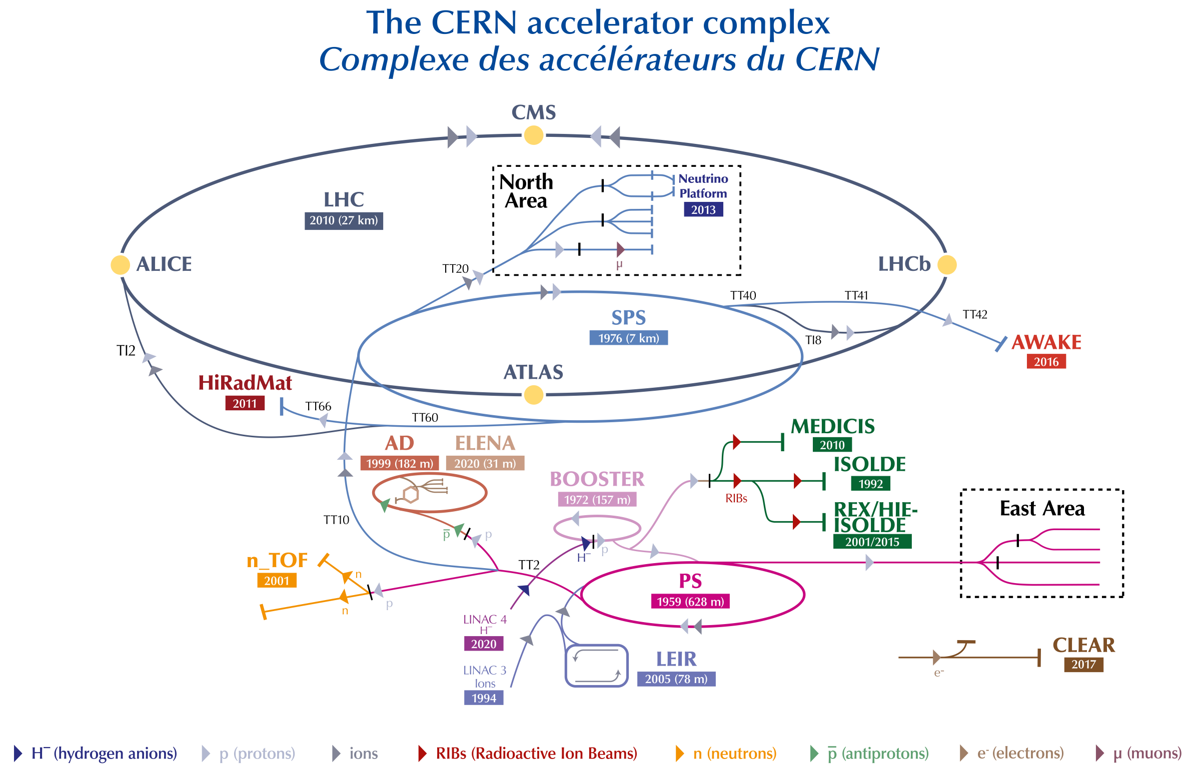
2.1. The Large Hadron Collider

The LHC [3] is a circular collider hosted by CERN, an international research laboratory located on the French-Swiss border near Geneva, founded in 1954. It is the largest and most powerful particle accelerator ever built, housed in a 26.7 km-long tunnel at a depth ranging from 45 to 170 m below ground level. Constructed within the tunnel previously used by the Large Electron Positron Collider (LEP), which operated from 1989 to 2000, the LHC consists of eight straight sections and eight curved arcs, with four main Interaction Points (IP).

At the LHC, collisions occur between proton beams and heavy ions, such as oxygen or lead. The collider is designed to achieve center-of-mass energies of up to 14 TeV for pp collisions, while heavy-ion collisions can reach up to 7 TeV for oxygen-oxygen (O-O) interactions and 5.5 TeV for lead-lead (Pb-Pb) interactions.

The LHC is the final stage of the CERN accelerator complex, shown in Fig. 2.1. To reach a beam energy of $\sqrt{s} = 6.5\,\text{TeV}$, protons undergo multiple acceleration stages:

- Protons are produced via hydrogen ionization and injected into the 36 m-long Linear Accelerator (LINAC 2), where they are accelerated up to 50 MeV.
- They are then transferred to the Proton Synchrotron Booster (PSB), a circular accelerator with a diameter of 50 m, reaching a centre of mass energy of 1.5 GeV.
- Subsequently, protons enter the Proton Synchrotron (PS), a circular accelerator with a circumference of 628 m, accelerating them to 25 GeV. At this stage, the



[redacted]

Figure 2.1: The CERN accelerator system: the acceleration steps for protons and ions are represented by the light and grey triangles. Many accelerator lines are present as well as many beamlines, from radioactive ion beams to neutrons, and muons [69].

protons are grouped into beams, structured into 2808 bunches, each divided in bunches with 25 ns time-separation and then split into 2808 bunches with a nominal number of protons per bunch of 1.15×10^{11}.

- The last acceleration step before the injection in the LHC is Super Proton Synchrotron (SPS), a 6.9 km-circumference accelerator where proton are injected after they reach the maximum energy at the PS. In SPS they reach the centre of mass energy of 450 GeV.
- Finally, the beams are injected into the LHC, composed of two rings where protons circulate in opposite direction. These beams are kept in separate vacuum chambers within the main arcs and share a common region at the IPs. Inside the LHC, protons are further accelerated up to 6.5 TeV per bunch[1].

The beam acceleration up to collision energy is performed by superconducting radio-frequency (RF) cavities located in a dedicated cavern. These cavities are tuned to

[1]Throughout this thesis, unless otherwise specified, numerical values refer to the data-taking period between 2016 and 2018, part of the so-called Run 2 (2015-2018), as these data have been analyzed in this work.

oscillate at 400 MHz, shaping the beams into proton bunches and distributing the timing signal to all LHC experiments. The CERN injector complex provides high-intensity bunches that meet the stringent requirements of the LHC in terms of transverse beam size and longitudinal characteristics. Beams are kept in two counter-rotating orbits within the LHC rings by 1232 superconducting dipole magnets powered by an 11 kA current. Each magnet weighs approximately 35 tons and measures 15 m in length. To maximize the proton density at the collision points, the beams are focused by 392 quadrupole magnets, whose length spans from 5 to 7 meters. Additional magnets, including sextupoles and octupoles, are placed along the accelerator to correct the field geometry. The LHC magnet system relies on well-established technology based on Nb-Ti Rutherford cables and operates at a temperature of 1.9 K, achieved using a superfluid He^4 cryogenic system. The magnets generate fields exceeding 8 T [3]. The proton beams complete 11,245 revolutions per second while traveling in separate beam pipes. The time interval between successive bunches is 25 ns, corresponding to a nominal bunch-crossing frequency of 40 MHz. However, due to missing bunches introduced during beam injection from the SPS to the LHC, the effective average crossing frequency is reduced to approximately 30 MHz. Under these conditions, electronic readout systems must be capable of operating at these frequencies to prevent the loss of potentially significant collision events.

2.1.1 The LHC experiments

Four main experiments are installed around the IPs, where proton bunches collisions occur. Compact Muon Solenoid (CMS) [70] and A Toroidal LHC ApparatuS (ATLAS) [71] are two general-purpose experiments. Their physics programs span a wide range of topics, including high-precision Standard Model (SM) measurements (focusing on Higgs boson, top quark, and electroweak physics), studies of CP violation (especially in B mesons), and searches for new physics such as vector-like quarks, supersymmetry, high-momentum signatures, and resonances.
The LHCb (*LHC beauty*) experiment [72] is a forward spectrometer designed for precision studies of CP violation in beauty and charm hadrons, flavor-changing neutral and charged currents, exotic and conventional hadron spectroscopy, electroweak physics tests, and direct searches for new particles.
Finally, ALICE (*A Large Ion Collider Experiment*) [73] is dedicated to studying heavy-ion collisions and the properties of the quark-gluon plasma.
In addition to these main experiments, several other experiments are positioned along the collision points:

- TOTEM (*TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC*) [74], which performs precise measurements of the pp interaction cross section and in-depth studies of proton structure using detectors placed on both sides of the CMS IP.
- LHCf (*LHC forward*) [75] focuses forward particles using two detectors positioned along the LHC beamline, 140 meters from the ATLAS IP.

- MoEDAL (*Monopole and Exotics Detector At the LHC*) [76], located near the LHCb experiment, whose main purpose is to search magnetic monopoles.
- Close to ATLAS there are *FASERν* and *SND@LHC*, two experiments oriented to neutrino physics study searches for light new particles [77–79].

2.1.2 Luminosity at the LHC

The *instantaneous luminosity* is one of the most important parameters for accelerating machines like the LHC. The event production rate for specific process, R, is defined as:

$$R = \mathcal{L} \cdot \sigma \tag{2.1}$$

where σ is the production cross section of the process, and $\mathcal{L}$ is the instantaneous luminosity. It is expresses in $\mathrm{cm^{-2}s^{-1}}$ and depends solely on the intrinsic characteristics of the accelerator. For a circular collider like the LHC, assuming a Gaussian beam distribution[2], the instantaneous luminosity depends on the machine parameters as follows:

$$\mathcal{L} = \frac{N_b n_1 n_2}{4\pi\epsilon_n} \cdot \frac{f_{rev}\gamma F}{\beta^*}$$

Here:

- N_b is the number of bunches.
- n_1 and n_2 are the number of particles per bunch. For protons in the LHC, $n_1 = n_2 = n_p$.
- f_{rev} is the bunch revolution frequency in the LHC.
- β^* is the value of the β function at the IP, and it determines how narrowly the particle beams are focused at that point.
- ϵ_n is the normalized transverse beam emittance.
- γ is the relativistic factor, accounting for the protons' relativistic energy.
- F is a geometric factor that corrects for the beam crossing angle at the collision points, θ_c. It depends on θ_c as well as on the longitudinal and transverse beam spread sizes, σ_z and σ_{xy}, which characterize the Beam Spot (BS), the luminous region where beam collisions occur.

During Run 2, the LHC achieved peak instantaneous luminosities of $2 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ for proton beams and $10^{27}\,\mathrm{cm^{-2}s^{-1}}$ for lead-ion beams at the CMS and ATLAS experiments.
Table 2.1 summarizes the design values of the beam parameters.

The instantaneous luminosity value decreases during an LHC run. The beam lifetime, defined as the time after which the beam intensity for luminosity production reaches $1/e$ of its initial value, ideally lasts over 80 h. However, in practice, the average

[2] This is a good approximation for most hadron colliders, as radiation and other effects that could alter the distribution are usually negligible.

Parameter	Description	Run 2 value
$\sqrt{s}$	centre-of-mass energy	13 TeV
N	number of bunches	2808
n_p	proton population per bunch	1.15×10^{11}
n_{Pb}	lead ions population per bunch	7×10^{7}
f_{rev}	revolution frequency	$11,245$ Hz
ϵ_n	transverse beam emittance	3.75 μm
β^*	β function at IP5	0.55 m
ϕ	beam crossing angle	285 μrad
σ_{xy}	BS transverse size	16.7 μm
σ_z	BS longitudinal size	7.55 cm

Table 2.1: LHC parameters during Run 2 pp collisions.

lifetime of stable beams[3] is $\sim 10\,$h. This reduction is due to continuous beam losses caused by proton interactions with residual gas that has escaped the vacuum of the beam pipe, the limited efficiency of correction and focusing systems leading to Coulomb scattering of protons traveling together or in different bunches at IPs, and general operational errors. Once the beam intensity decreases to $1/e$ of its nominal value, the beam is dumped, and a new cycle begins.

During a certain data taking period T, the number of produced events for a given process is obtained by integrating over time the Eq. 2.1:

$$N = \int_0^T R\,dt = \sigma \int_0^T \mathcal{L}\,dt = \sigma L \tag{2.2}$$

where L is the *integrated luminosity*, which represents the total luminosity delivered to an experiment over the time interval T. It allows estimating the number of events for both interesting process and of its backgrounds[4] are produced in an experiment over a certain time.

The *delivered luminosity* is the total amount of potential collision data provided by the collider to an experiment and depends only on machine parameters. Conversely, the relevant quantity for analysis is the *recorded luminosity*, which refers to the fraction of the delivered luminosity that is effectively collected and declared good for data analysis.when all subdetectors, triggers, Data Acquisition system (DAQ), and physics objects reconstruction show the expected performance.
Figure 2.2 illustrates the instantaneous luminosity (upper panel) and the delivered and

[3] Beams are defined as "stable" when they are aligned, squeezed, focused, and finally steered to collide head-to-head.

[4] Throughout the thesis, processes that mimic the searched signature but are not of primary interest will be referred to as backgrounds, while the process of interesty is defined as the signal.

recorded integrated luminosity (lower panel) during Run 2 data-taking period at the CMS experiment.
Section 2.5 provides a description of the luminosity measurement techniques adopted at CMS.

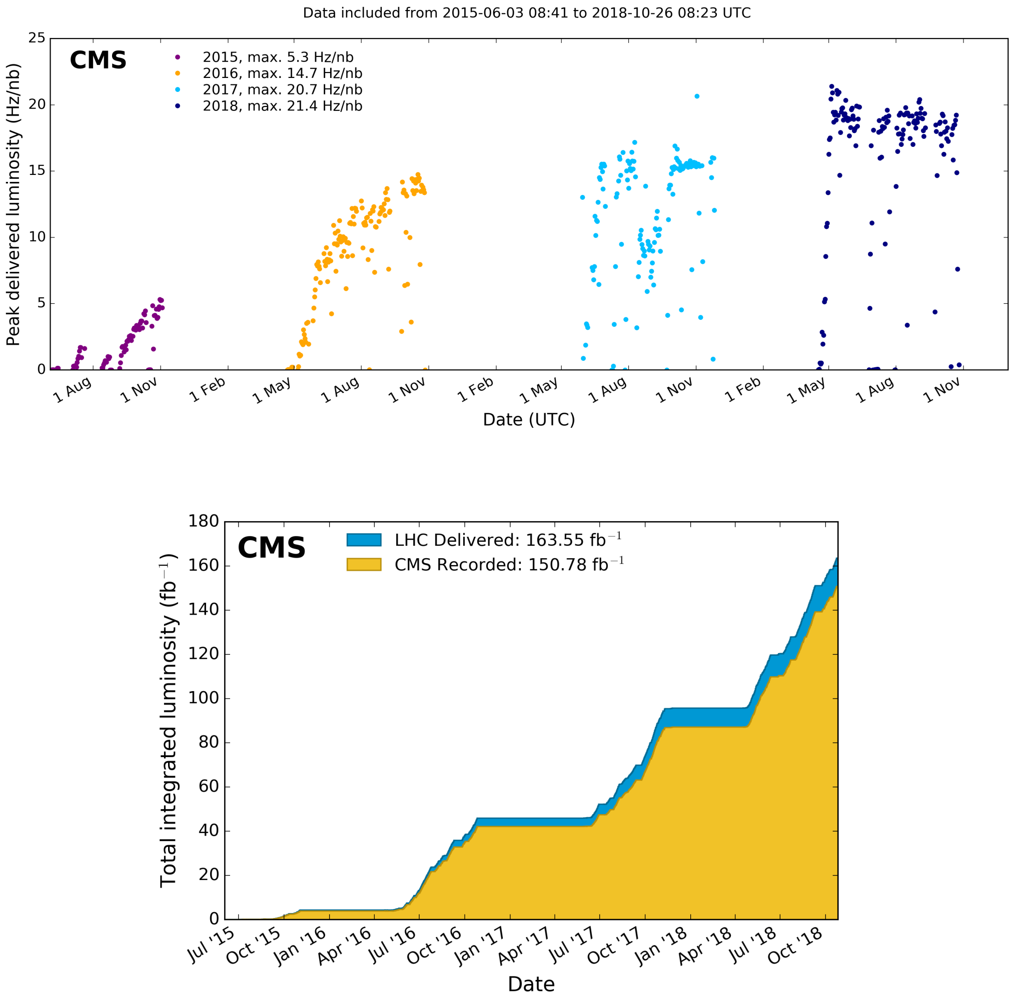


Figure 2.2: LHC performances during Run 2 at $\sqrt{s} = 13\,\text{TeV}$. Upper panel: Peak delivered instantaneous luminosity VS day. Lower panel: Integrated luminosity VS year. In the latter, the blue histogram corresponds to the delivered integrated luminosity from the LHC while yellow one corresponds to the luminosity recorded by the CMS detector under stable beam conditions [80].

2.1.3 Pileup at the LHC

At high luminosities, essential for precision measurements and searches for rare or Beyond Standard Model (BSM) phenomena, the number of interactions per bunch crossing increases significantly. The multiple interactions arising during a bunch crossing event are collectively referred to as *pile-up (PU)*; when they originate from different bunch crossings, they are categorized as out-of-time PU. The PU represents the average number of simultaneous interactions per bunch crossing and, in the case of pp collisions, is given by:

$$< \text{PU} >= \frac{\mathcal{L} \cdot \sigma_{inel}^{pp}}{n_p f_{rev}}$$

where σ_{inel}^{pp} represents the cross section for inelastic pp collisions.

To efficiently identify the primary hard interaction associated with a given event amid PU interactions, the most energetic collision is selected. This method assumes that the probability of multiple pp interactions producing significant physics processes in the same bunch crossing is negligible.

However, the PU implies the overlap of signals of several interactions in the detector. A precise understanding and modeling of this phenomenon is crucial, as it degrades the performance of the CMS detector in reconstructing physics objects (Chapter 3). Consequently, dedicated techniques are developed to mitigate the detrimental effects of PU. During Run 2, an average of 34 pp interactions per bunch crossing was reconstructed, with peak values exceeding 50, as illustrated in Fig. 2.3.

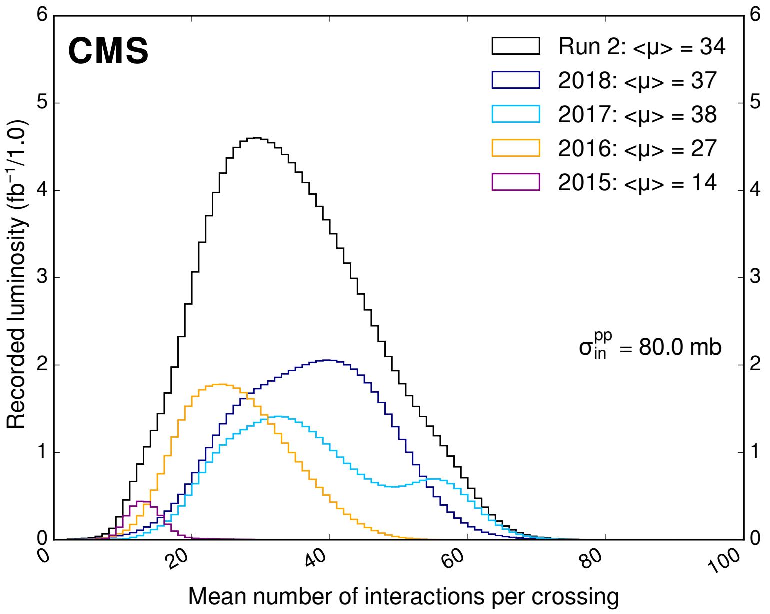


Figure 2.3: Distribution of the average number of interactions per crossing (pileup) for pp collisions in 2015 (purple), 2016 (orange), 2017 (light blue), and 2018 (navy blue). The overall mean values and the minimum bias cross section is also shown. These plots use only data where all CMS sub-detectors were flagged to be ok for any kind of usage in physics analysis, and the "LHC standard" values for the minimum bias cross sections, which are taken from the theoretical prediction from Pythia (Section 5.1) and should be used to compare to other LHC experiments [80].

2.1.4 LHC Operations

The first pp collisions at the LHC occurred on September 10, 2008, at a center-of-mass energy of $\sqrt{s} = 7$ TeV, marking the beginning of Run 1. This energy was later increased to $\sqrt{s} = 8$ TeV, and Run 1 lasted until early 2013. After a two-year shutdown (LS1), Run 2 started in 2015 with pp collisions at $\sqrt{s} = 13$ TeV. The data-taking period for Run 2 officially ended on October 24, 2018, initiating another long shutdown (LS2).

On July 5, 2022, ten years after the announcement of the Higgs boson discovery, the LHC resumed pp collisions at $\sqrt{s} = 13.6$ TeV, marking the beginning of Run 3. This phase will continue until the end of 2025, aiming to achieve an integrated luminosity of 300 fb^{-1} by its conclusion.

Following Run 3, a new long shutdown (LS3) will take place, leading into the HiLumi-LHC (HL-LHC) era. The operational phases are defined as Phase 1 (LHC), covering up to the end of Run 3, and Phase 2 (HL-LHC), for subsequent runs. The baseline peak luminosity for HL-LHC is projected to be $5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with an ultimate luminosity goal of approximately $7.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The total integrated luminosity for the entire HL-LHC operation is expected to reach 3000 fb^{-1} under the baseline scenario and 4000 fb^{-1} in the ultimate scenario.

Before the HL-LHC Phase 2 begins, significant upgrades are planned for all experiments. Major upgrades for the LHCb and ALICE experiments were implemented between Run 2 and Run 3, whereas substantial upgrades for CMS and ATLAS will occur during LS3, in preparation for HL-LHC Run 4. The LHC machine upgrades for Phase 2 include the installation of new superconducting quadrupole magnets at the ATLAS and CMS IPs, along with compact superconducting cavities ("crab cavities") designed to enhance luminosity performance.

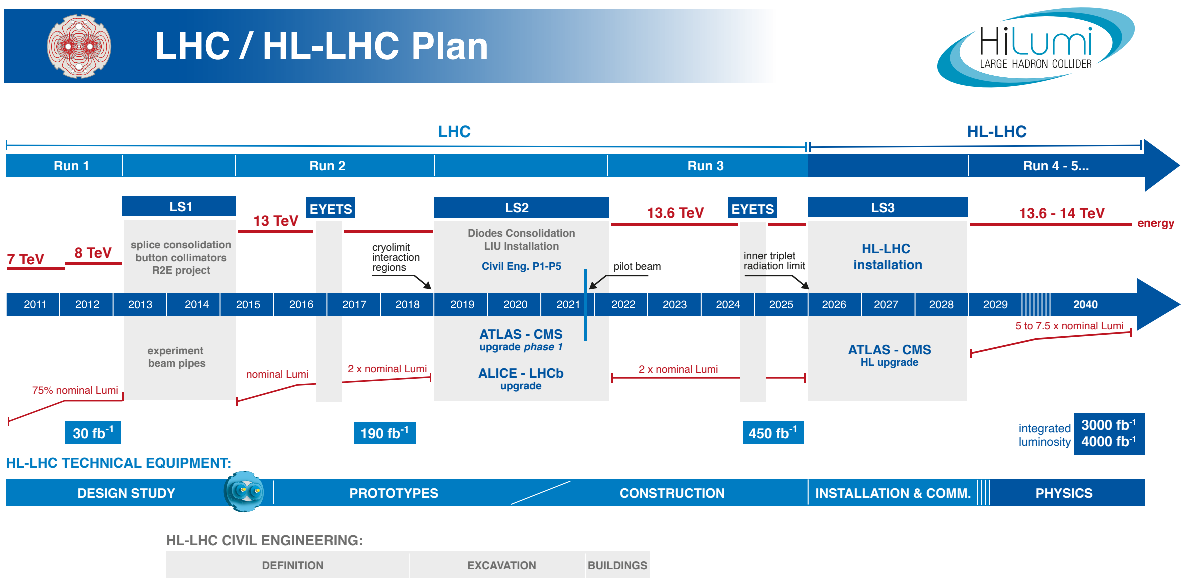


Figure 2.4: HL-LHC schedule [81].

2.2. The CMS Experiment

The CMS experiment is a general-purpose detector with cylindrical symmetry, located at the LHC collision point IP5 in Cessy (FR) [70]. As of 2024, the CMS collaboration is one of the largest scientific collaborations in history, bringing together approximately 6107 members from 256 institutions across 58 countries.

In 2012, the CMS and ATLAS collaborations announced the discovery of the Higgs boson, a milestone in modern physics that demonstrated the capabilities of these experiments. Since then, Higgs physics has become a central research focus in the CMS physics programme, including studies of both single and double Higgs production. Moreover, the large dataset expected to be collected during Phase 2 may enable direct measurements of the Higgs boson's self-coupling.

The advanced technologies of the CMS subdetectors and its sophisticated two-level trigger system (Section 2.4), combined with state-of-the-art software tools, allow for precise particle reconstruction and identification. These capabilities enable high-precision tests of the SM, Higgs boson studies, and the exploration of BSM theories, such as Warped Extra Dimensions theory (WED),the theory at the basis of the search conducted in this thesis, searches for Super symmetry (SUSY), Vector-Like Quarks and Long-Lived Particles. Moreover, CMS conducts competitive B-physics analyses, compared to dedicated experiments such as LHCb and Belle II, and heavy-ion physics studies.

2.2.1 CMC coordinate system and useful quantities

The CMS experiment adopts a coordinate system centered at the nominal collision point, with the y-axis directed upward and the x-axis radially pointing inward to the LHC ring centre. Consequently, the z-axis extends along the beam direction towards the Jura mountains from the IP [82].
Given the detector cylindrical shape, a cylindrical coordinate system (r, θ, ϕ) is convenient to use. The azimuthal angle ϕ is measured from the x-axis in the transverse plane and the polar angle θ is measured from the z-axis. A schematic view of these coordinate systems is shown in Fig. 2.5.

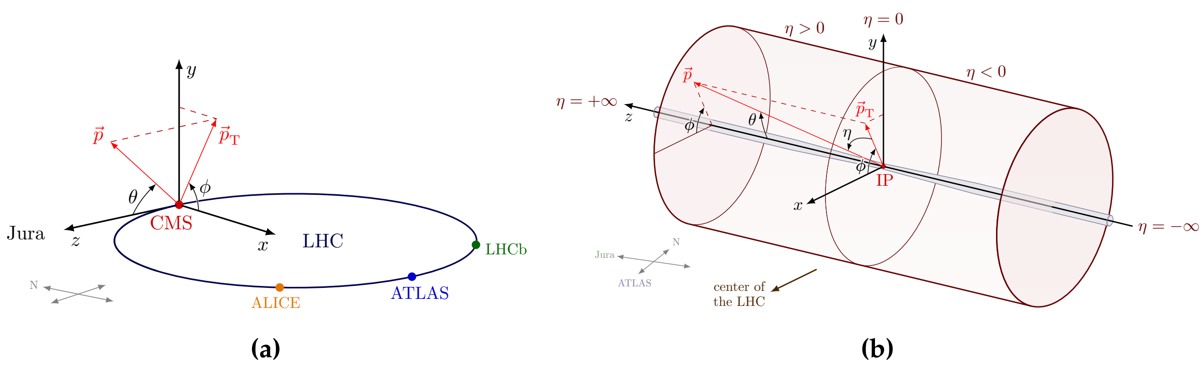


Figure 2.5: Scheme of CMS cartesian (Fig. 2.5a) and cylindrical (Fig. 2.5b) coordinate systems.

Momentum components

The momentum of a particle in the cartesian representation consists of a longitudinal component (longitudinal momentum component (p_Z)) along the beam axis and a transverse component transverse momentum component (p_T), defined as:

$$p_T = \sqrt{p_x^2 + p_y^2}$$

Differences in p_T are invariant under boosts along the z-direction.

Pseudorapidity

The polar angle θ can be conveniently expressed in terms of the pseudorapidity η:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.3}$$

The use of η instead of θ is convenient since both η and differences in η ($\Delta\eta$) are invariant under a Lorentz boost along the z-axis and at the same time both are dependent only from kinematics. In fact, pseudorapidity can also be defined in terms of the particle momentum:

$$\eta = \frac{1}{2}\ln\left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z}\right), \tag{2.4}$$

For ultra-relativistic particles ($E \approx |\mathbf{p}|$),which is a valid approximation at LHC energies, pseudorapidity closely aligns with rapidity $\mathcal{Y}$:

$$\mathcal{Y} = \frac{1}{2}\ln\left(\frac{E + p_z c}{E - p_z c}\right) \tag{2.5}$$

where c is the speed of light in vacuum.
The pseudorapidity η is commonly used to identify three detector regions relative to the beamline: the *barrel region* ($|\eta| < 0.9$), the *overlap region* ($0.9 < |\eta| < 1.2$), and the *endcap regions* ($|\eta| > 1.2$). A $r - z$ view of one quadrant of the CMS detector is shown in Fig. 2.6 with the red lines denoting the $\eta(\theta)$ coordinates.

Lorentz invariant distance

As for pseudorapidity, also azimuthal angle differences ($\Delta\phi$) are invariant under a Lorentz boost along the z-axis. Thus, $\Delta\eta$ and $\Delta\phi$ can be exploited to define Lorentz invariant distance between two particles in the detector:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{2.6}$$

Also in this case this definition is convenient, especially thanks to the cylindrical symmetry This variable is also used to define a longitudinal boost invariant cone aroung a single particle in order tu study nearby activity around it.

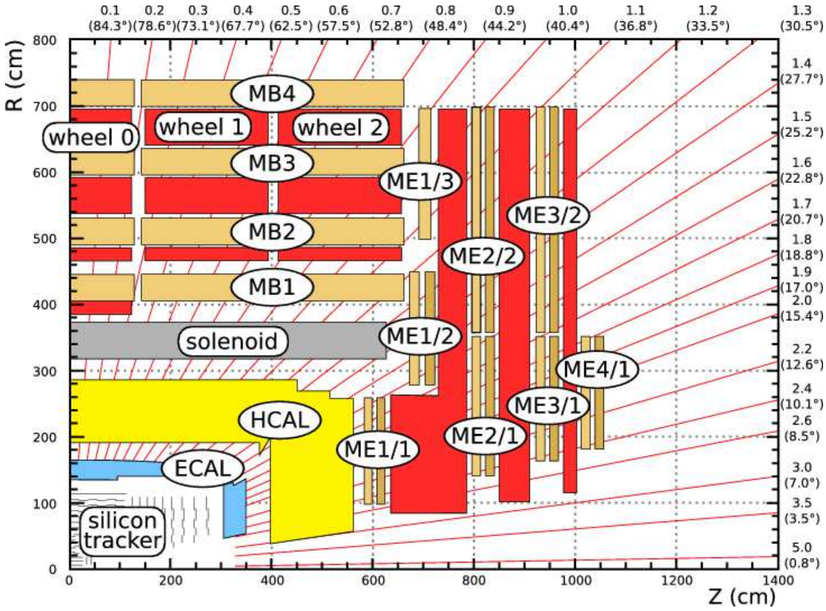


Figure 2.6: Layout of a CMS quadrant in the $r-z$ plane. The red lines denote the corresponding values of η and θ. Image took from [83].

Beam Spot (BS)

The luminous region near to the coordinates origin, where two proton beams intersect each other, is defined as BS. A typical width value of this region is 16 µm and 100 µm along x and y directions respectively, while its length in z direction is approximately 15 cm. The BS location plays an important role during the event reconstruction as it is the origin for all pp interaction products.

2.3. The CMS detector components

The CMS detector features a high-field, large-bore superconducting solenoid, which surrounds a silicon tracker, a lead-tungstate scintillating-crystal electromagnetic calorimeter, a brass-scintillator hadron calorimeter, and muon detectors. Forward sampling calorimeters extend the detector coverage to high pseudorapidity values ($|\eta| \leq 5$), ensuring hermeticity. The detector has a length of 21.6 m, a diameter of 14.6 m, and a total weight of 14 000 t.

Designed to handle high event rates and short bunch crossing times, the CMS detector achieves precise momentum resolution for all charged particles, with a particular focus on muon identification. The strong magnetic field, generated by superconducting technology, provides the necessary bending power for accurate momentum measurements. Additionally, the detector delivers excellent electromagnetic energy resolution, enhancing photon and electron identification. The finely segmented hadron calorimeter enables precise measurements of missing transverse energy. By combining calorimetric data with tracking information, the detector effectively distinguishes between neutral and charged hadronic components.

All detector components and front-end electronics, especially those near the interaction region, are designed to withstand high radiation levels, ensuring minimal performance degradation throughout the LHC operation. These features collectively enable the CMS detector to maintain high efficiency and precision under the extreme conditions of the LHC [82].

2.3.1 Tracker

The CMS tracker [84,85] is a cylindrical structure measuring 5.8 m in length and 2.6 m in diameter, with a layered structure incorporating different arrangements of silicon sensors to optimize particle tracking. It is designed to precisely measure trajectories and momenta of charged particles in high-energy pp collisions, for particles with p_T between 1 GeV and 100 GeV and within $|\eta| < 2.5$. The layout of the original inner tracking system is shown in Fig. 2.7.
The entire tracking system employs silicon technology, which provides high granularity and detection efficiency, with a low material budget to minimize multiple scattering effects. Silicon detectors also feature fast response times and radiation tolerance, making them suitable for high-speed particle tracking in a high-density environment. The modular structure of the tracker facilitates scalability and maintenance. The tracker is placed within the superconducting solenoid which produces a nearly uniform magnetic field of 3.8 T (Section 2.3.5), essential for p_T measurements of charged particles by bending their trajectories. Although the tracker is equipped to handle high particle flux ($\sim 10^7$ particles per second at a radius of 10 cm), challenges such as track overlap and primary and secondary vertices identification require sophisticated reconstruction algorithms to achieve excellent precision.
The tracker system covers the region up to $|\eta| < 2.5$ with a radius $r < 1.2$ m around the beam pipe and for $|z| < 2.7$ m. The innermost section of the tracker consists of silicon pixel sensors, while the outer regions have silicon detectors aranged in strips. Both sections are further divided into barrel and endcap regions. The high level of precision makes silicon detectors ideal for placement in the region closest to the beam pipe, wher there are the highest particle densities.
The pixel detector technology offers low occupancy, high spatial resolution, allowing to perform an accurate vertex reconstruction thanks to their extremely fine granularity. It contains ~ 66 millions of pixels with an area $100 \times 150\,\mu m^2$, and for a total area of about $1\,m^2$ [82]. Its readout consists in approximately 16000 chips, which are bump-bonded to the detector modules. Double-sided modules are built from two single-sided modules glued back-to-back at a stereo angle of 100 mrad between the strips to enhance the spatial resolution by providing a measurement in $r - \phi$ and $r - z$ coordinates [86]. The passage of a charged particle in the pixel detector produces a signal in the neighbouring pixels, which is used to reconstruct hits. The shapes of these clusters are analyzed and fitted to achieve a precise hit position, typically with a resolution on the order of 10 µm in $r - \phi$ plane (the transverse direction) and $20 \div 40$ µm in z (the longitudinal direction), depending on η.

The strip detectors cover a larger area beyond the pixel detector. Their design

reduces the number of read-out channels while maintaining good resolution, as the occupancy decreases with the particle flux following a $1/r^2$ dependence. They consist in more than 15,000 modules mounted on carbon-fiber structures. Strip modules can be single- or double-sided, containing about 9.6 million silicon sensors in total [70]. The strip modules feature different geometries [87,88].The strip hit resolutions vary from 23 to 52 µm in the $r - \phi$ plane and 23 µm in the z direction, depending on the distance from the beam axis and η.
The tracker modules are housed inside a temperature controlled outer support tube. To mitigate radiation damage, the pixel and strip detectors are operated at temperatures of approximately −15 °C and −20 °C, respectively. An efficient cooling system is used to dissipate the heat generated by the on-board electronics, ensuring optimal performance and longevity of the detectors.
The pixel and the strip detectors are described in the following.

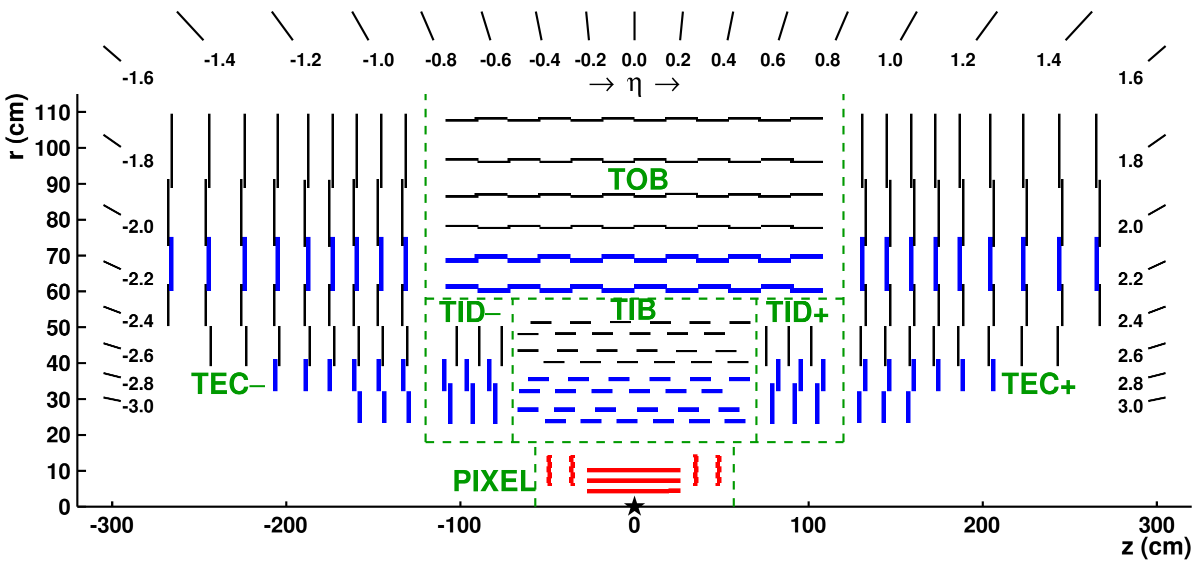


Figure 2.7: Schematic view through the CMS tracker in the $r - z$ plane; it is symmetric about the horizontal line $r = 0$, so only the top half is shown here. The tracker centre, corresponding to the approximate position of the pp collision point, is indicated by a star. Green dashed lines help to understand which modules belong to each of the named tracker subsystems. Strip tracker modules that provide 2D hits are shown by thin, black lines, while those permitting the reconstruction of hit positions in 3D are shown by thick, blue lines. The latter actually each consist of two back-to-back strip modules, in which one module is rotated through a “stereo” angle. The pixel modules, shown by the red lines, also provide 3D hits. Within a given layer, each module is shifted slightly in $r - \phi$ or z with respect to its neighbouring modules, which allows them to overlap, thereby avoiding gaps in the acceptance. Image taken from [89].

Pixel Trackers

- **Pixel Barrel (BPIX)**: The innermost section of the CMS Tracker features three concentric pixel layers located at radii of 4.4, 7.3, and 10.2 cm. These layers consist of silicon pixel sensors approximately 285 µm thick. The BPIX includes 768 modules, ensuring high granularity and hermeticity in the $r - \phi$ plane.

- **Pixel Forward (FPIX)**. At each endcap there are two disks extending from 6 to 15 cm in radius, positioned at $|z| = 34.5$ cm cm and $|z| = 46.5$ cm arranged in a turbine-like geometry with 24 blades rotated of 20 °C with respect to each other, resulting in a total of 672 pixel modules with 7 modules in each blade.

Phase-1 pixel detector The original pixel tracker has been completely replaced during the technical stop at the end of 2016, with a completely new system able to sustain an instantaneous luminosity of $2 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, and high PU environment [90]. Fig. 2.8 shows a comparison between the old pixel detector and the new one; it has an additional layer per region, for a total of four disks in the barrel region and three in the endcaps. Moreover, the new pixel detector is closer to the beam pipe, and it is designed to be lighter (together with the support structure) than the previous one, reducing at same time the material budget. All these characteristics improved the track reconstruction, thanks to the addition of redundancy to mitigate hit losses [91].

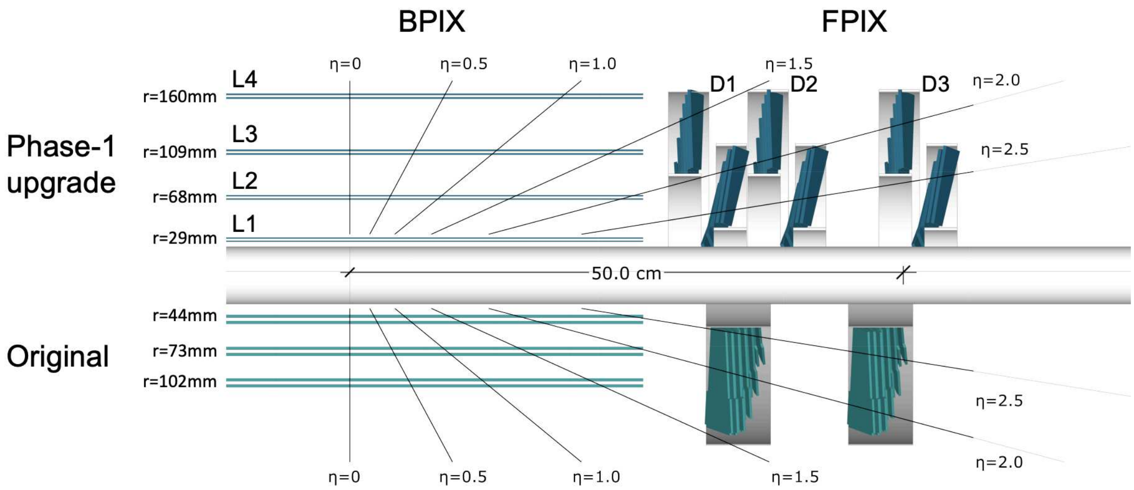


Figure 2.8: Layout of the CMS Phase-1 pixel detector compared to the original detector layout, in longitudinal view [90].

Strip Trackers

- **Tracker Inner Barrel (TIB)**: Located between 20 and 55 cm in radius and extending to $|z| < 65$ cm, the TIB features four layers of silicon microstrip detectors. These sensors are 320 µm thick with strip pitches ranging from 80 to 120 µm. The first two layers use stereo modules for $r - \phi$ and $r - z$ planes.
- **Tracker Outer Barrel (TOB)**: Positioned at $r > 55$ cm and $|z| < 110$ cm, the TOB consists of six layers of 500 micrometer thick silicon sensors with strip pitches of 120 to 180 µm. Due to lower radiation levels, these layers provide stereo measurements.

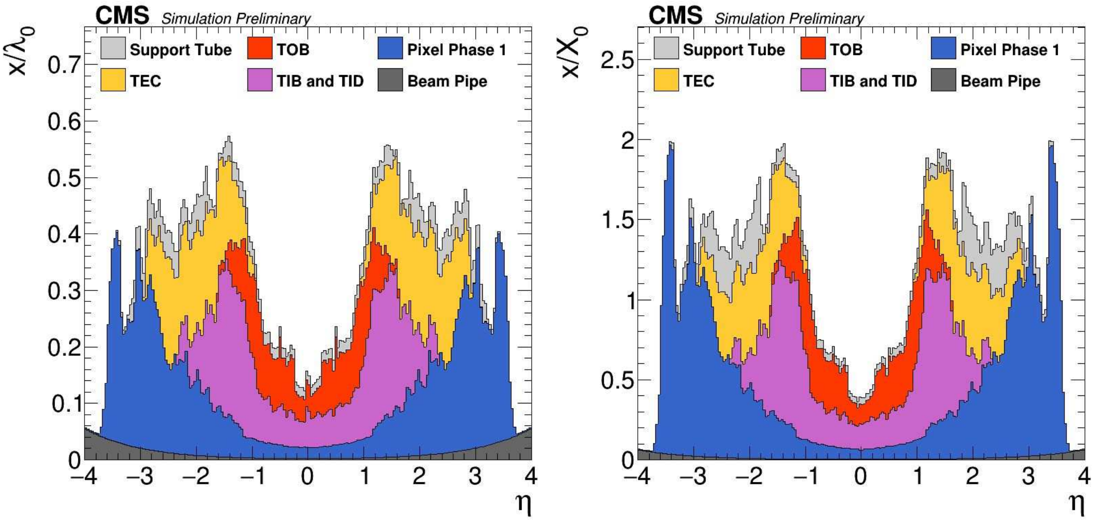


Figure 2.9: Material budget of the Phase-1 tracker for a charged particle originating from the interaction point, shown as a function of η. The left panel represents the material budget in terms of radiation length (X_0), while the right panel shows it in nuclear interaction length (λ) [92].

- **Tracker Endcap (TEC)**: This subsystem comprises nine disks on each side, extending from $120 < |z| < 280$ cm. Modules are arranged in rings centered on the beam line, with strips oriented toward it. Sensor thickness varies from 320 µm for the innermost rings to 500 µm for the outer ones, with stereo modules in some of the rings; the pitch size between layers spans from 96 to 143 µm.
- **Tracker Innter Disks (TID)**: Positioned between the TIB and TEC, the TID features modules arranged in rings centered on the beam line. The first two rings are in a stereo configuration. The sensor thickness is 320 µm, with a pitch size between layers from 97 to 143 µm.

Tracker Performance

Despite operating under high instantaneous luminosities, the tracker performs efficiently, meeting the stringent requirements of the physics program. While crucial for precision measurements, the material composition of the tracker introduces effects such as multiple scattering, energy loss, photon conversions, and electron bremsstrahlung, which complicate trajectory and momentum measurements. Therefore, optimizing the material budget is essential for enhancing detector performance.

The material budget is typically expressed in terms of radiation length (X_0) and nuclear interaction length (λ). The thickness of the silicon sensors traversed by a charged particle varies with pseudorapidity (η), as illustrated in Fig. 2.9. At small $|\eta|$, the material budget is approximately 0.35 X_0, increasing to 1.8 X_0 in the transition region between the barrel and endcap, and then decreasing to 1.1 X_0 at $|\eta| \simeq 2.5$.

The combination of silicon pixel and strip detectors enables precise reconstruction of charged particle trajectories (tracks) by detecting energy deposits in the sensitive regions of the tracker modules (hits).

2.3.2 The Calorimeters

The inner tracker is surrounded by the calorimetric systems: the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) [93, 94]. Their primary goal is to accurately measure the energy and the direction of electrons, photons, and hadrons produced in pp collisions.
Both calorimeters were designed to meet compactness requirements, as they are situated within the superconducting solenoid. At the same time, they are required to reach high granularity, fast response times, and resistance to radiation.
The ECAL is designed to provide precise measurement of incoming electrons and photons energies and directions, while hadrons typically traverse the ECAL volume without being stopped.
To handle hadrons, the HCAL is designed to absorb their energy and measure it. It plays a critical role in the reconstruction of final states involving hadronic jets or non-interacting particles like neutrinos. Its hermeticity and full geometrical coverage enable the determination of missing transverse energy (MET, defined in Section 3.3.6) in events. Additionally, HCAL provides independent energy measurements that enhance the identification of muons, electrons, and photons as detected by the ECAL and the muon system (Section 2.3.6).
Due to their fast response times, both the ECAL and HCAL are crucial for triggering purposes, especially for triggers targeting the identification of high-energy objects such as electrons, photons, and jets, which may be a signature of potentially interesting physics events.
In the following, detailed descriptions of ECAL (Section 2.3.3) and HCAL (Section 2.3.4) are presented.

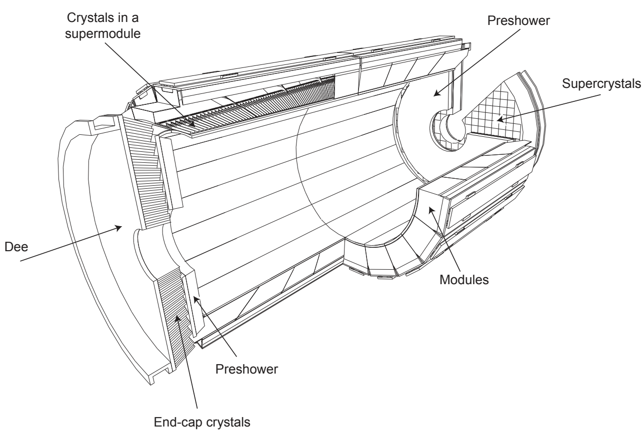
2.3.3 The Electromagnetic Calorimeter (ECAL)

The CMS ECAL was designed to detect the Higgs boson decay channel $H \to \gamma\gamma$, which was crucial for its discovery. To achieve accurate measurements of the invariant mass of the photon pairs, both energy and angular resolutions are essential, particularly for distinguishing closely located photons and reducing background from $\pi^0 \to \gamma\gamma$ decays.
The ECAL is a homogeneous and hermetic calorimeter covering a region up to $|\eta| < 3$, composed of lead tungstate ($PbWO_4$) scintillating crystals, 61200 in the barrel and 7324 in each endcap. When photons and electrons interact with the crystal material, they produce a particle shower, depositing their energy primarily through ionization and excitation of the material. This process generates scintillation light, and signals from individual crystals are used to convert this light into energy measurements. Adjacent energy deposits are grouped into clusters, representing the total energy from a single particle interaction.
Lead tungstate was selected for its high density ($\rho = 8.28\,\mathrm{g/cm^3}$), small radiation length ($X_0 = 0.89\,\mathrm{cm}$) and Molière radius ($\lambda_M = 2.2\,\mathrm{cm}$), along with its rapid light emission (80% of light in $\sim 25\,\mathrm{ns}$) and highly radiation hard (up to 10 Mrad). The

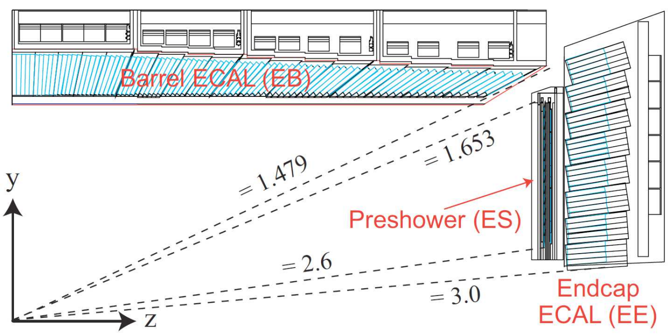
small Molière radius provides good two-shower separation (particularly in high PU environment) by reducing the area over which the energy is summed. The $PbWO_4$ light yield is $30\gamma/\text{MeV}$ necessitates the use of high-gain fast and radiation tolerant photodetectors, wchich can operate inside the CMS magnetic field and are not sensitive to particle traversing them. Silicon Avalanche Photodiodes (SiAPDs) are employed in the barrel region, while Vacuum Photodiodes (VPTs) are used in the endcaps.
The ECAL consists of several subsystems: the ECAL Barrel (EB), the ECAL Endcaps (EE) at the two sides, and a sampling ECAL preshower (ES) in front of EE with high granularity suited for discriminating neutral pions from prompt photons. The layout of the ECAL is shown in Fig. 2.10.

- The **ECAL Barrel (EB)** has an internal radius of 1.29 m and is divided in 36 supermodules, each long half of the barrel length, covering up the region $|\eta| < 1.48$. The crystals are 23 cm, corresponding to 25.8 X_0, with a superficial area of $21.8 \times 21.8\,\text{mm}^2$, and are tilted at 3° ;
- The **ECAL Endcaps (EE)** cover the range $1.48 < |\eta| < 3.0$ and are located at $|z| = 3.154$ m. In each endcap there are two aluminum semicircular plates which include structural units of 5×5 crystals, called "supercrystals". They are 22 cm long, corresponding to $\sim 24.7\,X_0$, and have a superficial area of $28.6 \times 28.6\,\text{mm}^2$.
- The **ES** is placed in front of the EE in the $1.653 < |\eta| < 2.6$ region, to enhance photon identification capabilities, and in particular to optimize the π^0 detection. It is a sampling calorimeter made by two planes of silicon strip detectors with 1.9 mm pitch and absorber lead disks of $2X_0$ and $3X_0$ depth, respectively.

The transverse granularity of $\Delta\eta \times \Delta\phi$ is 0.0175×0.0175 in the barrel, while in the endcaps the value is the same of the barrel for the first 7 towers of the EB, decreasing to of $\Delta\eta \times \Delta\phi \simeq 0.05 \times 0.05$ for the last 5 towers.



(a)



(b)

Figure 2.10: Fig. 2.10a: Layout of the CMS electromagnetic calorimeter, showing the barrel supermodules, the two endcaps and the preshower detectors [95]. Fig. 2.10b: Geometric view of one quarter of the ECAL [96].

The ECAL energy resolution for electrons up to energies between 20 and 500 GeV in beam tests can be expressed as [93]:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E\,(\text{GeV})}} \oplus \frac{b}{E\,(\text{GeV})} \oplus c \tag{2.7}$$

The three contributions correspond to the stochastic fluctuations (a), electronic noise (b), and constant term (c), respectively. The stochastic term includes contributions from the shower containment, the number of photoelectrons and the fluctuations in the gain process. The noise term of corresponds to a single channel noise of about 40 MeV. The constant term, which dominates the energy resolution for high-energy electron and photon showers, depends on non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter, single-channel response uniformity and stability. Estimations ofthe contributions of these terms to the energy resolution based on the beam testresults and simulations is shown in Tab. 2.2.

Contributing term	Barrel ($\lvert\eta\rvert = 0$)	Endcap ($\lvert\eta\rvert = 2$)
stochastic term, a	2.7%	5.7%
Total noise term, b (MeV)	210 MeV	915 MeV ($E_T = 245$ MeV)
Total constant term, c	0.55%	0.55 %

Table 2.2: Contribution to the energy resolution in barrel and endcap, assuming constant luminosity operation. The values for the noise term correspond to the energy reconstructed in a 5×5 crystal array [93].

2.3.4 The Hadronic Calorimeter (HCAL)

The ECAL is surrounded by the HCAL, whose purpose is to measure the energy of charged and neutral hadrons, as they typically pass through the ECAL, where only $\sim 30\%$ of their energy is deposited, before being absorbed in the HCAL [97,98].
The HCAL is a sampling calorimeter, alternating between absorber and active material layers, and consisting of four sub-detectors: the the HCAL Barrel (HB), HCAL Endcap (HE) and Hadron Forward (HF) are placed inside the magnet; as the CMS solenoid coil restricts the available space immediately after the ECAL, the HCAL system features also the Hadron Outer (HO) to improve the hadron shower containment for high energy hadrons, placed outside the magnet coil.
Its almost hermetic design and shower containment are driven by the need to achieve a precise measurement of the Missing Transverse Energy (MET) (Section 3.3.6), a good energy resolution for hadronic jets (Section 3.3.4) and to minimize the muon misidentification, strongly reducing the punch-through hadrons in the muon chambers (Section 2.3.6).
Figure 2.11 shows a schematic layout of one quadrant of the CMS detector where HCAL subsystems are indicated.

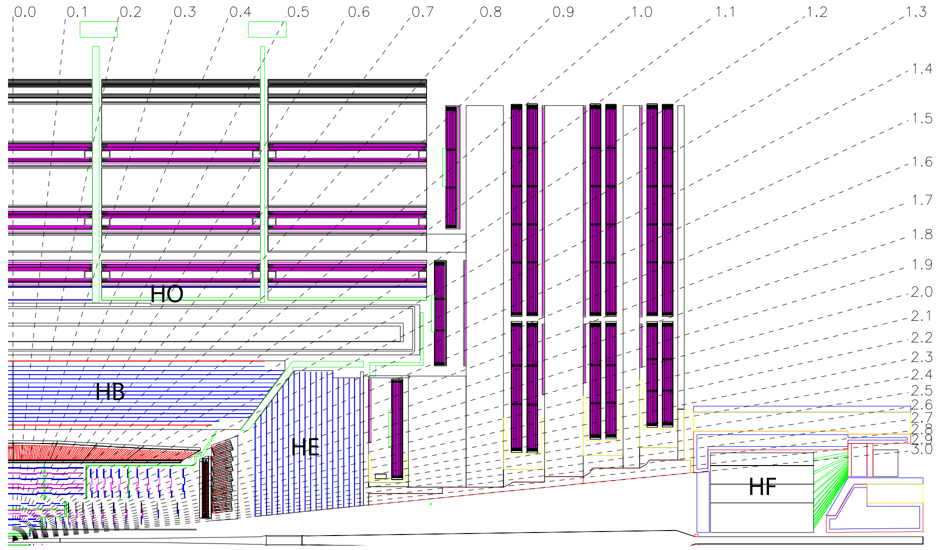


Figure 2.11: Longitudinal view of one quadrant of the CMS in the $r - z$ plane. The HB, HE, HO and HF components of the HCAL detector are highlighted. The tracker and ECAL sub-detectors are visible in the inner part of the detector and the muon chambers are visible in the outer part [70].

- **The HCAL Barrel (HB)** covers the $|\eta| < 1.4$ region and extends radially from $r = 1.77$ m up to $r = 2.96$ m; 18 identical wedges, each segmented in four azimuthal angle (ϕ) sectors, shape the two parts in which the barrel is divided. The active medium in the scintillator is composed of 70000 tiles, connected to wavelength shifter fibres that collect the emitted light. The tiles placed in the same ϕ region are grouped in scintillating units, called tray. Scintillators, with thickness $3.7 - 9$ mm,interspace the absorbing layers. The absorber consists of a front steel plate, 40 mm thick, 8 brass plates 50.5 mm thick plus other 6 brass plates 56.5 mm thick and finally a 75 mm thick back steel plate. The two steel plates in the front and the back of the HB are used to ensure structural strength. The effective thickness in units of nuclear absorption length, λ_I , seen by a hadron entering in the barrel grows as $1/\sin(\theta)$, reaching $10.6\lambda_I$ at $|\eta| = 1.3$.
- The **Hadron Outer (HO)** covers the same η region as the HB, located outside the magnet coil, which uses the coil as an additional absorber to catch the tails of hadronic showers escaping from the HB and avoids the misidentification of muons. It contains 10 mm thick scintillators and is divided into five η regions, each called ring and covering 2.54 m along z. The central ring has two layers of scintillators at radial distances of 3.82 m and 4.07 m respectively, divided by a 19.7 cm thick piece of iron (the tail catcher iron). All other rings have a single scintillator layer at a radial distance of 4.07 m. The additional absorber material of the HO corresponds to an interaction length of about $1.4\lambda_I/\sin(\theta)$. The HO allows increasing the minimal interaction length of the HCAL at $\eta = 0$ to $11.8\lambda_I$.
- **The HCAL Endcap (HE)** extends the coverage to $1.3 < |\eta| < 3$, and is divided into two sections: eight internal and five external towers with different granularities. Its hermetic structure consists of 79 mm thick brass plates with 9 mm gaps to

accommodate the scintillators, providing a total interaction length of $\approx 10\lambda_I$.

- **The Hadron Forward (HF)** is located at $|z| = 11.2\,\text{m}$ and covers the extreme forward regions to $3.0 < |\eta| < 5.0$. It is placed outside the muon chambers. It is designed to resist to the very high particle fluxes in an high luminosity environment, which in the region $4.5 < |\eta| < 5.0$ results in a radiation dose that approaches 100 Mrad/year for the design LHC operating luminosity. The HF is composed of a 1.65 m long cylindrical steel absorber ($\approx 10\lambda_I$), and the active part uses radiation hard quartz fibres made of a fused silica core and polymer hard-cladding. In the transverse direction, the HF forms a ring around the beam line with an inner radius of 12.5 cm and an outer radius of 1.30 m. The HF photodetectors, readout system, and electronics were upgraded at the end of 2017 [99]. The HF detector initially used photomultiplier tubes to collect the Cherenkov light, which have been replaced by multi-anode tubes to reduce the rate of anomalous signals; thanks to the introduction of precision timing measurements and longitudinal depth segmentation, the pattern recognition capabilities in the HCAL were improved, granting an additional handle for background rejection.

The HCAL provides significant granularity in the $\eta - \phi$ plane, with a typical tower size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$, equivalent to the area of a 5×5 array of the ECAL crystals, which ensures precise spatial resolution for energy measurements. In the forward regions ($|\eta| \geq 1.7$), the granularity is reduced to 0.175×0.175 due to the large area and extreme particle flux. In the following a more detailed description of HCAL sub-detectors is given.

The combined ECAL and HCAL energy resolution is parametrized with the relation:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E\,(\text{GeV})}} \oplus b$$

where, in analogy to Eq. 2.7, a accounts for the stochastic term, and b for effects that are non-linear with the energy. Both parameters are obtained in beam-test measurements, yielding $a = 85\%$ ($a = 198\%$) and $b = 7\%$ ($b = 9\%$) in the barrel (endcap) region. Despite the larger stochastic term in the forward region, the energy resolution remains similar, as jets tend to be much more energetic in this region [100].

2.3.5 Superconducting Solenoid

One of the key features of the CMS detector is the superconducting solenoid, which generates a strong magnetic field of 3.8 T [101]. This field is crucial for precise charge and momentum measurements of particles produced in collisions.

The CMS magnet consists of a superconducting solenoid surrounded by an iron return yoke. It uses a high-purity aluminum-stabilized conductor and employs indirect cooling via thermosyphon, along with complete epoxy impregnation—technologies previously demonstrated in the ALEPH and DELPHI experiments at LEP and H1 at HERA. To handle the high magnetic field, current, and stored energy, a four-layer winding design was adopted. This design incorporates a novel conductor with a larger

cross-section, capable of withstanding an outward pressure (hoop stress) of 64 atm. The Rutherford-type cable is co-extruded with pure aluminum to act as a thermal stabilizer.
The solenoid itself is composed of four layers of NbTi superconducting wire. With a length of 12.5 m and a diameter of 5.9 m, the solenoid provides a uniform magnetic field across the central region of CMS, which is crucial for consistent particle tracking. A cryogenic system cools the solenoid to an operational temperature of approximately 4.5 K using liquid helium. The cooling is essential to allow the magnet to carry the large electrical currents (18.5 kA) required to generate the 3.8 T field, as shown in Fig. 2.12, without significant resistance.

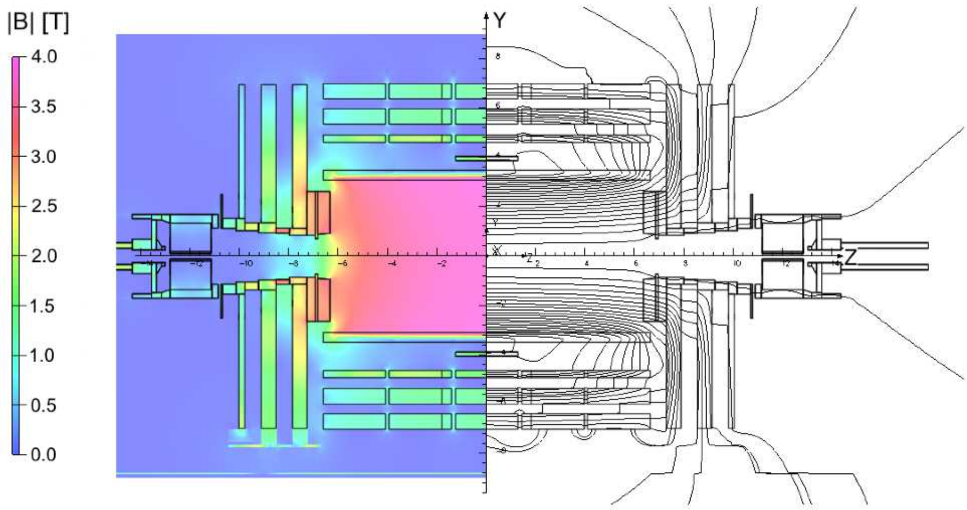


Figure 2.12: Map of the predicted $|B|$ field (left) and field lines (right) for a longitudinal section of the CMS detector by a magnetic field model at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb [102].

The solenoid features a cold bore diameter of 6.3 m, housing the inner tracking system, ECAL, and most of the HCAL. This design minimizes interference from magnetic components, preserving the precision of energy measurements.
To reduce radiation and interaction length while maximizing space for the inner tracking system, electromagnetic calorimeter, and most of the hadronic calorimeter, the solenoid employs a relatively thin coil design with a radial thickness of 0.312 m. This design optimizes space utilization without compromising performance.
The magnet operates with an high energy density of about 11.6 kJ/kg. To maintain stable operation and prevent damage to the coil, the cryogenic system must meet stringent requirements, ensuring a temperature margin of 1.8 K above the operating temperature.
The magnetic field loop is closed by a massive iron yoke weighing approximatively 107 t. This yoke, designed for structural support, consists of five barrel wheels and six endcap disks, which also host the muon chambers. The return field, around 1.8 T, directed opposite to the main field within the solenoid, helps improve the momentum resolution of muons by bending their trajectories in the opposite direction as they pass through the tracking and muon detection regions (Section 2.3.6). This dual bending

effect enhances the overall precision of muon charge and momentum measurements.

2.3.6 Muon System

The muon system is designed to identify muons and measure their charge and transverse momentum (p_T) in the range from 1 GeV to 1 TeV. In this range, muons traverse the entire detector volume, while most other interaction products are absorbed by the ECAL and HCAL. For this reason, the muon system is located outside the superconducting solenoid, making it the outermost component of the CMS detector [103].
Beyond muon identification and momentum measurement, the muon system also plays a crucial role in triggering thanks to its excellent time resolution, enabling fast muon detection and precise bunch crossing identification. The system [70, 102] covers the pseudorapidity range $|\eta| < 2.4$ and consists of gaseous detectors embedded within the layers of the steel flux-return yoke. This arrangement allows a traversing muon to be detected at multiple points along its trajectory. Three types of gas ionization chambers form the core of the CMS muon system: Drift Tube Chambers (DT), Cathode Strip Chambers (CSC), and Resistive Plate Chambers (RPC). Together, they achieve a muon reconstruction, identification, and isolation efficiency greater than 95% [104]. Additionally, in 2017, supplementary Gas Electron Multipliers (GEM) were installed to enhance performance in the forward region. The layout of these detectors in one quadrant of the CMS detector is shown in Fig. 2.13.

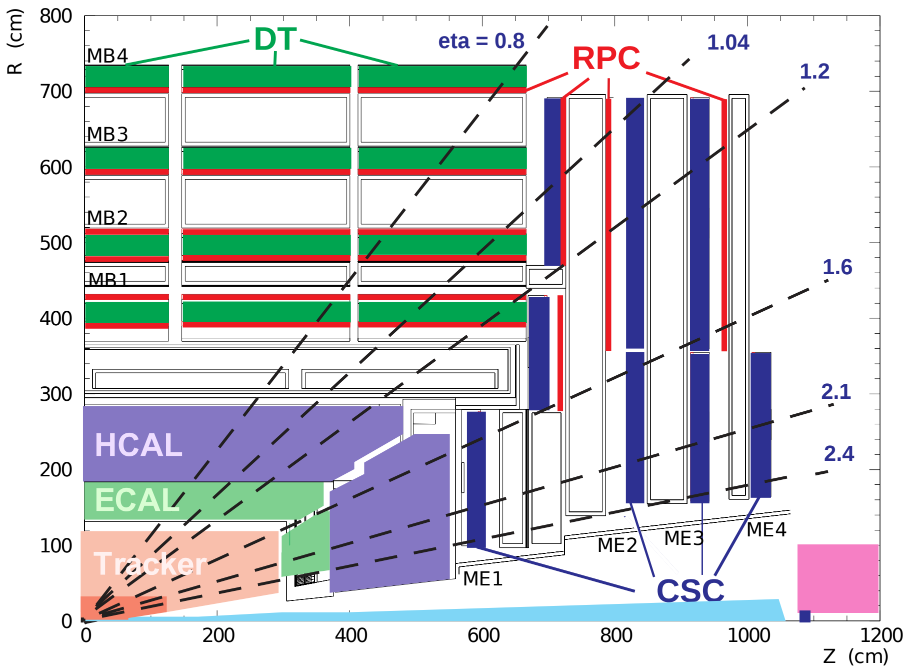


Figure 2.13: Schematic view of one-quarter of the CMS detector. The DT, CSC and RPC stations are shown in green, blue and red, respectively [102].

Drift Tube Chambers (DT)

In the barrel region, where the neutron-induced background is minimal, the muon rate is low, and the 3.8 T magnetic field is uniform and mostly contained in the steel yoke, drift chambers with standard rectangular drift cells are used. These chambers, known as **Drift Tubes (DTs)**, cover the pseudorapidity range $|\eta| < 1.2$. The DT system consists of 250 chambers arranged in 5 wheels along η. Each wheel is divided into four layers ("stations") interspersed among the flux return plates, positioned at distances of 4.0, 4.9, 5.9, and 7 m from the beam axis. In the $r - \phi$ plane, the DTs are arranged in 12 sectors, each spanning a 30° azimuthal angle.
Each of the first three stations contains 8 layers of chambers, arranged in two groups of 4 to measure the muon coordinate in the $r - \phi$ bending plane. Additionally, 4 layers provide a measurement in the z direction along the beamline. The fourth station does not include z-measuring layers. To maximize angular resolution, the two sets of four layers in each station are placed as far apart as possible.
To eliminate dead zones in efficiency, the drift cells within each chamber are staggered by half a cell width relative to their neighbors. This arrangement enhances time resolution by enabling the use of simple meantimer circuits for efficient standalone bunch-crossing identification.
The number of chambers per station and their orientation were optimized to ensure high efficiency in linking muon hits across different stations into a single track while effectively rejecting background hits. Each DT chamber consists of rectangular drift cells measuring $1.3 \times 4.2\,\text{cm}^2$ with a length of 2.4 m. These cells are filled with a gas mixture of Ar (85%) and CO_2 (15%). A uniform electric field and constant drift velocity of approximately 55 μm/s are maintained by placing electrodes at the top and bottom of each cell, while cathodes are positioned along the sides. The position of a muon traversing the chamber is determined by measuring the drift time to the anode wire.
Each drift cell achieves a spatial resolution of approximately 200 μm, leading to an overall resolution of about 100 μm in the $r - \phi$ bending plane and 150 μm in the z direction for an entire station.
All DTs are coupled with one or two RPCs, as described in Section 2.3.6. High-p_T muons can traverse up to 6 RPCs and 4 DTs, producing up to 44 measured points in the DT system. These measurements are crucial for accurately reconstructing muon tracks, particularly for high-p_T muons, which are more likely to have small pseudorapidity values.

Cathode Strip Chambers (CSC)

The endcap regions, where the background rates are higher than in the barrel and the magnetic field is non-uniform, are equipped with CSCs.
A CSC consists of 6 layers of anode wires and cathode strips, filled with a mixture of Ar (45%), CO_2 (50%), and CF_4 (10%), each measuring the muon position in 2 coordinates. The gas is ionized as the muon passes through the chamber, and the induced signals are combined to provide a precise measurement of the muon position.

Each endcap section consists of 4 CSC stations perpendicular to the beam and interlayered between the magnet flux return plates. In each CSC station, 6 layers of approximately perpendicular anode wires and cathode strips provide efficient muon tracking and robust pattern recognition for background rejection. Both cathodes and anodes are readout delivering $r - \phi$ and η measurements, respectively. The anodes also provides good time resolution, which is used to identify the beam-crossing time of a muon. Cathode strips are positioned radially and provide a precision measurement in the $r - \phi$ plane. Anode wires provide a measurement in the radial direction. To ensure continuous muon detection, most CSCs overlap. The CSC stations have $r - \phi$ resolution of about $75 \div 150\,\mu\text{m}$ and a z resolution of about $200\,\mu\text{m}$. The anodes also provides good time resolution ($\approx 3\,\text{ns}$), which is used to identify the beam-crossing time of a muon.
Additional RPCs (described in next section) are deployed in the endcap regions to resolve any ambiguities in the CSC measurements.

Resistive Plate Chambers (RPC)

The RPCs are double-gap chambers operating as gaseous detectors in avalanche mode, in an high electric field achieved using High Pressure Laminate (HPL, commonly known as Bakelite) electrodes with a high bulk resistivity. They are formed of resistive 2 mm thick layers interlaid with a mixture of gas: 95.2% of freon ($C_2H_2F_4$) 4.5% of isobutane $i - C_4H_{10}$ used as a quencher gas to reduce streamer formation, and 0.3% of sulfur hexafluoride (SF_6) to control the background electrons. The RPCs are installed in both the barrel and the endcap regions. When traversed by a muon, an electron cascade is triggered by the high electric field inside the volume and is read out using strips oriented along z in the barrel and r in the endcap regions. There are 4 RPC regions in the barrel and 3 in the endcaps.
Although the spatial resolution is coarser with a strip pitch of $\simeq 1\,\text{cm}$ and no segmentation in the orthogonal direction, these detector provide further information for tracks ambiguity resolution. In addition, being faster than the DTs and CSCs, with a perfect time resolution better than $\approx 1.5\,\text{ns}$, RPCs ensure a robust bunch crossing identification and efficient triggering.

Gas Electron Multiplier (GEM)

They introduced in the endcap region $1.6 < |\eta| < 2.4$ at the end of 2017 to increase the muon system redundancy in these regions. The technology used is the GEM [105], which can accomplish the same physics goals of the RPCs in a high radiation environment.

2.4. The CMS Trigger and Data acquisition system (TriDAS)

At the design luminosity of $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$, the LHC bunch crossing frequency is 40 MHz as beams are spaced by 25 ns. This corresponds to a collision rate of approximately 100 million events per second from proton-proton (pp) interactions. This collision rate is many orders of magnitude higher than what any data acquisition system can record, in particular for CMS it can afford a maximum of 1 kHz. Moreover, while the total pp cross-section is about 80 mb, many processes of interest, such as Higgs boson production, have significantly smaller cross-sections. Even with increased data storage capacity, filtering the most relevant events is crucial to reduce computational costs and optimize data analysis.
To handle this enormous event rate, a robust DAQ and trigger system is employed to preselect events in real-time, determining whether to store or discard each event. The CMS trigger system must be highly efficient at selecting potentially interesting physics signals, which occur much less frequently than non-interesting events. This is achieved through a two-level trigger system: the **Level 1 Trigger (L1T)** [106], which reduces the rate to a maximum of 100 kHz, and the **High Level Trigger (HLT)** [107], which further reduces it to approximately 1 kHz.
The **DAQ system** [108] is designed to efficiently manage the enormous data volume generated while ensuring stable operation under extreme high-luminosity conditions. It works in close collaboration with the trigger system to manage and process the enormous amounts of data generated.
The L1T is a hardware-based trigger that employs custom electronics to perform event selection using information from the calorimeters and muon systems. Since it does not use the full detector information, it must operate with high efficiency while maintaining a low false rejection rate. Events accepted by the L1T are forwarded to the HLT. It has access to complete detector information and employs sophisticated reconstruction algorithms, which are simplified and optimized versions of offline event reconstruction, to further refine event selection. These algorithms run on a dedicated computing farm consisting of approximately 22,000 CPU cores, allowing for rapid event processing and filtering.
Once an event is selected by the L1T, the data is transferred to the DAQ system, which collects data fragments from all subdetectors and assembles them into a complete event. This real-time process, known as **event building** (EVT), ensures that fully reconstructed events are available for further processing in the HLT.
After HLT processing, which reduces the event rate to approximately 1 kHz, the DAQ system transfers the selected events to long-term storage for detailed offline analysis. Additionally, the DAQ system regulates data flow to prevent buffer overflows. If the incoming data rate exceeds the processing capacity of the HLT or storage, the DAQ can temporarily buffer the data or implement predefined strategies to control the event flow.

2.4.1 The Level-1 Trigger (L1T)

As particles collide inside the CMS detector, the subdetectors generate raw data, which are digitized and temporarily stored in the front-end electronics of each subdetector. The L1T system exploits information coming from ECAL, HCAL and muon systems, receiving a reduced version of raw data from the front-end electronics in real-time to quickly analyze a simplified version of the event and apply the selection, using information such as energy deposits or hit patterns. The L1T has $\sim 3.8\,\mu s$ after each beam crossing to decide whether to keep or reject an event.
At the L1 stage, the event undergoes to three steps in sequence: local, regional, and global.

- **Local Trigger**. At this stage, the calorimeters trigger takes energy deposits from ECAL and HCAL and combines them into "trigger towers"[5]. Muon triggers use track segments and hit patterns from all the muons subsystems. These signals are defined as "Trigger primitives" (TP).
- **Regional Trigger**: TPs are further processed at the regional level, where calorimeter electron/photon candidates are defined (Regional Calo Trigger) and combined DT/CSC tracks are identified assigned with physical parameters, as the p_T (Regional Muon Trigger).
- **Global Trigger**: The Global Calorimeter Trigger (GCT) and Global Muon Trigger (GMT) combine regional information to form trigger objects like isolated e/γ candidates (since no information from the tracker systems is available at this stage, electrons and photons have almost the same experimental signature and are reconstructed as e/γ objects), jets, hadronically decaying taus (τ_h), and MET. These information are sent to the L1 Global Trigger (GT), which decides if the event should be accepted based on predefined algorithms.

Each trigger algorithm checks specific conditions, like energy thresholds, and if the event satisfies at least one condition, the L1T accepts the event, forwarding it to the HLT for further analysis.

2.4.2 High Level Trigger (HLT)

Events accepted by the L1T are transmitted to the DAQ system, which collects data fragments from all subdetectors. The DAQ's back-end electronics then assemble these fragments into complete events before passing them to the HLT for further event rate reduction, decreasing the rate from the maximum L1T accept rate of 100 kHz to an output rate of around 1 kHz, with a latency of approximately $350 \div 1000$ ms. This process is managed by event builder units within the back-end system, which oversee the flow of data from the front-end electronics to the HLT.

[5] In ECAL each trigger tower is composed of a group of 5x5 individual crystal cells in the barrel region. In HCAL the trigger towers correspond to specific regions that match the ECAL geometry, with each HCAL tower covering the same pseudorapidity and azimuthal angle range as a group of ECAL crystals.

The HLT primary goal is to identify and retain only the events that are potentially interesting for physics analysis while discarding the rest. It operates using an Event Filter Farm (EVF), a computing cluster composed of thousands of processor cores. The EVF includes filter-builder units responsible for assembling event fragments into complete events and filtering them. These units perform raw data unpacking, event reconstruction, and trigger filtering.
During data-taking, the HLT reconstructs physics objects in real time, including jets, muons, electrons, τ_h, and MET, collectively referred to as *online objects*. Specific identification criteria are then applied to these objects to select relevant events for physics analyses. The reconstruction and identification processes utilize simplified versions of the algorithms used for offline reconstruction, ensuring efficiency while maintaining high-quality selection.
The HLT execution starts with local reconstruction and selection of data from the calorimeter and muon systems, using Level 1 (L1) objects as seeds; this initial step reduces the event rate before the more CPU-intensive process of reconstructing complete tracks in the tracker, which depends on the outcome of the first step.

HLT menu and paths structure

The HLT operates through "HLT paths" organized in a detailed menu. Each path consists of sequences of reconstruction and selection steps that refine physics objects based on events that have already passed the L1T. This method helps reduce the data rate while retaining significant events for analysis. Each path is designed for specific event signatures relevant to studies like searching for top quarks, Higgs and vector bosons. To optimize computational resources, not all HLT paths are executed for every event. Instead, execution is guided by the results of the L1T, allowing multiple L1 algorithms to seed a single HLT path, ensuring efficiency while preserving sensitivity to rare event signatures.

Event Selection and Trigger Information Each HLT path can be in one of four states: it can run, accept, reject an event, or encounter an error. These states are encoded into two bits (since four states can be represented with 2 bits), stored in the trigger bits of the data files. For data analysis purposes, to ensure an event has passed through a specific trigger path, the corresponding trigger bit must indicate the "accepted" status.

Real-Time and Offline Reconstruction The HLT performs real-time reconstruction of physics objects, which are referred to as *online objects*. Due to the limited time window for decision-making ($350 \div 1000\,\text{ms}$), the HLT uses fast, simplified algorithms that prioritize speed over precision. However, once the events are stored for offline analysis, the objects are reprocessed using more accurate algorithms that incorporate data from all CMS subdetectors, resulting in what are known as *offline objects*. These are more finely reconstructed with complex algorithms prioritizing accurancy over speed, and used for physics analyses.

Triggers prescaling The CMS trigger system uses the prescaling rechnique to manage triggers with high event rates and prevent memory saturation during data acquisition. Trigger paths with low thresholds often select a large number of events, so a prescale factor is used to randomly discard some of these events. This technique allows for the collection of representative samples of less rare but still interesting events while maintaining manageable data rates. It is especially beneficial for frequently firing triggers that do not require high precision, thereby conserving resources for rarer, critical events. However, in the search described in this thesis, no prescaling is applied.

Trigger paths in the $X \to HH \to b\bar{b}\tau\tau$ analysis

The trigger paths employed in the search presented in this thesis are specifically designed to detect the τ leptons coming from one of the two Higgs bosons ($H \to \tau^+\tau^-$). These paths require the presence of an electron, muon or hadronic tau; other included paths select a lepton pair made by either an electron or a muon paired with hadronic tau or hadronic tau pairs. Each of these triggers must meet certain criteria, including kinematic thresholds, quality of particle identification, and isolation requirements. Due to the nature of the $\tau\tau$ decay, which involves the production of multiple neutrinos, up to four in total, some of the energy from the decay is undetected. For this reason, an additional trigger path requiring a certain amount of the MET is employed. This MET threshold helps capture events where significant energy has escaped detection along with the neutrinos. The various trigger paths used in this analysis ensure comprehensive coverage and help maximize the detection of a potential $H \to \tau^+\tau^-$ decays. Detailed descriptions of each of these trigger paths and their specific criteria are provided in Section 4.4 of the thesis.

2.4.3 Software and Computing in CMS

The CMS experiment relies on a globally distributed computing infrastructure, the Worldwide LHC Computing Grid (WLCG), to manage the vast data volume efficiently. This system, structured in a tiered architecture, ensures efficient data handling and accessibility for CMS collaborators worldwide (Fig. 2.14).

A key concept in CMS data organization is the **Primary Dataset**, which groups events based on the trigger paths fired during data acquisition. Events selected by the HLT are initially stored in RAW format, containing unprocessed detector signals. These RAW events are sent to Tier-0 (T0) at CERN, where they are categorized into Primary Datasets and processed into RECO (reconstructed) and AOD (Analysis Object Data) formats. RECO data apply reconstruction algorithms to identify high-level physics objects, while AOD is a streamlined format optimized for analysis. To further reduce data size while preserving analysis-relevant information, additional formats such as MiniAOD and NanoAOD are produced, facilitating efficient storage and analysis workflows.

These datasets are stored on tape and distributed to Tier-1 (T1) centers for further processing and analysis. T1 centers, located in various collaborating countries, handle

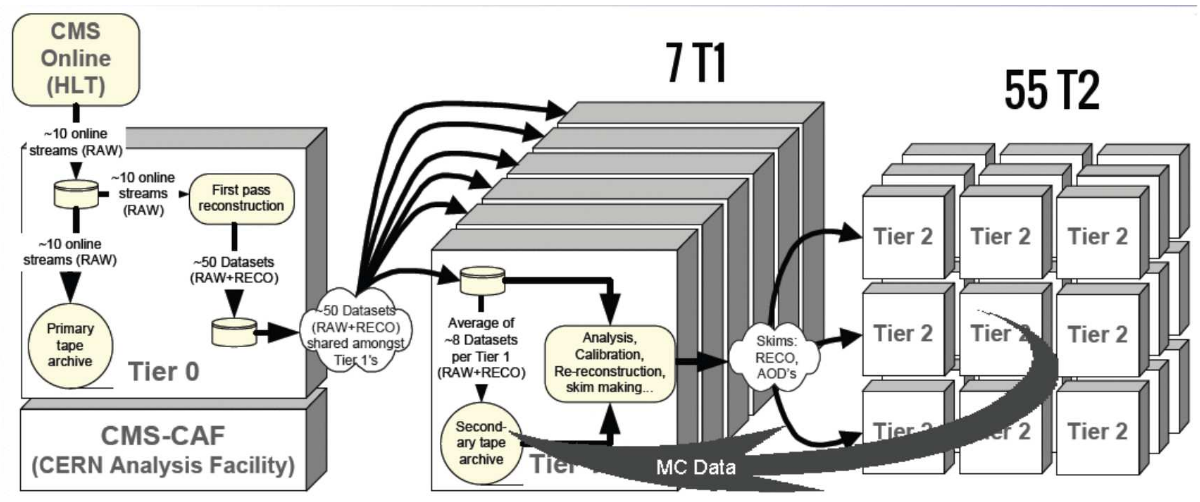


Figure 2.14: The CMS data flow within the tiered computing system. Physics data are grouped into Primary Datasets at Tier-0 and distributed across the WLCG for analysis and storage.

dataset calibration, reconstruction, and long-term secure storage. They also manage Monte Carlo (MC) simulation samples, which are produced at Tier-2 (T2) centers. T2 facilities, primarily located at universities and research institutes, provide the computing resources needed for simulation, calibration, and user-driven data analysis. The Primary Dataset structure ensures efficient handling of events relevant to specific physics analyses, facilitating their targeted processing and distribution across the computing tiers.

CMS Software Framework

The CMS software, `CMSSW`, has been developed to handle all aspects of data processing and analysis within CMS. Written in `C++` and Python, it provides a modular and flexible framework centered around the Event Data Model (EDM), which organizes event information into a `C++` object container.
`CMSSW` integrates with the Primary Dataset system by enabling efficient processing of RAW, RECO, and AOD formats, and storing outputs in ROOT files for further analysis. Its object-oriented design ensures scalability and adaptability to the evolving computational and physics needs of CMS.

2.5. Luminosity measurements at CMS

The luminosity measurement [109] is performed during special calibration runs, called *Van der Meer (VdM) scans* [110], characterized by dedicated LHC setup with well separated beam and using specific subsystems called "luminometers". In general, the measured rate from the luminometers is proportional to the instantaneous luminosity via the σ_{vis} proportionality constant:

$$R = \mathcal{L}_{\text{inst}}\sigma_{\text{vis}} \tag{2.8}$$

Where σ_{vis} is called the *visible cross section*. The purpose of VdM scans is to measure σ_{vis}.

During VdM scans the two beams are separated and then moved across each other. In head-on collisions, when there is no crossing angle and the beams collide with zero relative separation, the istantaneous luminosity for a single colliding bunch i, $\mathcal{L}^i_{\text{inst}}$, and can be expressed as:

$$\mathcal{L}^i_{\text{inst}} = \frac{n^i_{p,1} n^i_{p,2} \text{f}_{\text{rev}}}{2\pi\Sigma_x\Sigma_y} \tag{2.9}$$

It depends on the beam overlap widths in the x and y directions, which are respectively Σ_x and Σ_y, on the number of protons in the bunch ($n^i_{p,1}$ and $n^i_{p,2}$), and on beam revolution frequency in LHC, f_{rev}.
The values Σ_x and Σ_y are obtained by conducting two separate scans that vary the beam separation and measure the rates in the x and y directions. The rates values are fitted with double-Gaussian functions and Σ_x and Σ_y correspond to the widths .

Then Eq. 2.8 can be used to derive the σ_{vis} once the beam overlap widths are estimated and σ_{vis} is obtained inverting the Eq. 2.8 and using $\mathcal{L}^i_{\text{inst}}$:

$$\sigma_{\text{vis}} = \frac{2\pi\Sigma_x\Sigma_y}{N_1 N_2 \text{f}_{\text{rev}}}\mathcal{R}_0 \tag{2.10}$$

Where $\mathcal{R}_0$ is the head-on collision rate. Knowing σ_{vis} and since Eq. 2.8 is valid under all different beam conditions, measuring the rate allows to determine the luminosity in regular physics fills, during which the beam overlap width cannot be measured with the necessary accuracy, and not only in VdM scans. Once the $\mathcal{L}_{\text{inst}}$ is obtained, the integrated luminosity can be computed integrating over time the instantaneous one, as σ_{vis} is a constant term.

In practice, luminometers often show a nonlinear relationship with instantaneous luminosity, PU, and external factors like the LHC filling scheme. To achieve accurate measurements, these nonlinearities must be corrected. Moreover, before fitting to extract Σ_x and Σ_y, it is essential to correct the raw measured rates for background in the detector. This background contribution is estimated independently by analyzing two special 5-minute periods during the VdM scan, when the beams were separated by $6\sigma_b$ (where σ_b is the beam size) in both the x and y directions, ensuring that only background sources, such as beam-induced background or detector noise, contribute.

2.5.1 Luminosity detection techniques at CMS

CMS employs specific detectors and techniques to monitor and measure the luminosity. They are both subdetectors of CMS and dedicated detectors put along the beam axis outside of CMS. The use of multiple luminometer provides independent and robust

luminosity measurements, and measure the same quantity in different acceptance ranges. Each luminometer computes the rate relying on specific quantities as hits, tracks, vertices.
The analysis of luminosity is conducted in two distinct ways. It can be performed online, in real-time, on a bunch-by-bunch basis, providing high precision in a short timeframe. Alternatively, it can be done offline using raw data that have already been collected, which typically allows for greater precision but involves analyzing data per fill over a longer duration.

Online measurements exploit the following techniques and tools:

- **Pixel Luminosity Telescope (PLT)** [111]: a silicon pixel detector installed at 1.75 m from the interaction point, at $|\eta| \simeq 4.2$ and is made of 16 "Telescopes", 8 at each side of CMS. Each telescope is made of 3 silicon pixel planes. The luminosity is measured by counting events in which all the planes register a hit, producing an electric pulse at the collision rate. This initial signal is processed using very fast custom hardware, ensuring excellent statistical precision and effective background suppression. The complete data is then read out at a lower rate.
- **Fast Beam Conditions Monitor (BCM1F)** [112], consisting of 4 half rings put on both sides of CMS interaction point after the PLTs ($\simeq$ 1.8 m). The adopted technology is made by poly-crystalline (pCVD) diamond sensors and silicon sensors and specific ASICs for readout, providing a response time of $\approx$ 6.25 ns.
- **The HF detector** is exploited too. Two algorithms are used: one based on the fraction of occupied towers (HFOC), while the other relies on the sum of the transverse energy (HFET).
- **DT measurements**, exploiting the muon stubs rate in the muon barrel track finder. This technique provides stable and linear integrated luminosity measurements, and therefore they can be used as complementary cross checks.
- **RAMSES**, a LHC detector used for radiation and environmental monitoring [113]. It is made of plastic ionization chambers coated with PE graphite, which detects photons in a specific energy range. It does not make bunch-by-bunch measurements, but as in the DT case, it is less affected by instability or nonlinearities, thus is a good cross check for integrated luminosity measurements.

The PLT and BCM1F exploit a separate DAQ system (BRILDAQ), made of special readout electronics capable of handling the 40 MHz collision rate and thus providing very fast response. Also the DT measurements are available in the BRILDAQ.
The HF has a special readout to perform luminosity measurements, installed on all the rings, even if only two HF rings are used for this purpose.
The online measurements allow for both instantaneous and integrated luminosity estimation: PLT, BCM1F and HF perform bunch-by-bunch scans, thus measuring the instantaneous luminosity, whereas DT and RAMSES give complementary information on integrated luminosity.

Offline measurements are performed on dataset using special trigger mode is to cope higher data rate, which lasts 5 bunch crossings, requiring collisions but not any specific event activity ("zero bias"). The following methods are employed:

- **Pixel Cluster Counting method (PCC)** [114], using the inner tracker with the eclusion of the first layer. Giving the smaller bandwith and the high precision achieved by the tracker, the PCC method provides stable measurements of integrated luminosity over "lumi-sections" time ($\approx 23\,\mathrm{s}$). Luminosity is evaluated considering the number of pixel clusters in the events.
- **Vertex Counting (VTX)**. This method is employed for low PU luminosity measurements, with negligible background, providing good profiles of beam shape overlaps. The disadvantage is that the statistics is low, since it uses the same dataset collected from the PCC method.

The offline methods provide only integrated luminosity measurements. In particular, the PCC method exhibits a very small dependence on PU and an excellent linear detector response with the increasing of the luminosity.

Chapter 3

Event reconstruction at CMS

The raw data collected by the Compact Muon Solenoid (CMS) detector in proton-proton (pp) collisions are processed into high-level objects, such as tracks, vertices, and calorimeter clusters. These high-level objects are then used to reconstruct physical objects, including electrons, muons, tau decays involving hadrons (also referred to as *hadronic* tau decays), jets, b-jets, and Missing Transverse Energy (MET), which are essential for this analysis. This chapter provides a brief overview of the algorithms responsible for converting subdetector signals into reconstructed physical objects and assesses their performance.

3.1. Detector objects reconstruction

Each particle produced in pp collisions at CMS leaves a unique signature in the subdetectors, as illustrated in Figure 3.1.
Muons (azure line in Fig. 3.1) lose only a small fraction of energy in the innermost detector layers and are detected in both the tracker and the muon chambers. Electrons (red line in Fig. 3.1) and photons (blue dashed line in Fig. 3.1) are absorbed within the Electromagnetic Calorimeter (ECAL) (Section 2.3.3); however, only electrons leave a visible trajectory in the tracker. Hadrons (green line in Fig. 3.1) pass through the ECAL, depositing a small fraction of their energy, and release the majority of their energy in the Hadronic Calorimeter (HCAL) (Section 2.3.4). The trajectories of their charged components are recorded in the tracker. Neutrinos escape CMS undetected, but their presence can be inferred from the transverse energy imbalance in the event.
Events accepted by the trigger system (Section 2.4) are reprocessed offline, using more sophisticated and slow algorithms. Then the Particle Flow (PF) algorithm [115], described in Section 3.3, combines the raw signals from the subsystems to reconstruct and identify high-level detector objects, like tracks in in the inner tracker and in the muon system, clusters in calorimeter and pre-shower, primary and secondary vertices produced in the event (Section 3.1.2) with maximum efficiency and low background.

The basic of the algorithms utilized for this purpose by both High Level Trigger (HLT) and offline processing are briefly described in the following sections.

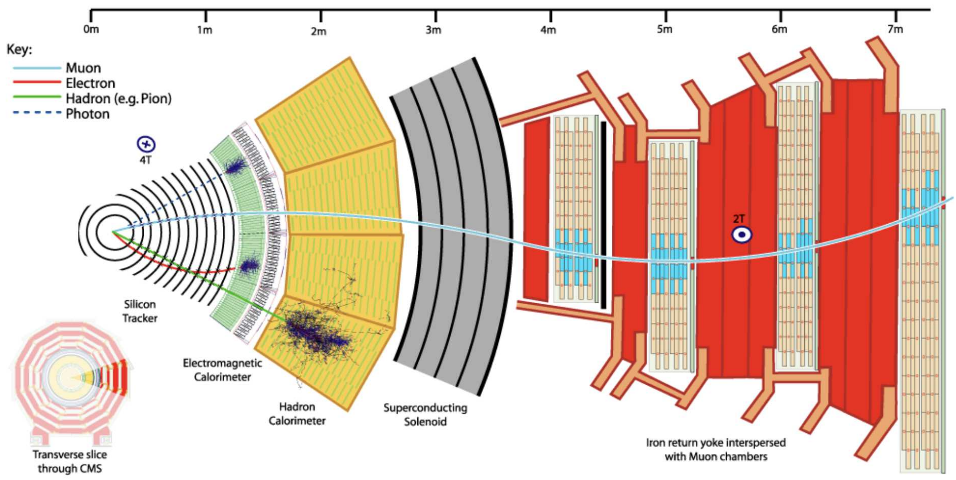


Figure 3.1: A sketch of particles passing through the CMS detector in a transverse slice of it, from the beam interaction region to the muon detector. The represented muon and charged pion are positively charged, whereas the electron is negatively charged [115].

3.1.1 Tracks

Within the tracker region, the nearly uniform magnetic field and minimal material allows the trajectory of a charged particle to be modeled as a helix. The five parameters used by CMS to determine track trajectory in this region are: d_0, z_0, ϕ, $\cot(\theta)$ and q/transverse momentum component (p_T). The last parameter is ratio between the charge of the track,q and its p_T defined at the impact point (x_0, y_0, z_0), the closest track trajectory point to the nominal beam axis $(z = 0)$; θ and ϕ are respectively the polar and the azimuthal angles of the track momentum vector at a given point point; d_0 is the impact parameter of the track in the transverse plane, defined through the coordinates of the impact point, $d_0 = -x_0 \cos\phi_0 + y_0 \sin\phi_0$. The longitudinal component of the momentum in the point of coordinates (x_0, y_0, z_0), p_z, is directly related to $\cot(\theta)$: $p_z = p_T \cot(\theta)$. The trajectories of charged particles and its momenta are defined as *tracks*. Figure 3.2 gives a visual picture of a part of the helix trajectory for a positive charged track and the variable used to define the track parameters.

Accurate track reconstruction is essential for achieving high efficiency in vertex reconstruction and precise momentum measurement. Track parameters are estimated using the spatial positions of the charge deposits (hits) created by a charged particle traversing the layers of silicon detectors. Due to high particle multiplicity within the tracker, reconstructing particle tracks is a computationally intensive task. To manage the combinatorial complexity, an iterative tracking approach is employed. The core concept of iterative tracking is to identify the initial iterations on finding the easiest tracks to identify—those with relatively high p_T and originating near

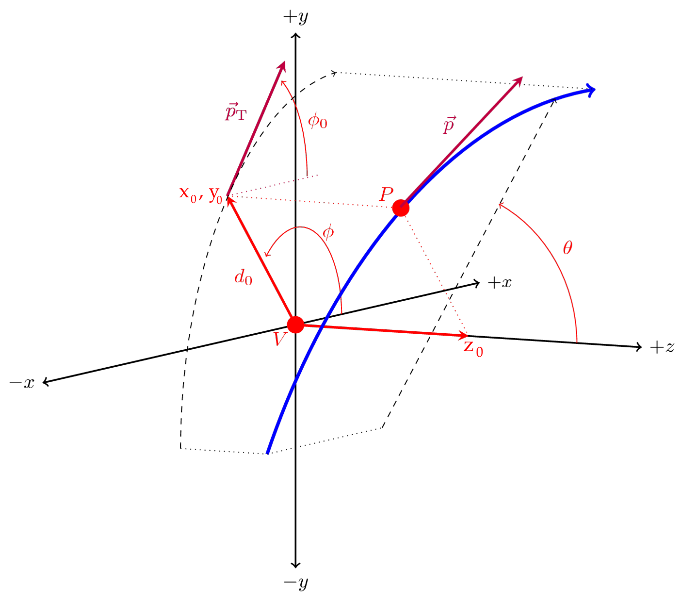


Figure 3.2: Visual picture of a part of the helix trajectory for a positive charged track and the variables used to define the track parameters. Readapted from [116]

.

the interaction region. After each iteration, hits associated with identified tracks are removed from consideration, thereby reducing combinatorial complexity and simplifying the subsequent iterations. This process allows later iterations to focus on more challenging track types, such as those with low p_T or significant displacement from the interaction region.

Tracks are initially roughly reconstructed at the HLT level using hit information from the tracker. The offline reprocessing begins with local reconstruction in the inner tracker, where signals exceeding a certain threshold generate hits. These hits represent estimated points of track intersections for charged particles within the silicon detector layers. A local orthogonal coordinate system (u, v), defined in the plane of each sensor, is employed to estimate the hit position and its uncertainties.

Subsequently, a transformation from the local to the global coordinate system is performed accounting for discrepancies between the assumed and actual detector positions, as well as surface deformations identified through the alignment procedure. The detector element uncertainty is then combined with the intrinsic uncertainty of the local hits.

During earlier reconstruction stages, some tracks may not be linked to actual charged particles and are classified as fake tracks. To mitigate the fake rate, quality requirements are imposed, evaluating parameters such as the number of hit layers, the final fit's χ^2 value, and the track's compatibility with a primary interaction vertex.

The final reconstruction steps are intended to find tracks that originate outside the Beam Spot (BS) and to recover tracks that were not reconstructed in earlier iterations.

Iterative tracking procedure

The tracking process is divided into six iteration steps. The first three iterations (0–2) reconstruct prompt tracks originating near the BS, while the remaining three iterations (3–5) reconstruct non-prompt tracks with origins outside the BS [89].
Within each iteration, the Combinatorial Track Finder (CTF) algorithm, an adaption of the combinatorial Kalman Filter (KF) [117,118], is used with configuration parameters tailored to each specific iteration. The CTF algorithm consists of four logical steps [89], which are applied to progressively refine and identify track candidates through each stage of reconstruction:

- **Seeding**: the initial trajectory parameters and uncertainties are determined using seeds formed from two or three hits in the pixel detector from different layers.
- **Pattern recognition/track finding**: Seed trajectories are extrapolated to all tracker layers to search for compatible outer hits, updating traversed material and track parameters iteratively. A backward search identifies duplicate tracks using a *trajectory cleaner*, which retains the best track based on the highest number of hits and the lowest χ^2 value.
- **Track fitting**: The trajectory is refined using “inside-out” and “outside-in” KFs. The first filter processes hits from the inner layers, increasing uncertainties, while the second smooths results from the first, yielding helix parameters for further extrapolation.
- **Track selection**: applies quality requirements to reject fake tracks which are not associated with a charged particle. These criteria are optimized in function of p_T, η, and the number of hit layers in the tracker.

Track reconstruction performance

The track reconstruction performance is evaluated in terms of tracking efficiency (the fraction of simulated charged particles that can be associated to the reconstructed tracks), fake rate (fraction of reconstructed tracks that are not associated with any simulated particle) and track parameters resolution.
Both the offline and HLT performances have been assessed using $t\bar{t}$ simulated samples corresponding to the 2016 − 2017 data, with an average pile-up (PU) of 35 [119,120]. Track reconstruction was updated to take advantage of the enhanced features of the upgraded pixel detector introduced at the beginning of 2017.
For high-purity quality tracks with $|\eta| < 2.5$ and $|d_0| < 3.5\,\text{cm}$ the efficiency exceeds 85%, as shown in Figure 3.3 (left). The fake rate as a function of p_T is shown in Figure 3.3 (right). The fake rate is lower across all p_T and η ranges in the 2017 simulations.

Fig. 3.4 shows the resolution of the transverse (d_0, left) and longitudinal (z_0, right) impact parameters as a function of η. In all cases tracking performance for the 2017 show that the upgraded pixel tracker significantly enhances the impact parameters resolution. Most inefficiency is due to the tracks produced by pions and kaons, which undergo elastic and inelastic nuclear reactions, and electrons, which loss energy

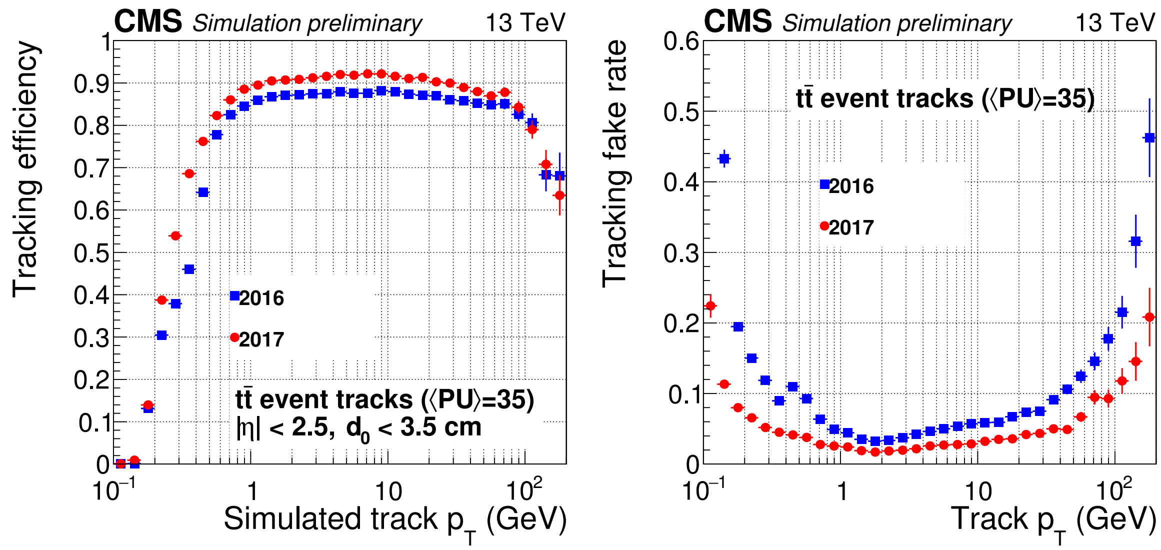


Figure 3.3: Track reconstruction efficiency (left) and fake rate (right) as a function of reconstructed track p_T for 2016 and 2017 detector configurations [119,120].

through bremsstrahlung [89]. The tracking performances at HLT are shown in Fig. 3.5.

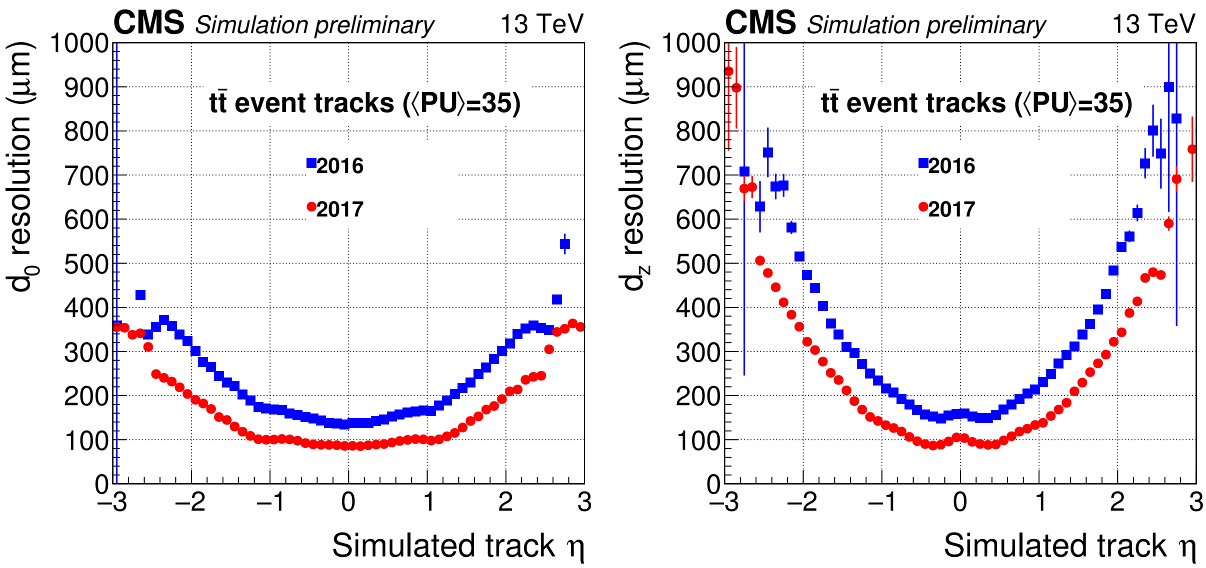


Figure 3.4: Track d_0 (transverse impact point, left) and d_z (longitudinal impact point,right) resolutions as a function of the simulated track η for 2016 and 2017 detector configurations [119,120].

The contributions to the total efficiency from the different tracking iterations are shown in the different colors. The initial three iterations are shown in shades of blue while the contribution of the doublet recovery iteration is shown in violet. The performance that would be achieved with a perfect detector is shown as a red line. The efficiency is robust against the presence of PU, decreasing only slightly with the number of additional interactions. The performance of the doublet-seeded recovery is also independent of the PU conditions.

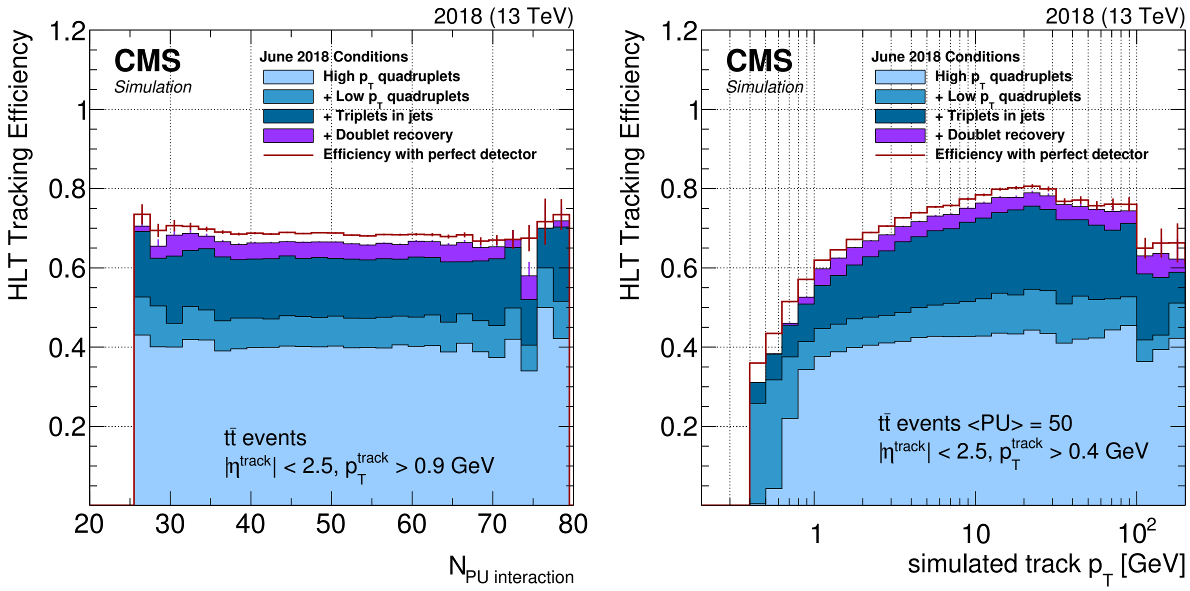


Figure 3.5: Tracking efficiency as a function of the number of PU interactions and simulated track p_T respectively in left and right [121,122].

3.1.2 Primary Vertex reconstruction

The precise reconstruction of all pp interaction vertices within an event allows for the clear distinction between the primary hard-scatter interaction vertex and the additional particles generated by PU collisions. Accurate identification of the Primary Vertex (PV), together with supplementary information of secondary vertices (coming from subsequent particle decays), plays a critical role in determining the properties of physical objects, such as hadronic jets originating from b-hadrons (Section 3.3.4), where the displacement between the primary and secondary vertices is one of the most relevant features for their identification.
The reconstruction is carried out using the available reconstructed tracks and is comprised of three steps [84,85,89]:

- **Tracks selection**: Tracks that meet specific quality criteria and are compatible with originating from the BS.
- **Clustering**: The selected tracks are then clusteredat in groups based on a possible common vertex at their point of closest approach to the center of the BS through a deterministic annealing algorithm [123].
- **Fitting position for each interaction vertex**. The position of each vertex is determined by fitting the associated tracks using an adaptive vertex fitter [124].

The primary vertices, which are spatially separated along the beam axis, are ranked according to the quadratic sum of the transverse momenta of their associated tracks, $\sum p_T^2$. The PV with the highest $\sum p_T^2$ is identified as the hard-scatter vertex, the one used in analyses, while the remaining vertices are classified as PU vertices.

Vertex resolution

For the data recorded by CMS from 2016 to 2018, the number of pp interactions (number of reconstructed vertices) per bunch crossing are reported in Figure 3.6 . The performances have been evaluated using $t\bar{t}$ simulated samples corresponding to the 2016 − 2017 data, featuring an average PU of 35. Similar results are obtained on data [119]. The transverse and the longitudinal PV resolution depends on the number of associated tracks, as shown in Fig. 3.6: in both case the resolution ranges from 10 to 100 µm. Overall, the 2017 detector configuration shows better performance than 2016 one.

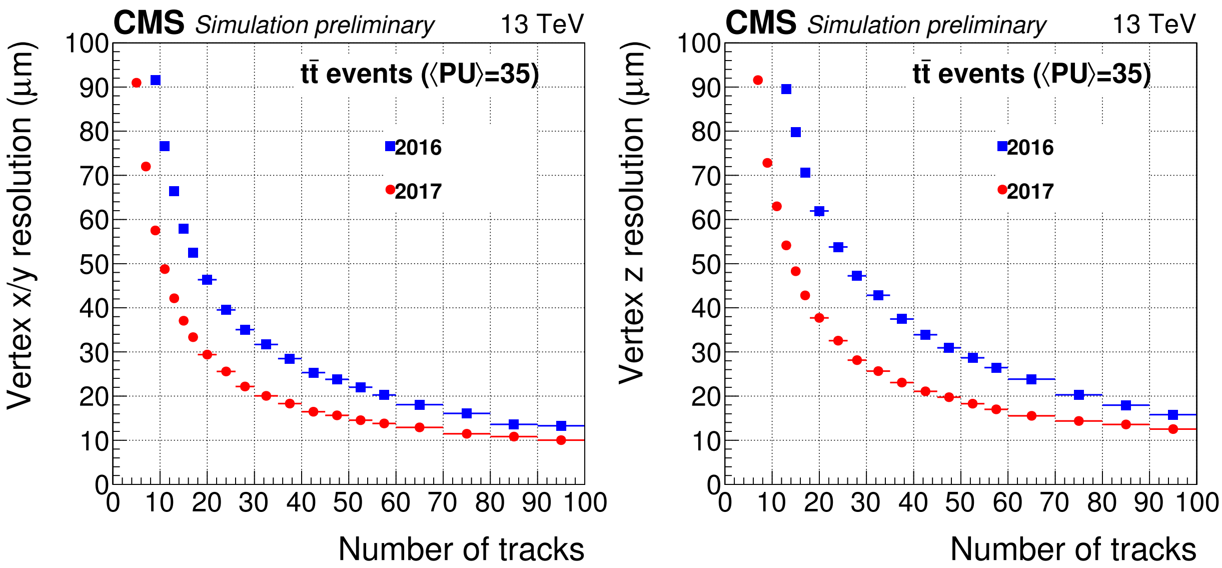


Figure 3.6: Vertex transverse (left) and longitudinal (right) resolution as a function of the number of tracks used in the vertex fit for the 2016 (blue) and 2017 simulated $t\bar{t}$ samples [119, 120].

3.1.3 Calorimeter clusters

Clusters are formed from energy deposits in the cells of the calorimeters (Sections 2.3.3 and 2.3.4). The clustering algorithm in the calorimeters serves four main purposes: detecting and measuring stable neutral particles, distinguishing them from charged hadron energy deposits, reconstructing electrons and their bremsstrahlung photons, and improving energy measurements for charged hadrons with poorly determined track parameters.

The clustering process is customized for particle-flow reconstruction to ensure high efficiency, even for low-energy particles, and to separate nearby energy deposits. It employs an iterative clustering technique that follows the typical lateral profile of the shower. Clustering is performed separately in each subdetector (ECAL barrel/endcaps, HCAL barrel/endcaps, and preshower layers), while in the Hadron Forward (HF), energy deposits from individual cells directly define clusters.

The process begins by defining a cluster seed, which is a calorimeter cell with an energy deposit exceeding a predefined threshold and higher than that of neighboring cells, which can be the four closest (sharing a side) or the eight closest (including those sharing only a corner). In the ECAL endcaps, the seed cells must also meet an additional energy threshold due to varying noise levels. After seeds are identified, clusters are grown by iteratively adding neighboring cells with energy above a noise-based threshold. One topological cluster may contain more than one seed. Therefore, to refine cluster reconstruction, an expectation-maximization algorithm is used based on a Gaussian-mixture model. This model assumes that the energy deposits in the cluster cells arise from several Gaussian distributions corresponding to the number of seeds. Key parameters of the model include the amplitude and the mean coordinates of each Gaussian, while the width is fixed based on the calorimeter type. The algorithm iteratively estimates the expected energy fraction in each cell due to each Gaussian and updates the model parameters based on maximum likelihood, adjusting these parameters until convergence is reached, at which point the final cluster positions and energies are assigned.

3.2. Physical Objects

Historically, at hadron colliders, object reconstruction in collision events was carried out by analyzing signals from specific subdetectors. This straightforward approach can be significantly improved by correlating data from all detector layers, allowing for the reconstruction of final state particles using comprehensive information from all subsystems. This method is known as the *Particle Flow (PF)* approach. This was first developed and implemented by the ALEPH experiment at Large Electron Positron Collider (LEP) [125], and it is currently driving the design of detectors for potential future colliders.

3.3. Particle Flow

The PF algorithm [115] in CMS aims to reconstruct and identify all stable particles produced in pp collisions by optimally combining information from all subdetectors. A key feature enabling this integration is the fine spatial granularity of the detector layers. Without this, coarse-grained systems may merge signals from different particles, reducing the ability to distinguish and reconstruct them. The segmentation of CMS subdetectors ensures efficient separation of particle signals.
The algorithm starts with the low-level information (hits in the silicon tracker, hits in the muon system and energy deposits in the calorimeters) to build the PF basic elements: tracks (Section 3.1.1), muon tracks (Section 3.3.2) and calorimeter clusters (Section 3.1.3). Each element undergoes standalone reconstruction for optimal precision. Charged-particle tracks are reconstructed iteratively in the inner tracker (Section 3.1.1), while muon tracks are first identified using pattern recognition and linear fitting in the Drift Tube Chamber (DT) and Cathode Strip Chamber (CSC) subsystems. These

are later combined with hits from Resistive Plate Chamber (RPC) and refitted using a Kalman filter [117,118] to form the so called standalone muon tracks (Section 3.3.2). Clustering algorithms (Section 3.1.3) reconstruct the position and energy of calorimeter deposits.

The PF basic elements serve as input for reconstructing higher-level detector objects, including primary and secondary vertices and fundamental physics objects: muons, electrons, photons, and neutral and charged hadrons. The algorithm first connects multiple PF elements originating by a single particle across the different subdetectors into composite structures, known as PF blocks, by forming geometric links.

These links are only created between neighboring elements. The quality of these connections is quantified using a link distance parameter. Once the PF blocks are formed, they are used to identify individual particles, known as PF candidates, which serve as the foundation for the reconstruction of higher-level objects for physics analyses, such as jets, hadronically decaying taus (τ_h) and MET.

Link algorithm

In the CMS detector, a single particle generates multiple PF elements across various subdetectors. A linking algorithm reconstructs the particle by connecting these elements, with accuracy depending on detector granularity and particle density. However, material interactions before the calorimeters and muon system can cause trajectory deviations and secondary particle production, limiting full reconstruction.

To optimize computation, the algorithm links only the nearest neighbors in the (η, ϕ) plane using a k-dimensional tree. A distance parameter quantifies link quality, and elements are grouped into PF blocks through direct or indirect connections.

Tracks are first extrapolated from its last recorded hit in the tracker, then it is extended to the ECAL preshower (ES) layers, the ECAL (at depths where a typical electron shower peaks), and the HCAL (at one interaction length). The track is linked to a cluster if its projected position falls within the cluster area, which is determined by the union of all its calorimeter cells in the ($\eta - \phi$ for HCAL and ECAL Barrel (EB), or $x - y$ plane for ECAL Endcap (EE) and ES). The cluster area is expanded by up to one cell size to account for detector gaps, shower maximum uncertainties, and multiple scattering effects in low-momentum particles. If multiple clusters are linked to a single track, or vice versa, the closest link is selected. Additional links between charged particle tracks and calorimeter clusters are established to take into account Bremsstrahlung by electrons. In this case, tangents to the tracks are extrapolated to the ECAL from the intersection points between the track and each of the tracker layers, if the extrapolated tangent position is within the boundaries of a cluster, as defined above, then the cluster is linked to the track as a potential Bremsstrahlung photon. Additionally, photon conversion is handled by linking two tracks associated with an electron-positron pair if their combined direction matches the extrapolated track tangent. Calorimeter cluster-to-cluster links connect HCAL to ECAL clusters and ECAL to ES clusters when the more granular cluster falls within the envelope of the less granular one. The link distance is also defined as the distance between the two

cluster positions, in the (η, ϕ) plane for an HCAL-ECAL link, or in the (x, y) plane for an ES link. The closest link is retained when multiple connections exist. ECAL clusters are linked to superclusters if they share at least one ECAL cell.
Charged-particle tracks can also be linked through a shared secondary vertex for nuclear-interaction reconstruction, provided the vertex includes at least three tracks and meets an invariant mass threshold. Finally, inner tracker tracks are linked to muon system tracks to form global muons (Section 3.3.2). A global fit is performed, and the two tracks are linked if the χ^2 of the fit is below a threshold. If multiple combinations satisfy this condition, the one with the lowest χ^2 is chosen.

PF Particle Reconstruction and Identification

The PF algorithm efficiently processes events due to the high granularity of the subdetectors, ensuring that each PF block contains only a few elements from one or a small number of particles. This design allows the logic of the algorithm to remain unaffected by the event's particle multiplicity, with computing time scaling linearly. The PF particles are reconstructed and identified using the PF blocks obtained from the linking algorithm. The reconstruction proceeds iteratively, and once a PF particle is identified, the corresponding elements are excluded from further iterations.

The reconstruction sequence follows a well-defined order. Muon candidates are identified first, and their associated tracks and clusters are removed (Section 3.3.2). In the second iteration, electron reconstruction and identification follows, combining data from the inner tracker and calorimeters (Section 3.3.3). Due to significant tracker material, electrons frequently emit bremsstrahlung photons, and photons often convert into e^+e^- pairs, leading to further bremsstrahlung emissions. The aim of the electron reconstruction is to collect the energy of all bremsstrahlung photons. Because of these similarities in tracking and energy deposition, photon and electron reconstruction are performed together.
At this stage, tracks with a p_T uncertainty exceeding the calorimetric energy resolution expected for charged hadrons are masked, reducing the rate of misreconstructed tracks at high p_T . On average, this requirement rejects about 0.2% of the tracks in multijet events, with only 10% of these being genuine high-momentum hadrons, whose energies are better measured in the calorimeters than in the tracker.
Following this, the remaining elements in the PF block are used to identify and reconstruct charged hadrons, neutral hadrons, and photons originating from parton fragmentation, hadronization, and decays in jets (Section 3.3.4). Hadrons undergoing nuclear interactions in the tracker material, which produce secondary particles, are reconstructed separately. If an incoming track is identified, it is used to refine the reconstruction but is otherwise ignored in subsequent steps of the PF reconstruction.
Finally, once all PF blocks have been processed and all particles have been identified, a global event description is obtained, and a post-processing step is performed to further refine the reconstructed event.

3.3.1 Leptons

Isolation

Distinguishing between **prompt leptons** produced in the primary hard interaction (e.g., in decays of W, Z, or Higgs bosons) and **non-prompt leptons** from the weak decay of particles produced during hadronization (e.g., decays of heavy-flavor hadrons) is essential. **Prompt leptons** are typically isolated, with minimal nearby energy deposits, as they emerge from the primary interaction. In contrast, **non-prompt leptons** are often surrounded by hadronic activity from the hadron's decay, leading to less isolation and larger energy deposits in their vicinity. **Lepton isolation** helps distinguish these by measuring the energy or momentum around the lepton.

When using all the PF candidates the isolation is usually quantified by estimating the total p_T of the particles inside an isolation cone of size ΔR around the direction of the lepton and is function of the sums of p_T of PF charged and neutral hadrons ($p_T^{charged}$ and $p_T^{neutral-had}$ respectively), of PF photons (p_T^{γ}) and PU corrections ($p_T^{pile-up}$). It is indicated with I_ℓ and is defined as:

$$I_\ell = \sum \mathrm{p_T^{charged}} + \max\left[0, \left(\sum \mathrm{p_T^{neutral}} + \sum \mathrm{p_T^{\gamma}} - p_T^{pile-up}\right)\right] \tag{3.1}$$

Where the $p_T^{pile-up}$ refers the neutral PU contribution, which depends from the nature of the lepton as in the isolation cone contribution from PU is different. In the case of the electron the neutral PU is the major contribution, e.g. from photons in the ECAL (Section 2.3.3), while muons primarily deal with charged PU particles, corrected using vertex association techniques and a smaller neutral energy contribution.
CMS uses the PF relative isolation variable I_ℓ^{rel} to assess the isolation for leptons like electrons, muons. This variable is defined as the ratio of the lepton isolation and the transverse lepton momentum p_T^ℓ:

$$I_\ell^{rel} = \frac{I_\ell}{p_T^\ell} \tag{3.2}$$

In the case of muon, the isolation is:

$$I_{PF}^{\mu,\mathrm{abs}} = \sum_{h^\pm,HS} p_T^{h^\pm} + \max\left[0, \left(\sum p_T^{h^0} + \sum p_T^{\gamma} - \Delta\beta \sum_{h^\pm,PU} p_T^{h^\pm}\right)\right] \tag{3.3}$$

The detrimental effect of PU, which spoils the isolation, is mitigated by considering only charged hadrons linked to the hard-scatter (HS) vertex, while neutral contributions from PU are estimated using charged hadrons from PU vertices. The factor $\Delta\beta = 0.5$ is applied to account for the neutral particles coming from PU vertices in a cone around the lepton. This factor approximates the ratio of neutral particles to charged hadrons produced in pp collisions. The corrected energy sum from neutral particles is limited to being positive or zero. Electron isolation is determined and validated using similar techniques, and the use of smaller exclusion cone to remove contributions from the electron bremsstrahlung. In this analysis, the PF relative isolation is employed to categorize muons and electrons as either "isolated" (Iso) or "non-isolated" (NonIso).

3.3.2 Muons

Muons with p_T above 5 GeV typically pass through the inner tracker, leaving hits, then traverse the calorimeters with minimal energy deposition, before reaching the muon chambers and leave additional hits. This allows high-p_T muons to be reconstructed independently in both the inner tracker and the muon system, improving the ability to distinguish muons from hadronic activity and providing more accurate measurements of their momentum, particularly for high-p_T muons.

Muon Reconstruction

Muon reconstruction begins at the Level 1 Trigger (L1T), where Trigger Primitives (TP) (Section 2.4.1) from CSCs, DTs, and RPCs generate muon candidates with a rough measurement of their p_T. At HLT the process employs full detector information for improved muon identification. The use of simplified versions of the offline algorithms allows to reduce computing time and resources. In the offline reconstruction, candidate muon tracks are initially reconstructed separately in the inner tracker (tracker muons) and the muon system (standalone muons). Based on these objects, three reconstruction approaches are used:

- The **standalone muon tracks** reconstruction relies solely on the muon chambers. It begins by constructing track segments in the DT and CSC chambers using pattern recognition and linear fitting techniques. The track segments are then combined with RPC clusters and refitted using a Kalman filter [117] to form standalone muon tracks (see also Section 3.3).
- **Tracker Muon Tracks** are reconstructed extrapolating tracks in the inner tracker with $p_T > 0.5$ GeV and $|\vec{p}| > 2.5$ GeV from the inner tracker to the muon system ("inside-out"). If the extrapolated track position aligns within the uncertainties with at least one muon segment, the corresponding track is identified as a **tracker muon**.
- **Global muon tracks**: each standalone muon track is matched to a tracker track if the parameters of both tracks, when propagated to a common surface, are compatible (the "*outside-in*" method). Once a match is established, the hits from both, the inner tracker and the standalone muon system, are combined. These combined hits are then used to perform a joint fit with the KF [117] resulting in the creation of a global muon track, which exploits information from both the tracker and the muon system to improve the overall track resolution and accuracy.

The high efficiency of the tracker track and muon segment reconstruction guarantees that $\sim 99\%$ of the muons produced within the geometrical acceptance of the muon system are reconstructed either as a global muon track or as a tracker muon track, or both. If global and tracker muons share the same track in the tracker, they are merged into one candidate. Tracker muons are efficient in areas of the CMS detector with less instrumentation and for low p_T muons, but they often only match segments in the innermost muon station, increasing the risk of misidentification due to hadron shower remnants. Global muon reconstruction, which incorporates standalone muon tracks, enhances efficiency for muons passing through multiple stations, thereby reducing misidentification rates. This method also improves the p_T measurement of global muons, particularly for those with p_T > 200 GeV. In contrast, standalone muon tracks have poorer momentum resolution and a higher presence of cosmic muons compared to global or tracker muons.

Muon identification

Muon Identification (ID) in CMS relies on several key variables to achieve the desired balance between efficiency and purity for different analyses. These variables assess the quality of the track fit (χ^2), the number of hits in the tracker or muon system, and the degree of matching between tracker tracks and standalone muon tracks. The muon segment compatibility is determined by propagating the tracker track to the muon system and evaluating how well it matches in terms of position and direction. The algorithm returns a score between 0 and 1, with 1 indicating a perfect match.
For further verification, a kink-finding algorithm breaks the tracker track into segments and compares them. A significant (χ^2) indicates a mismatch between the track segments. Other variables, such as compatibility with the primary vertex (the point of origin of the muon), are also considered.
CMS uses various identification strategies depending on the analysis objectives:

- **Loose Muon ID** is optimized to identify prompt muons from the primary vertex and those from light/heavy flavor decays. Loose muons are selected by the PF algorithm (Section 3.3) and are either tracker or global muons.
- **Medium Muon ID**, designed to identify for prompt muons and those from heavy flavor decays. Medium muon is a loose muon satisfying stricter track quality requirements, such as using hits from more than 80% of inner tracker layers. Depending on whether the muon is a tracker or global muon, segment compatibility and track fit quality constraints vary, targeting an overall efficiency of 99.5% for muons from W and Z bosons.
- **Tight Muon ID**: aims to suppress muons from decays in flight and hadronic punch-through. A tight muon is a loose muon with hits in at least six tracker layers (including one pixel hit) and must be reconstructed as both a tracker and global muon. Segment matching in at least two muon stations is required for the tracker muon. The global muon fit must have χ^2/N_{dof}. Stringent vertex compatibility conditions are applied (e.g., tight muon transverse impact parameter $|d_{xy}| < 0.2\,\text{cm}$ and longitudinal impact parameter $|d_z| < 0.5\,\text{cm}$).
- **Soft Muon ID**, designed to tag low-p_T muons, particularly in B-physics. A soft muon is a tracker muon which must meet strict tracker track purity and segment matching criteria, but looser requirements for compatibility with the primary

vertex (e.g., $|d_0| < 0.3$ cm and $|z_0| < 20$ cm). Developed specifically for B-physics studies, this ID is loosely compatible with the primary vertex.

- **High-p_T Muon ID**, optimized for $p_T > 200$ GeV muons. High momentum muons must pass the same requirements as tight muons, but the χ^2 requirement on the global muon fit is removed to account for radiative effects at high p_T. Instead, a relative p_T uncertainty of $\sigma(p_T)/p_T < 30\%$ s applied to ensure accurate momentum measurement.

These muon identification categories provide a flexible framework for selecting muons in various CMS physics analyses, each optimized for different physics goals. In this thesis the Medium Muon ID for background rejection (third lepton veto, see Section 4.5) and the Tight Muon ID for signal selection (Section 4.6.1) are used.

Muon efficiency

The total efficiency for muons is measured using the Tag and Probe (TnP) method in the $Z \to \mu\mu$ simulated samples, beginning with tracker tracks as probes. The total efficiency is computed by factorizing it into four components [104]:

$$\epsilon_\mu = \epsilon_{\text{track}} \times \epsilon_{\text{reco+ID}} \times \epsilon_{(\text{ISO}\,|\,\text{ID})} \times \epsilon_{\text{trg}} \tag{3.4}$$

The ϵ_{track} is the traking efficiency, computed on tracker track muons. The second term comprises the reconstruction and ID efficiency, including the matching of this muon to the tracker track. The $\epsilon_{(\text{ISO}\,|\,\text{ID})}$ is the isolation efficiency computed relatively to a fixed ID. The last term is the trigger efficiency. The application of Equation 3.4 is dependent on the specific needs of each analysis. If an analysis does not require isolation $\epsilon_{(\text{ISO}\,|\,\text{ID})}$ is removed from the equation and ϵ_{trg} is computed relative to reconstructed muons without an isolation requirement.

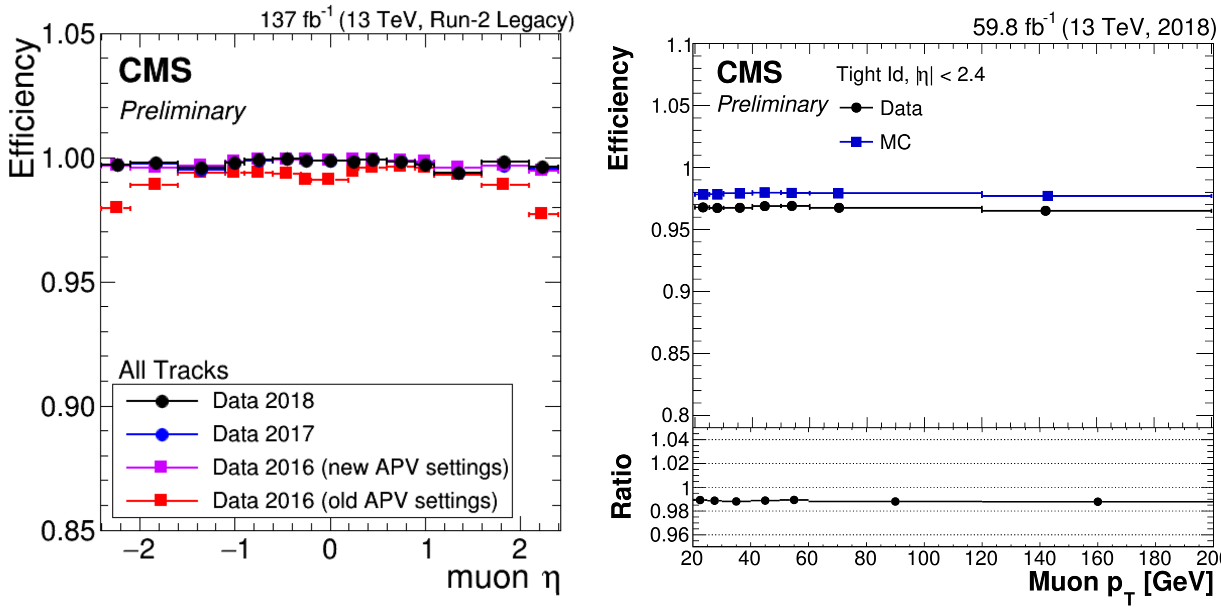
Figure 3.7a shows the ϵ_{track} as a function of muon η for the full Run 2 data divided by the data-taking periods. The efficiency is shown for tracker and global muon tracks collections: a good agreement of the tracking efficiency is observed for 2018, 2017 and 2016 after the new APV setting.

Tight and loose Isolation (Iso) Working Points (WP) are defined to achieve efficiencies of 98% and 95%, respectively [126].

The analysis presented in this thesis targets prompt isolated muons requiring a Tight ID and a Tight isolation for muon candidates. Criteria for muon Tight Iso will be defined in Section 4.6.1.

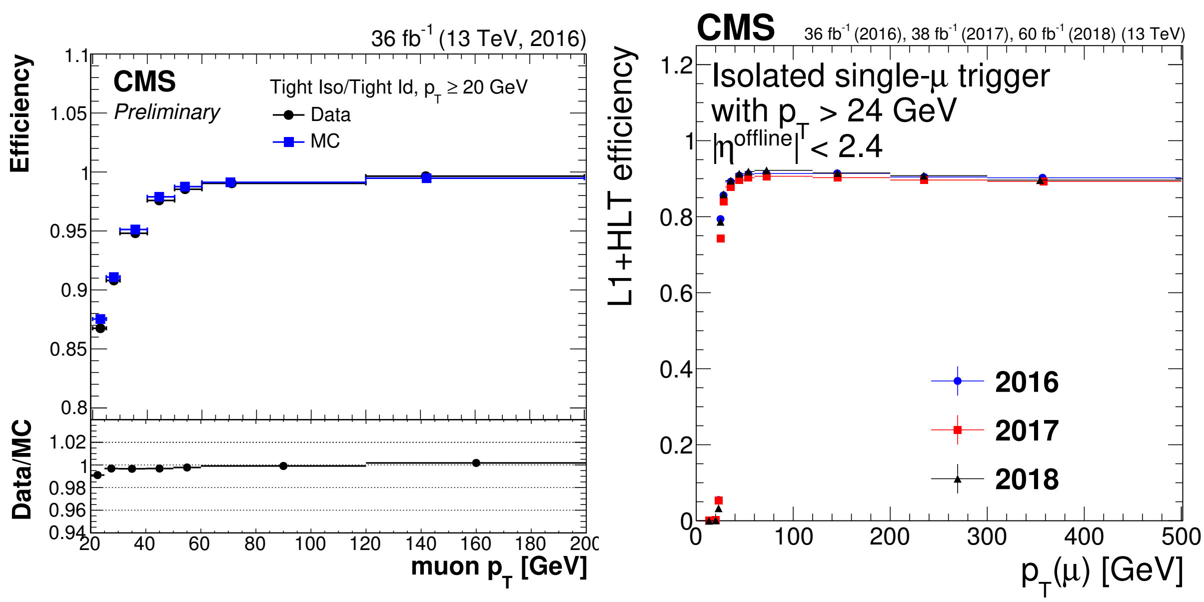
An example of the efficiency for Tight ID muons $\epsilon_{\text{reco+ID}}$ as a function of p_T is shown in Fig. 3.7b for 2018 data and simulated samples [127], while an example of the combined efficiency $\epsilon_{(\text{ISO}\,|\,\text{ID})}$ as a function of p_T muon is shown in Fig. 3.8a for 2016 data and simulated samples [126].

In the analysis, one of the trigger path used requires a single muon with $p_T >$ 24 GeV with a muon candidate satisfying the criteria of Tight ID and Tight Iso (Section 4.6.1). One example of the trigger efficiency for this trigger path with full Run 2 data sample of $Z \to \mu\mu$ [129] is shown in Fig. 3.8b.



(a) (b)

Figure 3.7: Fig. 3.7a: Full Run 2 muon tracking efficiency as a function of pseudorapidity η for global and tracker tracks [128]. Fig. 3.7b: Tight lepton ID efficiency as a function of p_T for 2018 data and Monte Carlo (MC) [127].



(a) (b)

Figure 3.8: Fig. 3.8a: Tight Isolation efficiencies vs muon p_T, with respect to Tight ID with 2016 data [126]. Fig. 3.8b: efficiency of combined L1T and HLT requiring isolated single muon with $p_T > 24$ GeV as a function of offline muon p_T for full Run 2 data [129].

3.3.3 Electrons

Electrons deposit almost all of their energy in the ECAL, and produce hits in the tracker layers. Electron reconstruction is performed by a multi-step process that exploits both the ECAL and the inner tracker information, whereas the identification employs both cut-based and machine learning based algorithms. The reconstruction and identifcation at HLT (described in Section 2.4.2) use similar techniques to offline ones, with the difference that algorithms running at HLT are limited by the CPU time (260 ms), thus are simplified and more conservative selection criteria are applied.

Electron Reconstruction

As an electron traverses the material before reaching the ECAL, it may undergo bremsstrahlung radiation, leading to the emission of photons that can subsequently convert into electron-positron pairs. Consequently, the electron arrives at the ECAL as a shower of multiple particles rather than a single one. Electron and photon reconstruction within PF exploits available PF blocks after muon reconstruction, combining inner tracker and ECAL information to improve momentum estimation.

The reconstruction algorithm begins with **cluster formation**, where ECAL crystals with energy above a threshold ($\sim$80 MeV in the EB, $\sim$300 MeV in the EE) are grouped into clusters [115]. The highest-energy cluster in a region serves as the **seed cluster**, requiring a minimum deposited transverse energy $E_T > 1$ GeV. Energy deposits may be shared between clusters, and a Gaussian shower profile determines the fractional assignment of energy. **Superclustering** follows, merging ECAL clusters within a defined geometric window around the seed to account for photon conversions and bremsstrahlung losses. The **Gaussian Sum Filter (GSF) track fitting algorithm** is applied subsequently. Due to its computational cost, seeding is performed first, using either ECAL-driven or tracker-driven methods. ECAL-driven seeding selects Super-Clusters (SC) based on energy and matches them to inner tracker seeds. Tracker-driven seeding matches generic tracks [1] to ECAL clusters using selection criteria and a Boosted Decision Tree (BDT). ECAL-driven seeding is highly efficient for high-energy, isolated electrons (>95% for $E_T > 10$ GeV), while tracker-driven seeding recovers low-p_T and nonisolated electrons ($\sim$50% for $p_T > 3$ GeV), improving efficiency in less precise ECAL regions (e.g., barrel-endcap transitions). The GSF tracking algorithm reconstructs electron trajectories from both types of seeds. The track parameters evaluated at each successive tracker layer are used by the KF algorithm to iteratively build the electron trajectory,applying a Bethe–Heitler distribution to model electron energy loss [130]. Once KF track candidates are formed, a GSF fit refines the parameters, approximating energy loss with Gaussian distributions [89]. A dedicated algorithm [131] identifies tracks from photon conversions into e^+e^- pairs. Reconstructed objects (ECAL clusters, SCs, GSF tracks, and conversion tracks) are imported into the PF algorithm, which links them into particle flow blocks. These blocks resolve into electron and photon

[1] In this case "generic tracks" refer to the collection of tracks (not specific to electrons) selected with $p_T > 2$ GeV, reconstructed from hits in the tracker through the KF algorithm [117,118].

objects based on a GSF track or an SC. A BDT classifier associates GSF tracks with ECAL clusters, incorporating track kinematics, SC observables, and matching variables. For tracker-driven electrons, only the BDT determines the association.

Electrons and photons present a unique challenge in the PF framework due to their composite nature, consisting of multiple clusters and tracks. Misidentifications, such as hadronic components in jets being reconstructed as e/γ objects, can affect jet energy measurements. A common "loose" selection helps distinguish electrons and photons from hadronic components. A BDT classifier, incorporating shower shape, isolation, and tracking information, applies separate models for isolated and nonisolated cases. Additional criteria, such as the energy-to-momentum ratio (E/p), hadronic energy fraction (H/E), and track-based variables, refine the selection. Photon candidates must be isolated and exhibit shower shapes consistent with genuine photons. Misclassified objects undergo further reconstruction, reinterpreting their elements as neutral or charged hadrons or nonisolated photons, ensuring robust electron and photon reconstruction within the PF framework [132].

Electron identification

Electrons are identified as primary charged-particle tracks linked to ECAL energy clusters. The primary background sources include electrons from photon conversions, jets misidentified as electrons, and electrons from semileptonic decays of b- and c-quarks. At CMS the electron ID is performed with either a cut based or a Multi Variate Analysis (MVA) technique. In this analysis, a BDT is used. Furhter details on MVA techniques and BDT are reported in Appendix A.1.3. This MVA technique exploits a variety of observables, including the amount of bremsstrahlung radiation, transverse energy density within an isolation cone, geometrical and momentum matching between the electron trajectory and associated clusters, and shower shape observables.
Two different MVA ID taggers are provided: the *Electron MVA-NoIso*, and the *Electron MVA-Iso*, which includes among the observables the three components of PF isolation, due respectively to charged hadrons, neutral hadrons and photons, in the training. In both cases the training is performed on Z + Jets MC samples under 2017 detector condition [132]. The MVA IDs output is a real number in the interval $[-1, 1]$ defined in such a way that the genuine electrons have a tendency to obtain higher output values, while the non-genuine electrons obtain lower output values with a bigger probability. They are used as a discriminant for classifying electron candidates, with a fixed threshold set to balance identification efficiency and misidentification probability. Three WPs are set for both Iso and NonIso MVA ID taggers: *wp90* and *wp80*, defining η regions where the signal efficiencies in each category are, respectively, about 90% and 80%, and *wpLoose*, with a high overall efficiency of 98%. Details on WP definition are reported in Appendix A.1.2.
In the analysis presented in thesis, the following MVA ID taggers and WPs are adopted: *MVA-Iso-WP80*, *MVA-NoIso-WP80* and *MVA-Iso-WP90* (Section 4.5 and Section 4.6.1).

Reco and ID Performances

In the left panel of Fig. 3.9 the reconstruction efficiency as a function of η is shown for 2017 data; the region $1.44 < |\eta| < 1.57$ corresponds to the transition region between the barrel and endcap regions of ECAL and is not considered in physics analyses. The right panel displays the performance of *MVA-NoIso* (red) and *MVA-Iso* (green) in terms of Receiver Operating Characteristic (ROC) curves (see Appendix A.1.9). A comparison with the sequential selection approach, where the BDT and isolation criteria are applied consecutively (blue), demonstrates that the BDT trained with isolation variables provides a significant advantage, particularly at higher selection efficiencies.

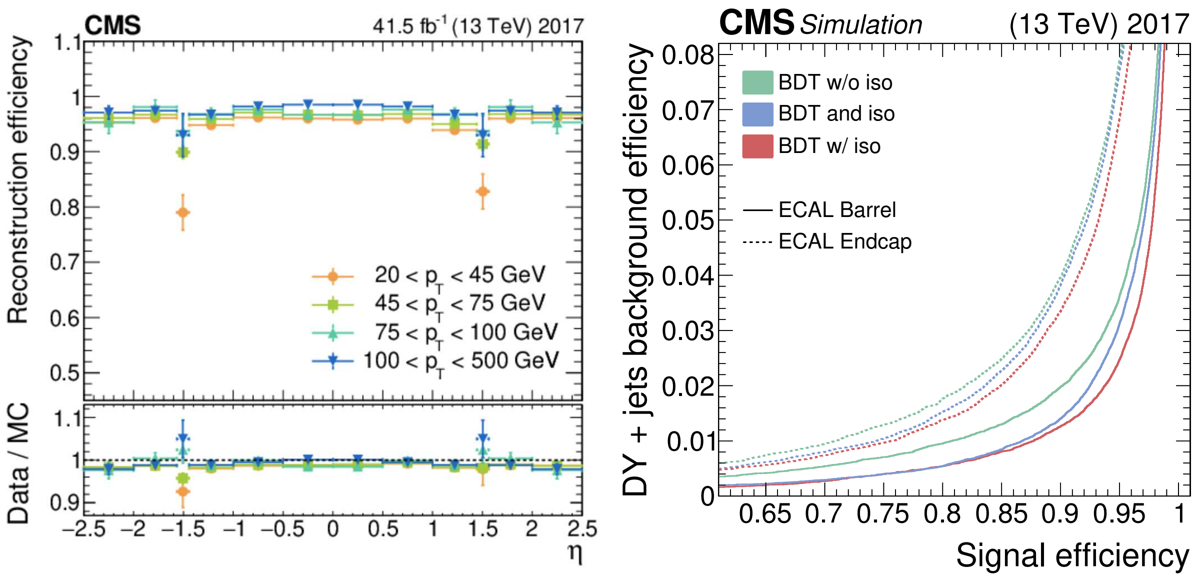


Figure 3.9: Left: Electron reconstruction efficiency vs η 2017 data. The vertical bars on the markers represent the combined statistical and systematic uncertainties. Right: ROC curves for the electron MVA ID trained with and without the isolation variables (red and green), compared to an optimized sequential selection using the BDT without the isolations plus a cut on the combined isolation (blue) [132].

3.3.4 Jets

In pp collisions, quarks and gluons are produced and quickly fragment due to confinement, leading to hadronization, which forms a collimated "spray" of hadrons. These unstable hadrons then decay into stable particles, which, along with those formed during hadronization, move toward the detector. This series of particles creates a concentrated energy deposit in the calorimeter, known as a hadronic *jet*. A visual representation of this jet in CMS is shown in Fig. 3.10.

Jet reconstruction

One of the main challenges in analyzing jets from hard collisions is recovering the initial energy, momentum, and potentially the type of partons produced. Through the

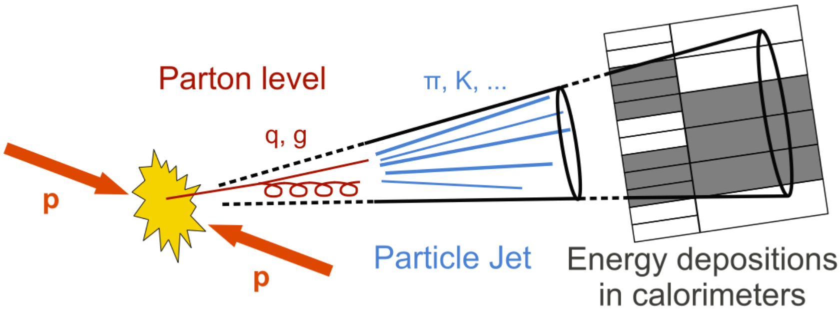


Figure 3.10: Sketch of a jet produced during a pp collision.

reconstruction of the jets the total energy and the momentum of these partons can be measured and using tagging algorithms the type of parton originating from the hard collision can be inferred.

At CMS jets are reconstructed by clustering PF candidates using the Anti-k_T (AK) algorithm [133, 134]. This algorithm iteratively combines nearby PF candidates based on a metric designed to produce jets with an approximately conical shape, clustered around the hardest particles in the event. The jet cone size is determined by the distance parameter R at which the algorithm operates. The AK algorithm is robust against infrared and collinear effects, meaning it is unaffected by soft radiation or collinear parton splitting.
To form a jet two quantities, the distances between pairs of entries (d_{ij}) and a parameter used to represent distance between an entry and the beam line (d_{iB}), are updated in each step and are defined as:

$$d_{ij} = \min(p_T^{-2}(i), p_T^{-2}(j)) \times \frac{\Delta R_{ij}^2}{R^2} \tag{3.5}$$

and

$$d_{iB} = p_T^{-2}(i) \tag{3.6}$$

If d_{iB} is smaller than $\min_j(d_{ij})$, entry i is promoted to reconstructed jets and excluded from the further iterations. Otherwise, entry i and j' = arg($\min_j(d_{ij})$) are merged into new entry i' and the algorithm proceeds to the next iteration. The algorithm stops when all jet-candidates are promoted to the reconstructed jets or when $\min(d_{ij}, d_{iB})$ is below a threshold d_{cut}.

Two types of AK reconstructed jets are identified in CMS, according to the cone size R: AK4 Jets with $R = 0.4$, and AK8 Jets with $R = 0.8$, which will be also referred as resolved and boosted jets respectively. AK8 jets are particularly important for a high mass particles (Higgs boson, W and Z, and Beyond Standard Model (BSM) high mass resonance) produced with high p_T, decaying into hadrons. The jets produced in these decays are collimated and often too close together to be resolved individually. In this analysis, both AK4 and AK8 have been used, based on the final state topologies being studied.
The algorithm for the jets reconstruction can be run also using only calorimeter information, resulting in "calorimeter jets", which are used in the triggers. In contrast, the

PF algorithm (Section 3.3) reconstructs jets by identifying individual particles—such as charged hadrons, neutral hadrons, photons, and leptons—using combined information from the calorimeters, tracker, and muon detectors. The reconstruction of jets using PF particles benefit from improved energy resolution, a response closer to the true jet energy, and enhanced mitigation of PU effects, as charged PF candidates not originating from the primary vertex are removed before clustering of jets by the use of two algorithms [135]: Charged Hadron Subtraction (CHS) and PileUp Per Particle Identification (PUPPI).

Pileup mitigation

To avoid the inclusion, during the clustering, of particles coming from PU interactions, various PU mitigation techniques are used at CMS, notably CHS and PUPPI, briefly described in the following [135].

Charge Hadron Subtraction (CHS) exploits tracker information to exclude charged particles linked to PU vertices from jets, but because the tracker has limited coverage in η, further jet energy corrections are required to account for charged PU outside the tracker acceptance and neutral PU everywhere. However, these corrections affect the jet four-momentum but do not modify its shape or sub-structure, making the CHS algorithm less suitable for AK8 jets, which have more complex internal structure.

PileUp Per Particle Identification (PUPPI) To address the CHS limitation, an alternative PU mitigation technique, PUPPI, is utilized [136]. PUPPI calculates the probability that each particle, charged or neutral, originates from the primary vertex in each event and adjusts particle energy accordingly. As a result, objects clustered from hadrons—such as jets, missing transverse momentum, and lepton isolation—are expected to be less affected by PU with PUPPI. Figure 3.11 illustrates CHS and PUPPI schematically.

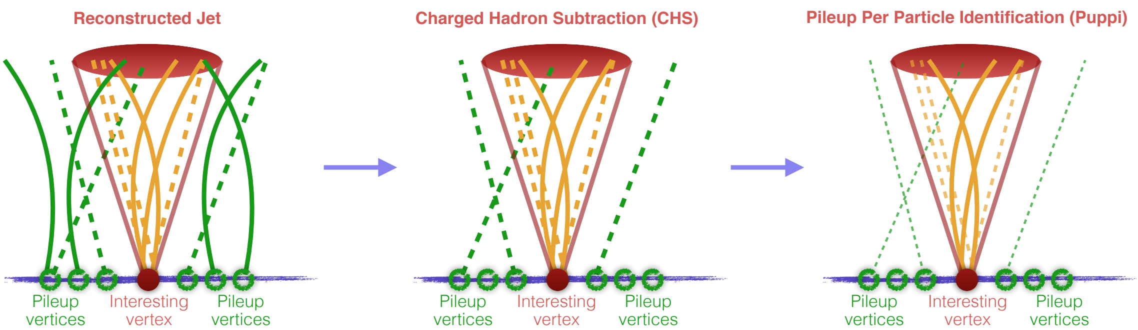


Figure 3.11: Sketch of PU suppression techniques. Solid (dashed) lines refer to charged (neutral) PF candidates. The weights applied by the PUPPI algorithm are represented by thin lines [137].

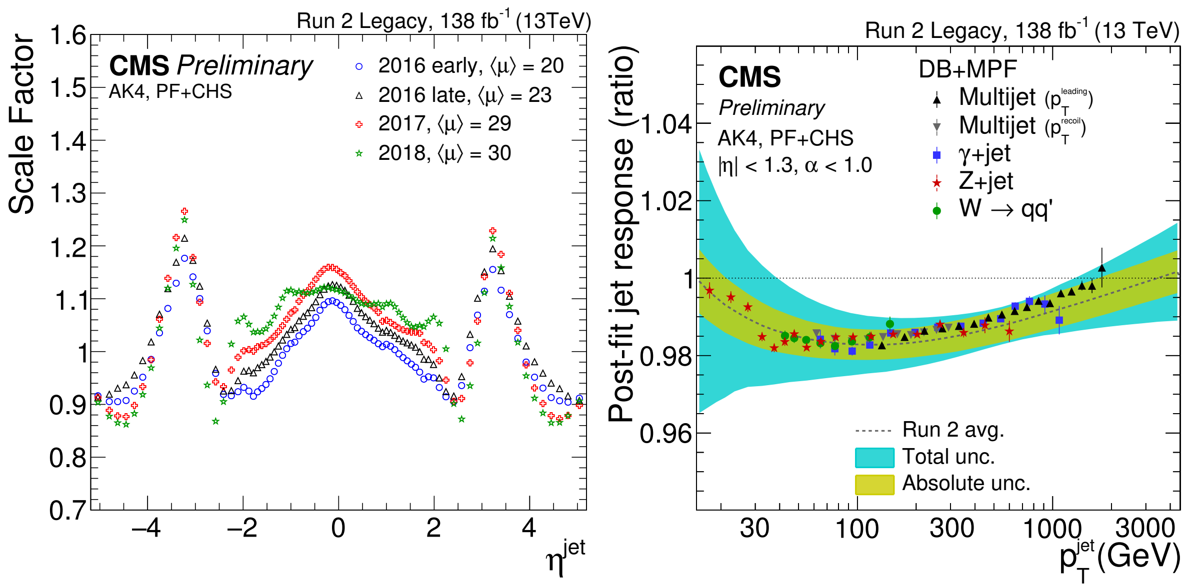


Figure 3.12: Left: Data-to-simulation offset correction in 2016-2018 for Legacy reconstruction. Right: Data-to-simulation comparison for the jet response dependence on p_T [138].

Jet energy corrections and resolution (JEC, JER)

The primary goal of jet reconstruction is to obtain accurate information about the properties of the original parton, such as its energy [138]. Various experimental effects, however, can lead to distortions in the measured jet energy. Additionally, a precise simulation of jet energy distribution is essential. Thus, Jet Energy Corrections (JEC) are applied to adjust the four-momentum of jets in both data and simulations. These corrections include two main components: the Jet Energy Scale (JES) and the Jet Energy Resolution (JER) corrections [139]. The JER correction involves a sequence of multiplicative factors applied to the jet four-momentum to account for various effects, including pile-up mitigation and adjustments for the non-linear detector response, particularly in real data. Discrepancies between data and simulation, as well as potential differences in the reconstructed energy of light jets (originating from u, d, s quarks and gluons) and heavy-flavor jets (originating from c and b quarks), are also corrected.
The JER correction improves the p_T resolution in simulations by smearing the reconstructed jet p_T [139]. If a reconstructed jet can be matched to a generator-level jet, a direct scaling is applied to match the jet p_T resolution to that observed in data. If no match is found, stochastic smearing is applied, using a Gaussian distribution centered at zero with a width derived from data to simulate the resolution effects.
Two examples of the JES estimation for the first step, the PU offset corrections (left), to the application of subsequent uncertainties and finally the absolute p_T residual corrections (right) are shown in Fig. 3.12

Jet identification

Even after applying CHS or PUPPI and the jet corrections, reconstructed jets may still contain PU and noise jets originating from calorimetric noise or leptons mis-reconstructed as jets. Identifying and rejecting these jets is essential for all CMS analyses targeting jets or MET in the final states, including this analysis. To suppress the noise and PU jet contribution, this analysis uses two specific identification techniques: the PU Jet ID and the Noise Jet ID [135, 140]. In the following, a brief description of each identification method adopted will be provided.

Noise Jet ID , or simply Jet ID, is a PF based jet discriminator aiming to discriminate between noise and physical jets [140]. It exploits jet energy fraction and multiplicity variables, which are sensitive to different sources of noise coming from ECAL and HCAL. Three PF jet ID Working Point (WP, see Appendix A.1.2) are defined: loose, tight and tight lepton veto. Details on the WP definitions are presented in [140]. In the analysis presented in this thesis, the jets are required to pass tight lepton veto WP, ensuring that jets originating from calorimetric noise are removed and the rejection of mis-reconstruction of lepton candidates. The efficiency and the background rejection of this discriminator is estimated via the TnP technique (see Appendix A.2) in dijet events, and it is computed for AK4 jets with CHS for PU mitigation. The jet criteria for all WPs allow to reject residual noise from the calorimeters, retaining $98 \div 99.9\%$ of genuine jets, i.e., jets initiated by genuine particles rather than detector noise. The background rejection is above $\approx 98\%$ for $|\eta| < 3$, which is the region of interest for the analysis [135, 140].

Pileup Jet ID PU jets can be divided in two categories: Quantum Chromodynamics (QCD)-like jets, referring to the PU Jets originating from a single PU vertex, and stochastic PU jets, originating from multiple different PU vertices. The PU mitigation techniques exploit tracking information about the primary vertex to remove the charged tracks associated to the PU vertices, redunging the QCD-like jets p_T of $\approx 2/3$. In order to remove the remaining ones, mostly constitued by stochastic PU jets, a multivariate technique is implemented, the PUJetID (PU Jet ID). It is based on a BDT trained with 12-15 input variables related to the jet and to the event. It is applied on CHS jets, since the PUPPI algorithm intrinsically suppresses PU jets better by rejecting more charged and neutral particles from PU vertices before jet clustering. In this analysis, an explicit requirement on the PU Jet ID to pass the Loose WP has been applied for jets with $p_T < 50\,\text{GeV}$; it ensures the 99% (95%) efficiency for $|\eta| < 2.5$ ($|\eta| \geq 2.5$), as shown in Fig. 3.13.

b-jets

The identification of jets from the hadronization of b-quarks, also known as b-tagging, is crucial for the analysis presented, as one of the two Higgs bosons decays into b-jet pairs. Discriminating b-jets from jets produced by light quarks or gluons relies on

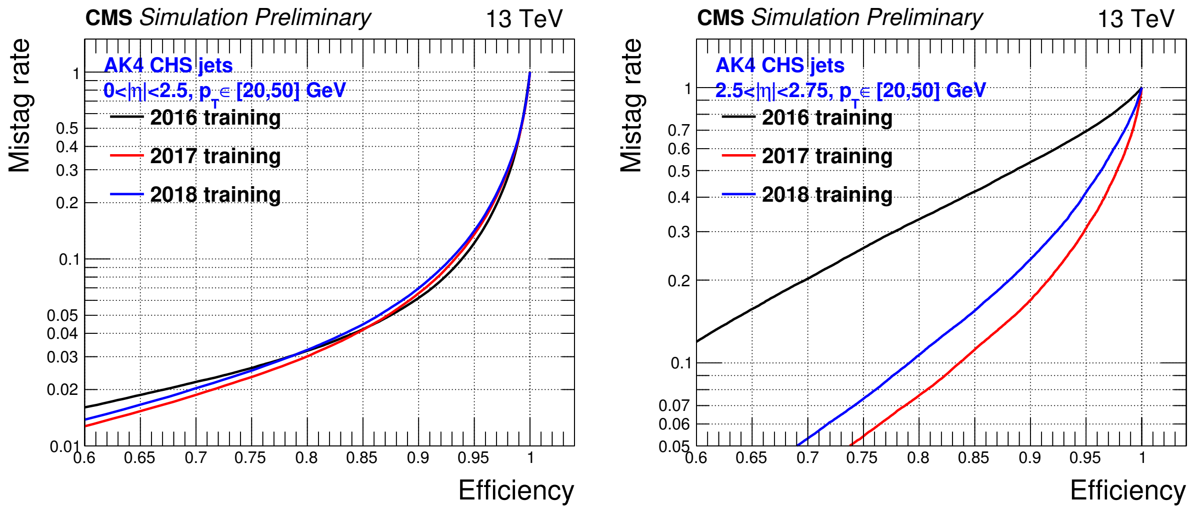


Figure 3.13: PUJetID mistag rate as a function of its efficiency for jets in different η ranges, for jet $p_T \in [20, 50]$ GeV in a Z plusjet event sample, for the 2016, 2017 and 2018 simulated samples. Image took from [141].

the distinctive properties of b-hadrons: their long lifetime (~ 1.5 ps) and relatively large mass (~ 5 GeV). The long lifetime leads to charged tracks with large impact parameters and secondary vertices displaced by several millimeters from the primary vertex, both of which are strong discriminants. A b-jet decay representation is shown in Fig. 3.14. Additionally, b-hadrons carry more of the parton momentum and produce collimated decay products with higher invariant mass. Around 25% of b-hadron decays are semi-leptonic, resulting in high-p_T leptons relative to the jet direction — another feature useful for b-jet tagging. In CMS, several machine learning algorithms have been developed to create powerful b-tagging tools. In the analysis presented in this thesis, the final state includes two b-quarks: the **DeepJet** b-tagger is used to classify AK4 jets originating from b-quark hadronization while the **ParticleNet-MD** tagger is applied to identify AK8 jets originating from bbar pairs produced in Higgs decays, which are detailed in Section 4.7. A summary of these algorithms and their performance is provided below.

The DeepJet algorithm, also known as **DeepFlavour** [142], is a multi-class flavour tagging algorithm, based on multivariate analysis and exploiting Deep Neural Network (DNN) (see Appendix A.1.6). DeepJet exploits the full information of all the jet constituents, in addition to charged and neutral particles, secondary vertices, and global event variables simultaneously. In total, it takes in input 16 (6) features of the 25 leading charged (neutral) PF Jet constituents. These features include track kinematics and fit quality information, secondary vertex displacement and relative uncertainties with respect to the primary vertex; in addition, 12 features of the vertices (for a multiplicity of maximum 4 candidates), including the decay length significance. Finally, 6 global variables containing jet-level information such as the jet kinematics, the number of tracks in the jet, as well as the number of secondary vertices in the

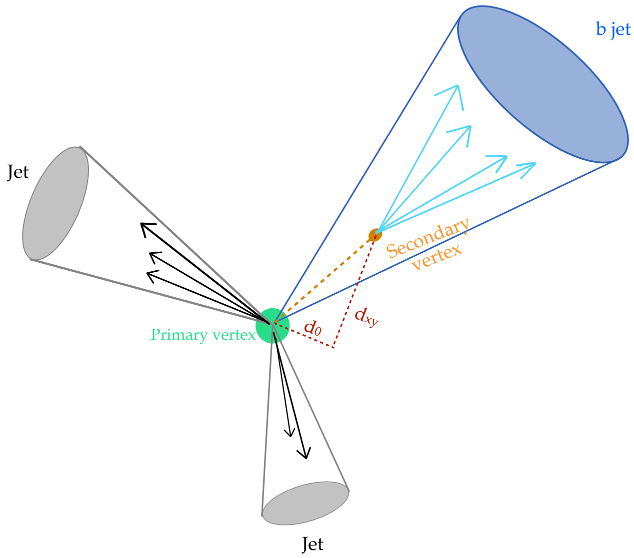


Figure 3.14: A basic 3-jet event where one of the reconstructed jets (blu cone toward upper right) comes from the b-quark hadronization. The secondary vertex is highlighted in orange. The other two jets are light jets or coming from different quarks hadronization processes; in all the cases the arrows correspond to the tracks inside the jet. The transversal and longitudinal impact parameters are highlited in red. This image is an original creation by the author.

jets are added too. The training is performed on a large dataset made by simulated events from QCD and top quark pair production. The output consists in 6 nodes, allowing it to perform b-tagging, c-tagging and quark/gluon tagging. In Fig. 3.15 the comparison between the deepJet and the previous algorithm used, *DeepCSV*, are shown. The DeepFlavour performs better than DeepCSV in all jet p_T range. Moreover, former jet-tagging algorithms, as DeepCSV had the quarks and gluon classes merged, as usually done in flavor tagging. For DeepJet though are individual classes. For the DeepJet tagger, three WPs are defined: loose, medium and tight, corresponding to misidentification rates of 10%, 1% and 0.1%. In the analysis, when performing event categorization, the Medium WP is required for resolved jets coming from the H $\rightarrow$ bb candidate.

The ParticleNet-MD [144] is a mass-decorrelated version of the ParticleNet, a machine learning algorithm based on a graph neural network (GNN see Appendix A.1.8) designed to identify hadronic decays of highly Lorentz-boosted particles like top quarks and W, Z, and Higgs bosons by classifying their decay modes (e.g. $X \rightarrow b\bar{b}$, $X \rightarrow c\bar{c}$, $X \rightarrow q\bar{q}$).
Unlike traditional methods that view jets as fixed structures, ParticleNet treats them as unordered point clouds of particles, characterized by features such as η, ϕ, p_T , and charge. A key aspect of ParticleNet is the EdgeConv operation, which builds dynamic graphs based on particle positions, enhancing representations by analyzing k-nearest neighbors in the $\eta - \phi$ plane. This approach captures spatial correlations while ensuring permutational invariance, allowing for a consistent representation

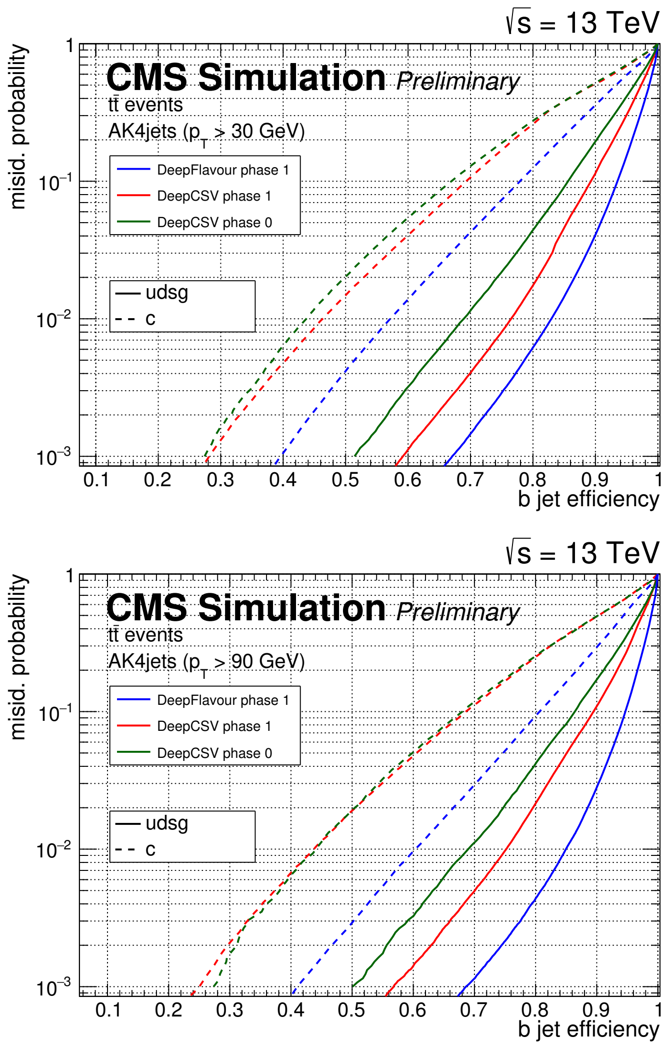


Figure 3.15: Performance of the DeepCSV and DeepFlavour b-jet identification algorithms expressed as ROC curve (see Appendix A.1.9), in terms of the b-jet misid probability and b-jet ID efficiency. The curves are obtained on simulated $t\bar{t}$ events using jets within $|\eta| < 2.5$ and with p_T >30 GeV (up) or p_T >90 GeV (down). The performances of DeepCSV with the 2016 detector (Phase 0) are also shown [143].

regardless of particle arrangement.
The algorithm processes two input lists for each jet: particle features and vertex information, treating each particle as a point in high-dimensional space. Through successive EdgeConv layers, ParticleNet constructs graphs that evolve spatial relationships among particles, producing new latent representations that capture the jet's structure. The ParticleNet tends to modify background jets mass distribution, making it similar to the signal one. Therefore also in this case a mass-decorrelated version of the tagger, the ParticleNet-MD algorithm has been introduced. In this way the tagger does not sculpt the jet mass distributions, thereby enabling the use of the jet mass in the background estimation. For each AK8 jet the algorithm outputs four probability-like scores: P($X \to b\bar{b}$), P($X \to c\bar{c}$), P($X \to q\bar{q}$),and P(QCD). In this thesis an ParticleNet-MD

discriminator, D_{bb}, is used to tag H boson candidate decay in $b\bar{b}$, defined as

$$D_{bb} = \frac{P(X_{bb}^{MD})}{P(X_{bb}^{MD}) + P(X_{QCD}^{MD})} \tag{3.7}$$

The D_{bb} WPs of High Purity (HP), Medium Purity (MP), and Low Purity (LP) are defined as $\mathrm{H} \to \mathrm{b\bar{b}}$ signal jets at efficiencies of 40%, 60%, and 80%, respectively, according to the common WP choices adopted in various CMS physics searches. Thec corresponding values of the D_{bb} discriminator are summarized in Tab. 3.1. Figure

Year	LP	MP	HP
2016-preVFP	0.9883	0.9737	0.9088
2016-postVFP	0.9883	0.9735	0.9137
2017	0.9870	0.9714	0.9105
2018	0.9880	0.9734	0.9172

Table 3.1: WPs defined at fixed efficiency values. The WPs are determined in four data-taking eras separately [145].

3.16 presents a comparison of the AK8 identification performance of ParticleNet-MD D_{bb} against QCD background, alongside previous taggers used by CMS. The Machine Learning (ML) based taggers, *ParticleNet-MD bbVSQCD*,*DeepDoubleBvL* and *DeepAK8-MD bbVSQCD*, significantly outperform the cut based one, *double-b*. Moreover, ParticleNet-MD outperforms above all the taggers.

3.3.5 Taus

Tau leptons have a mean lifetime of about 2.9×10^{-13}s, thus they decay before reaching the detector, away from the primary interaction point creating a secondary vertex. The decays are categorized as either leptonic or hadronic. In leptonic decays (τ_ℓ), the τ lepton decays into an electron or muon along with two neutrinos (one from the τ itself and one associated with the electron or muon), accounting for approximately 35.2% of cases. These decays are reconstructed as an electron (τ_e) or muon (τ_μ), though they are challenging to differentiate from other electron and muon sources. Additional selection criteria are applied to select high p_T, prompt taus, as detailed in Section 3.3.2 for τ_μ, and in Section 3.3.3 for τ_e. Almost all of the remaining ($\sim$ 64%) are semi-leptonic τ decays, where the τ is reconstructed as a hadronic jet. These are often referred to as hadronic τ decays (τ_h).
The hadronic decay modes of Tab. 3.2 are reconstructed at using the algorithms Hadron-Plus-Strips (HPS) [146, 147] for reconstruction and DeepTau [148] for identification, as explained below.

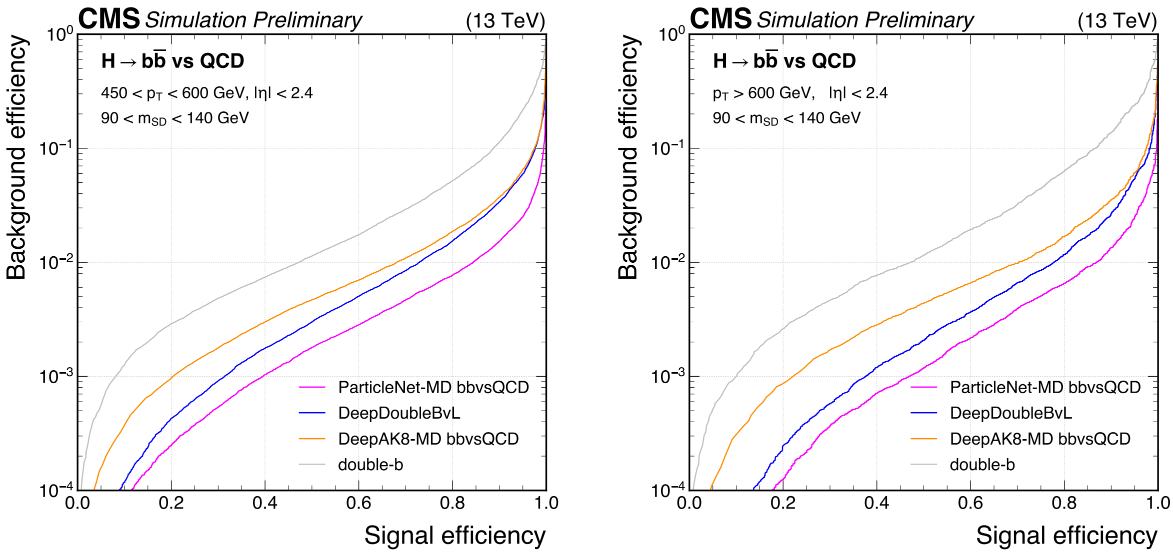


Figure 3.16: Comparison of the performance of the $H \to b\bar{b}$ identification algorithms in terms of ROC curves (see Appendix A.1.9) for $H \to b\bar{b}$ signal jets versus the inclusive of QCD jets background, using simulated events in the 2018 data-taking conditions. Performance is shown in the 450 < p_T < 600 GeV (left) and p_T > 600 GeV (right) regions [145].

Decay mode	Resonance	$\mathcal{B}$ (%)	
Leptonic decays		35.2	
$\tau^- \to e^-\bar{\nu}_e\nu_\tau$			17.8
$\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$			17.4
Hadronic decays		64.8	
$\tau^- \to h^-\nu_\tau$			11.5
$\tau^- \to h^-\pi^0\nu_\tau$	$\rho(770)$		25.9
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	$a_1(1260)$		9.5
$\tau^- \to h^-h^+h^-\nu_\tau$	$a_1(1260)$		9.8
$\tau^- \to h^-h^+h^-\pi^0\nu_\tau$			4.8
Other			3.3

Table 3.2: List of τ lepton principal decays with their branching ratios in percentage. Values taken from the updated PDG in [149].

τ_h reconstruction

Hadronically decaying taus are first reconstructed at trigger level: in the L1T τ_h reconstruction calorimeter data organized into trigger towers are used to identify tau candidates. At the HLT the τ_h reconstruction is organized in two level, exploiting calorimeter jets around Level 1 (L1) seeds and also pixel track based isolation for the first stage (L2) and then including the PF tau candidates at the next stage (L3), applying also addintional cuts to enhance the purity of tau selection. The offline reconstruction of hadronic tau decays of Table 3.2 starts taking in input the PF AK4 jets. For each jet the HPS algorithm combines PF charged hadrons within the cone with associated "strips" in ECAL from neutral pions in the cone.

The π^0 mesons promptly decay into pairs of photons, which have a high probability of converting into e^+e^- pairs as they traverse the tracker material. The large magnetic field of the CMS solenoid leads to a spatial separation of the e^+e^- pairs in the η-ϕ, plane, spreading the energy deposit in the ECAL due to neutral pions in a region narrow in η and wide in ϕ, referred to as a "strip".
The strip reconstruction starts by centering a strip on the most energetic electromagnetic particle within the PF jet. The algorithm then searches for other electromagnetic particles within a window of size $\Delta\eta$ and $\Delta\phi$ parameterized by the p_T of the clustered PF candidates, varying from 0.05 to 0.15 and from 0.05 to 0.3 respectively, centered on the strip center. If other electromagnetic particles are found within that window, the most energetic one gets associated with the strip and the strip four-momentum is recalculated. The procedure is repeated until no further particles are found that can be associated with the strip. Strips satisfying $p_T > 1\,\text{GeV}$ are finally combined with the charged hadrons to reconstruct individual τ_h decay modes.
Multiple τ_h hypotheses are considered per each jet. Each hypothesis includes one or three charged particles and 0–2 strips, and corresponds to the expected signature from one of the τ_h decay modes listed in Tab. 3.2: $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$, $h^\pm h^\pm h^\mp$ and $h^\pm h^\pm h^\mp$ π^0 [148].
The four-momentum of each τ_h candidate is calculated by summing the four-momenta of the associated charged hadrons and π^0 candidates in a specific decay mode. To choose between possible combinations of strips and charged hadrons within a jet, tche last step of reconstruction defines the selection of one reconstructed τ_h candidate among all the possible ones. The following quality criteria are applied to all the reconstructed τ_h candidates for a jet:

- **Mass Compatibility**: The mass of the reconstructed τ_h candidate must align with expected resonance masses:
  - For the $h^\pm\pi^0$ mode, it should match the $\rho(770)$ resonance.
  - For the $h^\pm\pi^0\pi^0$ and $h^\pm h^\mp h^\pm$ modes, it should match the $a_1(1260)$ resonance.
  - For the $h^\pm h^\mp h^\pm\pi^0$ mode, compatibility with the $\rho(1450)$ resonance is required.
- **Charge Assignment**: The charge of the τ_h candidate should correspond to ±1 unless a charged hadron is missed; in such cases, the charge is determined based on the higher p_T of the detected charged hadrons.
- **Signal Cone Criteria**: All reconstructed particles (charged hadrons and π^0) must be located within a defined signal cone relative to the τ_h momentum, constrained between 0.05 and 0.1.

Among the reconstructed τ_h candidates that meet these criteria, the candidate with the highest p_T is selected for further analysis. This candidate can then be discriminated against quark or gluon jets, electrons, and muons, which is crucial for isolating τ_h signals in high-energy collision data. The charged hadron candidates inside the AK4 used are required to be compatible with originating from the hard primary vertex and with a $p_T > 0.5\,\text{GeV}$.

All particles in an $\eta - \phi$ cone of radius $\Delta R = 0.5$[2] around the jet axis are considered for the next steps of the τ_h reconstruction.

τ_h reconstruction performances

Fig. 3.17 displays the fraction of reconstructed decay modes for generated τ_h candidates with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$. The HPS algorithm successfully reconstructs most τ_h decays in targeted modes, with non-reconstruction rates of 11% for $h^{\pm}$ and 25% for $h^{\pm}\pi^0$. The overall reconstruction efficiency is about 90%, mainly constrained by the accurate reconstruction of charged hadron tracks. For decay modes with no missing charged hadrons, charge assignment accuracy is high, reaching 99% for average $Z \to \tau\tau$ events, 98% at $p_T \approx 200\,\text{GeV}$, and 92% at $p_T \approx 1\,\text{TeV}$. However, in cases with missing charged hadrons, recovery rates drop to 19% of the $\tau^- \to h^-h^+h^-\nu_\tau$ decays and 13% of the $\tau^- \to h^-h^+h^-\pi^0\nu_\tau$ decays, with charge assignment accuracy decreasing to around 70% making these events unsuitable for analyses with charge requirements, like in this thesis.

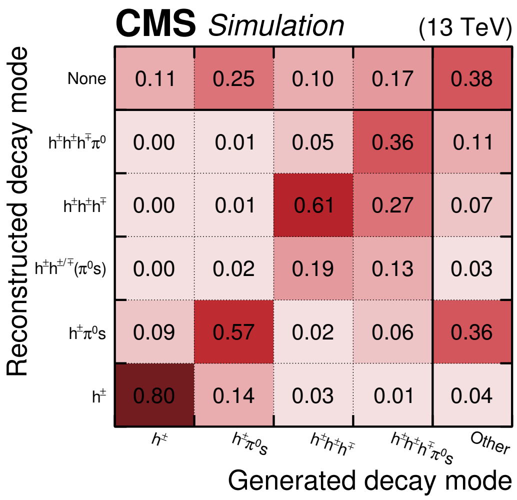


Figure 3.17: Decay mode confusion matrix. For a given generated decay mode, the fractions of reconstructed τ_h in different Decay Mode (DM)s are given, as well as the fraction of generated τ_h that are not reconstructed. Decay modes with the same numbers of charged hadrons and one or two π^0 are combined and labelled as "π^0s" [148].

τ_h identification

As for the b-jets, the τ_h identification (and its reconstruction) plays a very important role in this thesis, since the final state investigated has two taus, coming from the decay of one of the two Higgs bosons.

The τ identification at CMS is performed via a multiclass identification algorithm, *DeepTau* [148], based on a convolutional deep neural network (DNN, see Appendix A.1.6). The algorithm takes as input the information from all reconstructed particles

[2]where ΔR is defined in Eq. 2.6.

in the proximity of a τ_h candidate, combining low-level features, which are the reconstructed PF particles (electrons, photons muons, charged and neutral hadrons) and high-level ones as the τ_h candidate four momentum, charge, and number of charged and neutral particles used for its reconstruction, the compatibility of the leading τ_h track to come from the primary vertex and also its probability to come be an electron, η and ϕ distributions of the energy reconstructed in the strips, PU in the event. Most of the input variables are standardized and truncated, to avoid outliers. Two grids in $\eta \times \phi$ around the τ_h candidate are defined, an inner and an outer one.Three different subnetworks are created that independently process the inputs from the high-level variables, the outer cells, and the inner cells. The training was performed including different simulated processes with taus in the final state to make the training as much as possible independent on the process containing the τ_h in the final state: production of Z and W bosons associated with jets, $t\bar{t}$ production, heavy gauge bosons decaying to leptons $Z' \to \ell\ell$ (with $\ell = e, \mu, \tau_h$), and the production through the strong interaction with final states containing several jets (QCD-multijet), since this process is the most relevant source of background for τ_h misidentification. The outputs from the three subnetworks are concatenated and passed through four fully connected layers with 200 nodes each and then to a final layer with four nodes that yield outputs. A softmax [150] activation function is applied to get the probabilities for each τ_h candidate to come from each of the four classes (y_e, y_μ, y_{jet}, and y_{τ_h}). At the end four outputs are provided in form of probability for the candidate to be a τ_h, an electron, a muon or a jet. From these probabilities, discriminators are defined: *DeepTauVSe*, *DeepTauVSmu* and *DeepTauVSjet*, which are respectively against electrons, muons and jets. The expression of each discriminator is given by:

$$D_\alpha(y) = \frac{y_\tau}{y_\tau + y_\alpha} \tag{3.8}$$

with $\alpha \in \{\text{jet}, \mu, \text{e}\}$. To evaluate the performances of the training and avoiding biases, efficiencies are computed on $H \to \tau\tau$ samples, which have not been included in the training, and for taus with $p_T \in [30, 70]$GeV. Based on the efficiency and the background rejection targets, several WPs are defined for each discriminator, going from VVVLoose to VVTight, which are summarized, for each discriminator, in Tab. 3.3. These target efficiencies for *deepTauVSJet* range from 40% to 98%, for *deepTauVSmu* they are between 99.5% and 99.95% and finally for the *deepTauVSmu* they are in the range between 60% and 99.5%.

	VVTight	VTight	Tight	Medium	Loose	VLoose	VVLoose	VVVLoose
D_e	60%	70%	80%	90%	95%	98%	99%	99.5%
D_μ	—	—	99.5%	99.8%	99.9%	99.95%	—	—
D_{jet}	40%	50%	60%	70%	80%	90%	95%	98%

Table 3.3: Target τ_h identification efficiencies for the different working points defined for the three different discriminators [148].

τ_h identification with DeepTau performance

The DeepTau discriminators performance is evaluated through ROC curves for two p_T ranges: $20 < p_T < 100$ GeV and $p_T > 100$ GeV. Simulated sample sources for each plot are indicated, with DeepTau working points marked as filled circles. For comparison, previous identification algorithms working points are shown with vertical bars representing statistical uncertainties.
The electron discriminator D_e achieves τ_h efficiencies exceeding 90%, with misidentification rates between 0.3% and 8%, as shown in Fig. 3.18. The muon discriminator, D_μ, can reject up to ten times more muons at 99% τ_h efficiency, increasing τ_h efficiency by about 0.5% for comparable background rejection. The performances are shown in Fig. 3.19. Finally, as shown in Fig. 3.20, the jet discriminator, D_{jet}, reveals higher efficiency for quark jets from W+jets processes compared to the $t\bar{t}$ samples, which contains more gluon and b-quark jets.
This discriminator reduces jet efficiency by a factor of over 1.8 compared to the previous method based on MVA, with further reductions at higher τ_h efficiencies.

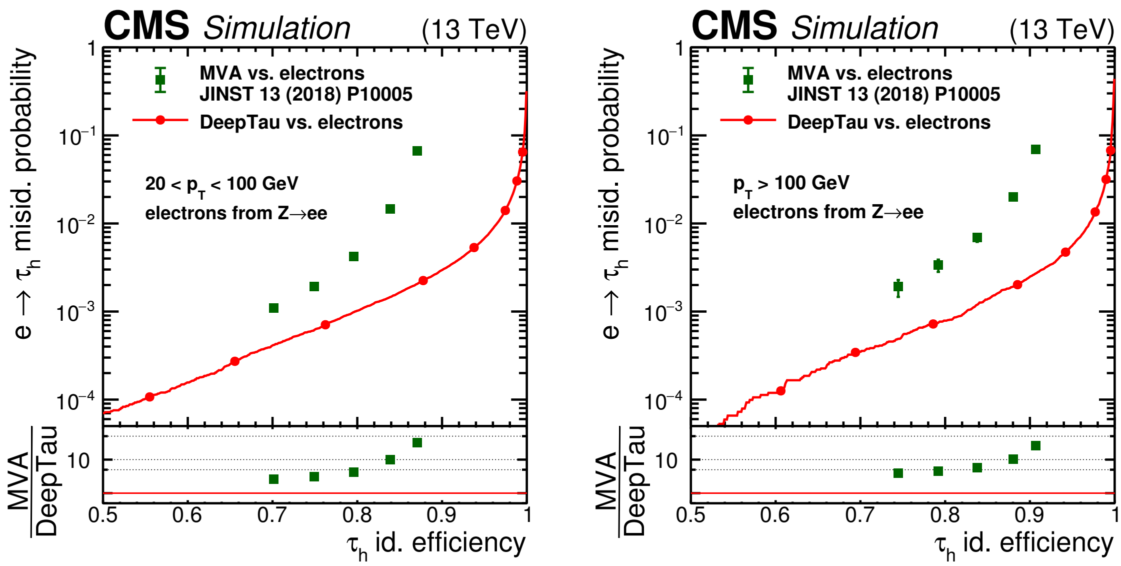


Figure 3.18: Electrons VS for genuine τ_h efficiency, for simulated τ_h to pass the MVA [151] and D_e discriminators. The τ_h candidates are reconstructed in one of the τ_h decay modes without missing charged hadrons [148].

3.3.6 Missing Transverse Energy (MET)

The missing transverse energy $\vec{p}_T^{miss}$ (also referred as MET or p_T^{miss} when indicating its magnitude) is calculated as the negative vector sum of the transverse momenta of all reconstructed PF particles [152]. MET is crucial in analyses because a substantial p_T^{miss} value often signals the presence of neutrinos or other elusive BSM neutral particles that do not interact with the detector.
In the search presented in this thesis, $X \to HH \to b\bar{b}\tau\tau$, p_T^{miss} is especially important due to the presence of neutrinos from the two τ decays and eventually from the

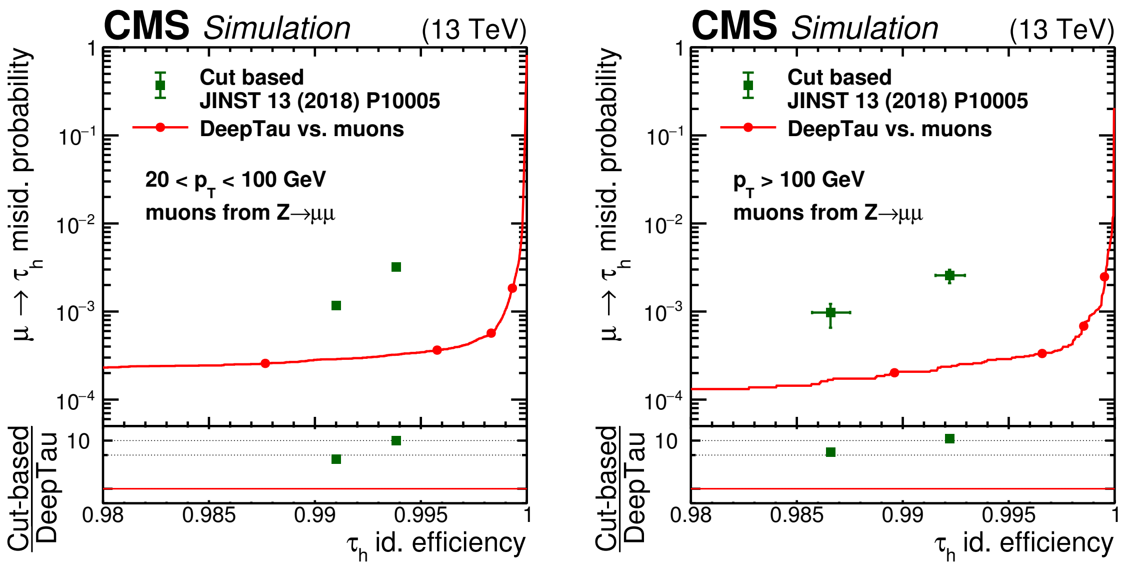


Figure 3.19: Muons VS genuine τ_h efficiency for simulated τ_h to pass the cutoff-based and D_μ discriminators [148].

hadronization of b-quarks into b hadrons, providing a distinctive signature for this decay process (Section 4.1). Since $\vec{p}_T^{miss}$ calculation exploits the transverse momenta of all the particles in the event, its accuracy is particularly sensitive to detector performance and PU effects, making it prone to anomalously large values in some events.
To ensure a good quality of the reconstructed $\vec{p}_T^{miss}$, jet energy corrections are propagated to the $\vec{p}_T^{miss}$ replacing the raw jet momenta, $\vec{p}_T^{miss,raw}$, in the $\vec{p}_T^{miss}$ computation with corrected jet momenta of all reconstructed PF particles:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss,corr}} = \vec{p}_{\mathrm{T}}^{\mathrm{miss,raw}} - \sum_{j=1}^{N_{\mathrm{jets}}} (\vec{p}_{T,j}^{\mathrm{corr}} - \vec{p}_{T,j}) \tag{3.9}$$

The sum is over jets with $p_T > 15\,\mathrm{GeV}$. To remove the overlap of jets with electrons and photons, jets with more than 90% of their energy associated to the ECAL are not included in the sum. In addition, if a muon reconstructed using the outer tracking system overlaps with a jet, its four momentum is subtracted from the four momentum of the jet, and the JES correction (Section 3.3.4) appropriate for the modified jet momentum is used in the $\vec{p}_T^{miss}$ calculation. To estimate its uncertainty, the uncertainties in the momenta of the all reconstructed objects are propagated to $\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ by varying the estimate of each PF candidate p_T and recomputing the $\vec{p}_T^{miss}$.
The JEC uncertainties are less than 3% for jets within the tracker acceptance and reach up to 12% for those outside. The JER uncertainties typically in the range $5 \div 20\%$. The uncertainty in the unclustered energy is evaluated based on the momentum resolution of each PF candidate, which depends on the type of the candidate. The largest contributions to the unclustered energy uncertainty are due to the PF neutral hadrons and PF candidates in the HF. As it depends on several objects, the total uncertainty on the $\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ measurement is expected to vary with the event topology.
The MET is sensitive also to the entire detector readout. For this reason dedicated

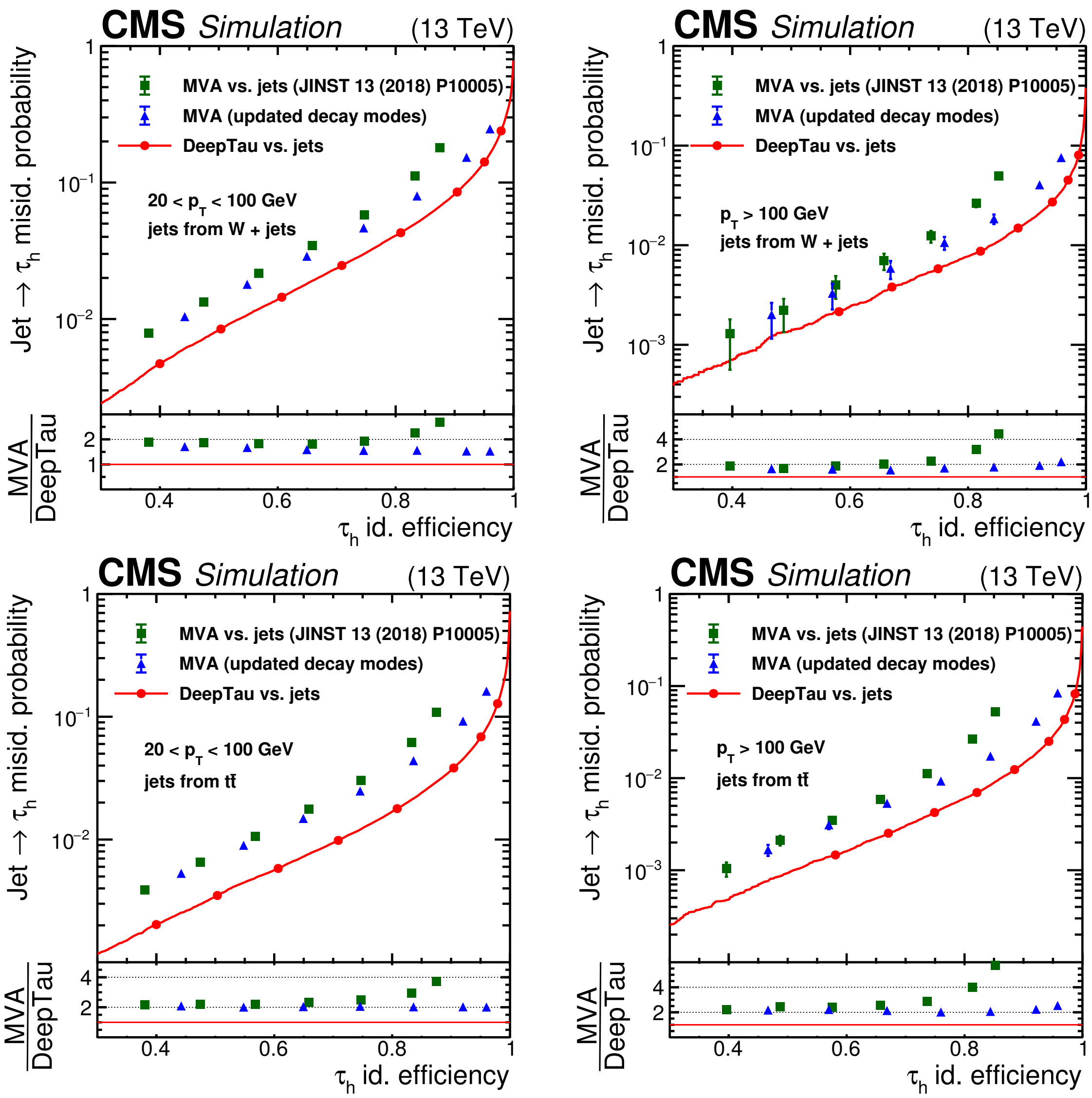


Figure 3.20: Efficiency for quark and gluon jets to pass different tau identification discriminators versus the efficiency for genuine τ_h, for simulated taus that pass the MVA [151] and D_{jet} discriminators. The upper two plots are obtained with jets from the W+jets simulated sample and the lower two plots with jets from the $t\bar{t}$ sample [148].

filters are applied to remove anomalous high-p_T^{miss} events, which can arise because of a variety of reconstruction failures or detector noise and inefficiency. The filters include:

- Calorimeter filters, based on the isolation, electronic signal shape and timing in HCAL, or on noisy crystals and non-functioning electronics in the ECAL.
- Beam halo filters,which remove events with energy deposits along a line with constant ϕ, sometimes matching the muon CSC signals attributed to beam halo particles.
- Reconstruction filters, which veto events with poorly reconstructed high energy charged hadrons or muons, by looking at tracks with low quality and very high-p_T.

MET reconstrution performance

The reconstruction performances of p_T^{miss} are usually measured in events with Z/γ + jets, with $Z \rightarrow \mu\mu$ or $Z \rightarrow ee$ where no genuine MET is expected. In such events the MET resolution is dominated by the hadronic activity, since the momentum resolution for leptons and photons is order 1%, compared to $5-20\%$ for the jet momentum resolution. The MET for $Z \rightarrow \mu\mu$ + jets events is shown in the upper left panel of Fig. 3.21, the peak value is at about 30 GeV. In the $Z \rightarrow \mu\mu$ + jets topology one can also measure the p_T^{miss} performance by comparing the momenta of the vector boson to that of the hadronic recoil system. The hadronic recoil system is defined as the vector p_T sum of all PF candidates except for the muons from the Z. The hadronic recoil vector is projected along the z-axis to evaluate the response, and the orthogonal component, as well as the difference between the recoil along the axis and the Z p_T provide a similar estimate of the p_T^{miss} resolution, as shown in the right and left lower panels of Fig. 3.21. The intrinsic resolution, after removing the PU contribution, is $\sim 10\,\text{GeV}$ for both the components of the recoil.

The $(\vec{\mathbf{p}}_\mathbf{T}^{\mathbf{miss}} - \vec{\mathbf{p}}_\mathbf{T}(\mu))$ quantity

For the first time in this search, a MET trigger has been introduced, inspired by earlier high-mass $b\bar{b}\tau^+\tau^-$ studies [153]. Unlike other triggers, centrally provided scale factors (SFs, see Appendix A.3 for the definition) for MET triggers are unavailable, necessitating their derivation as part of this analysis. Details of this procedure are described in Section 6.1.7.

Evaluating the efficiency of MET triggers is particularly challenging because MET includes contributions from all objects in the event by its definition. Consequently, there is no independent dataset (2.4.3 for dataset definition and Section 4.2 for the datasets used in this search) that serves as a non-overlapping reference point to define the denominator for efficiency calculations. Such a reference dataset, often termed "orthogonal", is critical for avoiding biases from overlaps with regions used for other purposes in the analysis, such as MC corrections or specific search regions.

To overcome this challenge, the contribution of muons is removed from the MET definition, introducing a modified quantity called MET-noμ, which is defined as:

$$\text{MET-no}\mu = \vec{p}_T^{miss} - \vec{p}_T(\mu). \tag{3.10}$$

This modified quantity, referred to as MET-noμ, matches the standard p_T^{miss} in scenarios where no muons are present in the $\tau\tau$ final state. Furthermore, the MET trigger used in this analysis specifically targets MET-noμ (Section 4.4), making it both a more appropriate and consistent choice for this study. This approach ensures that the MC correction calculations remain statistically robust and free from phase-space overlap biases.

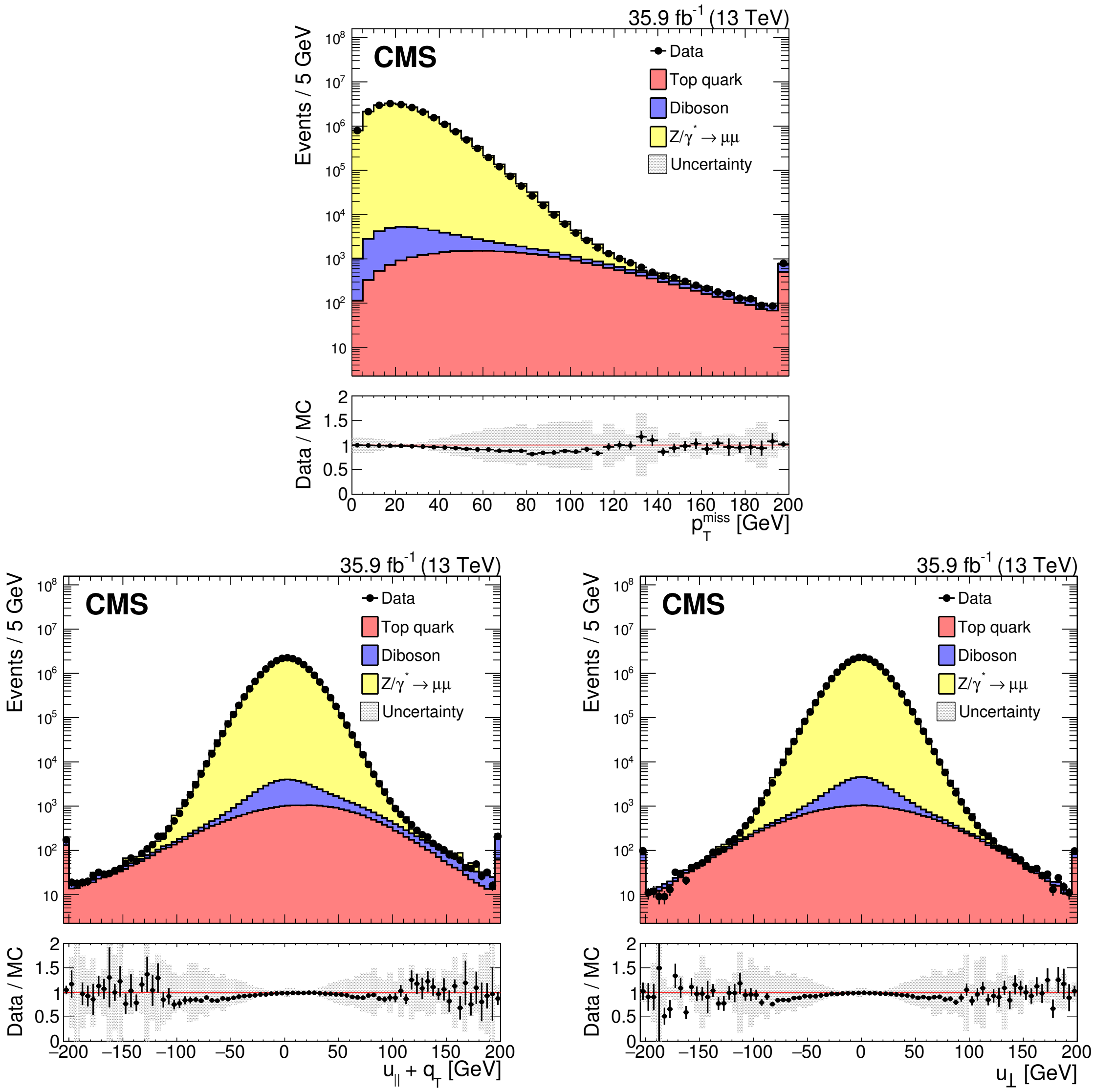


Figure 3.21: Distributions of p_T^{miss} in $Z \to \mu\mu$ (upper panel), parallel (lower panel, left) and transverse components (lower panel, right) of the hadronic recoil, in data and simulation. [152]

Chapter 4

The $X \to HH \to bb\tau\tau$ analysis strategy

The study of Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel necessitates the ability to identify and reconstruct various types of final state objects and use them effectively to select signal-like events. This involves the reconstruction of $H \to \tau^+\tau^-$ and $H \to bb$ decays and the strategic use of their properties to distinguish signal events from background ones. These processes and techniques collectively define the analysis strategy, which serves as the cornerstone of the $HH \to bb\tau\tau$ search.
This chapter presents the analysis strategy introducing first the experimental challenges associated with the $b\bar{b}\tau^+\tau^-$ decay channel and the distinctive signatures of signal processes and the primary sources of background. Section 4.1 introduces the final states explored in this analysis, highlighting the particles of interest and their experimental signatures. Sections 4.2 and 4.3 provide an overview of the collision data utilized in the analysis and the Monte Carlo (MC) simulated processes used to model the data composition. These considerations put the basis for the analysis strategy, which is detailed in subsequent sections. The event selection workflow, applied to both experimental data and simulations, is designed to maximize signal sensitivity while filtering out non interesting events. It is organized into four key elements: trigger requirements, objects selections, event categorization, and the definition of phase space regions optimized to maximise the signal presence (or other processes in some cases). Each step builds upon the previous one, employing increasingly sophisticated criteria to reject non-interesting events. Trigger requirements serve as the initial level of event selection, ensuring relevant events are stored for further offline analysis. These are described in Section 4.4. In addition to triggers, the selection process includes quality cuts that impose strict criteria on the reconstructed final state objects, discussed in Section 4.5. These refined objects are then used to form $H \to \tau^+\tau^-$ and $H \to bb$ candidates, detailed in Sections 4.6 and 4.7 respectively. After applying the full selection workflow, the remaining events are classified into distinct channels, categories, and phase-space regions exploiting the reconstructed masses of the $H \to \tau^+\tau^-$ and $H \to bb$ candidates, as well as their kinematic properties. The motivations and definitions of these subdivisions are presented in Section 4.8.

Note: Throughout the thesis, processes that mimic the searched signature but are not of primary interest will be referred to as **backgrounds**, while the process of interesty is defined as the **signal**.

4.1. The HH $\to bb\tau\tau$ final state

In the CMS detector, the $b\bar{b}\tau^+\tau^-$ final state particles cannot be directly observed. This is because b-quarks undergo hadronization, forming hadronic jets (as described in Section 3.3.4), while τ leptons decay into neutrinos and lighter particles, such as leptons (muons or electrons) or quarks. Consequently, their reconstruction relies on identifying muons, electrons, or hadronic taus, described in Sections 3.3.2, 3.3.3, and 3.3.5, respectively. Additionally, the presence of Missing Transverse Energy (MET), discussed in Section 3.3.6, is a key feature due to neutrinos produced in the decays of τ leptons and during b-quark hadronization. Thus, even if this Higgs bosons pair (HH) final state is more clean with respect to fully hadronic ones (as discussed in Section 1.7.1), it poses non trivial experimental challenges, such as reconstructing tau leptons and identifying hadronic jets coming from b quarks, so the use of precise particle reconstruction and identification techniques is crucial. The possible $\tau^+\tau^-$ final state topologies are: $\tau_e\tau_e$, $\tau_e\tau_\mu$, $\tau_e\tau_h$, $\tau_\mu\tau_\mu$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$, with $\tau_\ell = \tau \to \ell\nu_\ell\nu_\tau$ ($\ell = \mathrm{e}, \mu$) and $\tau_h = \tau \to \nu_\tau +$ hadrons representing leptonic and hadronic decays respectively. In the following, the channels will be indicated as ee, $e\mu$, $e\tau_h$, $\mu\mu$, $\mu\tau_h$, and $\tau_h\tau_h$respectively.
The X $\to$ HH $\to$ $b\bar{b}\tau\tau$ search presented focuses on a subset of these final states, referred to as ***channels***, where one of the two τ leptons decays hadronically (τ_h): $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$, covering approximately 88% of all $\tau^+\tau^-$ decays.
Fully leptonic final states are excluded due to their lower branching fractions and higher background levels. However, they are used for different purposes such as data/MC agreement studies and the estimate of some MC corrections.

4.2. Datasets

This search uses proton-proton (pp) collisions data at $\sqrt{s} = 13\,\mathrm{TeV}$ recorded by the CMS experiment between 2016 and 2018, for a total integrated luminosity of $\sim 138\,\mathrm{fb}^{-1}$. Data and simulations are split in 4 *eras*: 2016preVFP and 2016postVFP[1], 2017 and 2018; the recorded luminosities in these eras are respectively 19.5, 16.8, 41.5, and $59.8\,\mathrm{fb}^{-1}$. For each era, only the pp runs certified good for physics are employed.
Data and simulations are provided in the form of **primary datasets** (Section 2.4.3), designed to efficiently record at the same time interesting processes characterized by different final state objects. This analysis uses four primary datasets, respectively

[1]The 2016 year is split in two eras due to the substantial change in detector conditions between them. In the 2016preVFP (also referred as 2016HIPM or 2016APV) era, the strip tracker had a lower signal-to-noise ratio and fewer hits on tracks due to saturation effects in the readout chip under high-luminosity conditions. This was mitigated in the 2016postVFP (also referred as 2016) era by changing the feedback preamplifier bias voltage (VFP) [154].

targeting one muon (`SingleMuon`), one electron (`SingleElectron`), one tau (`Tau`) or the presence of MET (`MET`).
Both MC and data are part of an important improvement in data quality after a reprocessing occurred during the Long shutdown 2 (LS2). This reprocessing, defined "Ultra-Legacy" (UL), includes enhanced Electromagnetic Calorimeter (ECAL) calibration and improved energy resolution, in order to achieve optimal performance [155].

4.3. Signal and backgrounds contributions

Data are compared with simulations of background processes. If the signal is present and the MC simulations are accurately modeled, statistical inference combined with the inclusion of systematic uncertainties can reveal a statistically significant difference between the observed data and the background-only simulation. This makes accurate modeling of both the background and signal processes essential for performing reliable statistical analyses and obtaining precise results. A detailed description of the MC simulations, including their production and cross sections, is presented in the next chapter.

Signal simulations play a crucial role in defining the analysis strategy. They are used to determine the trigger paths to include, establish selection criteria for the relevant objects to maximize signal acceptance while rejecting irrelevant events and define the phase-space regions most sensitive to the signal.

Background processes are divided into "irreducible" and "reducible" types. Irreducible backgrounds arise from processes that produce the same final state as the signal, while reducible backgrounds include processes with different final states that can mimic the signal due to misidentification of objects. Although particle Identification (ID) taggers and event selection strategies target very low misidentification rates, some processes, such as Quantum Chromodynamics (QCD) multijet production, have a production cross section orders of magnitude larger than the signal, contributing significantly to the overall background.

The primary source of irreducible background is the top-quark pair ($t\bar{t}$) production, particularly in the decay channel $t\bar{t} \to W^{\pm}W^{\mp} \to bb\ell^{\pm}\ell^{\mp}\nu\bar{\nu}$, with $(\ell = e, \mu, \tau)$. This background, described in Section 5.2.2, impacts all channels and requires a dedicated phase-space region to validate the data/MC agreement and ensure accurate simulation, as reported in Section 4.8.3.
The second most relevant irreducible background, discussed in Section 5.2.3, is the Drell Yan (DY) $Z \to \ell^{\pm}\ell^{\mp}$ + jets, particularly the ones coming from the process $Z + bb \to \ell^{\pm}\ell^{\mp} + b\bar{b}$. The DY background modelling relies on MC simulations and is validated in a specific control region for data/MC validation, described in Section 4.8.3.
The most significant reducible background arises from QCD multijet processes, where quarks and/or gluons jets are misidentified as hadronically decaying τ leptons or b-jets. This background particularly impacts in the $\tau_h\tau_h$ channel, while its contribution to other channels is smaller due to stringent selection criteria for light leptons that

enhance rejection rates. Unlike other backgrounds, QCD background modelling relies on a data-driven technique, detailed in Section 5.4.
Another reducible background includes W boson production in association with jets (in Section 5.2.4). Finally, minor contributions, both reducible and irreducible, from rarer processes are also considered (see Sections 5.2.5-5.2.10).

4.4. Trigger selection

Trigger requirements are applied to both data and MC simulations. The adopted High Level Trigger (HLT) trigger paths for each era are summarized in Tab. 4.1, Tab. 4.2, and Tab. 4.3. For each event passing a specific trigger path, both online and offline objects must meet certain criteria. Online objects selections are typically indicated by the trigger path names, offline objects cuts are detailed in Section 4.6. Additionally, there must be a match between online and offline objects, as discussed in Section 2.4.2.

Trigger	Channel	HLT path name
diTau	$\tau_h\tau_h$	HLT_DoubleMediumCombinedIsoPFTau35_Trk1_eta2p1_Reg*[2]
		HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_Reg*[2]
SingleTau	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$	HLT_VLooseIsoPFTau120_Trk50_eta2p1 OR HLT_VLooseIsoPFTau140_Trk50_eta2p1
MET	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
SingleMu	$\mu\tau_h$, $\mu\mu$	HLT_IsoMu24 OR HLT_IsoTkMu24
CrossMuTau	$\mu\tau_h$	HLT_IsoMu19_eta2p1_LooseIsoPFTau20 OR HLT_IsoMu19_eta2p1_LooseIsoPFTau30 OR HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1
SingleEle	$e\tau_h$, ee	HLT_Ele25_eta2p1_WPTight_Gsf
CrossEleTau	$e\tau_h$	HLT_Ele24_eta2p1_WPLoose_Gsf_LooseIsoPFTau20 OR HLT_Ele24_eta2p1_WPLoose_Gsf_LooseIsoPFTau20 OR HLT_Ele24_eta2p1_WPLoose_Gsf_LooseIsoPFTau20_SingleL1

Table 4.1: HLT trigger paths for 2016 and 2016 HIPM eras.

4.5. Event selection

To enhance the signal presence and minimizing the background contributions, some requirements are applied to the events and the reconstructed objects. The selection flow is inherited from the non-resonant HH $\to bb\tau\tau$ analysis [156] but it has been optimized to maximise sensitivity to the resonant signals. The event selection workflow follows these steps:

[2]The HLT trigger paths labeled with the * symbol are dependent on the run number but it is omitted for simplicity.

Trigger	Channel	HLT path name
diTau	$\tau_h\tau_h$	HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg OR HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg OR HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg
SingleTau	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$	HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1
MET	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
SingleMu	$\mu\tau_h$, $\mu\mu$	HLT_IsoMu24
CrossMuTau	$\mu\tau_h$	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1
SingleEle	$e\tau_h$, ee	HLT_Ele32_WPTight_Gsf
CrossEleTau	$e\tau_h$	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau30_eta2p1_CrossL1

Table 4.2: HLT trigger paths for 2017 era.

Trigger	Channel	HLT path name
diTau	$\tau_h\tau_h$	HLT_DoubleMediumChargedIsoPFTauHPS35_Trk1_eta2p1_Reg* [2]
		HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg*[2] OR HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg*[2] OR HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg*[2]
SingleTau	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$	HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1
MET	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
SingleMu	$\mu\tau_h$, $\mu\mu$, $e\mu$	HLT_IsoMu24
CrossMuTau	$\mu\tau_h$	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau(HPS)27_eta2p1_CrossL1
SingleEle	$e\tau_h$, ee, $e\mu$	HLT_Ele32_WPTight_Gsf
CrossEleTau	$e\tau_h$	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau(HPS)30_eta2p1_CrossL1

Table 4.3: HLT trigger paths for 2018 era.

I. An initial filter is applied to data samples to select the luminosity sections[3] corresponding to runs considered good for physics analyses.

II. A set of filters for MET noise cleaning is applied to both data and simulations, following centrally distributed recommendations reported in Sect 7.7.5 of [157].

III. A general pre-selection is first applied to the leptons, followed by the trigger selection (see Section 4.4) and more stringent offline requirements. Further details of lepton selection process are discussed in Section 4.6.1.

IV. The next step is the $H \to \tau^+\tau^-$ candidate identification, performed via the $H \to \tau^+\tau^-$ pair assessment algorithm, described in Section 4.6.2. At this stage

[3]A lumisection is a time interval of 23.3 s serving as a unit for recording luminosity, monitoring data quality, and organizing the vast amount of collision data produced during each run. The 23.3 s duration of a is chosen to balance the need for precise luminosity measurements and practical data management, providing sufficient granularity for monitoring collision data and detecting anomalies while remaining manageable for storage and processing.

events are separated in different channels according to the final state leptons coming from the $\mathrm{H} \to \tau^+\tau^-$ decay.

V. Events with additional electrons or muons passing the requirements reported in Tab. 4.4 are discarded. This condition is defined as *third lepton veto*.

VI. Subsequently, the presence of at least either two AK4 resolved jets or one AK8 merged jet meeting specific selection criteria, is required to identify the $\mathrm{H} \to \mathrm{bb}$ candidate, as outlined in Section 4.7.

	p_T (GeV)	$\|\eta\|$	$\|d_z\|$	$\|d_{xy}\|$	ID
e	> 10	< 2.5	< 0.2	< 0.045	`MVAIsoWP90`
μ	> 10	< 2.4	< 0.2	< 0.045	`(TightID OR MediumID)`

Table 4.4: Third lepton veto selection. The electron and muon ID are introduced in Sections 3.3.3 and 3.3.2.

During the entire selection, the isolations of electrons and muons are the **Particle Flow (PF) relative isolation** (`PF RelIso03`), defined in Eq. 3.2, whereas hadronically decaying tau isolation relies on the **DeepTauVSJet tagger**, introduced in Section 3.3.5.

4.6. The $\mathrm{H} \to \tau^+\tau^-$ candidate reconstruction

4.6.1 Lepton selection

The requirements imposed on offline leptons can be divided into three groups: *general pre-selection*, *trigger-dependent selection* and *channel-dependent requirements*. These are applied sequentially.

The general pre-selection consists in the application of "general" quality cuts to leptons, summarized in Tab. 4.5. These criteria are based on standard recommendations from the CMS collaboration, optimized for analyses targetting prompt leptons, and are independent of both the data-taking era and channel.

The trigger-dependent selection depends on applied triggers and eras. Each reconstructed offline object matched with its corresponding online trigger object (within $\Delta R < 0.5$) and satisfy the η and transverse momentum component ($\mathrm{p_T}$) thresholds specified by the HLT trigger path fired in the event:

- $p_T^{offline} > p_T^{online} + p_T^{offset}$,
- $|\eta| < |\eta^{online}|$, if the trigger path includes η requirements.

Here, $p_T^{offline}$ and p_T^{online} represent the $\mathrm{p_T}$ of the offline object and the $\mathrm{p_T}$ threshold of the online trigger object, respectively. The offset (p_T^{offset}) depends on the lepton

	p_T (GeV)	$\|\eta\|$	$\|d_z\|$	$\|d_{xy}\|$	Iso	ID
e	> 10	< 2.5	< 0.2	< 0.045	PF RelIso03 < 0.3	MVAIsoWP80 OR MVANoIsoWP80
μ	> 15	< 2.4	< 0.2	< 0.045	PF RelIso03 < 0.3	TightID
τ_h	> 20	2.3	< 0.2	-	VSjet $\geq$ VVVLoose WP	VSe $\geq$ VLoose WP VSmu $\geq$ Tight WP

Table 4.5: General selection applied to offline leptons before the $H \to \tau^+\tau^-$ candidate selection. The electron, muon and hadronic tau ID are introduced in Sections 3.3.3, 3.3.2 and 3.3.5.

type: it is 2 GeV for electrons and muons and 5 GeV for τ_h. The values are chosen conservatively based on the trigger turn-on curves, ensuring triggers are fully efficient. For example, in 2018 the events fired by the
HLT_DoubleMediumChargedIsoPFTauHPS**35**_Trk1_**eta2p1**_Reg trigger require *online* taus with $p_T > 35$ GeV and $|\eta| < 2.1$ matched with *offline* taus with $p_T > 40$ GeV and $|\eta| < 2.1$.
Since also the MET trigger is included, some cuts on MET-noμ (defined in Section 3.3.6) are also applied at this stage. The offset value for the p_T^{miss} trigger is 10 eV.
The lepton and MET-noμ trigger based cuts are summarized in Tab. 4.6a, where the first column lists the trigger names, with the corresponding trigger paths reported in Tables 4.1, 4.2, and 4.3. For some trigger path, the p_T thresholds are dependent on the data-taking era, and these are provided in Tab. 4.6b. These era-specific and trigger-driven requirements define the so-called ***validity regions***.

	$p_T(\ell_1)$ (GeV)	$\|\eta(\ell_1)\|$		$p_T(\ell_2)$ (GeV)	$\|\eta(\ell_2)\|$
SingleEle	p_T (e)	2.5 (2016) 2.1 (2017, 2018)	OR	p_T (e)	2.5 (2016) 2.1 (2017, 2018)
CrossEleTau	> 25	< 2.1	AND	> 35	< 2.1
SingleMu	> p_T (μ)	< 2.4	OR	> p_T (μ)	< 2.4
CrossMuTau	> 22	< 2.1	AND	> 25	< 2.1
Di-tau	> 40	<2.1	AND	> 40	<2.1
SingleTau	> p_T (τ_h)	< 2.1	OR	> p_T (τ_h)	< 2.1
MET	MET-noμ > p_T (MET)				

(a)

p_T (GeV)	2018	2017	2016
e	33	33	26
μ	26	29	26
τ_h	190	190	130
MET	150	150	130

(b)

Table 4.6: Tab. 4.6a Offline lepton/MET-noμ (defined in Section 3.3.6) requirements based on trigger selection. The electron, muon and hadronic tau ID are described in Sections 3.3.2, 3.3.3, and 3.3.5. The validity regions names correspond to the trigger names in the first column of trigger tables (Tab. 4.1, 4.2 and 4.3). Tab. 4.6b: era-dependent p_T /MET-noμ values used in Tab. 4.6a.

The channel-dependent requirements are applied when events are split in channels depending on the $\mathrm{H} \to \tau^+\tau^-$ final state leptons types and are reported in Tab. 4.7.

4.6.2 The $\mathrm{H} \to \tau^+\tau^-$ pair assessment

In addition to the lepton selection, the possible $\mathrm{H} \to \tau^+\tau^-$ candidates are defined by selecting all the pairs of leptons passing the previous selections and separated by $\Delta R \leq 0.4$, where ΔR is defined in Eq. 2.6. The final $\mathrm{H} \to \tau^+\tau^-$ candidate is then chosen with the "pair assessment" algorithm inherited from the $\mathrm{H} \to \tau^+\tau^-$ analysis [158]. This algorithm evaluates candidate pairs legs[4] based on the following criteria, listed in decreasing priority order:

I. **First leg isolation**: The candidate pair with the most isolated first leg is chosen.

II. **First leg $\mathbf{p}_T$** : if two or more pairs have the same first leg isolation, the pair having the first leg with highest p_T is selected.

III. **First leg** $|\eta|$: in the case where two or more pairs have the first leg with same isolation and p_T the pair having the first leg with highest $|\eta|$ is considered.

IV. **Second leg assessment:** if there is still ambiguity among two or multiple pairs, the same sequence of criteria is applied to the second leg of the pair.

Events not passing the pair assessment algorithm are discarded. Once the $\mathrm{H} \to \tau^+\tau^-$ channel is defined, channel-specific selection requirements are applied. These are summarized in Tab. 4.7.

4.7. The $\mathrm{H} \to \mathrm{b\bar{b}}$ candidate definition

The $\mathrm{H} \to \mathrm{bb}$ decay is reconstructed requiring at least either two AK4 "resolved" jets or one AK8 "merged" jet passing the selection criteria listed in Tab. 4.8. The resolved jets passing this selection are assigned a score by the *HH-bTag* tagger, a dedicated machine learning based algorithm aiming to identify the jets coming from the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay. The HH-bTag algorithm was developed in the context of the non-resonant analysis [156] and has been retrained for this analysis and optimized for the resonant $\mathrm{X} \to \mathrm{HH} \to \mathrm{b\bar{b}}\tau\tau$ search; details on the HH-bTag algorithm are reported in Appendix B. The two resolved jets with the highest HH-bTag score are used to form the the $\mathrm{H} \to \mathrm{bb}$ candidate and are saved for further analysis.

The definition of the $\mathrm{H} \to \mathrm{bb}$ candidates using either two resolved or one boosted jet are concurrent. This will be discussed in detail in next section, but it is important to remark that the events always have only *one* $\mathrm{H} \to \mathrm{bb}$ candidate, either resolved or boosted.

[4]The first leg refers to a specific lepton in the pair, which is chosen based on the channel. For $e\tau_h$and $e\mu$, the first leg is the electron. For $\mu\tau_h$it is the muon, and for ee, $\mu\mu$ and $\tau_h\tau_h$ one either of the two $\mathrm{e}/\mu/\tau_h$ respectively can serve as the first leg.

Channel	Leg number	Iso	ID
$e\tau_h$	1 (e)	PF RelIso03 < 0.15	MVAIsoWP80
	2 (τ_h)	-	
$\mu\tau_h$	1 (μ)	PF RelIso03 < 0.15	-
	2 (τ_h)	-	
$\tau_h\tau_h$	1 (τ_h)	VSJet $\geq$ Medium	-
	2 (τ_h)	-	
ee	1 (e)	PF RelIso03 < 0.15	MVAIsoWP80
	2 (e)	-	MVANoIsoWP80
$\mu\mu$	1 (μ)	PF RelIso03 < 0.15	-
	2 (μ)	-	
$e\mu$	1 (e)	-	
	2 (μ)	-	

Table 4.7: Channel-dependent leptons requirements applied after the $H \to \tau^+\tau^-$ candidate selection.

	p_T (GeV)	$\|\eta\|$	ΔR(jet,leptons)	Other
AK4 jets	> 20	< 2.5	≥ 0.5	JetID = 2 p_T >50 GeV OR PuJetID
AK8 jets	> 250	< 2.5	≥ 0.8	$m_{softdrop} > 30$ GeV

Table 4.8: Jet selection applied before the $H \to bb$ candidate selection. The JetID and PuJetID respectively refer to the Noise and pile-up (PU) Jet identification algorithms introduced in Section 3.3.4, whereas the the soft drop mass (m_{softdrop}) is a reliable estimate of the mass for AK8 boosted jets [159].

4.8. Event categorization

After the selection the surviving events are divided into various subsets based on the phase space of different observables. The following definitions for event classification are adopted:

- **Channels** refer to the $\tau\tau$ final state, categorized based on the leptons types. The channels used to extract the signal are $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$, while ee, $e\mu$, $\mu\mu$ are used to validate MC/data agreement.
- **Categories** classify events based on the $b\bar{b}$ candidate jets and their b-tagging discriminants. The most general category is referred to as the baseline category. Depending on the number of resolved b-jets or the presence of an AK8 $H \to bb$

jet, events are further classified into Resolved 2b category (res2b), res2b, or boosted categories. Details are provided in Section 4.8.1.

- **Trigger application regions** are defined to handle trigger-related MC corrections. These regions are non-overlapping (orthogonal) subsets of events, classified combining different validity regions, as described in Section 4.8.2.
- **Signal and Control regions** are determined by combining channel definitions with specific requirements on the invariant mass of lepton pairs and $b\bar{b}$ candidates. Control Regions (CR) are insensitive to the signal and are used for tasks such as validating MC/data agreement and in some cases estimating MC corrections. The Signal Region (SR) corresponds to the phase space sensitive to the signal, where the final fit is performed. Detailed definitions are provided in Section 4.8.3.
- **QCD regions** are defined to estimate the contribution of the QCD multijet background using a data-driven method. These regions are categorized based on the second lepton leg isolation and charge-sign conditions. Both the regions and the method are described in Section 5.4.

4.8.1 Categories

The jets passing the selection discussed in Section 4.7 are used to define **categories** for the event classification. Initially the events belong to the most general category, the ***baseline category***. From there, events are categorized in the **res2b**, **boosted** and **Resolved 1b category (res1b) categories**, based on the following requirements:

I. ***Resolved-2b (res2b) category***: events are first checked to see if both AK4 resolved jets associated with the H → bb candidate pass the *DeepFlavour b-jet tagger* at the Medium working point (WP, see Appendix A.1.2), corresponding to a 1% misidentification rate as defined in the b-jet paragraph of Section 3.3.4. Events meeting this criterion are assigned to the ***res2b category***.

II. ***Boosted category***: events not meeting the above requirement are checked for the presence of at least one AK8 jet. If an AK8 jet is found, the event is classified as a *boosted candidate*. To promote the event to the ***boosted category***, the AK8 jet must pass the Loose Purity (LP) Working Point (WP) of the *pNetMD Jet Tagger*, D_{bb}, defined in Eq. 3.7. If multiple AK8 jets satisfy this condition, the jet with the highest D_{bb} value is selected. Boosted candidate events that fail the *pNetMD* requirement are discarded.

III. ***Resolved-1b (res1b) category***: events that do not fall into either of the above categories are checked for the presence of exactly one AK4 jet passing the Medium WP of the *DeepFlavour b-jet tagger*. If this condition is met, the event is assigned to the ***res1b category***.

IV. The non-boosted candidate events containing two resolved jets (independently on their DeepFlavour tagger values) fall into the ***inclusive* category**. This category is not considered for the signal extraction.

Events that do not meet the criteria for any category are discarded. The event categorization process is summarized in Fig. 4.1.

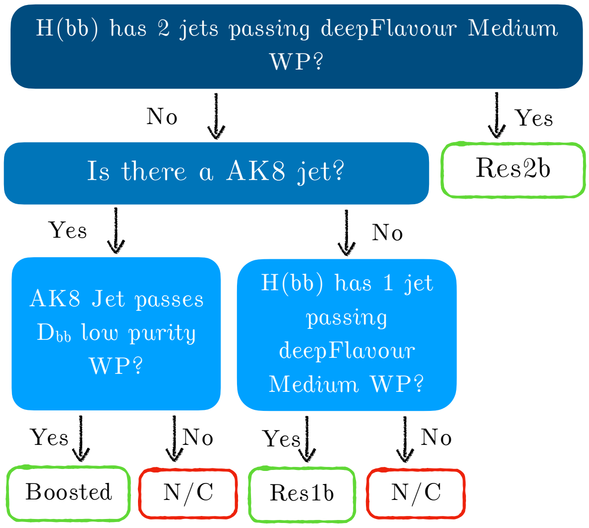


Figure 4.1: Event categorization scheme. In this scheme N/C stands for Not Condisered for the signal extraction process.

4.8.2 Trigger application regions

Differences in trigger efficiencies between data and MC simulations must be taken into account to ensure statistical accuracy and reliable signal extraction. These differences arise from factors such as hardware performance, software algorithms, and detector modeling. In CMS, Scale Factors (SF) are applied as multiplicative corrections to align MC trigger efficiencies with those observed in data. A more detailed overview of the SFs definitions and usage in HEP is provided in Appendix A.3. In general, the SFs correct these discrepancies to maintain the consistency of physics analyses, as will be explained in Section 6.1 The requirements applied to offline leptons in Tab. 4.6 define the **validity regions**, the phase-space regions within which SFs are centrally provided for individual trigger paths[5]. When two or more trigger paths are fired by the offline objects (e.g., τ_h, electron, muon or MET) in an event, there are two statistically valid approaches to handle the phase-space overlap [160] and correctly apply the associated SFs: either re-evaluating the efficiencies and SFs accounting for the overlap or defining orthogonal regions to apply the SFs independently. In this analysis, the first approach is adopted to manage the overlap between single-electron/muon trigger paths and electron/muon + τ_h trigger paths, defining the "Legacy[6] region". The triggers SFs evaluation for MC samples is described in Section 6.1.7. In addition, two trigger paths targeting high-p_T taus (SingleTau) and MET have been introduced in this search, and the motivation are discussed in Appendix C. To handle their SFs correctly, the

[5]Outside each validity region, SFs are not available for the corresponding trigger path.

[6]The term "Legacy" refers to the trigger paths and their associated application regions as defined in the non-resonant counterpart of this analysis [156].

second approach is applied, defining three orthogonal ***application regions***: the **Legacy region**, the **Single Tau region**, and the **MET region**. These regions are constructed by combining one or more **validity regions** with offline selection and triggers summarized in Tab. 4.6. This approach prevents double counting and ensures efficient and accurate SF application throughout the analysis. A summary of the application regions is presented in Tab. 4.9

Region Name	Channel	Trigger(s)	Definition
Legacy region	$e\tau_h$	`HLT_SingleEle OR HLT_eTau`	SingleEle `OR` CrossEleTau
	$\mu\tau_h$	`HLT_SingleMu OR HLT_muTau`	SingleMu `OR` CrossMuTau
	$\tau_h\tau_h$	`HLT_diTau`	Di-tau
	ee	`HLT_SingleEle`	SingleMu
	$\mu\mu$	`HLT_SingleMu`	SingleMu
	$e\mu$	`HLT_SingleEle OR HLT_singleMu`	SingleEle `OR` SingleMu
SingleTau Region	$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$	`HLT_SingleTau`	`NOT` in Legacy region
MET Region	$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$	`HLT_MET`	`NOT` in Legacy region `AND NOT` in SingleTau region

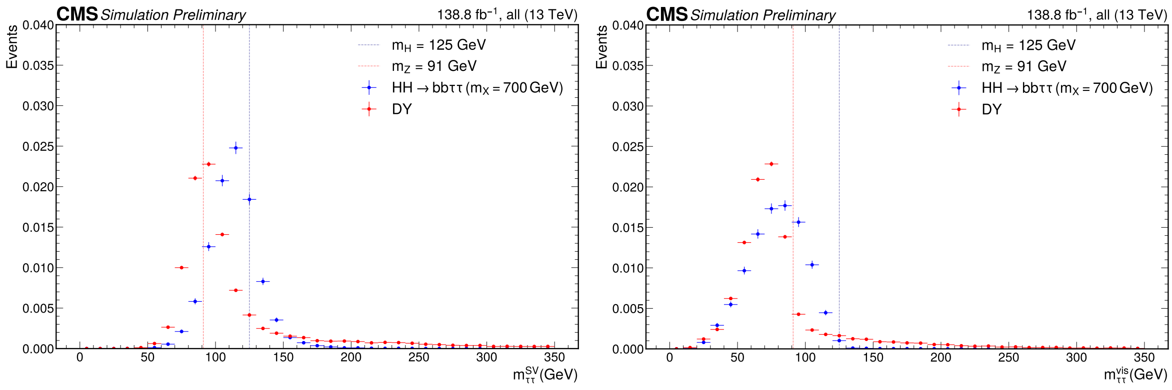
Table 4.9: Trigger application regions definition. The orange colored labels correspond to the validity regions requirements reported in Tab. 4.6.

4.8.3 Data and simulation agreement checks

To define a phase space sensitive to the signal while also enabling validation of statistical methods and data/MC agreement in regions not sensitive to the signal, events are divided into multiple orthogonal regions: a **signal region (SR)** and several **control regions (CRs)**. These regions are determined based on channels and constraints on the invariant masses of $\mathrm{H} \to \tau^+\tau^-$ and $\mathrm{H} \to \mathrm{bb}$ candidates.

In the case of the $\mathrm{H} \to \tau^+\tau^-$ candidate, due to the presence of two neutrinos from the τ decays in the final state, the invariant visible mass, $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$, calculated using the visible τ products, does not peak at the correct position and does not allow for the discrimination of the $h \to \tau_h\tau_h$ signal from the background, particularly in the case of $Z \to \tau_h\tau_h$. In order to improve the mass resolution, the Compact Muon Solenoid (CMS) collaboration developed the *SVFit* algorithm [161], which exploits the information of $\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the visible products of the τ decays, is used.
SVFit is a likelihood-based algorithm that reconstructs the invariant mass of the leptons pair coming from the $\mathrm{H} \to \tau^+\tau^-$ candidate, $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$, recovering the missing transverse energy information on an event-by-event basis. Further details on the algorithm and performances are reported in Appendix D. Fig. D.4 shows the $\mathrm{H} \to \tau^+\tau^-$ candidate pair distributions of the invariant visible mass, $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$(Fig. 4.2b) and the invariant mass reconstructed by *SVFit*, $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$(Fig. 4.2a) for signal $\mathrm{X} \to \mathrm{HH} \to \mathrm{b\bar{b}}\tau\tau$, in the spin-0

and $m = 700\,\text{GeV}$ hypothesis (in blue) and $Z \to \tau\tau$ plus jets (in red) events. The same distributions comparison for the other signal channels are reported in Appendix D. The improvement in discrimination power due to *SVFit* is clearly visible.



(a) **(b)**

Figure 4.2: Distributions of *SVfit* invariant mass,$m_{\tau\tau}^{(\text{SVFit})}$(Fig. 4.2a) and visible invariant mass $m_{\tau\tau}^{(\text{vis})}$(Fig. 4.2b) of the τ pair candidates in the $\tau_h\tau_h$ channel, for $X \to HH \to b\bar{b}\tau\tau$ (for spin 0 and $m_X = 700\,\text{GeV}$, in blue) and $Z \to \tau\tau$ + Jets ("DY", in red). Distributions are renormalized to the unit area.

The $H \to \tau^+\tau^-$ candidates are reconstructed via the $m_{\tau\tau}^{(\text{SVFit})}$ in signal channels ($e\tau_h$,$\mu\tau_h$ and $\tau_h\tau_h$), while for the other channels (which are used only for CR definition) the $m_{\tau\tau}^{(\text{vis})}$is used, calculated using the visible τ decay products as these channels are needed only for data/MC checks and the algorithm running the $m_{\tau\tau}^{(\text{SVFit})}$production is computationally heavy.
For $H \to bb$ candidates in the resolved categories the invariant mass of the resolved jet pair seected in Section 4.7 is used, whereas for the boosted categories the soft drop mass (m_{softdrop}) [159], a reliable estimate of the mass for AK8 merged jets [159], is used; in both cases the $H \to bb$ candidate mass is indicated with m_{bb}.
The SR is optimized for signal sensitivity, whereas the CRs are used to validate data/MC agreement for key backgrounds such as DY and $t\bar{t}$. Additional CRs are introduced in subsequent chapters, tailored for specific tasks such as estimatingQCD contributions, computing MET trigger and ParticleNet SFs.

In the plots presented throughout this chapter showing data/MC distributions, the data are displayed as points with error bars representing their uncertainties. The filled histograms indicate the different contributions, with the following color code applied (from top to bottom):

- **QCD**: Orange (reference).
- **DY**: Light Blue (reference).
- **$t\bar{t}$**: Dark Green (reference).
- **W+Jets**: Light Pink (reference).
- **Single top (ST)**: Cyan (reference).
- **Single Higgs (H)**: Yellow (reference).

- **Diboson (VV)**: Dark Violet (reference).
- **Triboson (VVV)**: Lavender (reference).
- **EWK**: Drably Olive (reference).
- **TTX**: Red (reference).

These plots are designed to facilitate a clear comparison between the observed data and the simulated contributions.

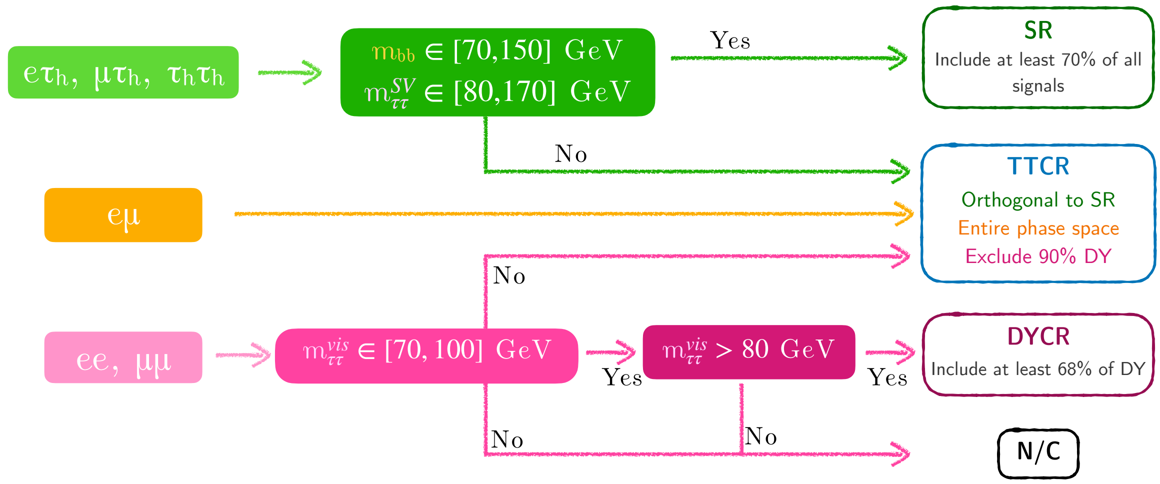


Figure 4.3: Regions definition scheme. In this scheme N/C stands for Not Condisered for the signal extraction process.

The signal and control regions, illustrated in Fig. 4.3, designed with uniform definitions, across all eras and channels, are summarized in Tab. 4.10. A consistent set of mass cuts is applied to both boosted and resolved categories, balancing simplicity and efficiency.

The Drell-Yan Control Region (DYCR) is defined to check the data/MC agreement for the DY simulations. To account for this background modelling (Section 5.2.3), three different DY MC sets have been included: the inclusive DY Z plus Jets processes, and the same processes binned in p_T and in jet multiplicity. To correctly treat those simulations which have phase space overlapping, a stitching procedure, described Section 5.3.1, is adopted. To check the correctness and the validity of the stitching technique, as well as the modelling of the DY MC sets, the Drell Yan Control Region (DYCR) is defined. It uses events in ee and $\mu\mu$ channels, which are by construction orthogonal to the SR with a $m_{\tau\tau}^{(vis)}$ range chosen to include at least 68% of the inclusive DY events. The DYCR conditions on channels, categories and $m_{\tau\tau}^{(vis)}$ are reported in Tab. 4.10.
The distributions of relevant observables in DYCR are reported in Appendix E. Some

Region	channels	categories	$m_{\tau\tau}$(GeV)	m_{bb}(GeV)
SR	$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$	res1b, res2b, boosted	$\in [80, 170]$	$\in [70, 150]$
DYCR	ee, $\mu\mu$	inclusive, res1b, res2b, boosted	$\in [80, 100]$	-
TTCR	ee, $\mu\mu$	res2b, boosted	$\notin [70, 100]$	-
	$e\mu$	res2b	-	-
	$e\tau_h$, $\mu\tau_h$	res2b, boosted	$\notin [80, 170]$	$\notin [70, 150]$
	$\tau_h\tau_h$	inclusive, res2b	$\notin [80, 170]$	$\notin [70, 150]$

Table 4.10: Mass ranges, channels and categories defining the signal regions and the control regions enriched in DY (DYCR) and $t\bar{t}$ (TTCR) background processes. For the ee and $\mu\mu$ channels the $m_{\tau\tau}$is the $m_{\tau\tau}^{(vis)}$whereas for the $e\tau_h$,$\mu\tau_h$ and $\tau_h\tau_h$ channels $m_{\tau\tau}$is $m_{\tau\tau}^{(SVFit)}$.

examples are shown in Figures 4.4 and 4.5, for leptons and b-jets/MET observables respectively.
The following considerations can be done:

- In res1b/res2b categories, the electrons p_T spectrum presents some fluctuations especially at high p_T, as shown in the upper panel of Fig. 4.4. These suggest that this background source can suffer of some mismodelling at high p_T even if dedicated MC set binned in p_T are included (Section 5.2.3). However, the fluctuations remain within $5 \div 10\%$ and can be accounted for once systematic uncertainties and MC statistical uncertainties are incorporated.
- There is a more relevant disagreement in the muons p_T spectra especially for $p_T \gtrsim 100\,\text{GeV}$), suggested by the lepton distributions for the $\mu\mu$ channel in the res1b and res2b categories respectively in the bottom right and left panels of Fig. 4.4. This is probably due to the same issue faced for electrons, however it is enhanced because the muon tight ID scale factors are less reliable at high p_T (Section 6.2.1).
- In the res1b and res2b categories the agreement is good in the entire p_T spectrum of the b-jets and in the part of the m_{bb} spectrum selected for the SR.
- The MET-noμ (Section 3.3.6) presents a good agreement in the inclusive category, whereas in the res1b/res2b categories (see Appendix E) the agreement tends to be worse, but still within $\approx 15\%$. This behaviour could be also related to the corresponding misbehaviour of the muons at high p_T as the muon ID tagger becomes inefficient.

Discrepancies between data and MC simulations are observed for high-p_T leptons, suggesting potential mis-modeling in the high-p_T DY samples. This persists even after applying the stitching technique to enhance statistics in the tails of distribution. However, the stitching method used in this study is a simplified version of more advanced techniques, as discussed in Section 2, which may introduce some inefficiencies. To mitigate this issue, systematic uncertainties associated with DY modeling and lepton taggers will be incorporated into the analysis (Section 6.2.1).

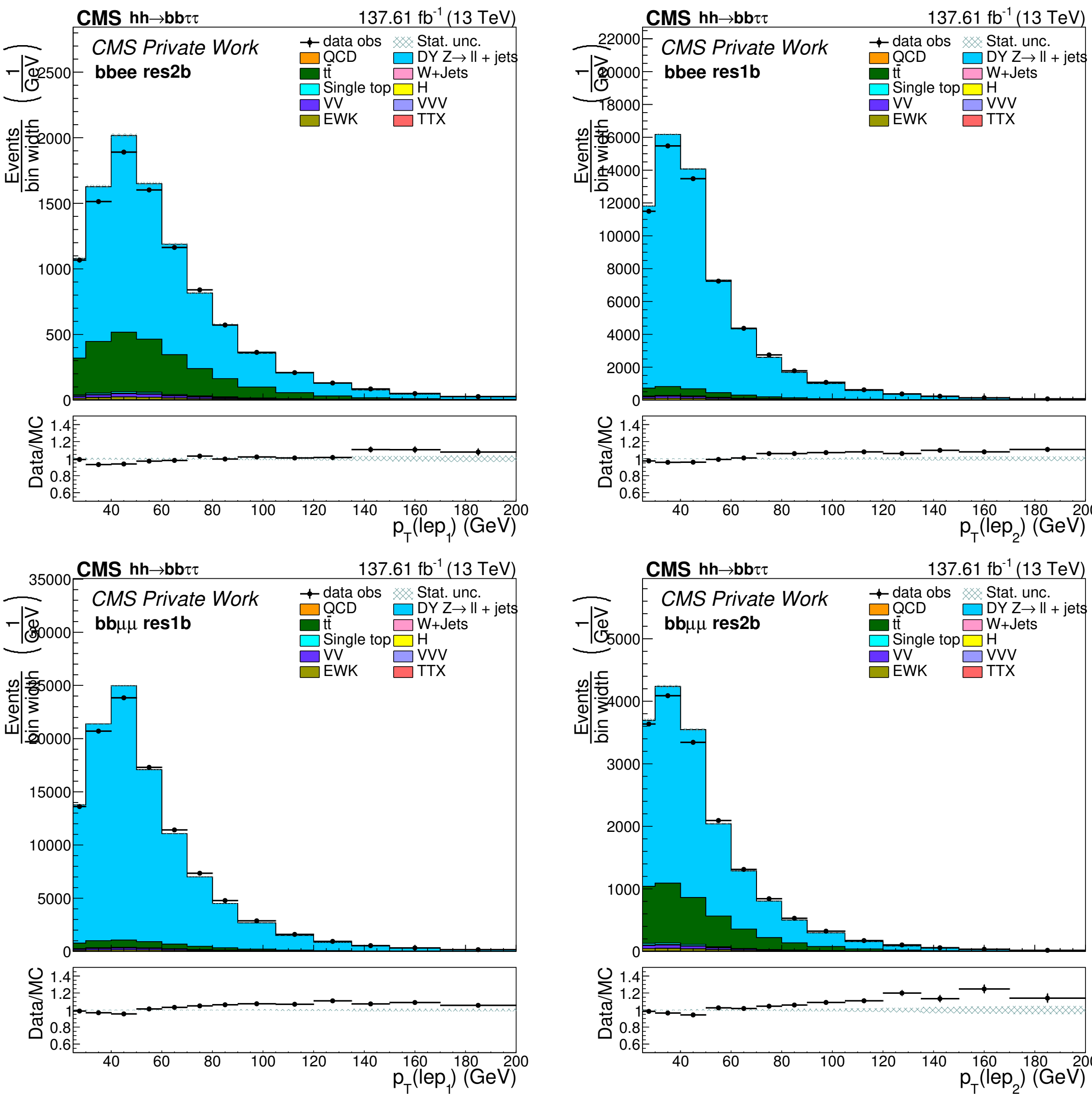


Figure 4.4: Leptons observables in the DYCR. Upper panel shows the first and second electron p_T in ee channel in the left and right respectively. The same distributions are shown in the lower panel for the $\mu\mu$ channel.

The $t\bar{t}$ control region (TTCR) is defined to perform data/MC agreement checks for the most important source of background: the top-antitop quark pair production. The reason of this CR definition is that the observed p_T spectra of top quarks in $t\bar{t}$ data were found to be significantly softer than predictions from MC simulations based on either Leading Order (LO) or Next to Leading Order (NLO) matrix elements interfaced with parton showers. This discrepancy was first noted in Run 1 and later confirmed in Run 2 by ATLAS [162]. This effect is believed to arise, at least partially, from the absence of higher-order corrections in the theoretical predictions. For instance, approximate Next-to-Next Leading Order (NNLO) predictions in perturbative QCD [163] demonstrated an improved agreement with data. Despite calculations at higher order [164–166] have further refined the theoretical description, though a residual discrepancy remains

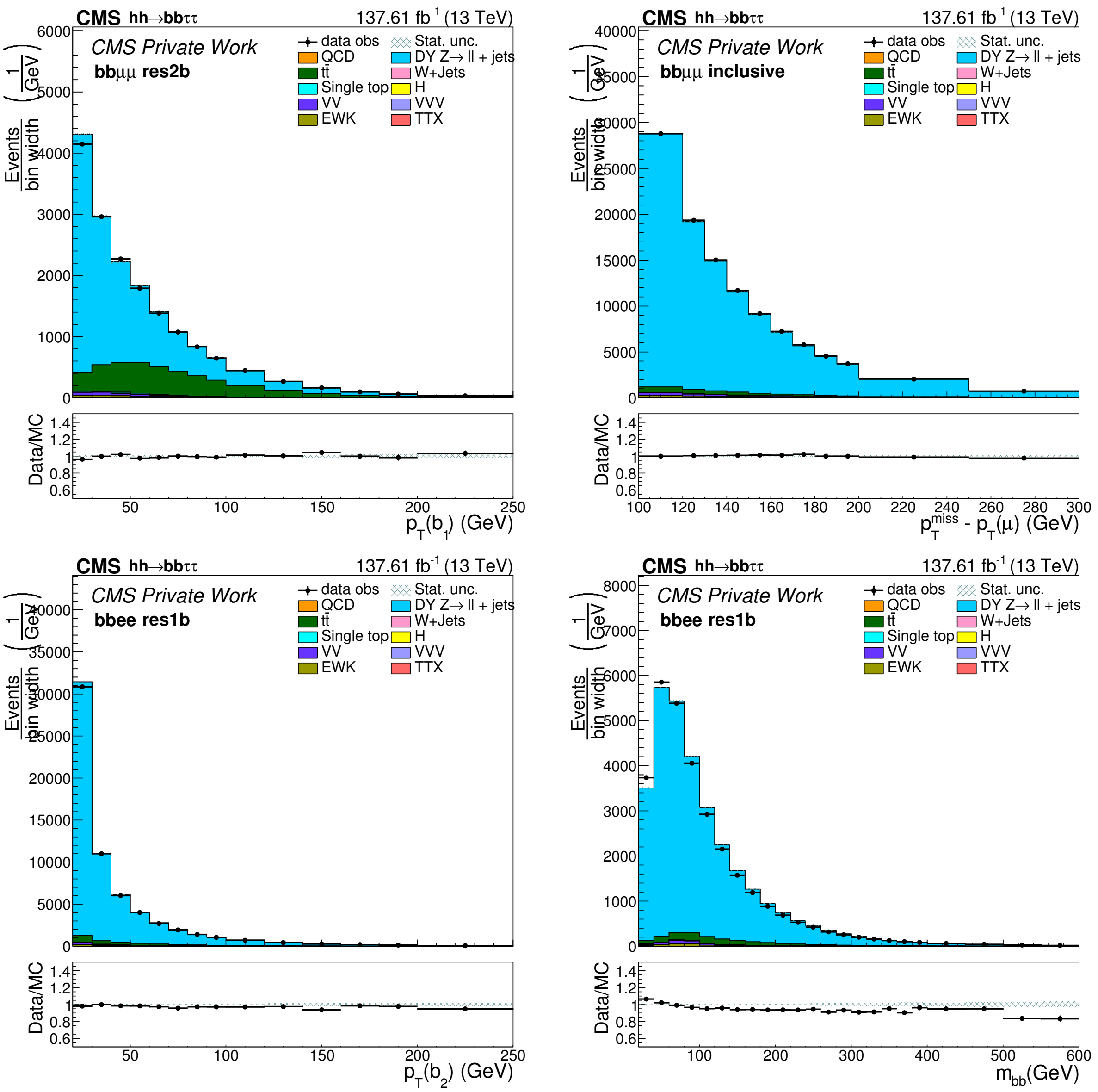


Figure 4.5: Jets observables and MET-noμ in the DYCR. Upper panel shows the first and first b-jet p_T and the MET-noμ (defined in Section 3.3.6) distributions in the $\mu\mu$ channel in the left and right respectively. In the lower panel the second b-jet and the mass of the resolved H $\to$ bb candidate are shown for the ee channel.

and its origin is not yet fully understood. Possible explanations include missing higher-order QCD and/or electroweak corrections, contributions from non-resonant production of $t\bar{t}$-like final states (e.g., $gg \to bb\mu\nu e\nu$), or potential unknown effects. The CMS recommendations for analyses involving $t\bar{t}$ backgrounds require to evaluate the impact of top-quark p_T mismodelling in a signal-depleted and $t\bar{t}$-enriched CR by assessing the data/MC agreement for key distributions. If the agreement falls within existing statistical and systematic uncertainties, no additional corrections or uncertainties are necessary. These motivate the definition of the t$\bar{\text{t}}$ Control Region (TTCR).

The requirements defining the TTCR are described in Tab. 4.10.

- The entire phase space in the $e\mu$ channel is analyzed, as it is already highly enriched in this process. This channel is not considered for the signal extraction and it not used in other control regions (Tab. 4.10), ensuring no overlaps with SR and DYCR. This allows the complete $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$range to be explored, enabling a clearer understanding of lepton behavior and $\mathrm{t\bar{t}}$ contributions in the mass range of $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$of the signal region, while remaining within orthogonal phase space.
- In the ee and $\mu\mu$ channel, not included in the signal search thus orthogonal to the SR, the TTCR is designed to exclude the 90% of the DY samples are excluded and to ensure the orthogonality also with the DYCR, $m_{\ell\ell} \notin [70, 100]$ GeV. In these channels, the $\mathrm{m}_{\tau\tau}$ range is sufficiently broad to explore portions of the $\mathrm{m}_{\tau\tau}$ region of interest for the SR, similarly to the $e\mu$ case.
- Finally, in $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels, the mass intervals ensure orthogonality with the SR while excluding by design the 95% of DY backgrounds and the 90% of the signals. These channels are used to validate the choice of the SR $\mathrm{m}_{\tau\tau}$ interval definition as the mismodelling of $\mathrm{t\bar{t}}$ contribution becomes relevant at high $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ values.

In Fig. 4.6, examples of the ee, $e\mu$, and $\mu\mu$ distribution in the TTCR (Tab. 4.10) are shown. The $\mathrm{m_{bb}}$ and $\mathrm{m}_{\tau\tau}$ spectra relevant for the SR, presented in the upper panel of Fig. 4.6, exhibit good agreement between data and MC simulations. Conversely, the lepton and b-jet p_T spectra show a slightly soft trend consistent with the known $\mathrm{t\bar{t}}$ p_T spectrum in MC issue. However, in this case, the deviations are relatively small ($\lesssim$ 10%), making additional reweighting or further corrective actions unnecessary to stabilize the effect.

When considering the signal channels, the TTCR remains orthogonal to the SR by inverting the cut on $\mathrm{m}_{\tau\tau}$. Furthermore, even though only the res2b category is used (Tab. 4.10), and thus only a fraction of the available statistics is considered for each channel, the region remains sufficiently enriched with $\mathrm{t\bar{t}}$ events to allow for an accurate modeling of the $\mathrm{t\bar{t}}$ contribution.
Considering the $e\tau_h$ and $\mu\tau_h$ channels, the b-jet and leptons distributions in Fig. 4.7 show an evident $\mathrm{t\bar{t}}$ mismodelling is evident. This was expected, being in the high $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ mass tails - though brings differences in data/MC are $\lesssim$ 10% also in these channels. However, these differences at high $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ values motivate the choice of applying the mass cuts to both maximise the signal acceptance and while avoiding larger mismodelling. Finally, it is interesting to observe the distributions of the leptons and b-jets observables in the $\tau_h\tau_h$ channel in this region, as they are enriched both in $\mathrm{t\bar{t}}$ in the res2b category and in QCD in the inclusive category. In particular, the QCD process uniformly contributes in all the $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ range, motivating the data/MC agreeement checks in this region. Therefore the events in TTCR and $\tau_h\tau_h$ channel can give insight on both the $\mathrm{t\bar{t}}$ background modelling and on the data-driven QCD estimation (discussed in the Section 5.4). In Fig. 4.8 two esemplificative distributions are shown for the res2b category in the top panels, whereas the lower panels provide two examples of inclusive category distributions of the first b-jet and lepton p_T. In the res2b category the difference between data/MC becomes more relevant at high $\mathrm{m}_{\tau\tau}$,

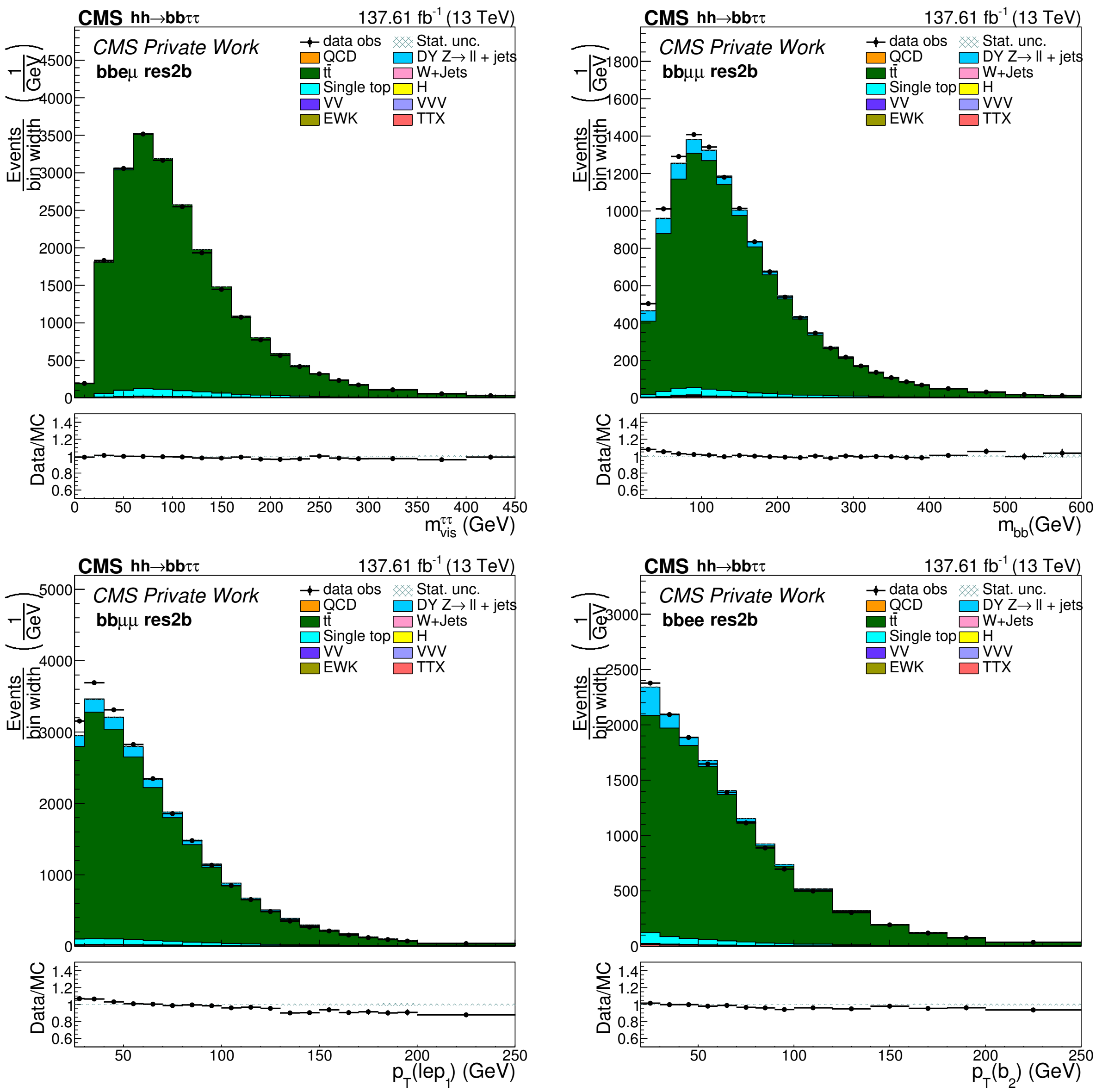


Figure 4.6: Distributions of different observables in the TTCR, in the res2b category. The invariant masses of the $H \to \tau^+\tau^-$ and $H \to bb$ candidates are shown in the upper panel for $e\mu$ and $\mu\mu$ channels respectively. In the lower panel the first lepton nd second b-jet are shown in the $\mu\mu$ and ee channels respectively.

however the agreement is still reasonable in the whole lepton p_T range.
In the QCD enriched inclusive category the data/MC agreement is good.
Other the distributions are shown in the Appendix F.

The Signal Region includes channels with one hadronic tau (τ_h) as second leg ($e\tau_h$, $\mu\tau_h$, or $\tau_h\tau_h$). It is defined by simultaneous conditions on $m_{\tau\tau}^{(SVFit)}$ and m_{bb}, optimized to maximize signal sensitivity, designed to cover at least the 70% of all signals and at the same time to avoide the inclusion of low-mass resonances and potential mismodeling in the high-mass tails, and to reducing background contamination. Two distinct mass intervals are defined based on the different categories: the *resolved* and *boosted* mass intervals.

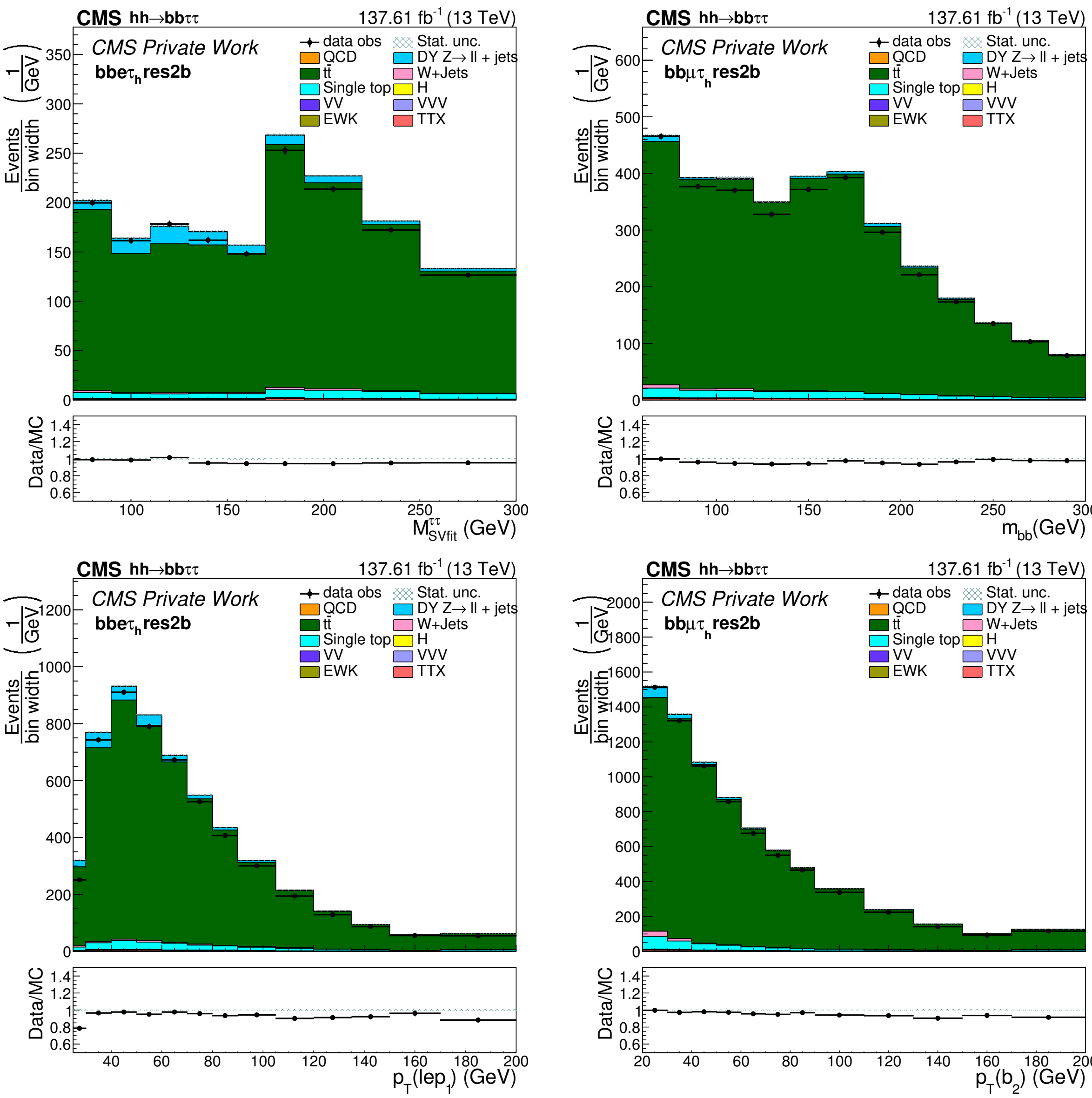


Figure 4.7: Distributions of different observables in the TTCR, in res2b category. The invariant masses of the $\tau\tau$ and $b\bar{b}$ candidates are shown in the upper panel for $e\tau_h$ (left) and $\mu\tau_h$ (right) channels respectively. In the lower panel the first lepton nd second b-jet are shown in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

The SR conditions on channels, categories, m_{bb} and $m_{\tau\tau}^{(SVFit)}$ are reported in Tab. 4.10. In Figures 4.15,4.16 and 4.17 some graphical examples of the squared cuts applied on $m_{\tau\tau}^{(SVFit)}$ and m_{bb} are shown, respectively for $e\tau_h$,$\mu\tau_h$and $\tau_h\tau_h$ channels. The right panels show some signal examples with different values of the resonance masses/spin, whereas the left panels show the same cuts on the main background contribution ($t\bar{t}$) showing the acceptance to the signals and at the same time the rejection power of these cuts for the most important background source.

The data/MC agreement check is performed also in the signal region, considering observables non sensitive to the signal. A complete set of the distributions is presented in Appendix G, and some examples are shown in Figures 4.9, 4.10 and 4.11 for the

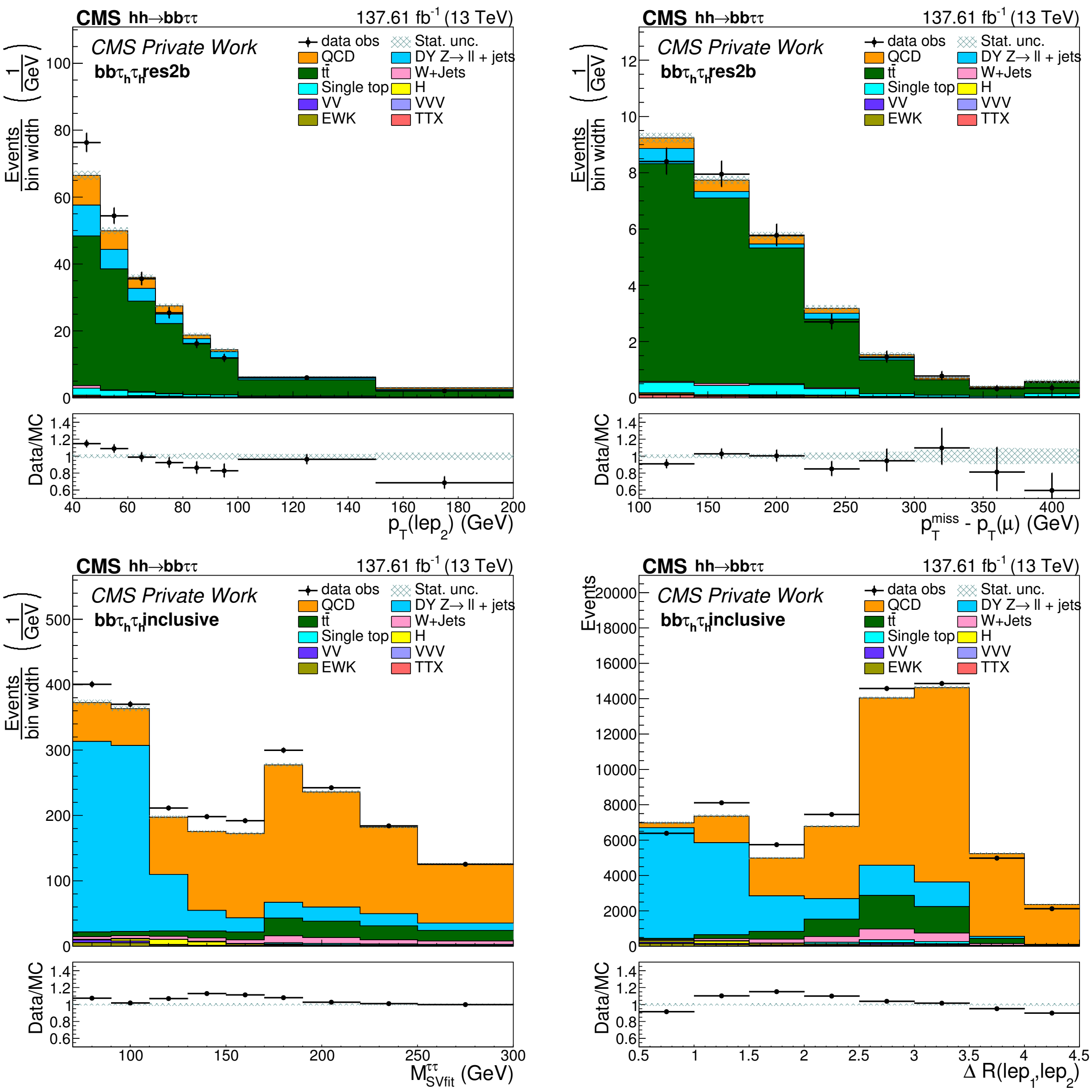


Figure 4.8: Distributions of different observables in the TTCR, in the $\tau_h\tau_h$ channel. The upper panel shows the second lepton p_T and $m_{\tau\tau}$ distributions in the res2b category, whereas in the lower panel the first b-jet and the $H \to bb$ invariant mass for resolved jets are shown in the inclusive category.

$e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels respectively. From these, it is possible to conclude that the overall data/MC agreement in signal channels is acceptable, since the fluctuations are within 5% far from the statistical error bands.

The $e\tau_h$ channel reveals a potential mismodeling at intermediate p_T values for electrons. This issue, previously observed in the TTCR - for instance, in the lower-left panel of Fig. 4.7 - generally remains within 5% (occasionally reaching up to 10%, as illustrated in the distributions provided in Appendix G). Despite this observation, there is no compelling evidence necessitating corrections to the electron energy scale or a specialized approach to address this mismodeling. Observables related to $H \to bb$ candidates, such as resolved and merged jets, and invariant mass distributions, display an overall good agreement between data and simulation. Similarly, distributions for the second lepton show consistent data/MC agreement.

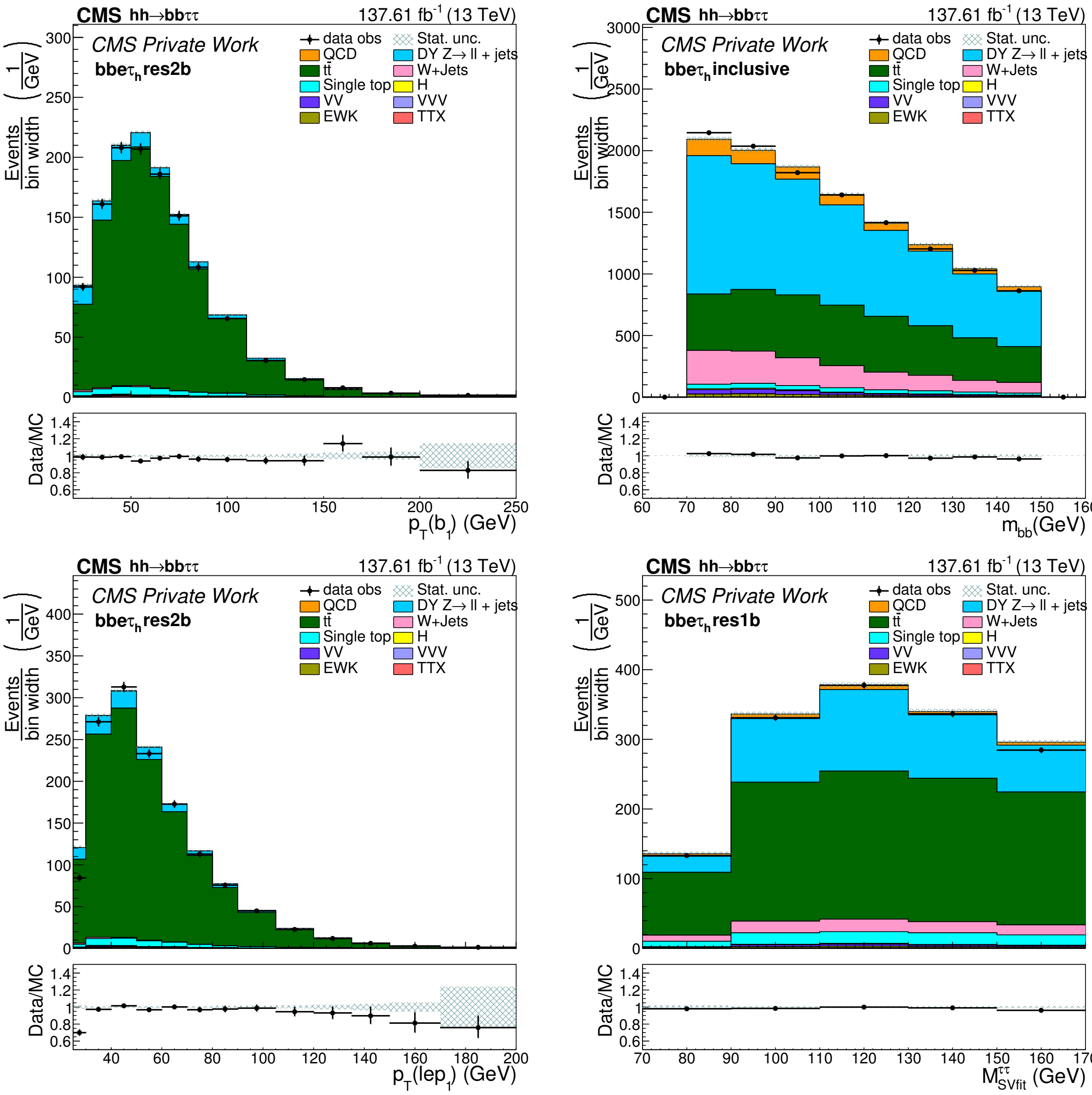


Figure 4.9: Distributions of different observables in the SR, in the $e\tau_h$ channel. In the upper panel the first b-jet and the H $\to$ bb invariant mass, whereas the lower panel shows the first lepton p_T and $m_{\tau\tau}^{(SVFit)}$ distributions in the signal sensitive categories.

In the $\mu\tau_h$ channel, within the relevant mass ranges and categories, the data/MC agreement is generally satisfactory. The MET-noμ (Section 3.3.6) distribution shows a slight trend consistent with the known $t\bar{t}$ p_T spectrum issues. However, these deviations remain within 5% (occasionally reaching 10%, as shown in Appendix G). Observables associated to resolved and boosted jets, and invariant mass distributions, show strong agreement between data and simulation, as do those for the second lepton.

The $\tau_h\tau_h$ channel, however, presents the most significant challenges due to its final state involving four hadronic jets. This channel is dominated by QCD background, making accurate modeling of this background as well as the choice of the DeepTauVSJet

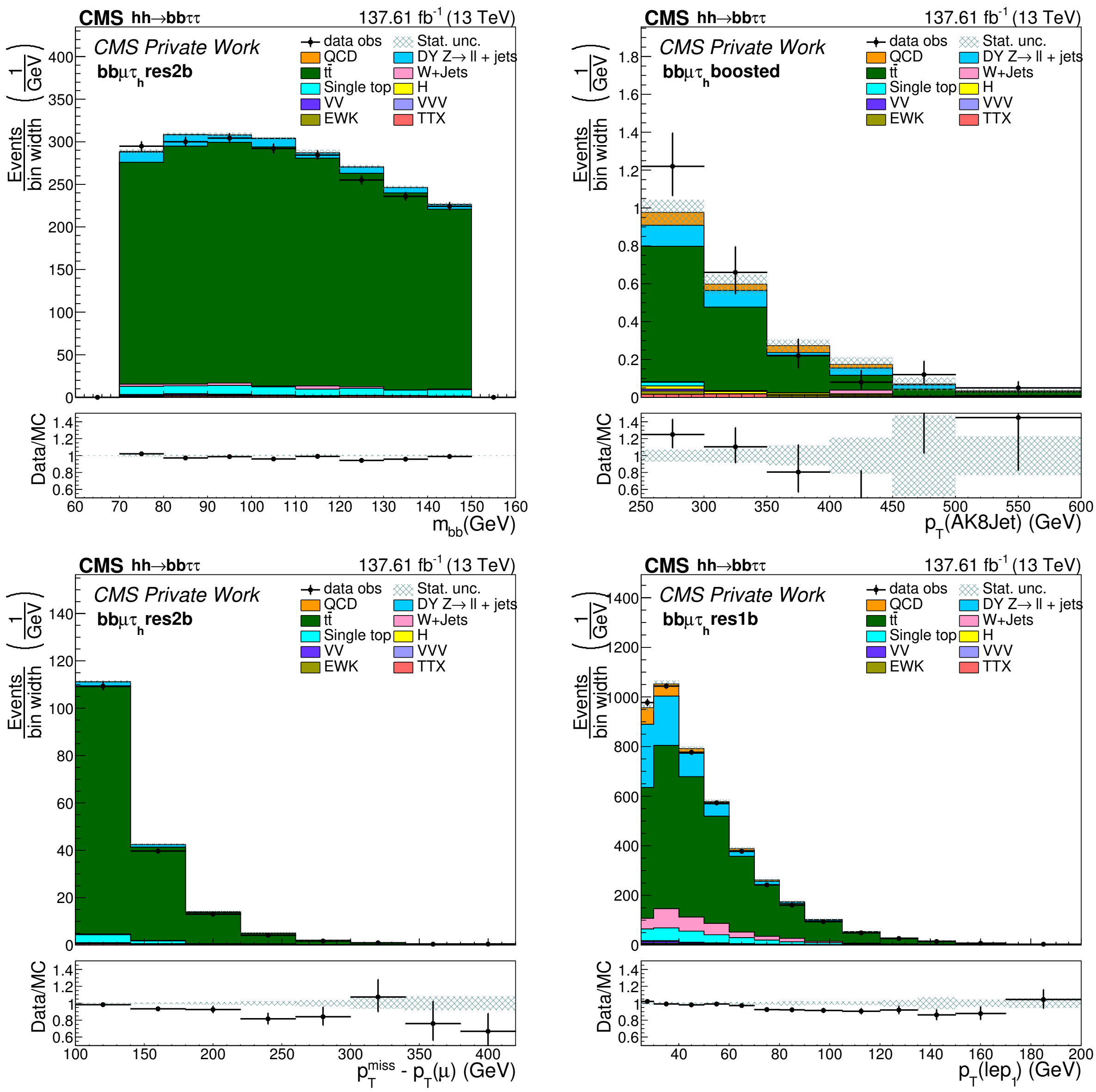


Figure 4.10: Distributions of different observables in the SR, in the $\mu\tau_h$ channel. The upper panel shows the resolved H $\rightarrow$ bb invariant mass and the boosted AK8 jet p_T , whereas the lower panel shows the MET-noμ (defined in Section 3.3.6) and the first lepton p_T distributions in the signal sensitive categories.

discriminator WP (see AppendixA.1.2 for WP definition) critical for this analysis. The shown distributions indicate discrepancies in data/MC agreement, particularly at high p_T, where the QCD contribution appears to be underestimated. To address this, dedicated corrections have been derived for the QCD contributions (Section 6.2.3) for the signal extraction.

4.8.4 The introduction of *DeepTauV2p5*

As discussed in Section 4.6, the *DeepTau* algorithm (Section 3.3.5) has been employed for tau selection, providing crucial discrimination against muons, electrons, and jets. In the event selection and H $\rightarrow \tau^+\tau^-$ candidate reconstruction, *DeepTauVSJet* score

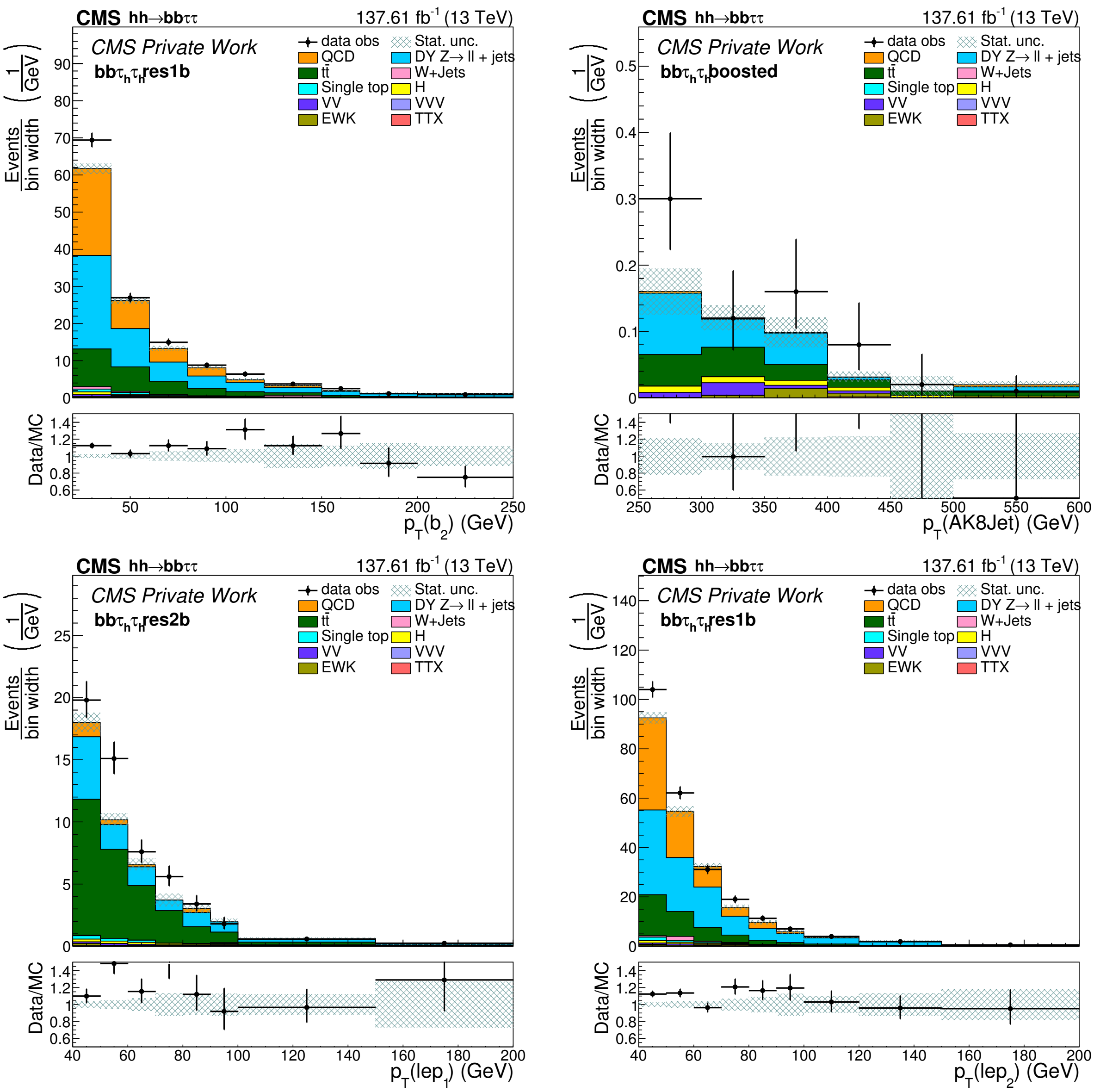


Figure 4.11: Distributions of different observables in the SR, in the $\tau_h\tau_h$ channel. The upper panel shows the second b-jet and the AK8 boosted jet transverse momenta, whereas in the lower panel the first lepton and second lepton p_T distributions are shown, in the signal sensitive categories.

has been specifically utilized to define tau lepton isolation, playing a central role in ensuring a high-purity selection of hadronic taus.

However, when performing data/MC agreement checks in both the signal and control regions for channels involving hadronically decaying taus, non-negligible discrepancies emerge between data and MC distributions, particularly in the $\tau_h\tau_h$channel. This channel presents the greatest challenge for event reconstruction due to the presence of multiple hadronic jets and p_T^{miss}, leading to substantial QCD multijet contamination. Accurate event reconstruction in this channel requires precise modeling of all MC contributions and a reliable QCD estimation strategy.

Several factors contribute to these discrepancies. The QCD background estimation method (Section 5.4) carries inherent uncertainties, particularly in high-purity tau

regions. Additionally, limitations in MC modeling of dominant background sources such as DY and $t\bar{t}$ introduce systematic mismatches. A crucial component influencing these differences is the performance of hadronic tau identification itself. Despite the advanced approach of *DeepTau*, residual discrepancies between data and simulations, coupled with the QCD background, indicate that refinements are necessary to achieve a more precise agreement in tau-related observables.
In fact, the most relevant data/MC+QCD mismatches are observed in τ_h-related distributions, as exemplified in the lower-left panel of Fig. 4.11. These deviations suggest a possible performance limitation of the *DeepTauVSJet* tagger, motivating the need for improvements. To address these challenges, the updated discriminator *DeepTau2018v2p5* has been introduced, offering enhanced tau identification capabilities. The same observables, now analyzed with *DeepTau2018v2p5*, are shown in Fig. 4.13, with a full set of distributions provided in Appendix H. One of the central goals of this thesis is to present, for the first time, a direct comparison between two versions of the *DeepTau* algorithm: *DeepTau2017v2p1*, the first deployed version, and the upgraded *DeepTau2018v2p5*. This preliminary study is of critical importance as it provides a first insight of the *DeepTau* developments using Run 2 data, before its full deployment for analyses which will use Run 3 data. This validation is fundamental for ensuring that *DeepTau2018v2p5* successfully enhances discrimination power, reduces systematic uncertainties, and improves data/MC agreement, ultimately strengthening future CMS analyses.
Given that the $\tau_h\tau_h$channel is the most affected by these changes, the comparison in this study focuses exclusively on this channel. However, a natural extension of this study would be to include all channels containing at least one hadronic tau, providing a more comprehensive assessment of the impact of the updated *DeepTau* algorithm.

DeepTau2018v2p5 introduces several key refinements, detailed in Appendix H, that significantly improve both performance and data/MC agreement. In particular, it features a more sophisticated treatment of high-purity τ_h candidates, effectively reducing discrepancies in critical kinematic regions. By benchmarking these improvements against *DeepTau2017v2p1*, this study provides essential insights into the evolution of tau identification algorithms and their pivotal role in advancing CMS physics analyses.

The following presents distributions of the same non-signal-sensitive observables previously shown for *DeepTau2017v2p1*, now obtained using *DeepTau2018v2p5* while applying the same event selection and reconstruction criteria. This direct comparison enables a rigorous assessment of the two *DeepTau* versions. The fully leptonic channels (ee, $e\mu$, and $\mu\mu$) are excluded from this analysis since the tau discriminator is not employed in their event selection and reconstruction.

The distributions are shown separately for the TTCR and SR, enabling a detailed evaluation of *DeepTau2018v2p5*'s impact in different event categories. The improved agreement observed in these comparisons underscores the importance of *DeepTau2018v2p5* in refining tau lepton selection and reducing systematic biases. These findings strongly suggest that integrating *DeepTau2018v2p5* into the analysis framework is a crucial step toward achieving a more precise and robust measurement of CMS processes involving hadronic tau decays.

Distributions in the TTCR

As illustrated in the following distributions (the full set is provided in Appendix H), the known $t\bar{t}$ issue is still evident in the TTCR, particularly in the subleading τ_h distribution. However, the data/MC agreement is significantly improved when employing *DeepTau2018v2p5* instead of *DeepTau2017v2p1*.

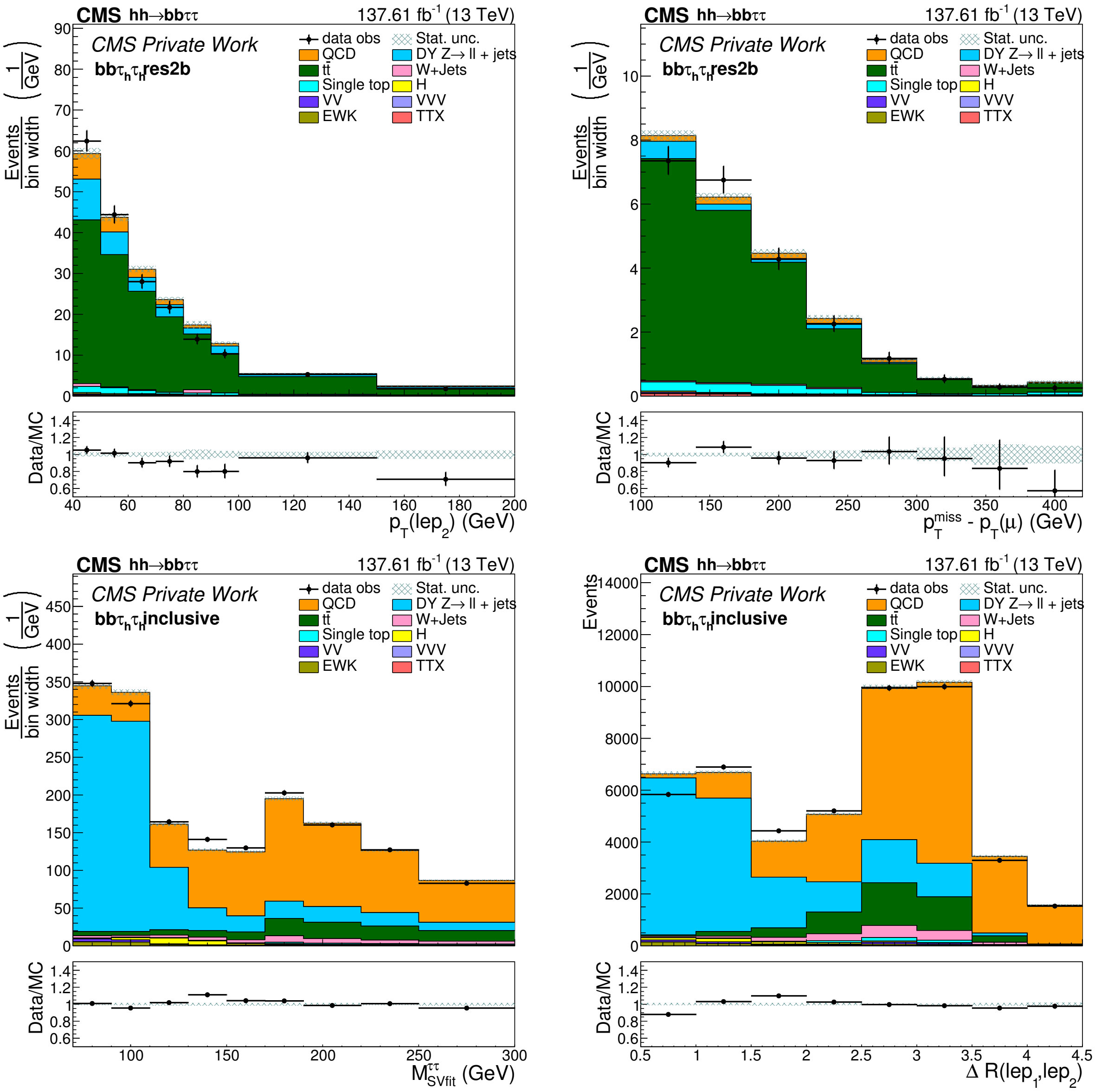


Figure 4.12: Distributions of different observables in the TTCR, in the $\tau_h\tau_h$ channel, with the new version of the *DeepTau* algorithm (*DeepTau2018v2p5*). The upper panel shows the second τ_h and the MET-noμ p_T in the $t\bar{t}$ enriched res2b category. The lowe panel shows the $m_{\tau\tau}^{(SVFit)}$ and the distance between the two τ_h leptons (ΔR) in the QCD enriched inclusive category.

This improvement becomes particularly evident when comparing Fig. 4.12 with Fig. 4.8 (corresponding to *DeepTau2018v2p5* and *DeepTau2017v2p1*, respectively). In the res2b category, the subleading τ_h p_T distribution shows a more accurate agreement, especially in the mid-low p_T range, still exhibiting the expected $t\bar{t}$ effect (p_T softer in data than in MC). Additionally, the MET-noμ distribution shows improved agreement across the full range, though some discrepancies persist at high values. Despite these improvements, the $t\bar{t}$ issue remains a compelling reason to exclude this region from the signal extraction phase space.

On the other hand, the inclusive category, enriched in QCD background, offers valuable insights into the QCD estimation (Section 5.4). The bottom distributions in Fig. 4.12, particularly $m_{\tau\tau}^{(SVFit)}$ and $\Delta R(\tau_{h,1}\tau_{h,2})$, further demonstrate the impact of using *DeepTau2018v2p5*. When compared with the corresponding distributions obtained with *DeepTau2017v2p1* in Fig. 4.8, the improvement in data/MC agreement, particularly when including QCD contributions, is evident, reinforcing the benefits of the updated *DeepTau* version.

Distributions in the SR Fig. 4.13 shows the same distributions of Fig. 4.11 obtained using *DeepTau2018v2p5* in the event selection and reconstruction. The presented distributions reveal persistent data/MC discrepancies, particularly at high p_T, where the QCD contribution appears to be underestimated. To mitigate this, dedicated corrections have been derived for the QCD contributions (Section 6.2.3) to improve the signal extraction.

A similar trend to the TTCR is observed in the SR, as shown by comparing Fig. 4.13 with Fig. 4.11 (corresponding to *DeepTau2018v2p5* and *DeepTau2017v2p1*, respectively). The following considerations can be done:

- In the res1b category, data/MC agreement is significantly improved across nearly all observables (see Appendices G and H for further comparisons). Only a few cases exhibit discrepancies similar to the previous selection.
- The boosted category, though limited in statistics, also shows signs of improved agreement.
- In the res2b category, discrepancies persist in most τ_h and b-jet observables, but their magnitude is reduced with *DeepTau2018v2p5*. This category is particularly impacted by QCD and $t\bar{t}$ backgrounds and applies b-jet discriminator cuts on both resolved jets, thus some remaining inefficiencies may arise from a combination of QCD estimation uncertainties, the $t\bar{t}$ issue, and the b-jet discriminator performance. These factors suggest that further refinements in these areas would be beneficial for future analyses.

In Fig. 4.14 other esemplificative observables are compared directly between the two approaches. Focusing on the key observables that define the SR for both resolved and boosted categories, the implementation of *DeepTau2018v2p5* consistently leads to better data/MC agreement. These results indicate that even if the selection and reconstruction techniques can still be refined, the adoption of *DeepTau2018v2p5* represents a valuable contribution to improve the tau leptons selection and discriminations against fake ones, potentially enhancing the sensitivity of the analysis.

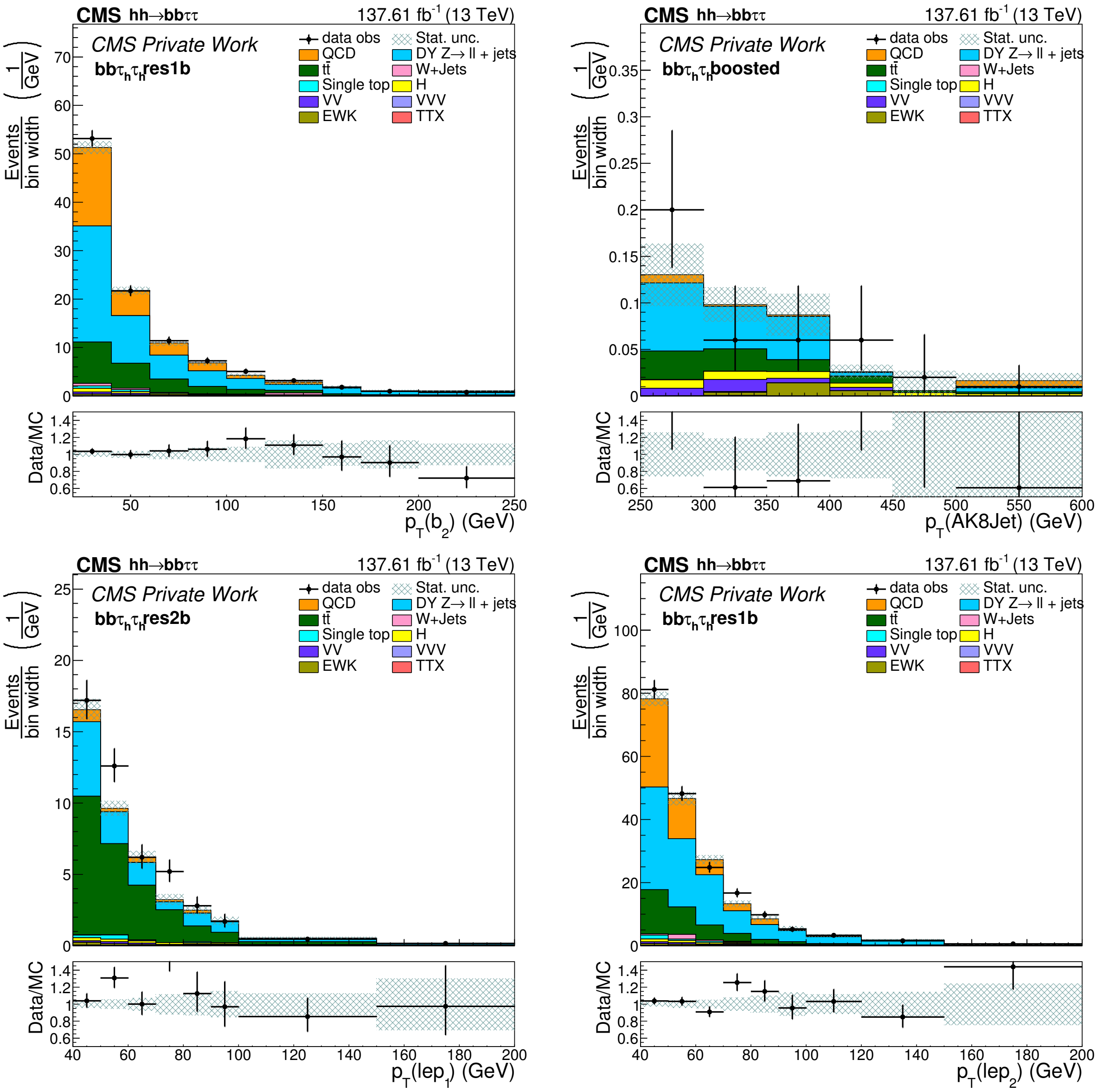


Figure 4.13: Distributions of different observables in the SR, in the $\tau_h\tau_h$ channel, with the new version of the *DeepTau* algorithm (*DeepTau2018v2p5*). The upper panel shows the second b-jet and the AK8 boosted jet transverse momenta, whereas in the lower panel the first lepton and second lepton p_T distributions are shown, in the signal sensitive categories.

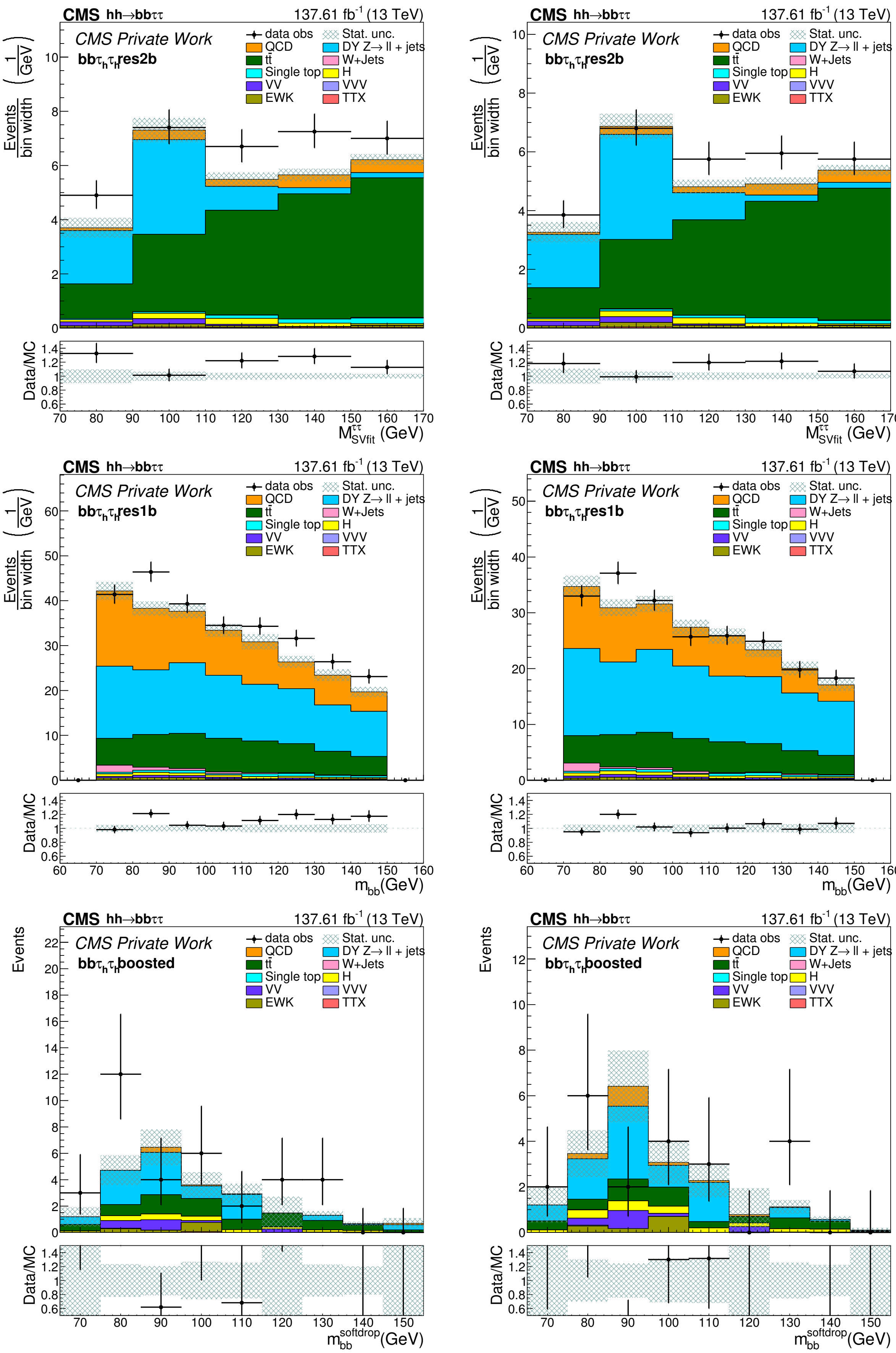
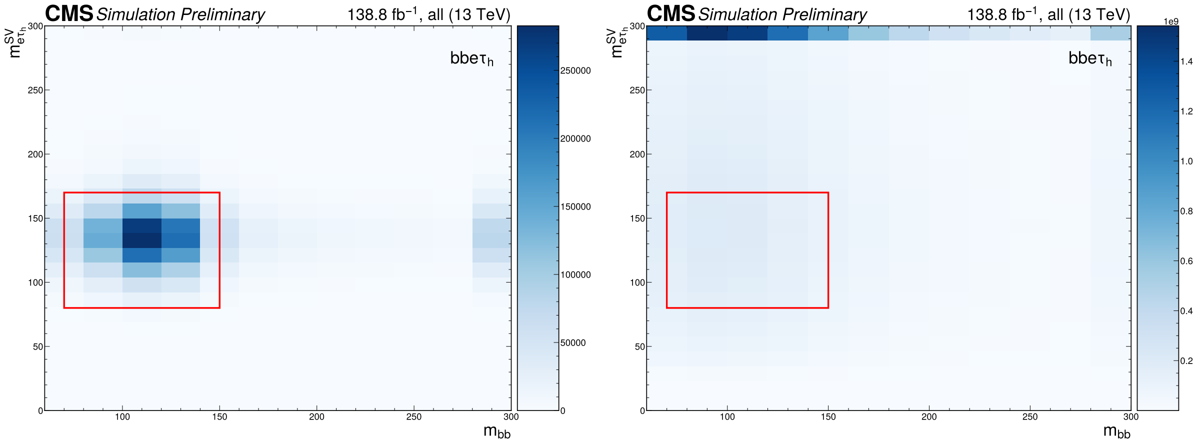


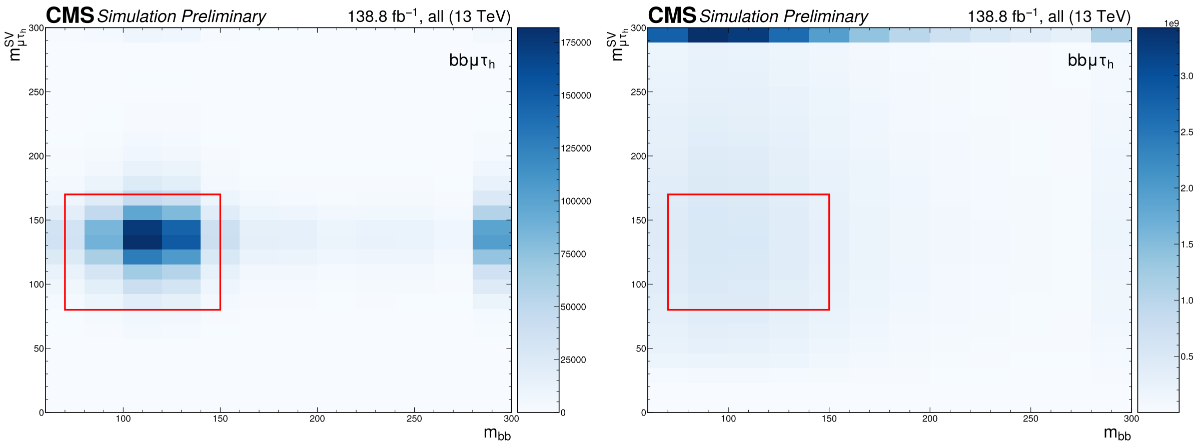
Figure 4.14: Distributions of different observables in the SR, in the $\tau_h\tau_h$ channel, with the new version of the *DeepTau* algorithm (*DeepTau2018v2p5*). The upper panel shows the second b-jet and the AK8 boosted jet transverse momenta, whereas in the lower panel the first lepton and second lepton p_T distributions are shown, in the signal sensitive categories.



(a) $m_X = 700\,\text{GeV}$, spin=0 (radion), $e\tau_h$.

(b) $t\bar{t}$ sample, $e\tau_h$.

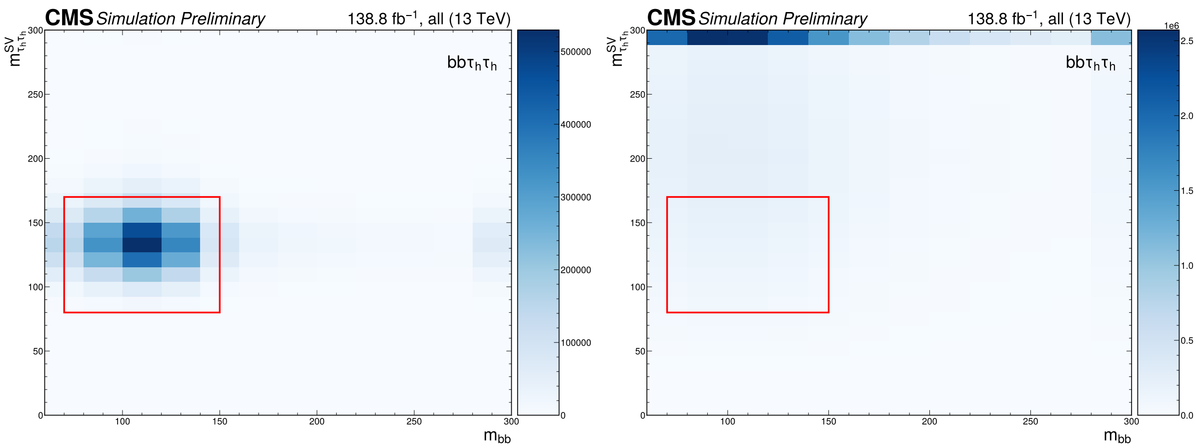
Figure 4.15: Examples of $m_{\tau\tau}^{(\text{SVFit})}$ vs m_{bb} distributions showing the mass intervals defined for signal region definition, for the $e\tau_h$ channel.



(a) $m_X = 1000\,\text{GeV}$, spin=2 (graviton), $\mu\tau_h$.

(b) $t\bar{t}$ sample, $\mu\tau_h$.

Figure 4.16: Examples of $m_{\tau\tau}^{(\text{SVFit})}$ vs m_{bb} distributions showing the mass intervals defined for signal region definition, for the $\mu\tau_h$ channel



(a) $m_X = 500\,\text{GeV}$, spin=0 (radion), $\tau_h\tau_h$.

(b) $t\bar{t}$ sample, $\tau_h\tau_h$.

Figure 4.17: Examples of $m_{\tau\tau}^{(\text{SVFit})}$ vs m_{bb} distributions showing the mass intervals defined for signal region definition, for the $\tau_h\tau_h$ channel

Chapter 5

Data composition modelling

Accurate modeling of the phenomenology of physical processes and their interaction with the CMS detector is fundamental to defining and optimizing the analysis strategy. Signal and most part of background processes rely on simulation using Monte Carlo (MC) techniques with state-of-the-art generators and precise theoretical calculations, with the exception of the Quantum Chromodynamics (QCD) multi-jets production, estimated via a data-driven method. The data composition modelling plays a critical role in optimizing the analysis strategy, comparing observed data to predictions, and searching for the presence of a signal. This chapter describes the modeling the physics processes which contribute to the amount of observed events in data, as well as the simulation tools and techniques used for this purpose. It outlines both simulation-based and data-driven techniques (e.g. in Section 5.4) employed to describe the data composition.
An overview of the simulation workflow and the main MC generators is presented in Section 5.1. Sections 5.2.1 - 5.4 describe the signal and background processes, emphasizing techniques like data-driven estimation of QCD processes and combining multiple simulation sets.

5.1. Event simulation and data formats in CMS

To prepare data and simulation samples for physics analyses at CMS, several processing steps are performed. Fig. 5.1 provides a schematic overview the processing chain for both recorded data and simulated samples.

The Generation (GEN) step uses one of the available event generators to simulate a specific physical process originating from beam collisions.

The simulated events considered in this analysis have been generated using two approaches: a **General-Purpose Generator**, which simulates both the event and the hadronization process, or a **Matrix Element (ME) Generator**, which generates events at the parton level paired with a general-purpose generator to handle the hadronization step. In the following, the generators used to produce the MC samples for this analysis are described:

- PYTHIA [167] is a general purpose event generator widely used in high-energy physics to simulate collisions involving electrons, protons, photons, and heavy ions at Leading Order (LO) accuracy. It incorporates theoretical models and algorithms for various physics processes, including hard and soft interactions, parton distribution functions, initial and final state parton showers, multiparton interactions, fragmentation, and particle decays.
- MAGRAPH5_AMC@NLO [168] is a ME generator designed for both Standard Model (SM) and Beyond Standard Model (BSM) studies. It provides tools for cross-section calculations, the generation of hard events and their matching with event generators. It combines the MAGRAPH5 and AMC@NLO generators, along with a variety of tools essential for event manipulation and analysis enabling the simulations at LO accuracy for any user-defined Lagrangian, and at Next to Leading Order (NLO) accuracy for QCD corrections in SM processes. Additionally, it can compute matrix elements at both the tree and one-loop levels.
- POWHEG [169–171], a ME generator aiming to perform accurate particle collisions simulation, combining NLO perturbative QCD calculations with parton shower algorithms to model both hard scattering processes and softer emissions.

Both POWHEG and MAGRAPH5_AMC@NLO are used in conjunction with general purpose generator PYTHIAto simulate both the hard interaction and subsequent parton showers and hadronization processes. The interaction of particles produced by collision with the Compact Muon Solenoid (CMS) detector is modeled using the GEANT4 toolkit [172,173].

The Simulation (SIM) step consists in simulating the response of the detector, electronics and trigger system, including also the pile-up. It provides an output similar to the actual data collected by the detector.

The Reconstruction (RECO) step is applied to both simulated and real data, reconstructs the events and serves as the final stage of sample preprocessing.

The format used for analysis, for both MC simulations and real data, is the NanoAOD [174]. These are ROOT [175] files containing a compact subset of GEN+SIM+RECO information, tailored for physics analyses. Further details on the CMS data format are reported in Section 2.4.3.

5.2. Modelling

The processes contributing to the data, which are modeled using MC simulations, are presented below. For each mechanism, the most relevant contributions in the data composition are highlighted. Along with the simulation details, the cross section for each process is also reported.

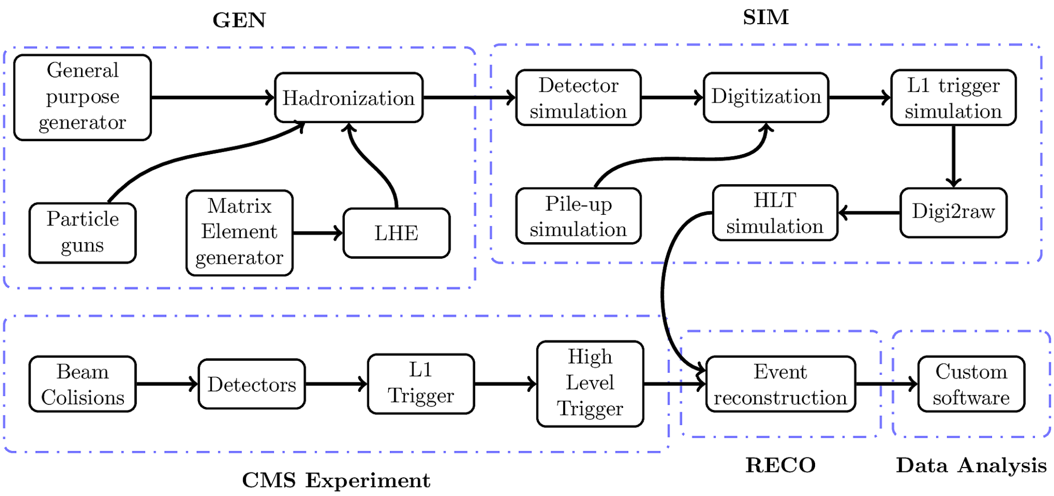


Figure 5.1: Schematic view of the processing chain in CMS for data and simulated events. The arrows describe the direction of the flow of information.

5.2.1 Signal

The resonant $\mathrm{X} \to \mathrm{HH} \to \mathrm{b\bar{b}}\tau\tau$ signal samples have been modeled with MAGRAPH5_AMC@NLO as ME generator and PYTHIA for showering and hadronization. The events are simulated at LO precision. Samples are generated under the assumption of a narrow width resonance, with a negligible width compared to the experimental detector resolution. Two set of signal simulation are considered, depending on the resonance spin hypotheses: spin-0, radion-like, and spin-2, bulk graviton-like particles. For each hypothesis, 25 mass points spanning from 250 GeV to 3 TeV are considered, for a total of 50 samples sets. The cross section considered for this process is normalized to 1 pb.

5.2.2 $\mathrm{t\bar{t}}$

One of the main background sources originates from $\mathrm{t\bar{t}} \to W^+bW^-\bar{b}$ process. This background relies on MC simulation produced with the ME generator POWHEG and PYTHIA at NLO precision.

The $\mathrm{t\bar{t}}$ simulations are divided in three groups to cover all the W bosons decays: the so-called fully leptonic decays, where both W bosons decay in a lepton plus neutrino ($W \to \ell\nu_\ell$, with $\ell = e, \mu, \tau$), the most relevant background for this analysis from $\mathrm{t\bar{t}}$ processes; the semi-leptonic decays, where one W decays in leptonplus neutrino and the other decays in a pair of quarks ($W \to qq'$); finally the fully hadronic decays, where both W decay in quarks pairs, the least important process among the $\mathrm{t\bar{t}}$ contribution. The last two processes are suppressed thanks to the high capability of *DeepTauVSJet* discriminator (Section 3.3.5) for the hadronic tau ID. The cross section considered for the inclusive process is $\sigma(\mathrm{t\bar{t}})$= 833.9 pb [176]. The Branching Ratio (BR) for the W boson hadronic and leptonic decays are respectively $\mathcal{BR}(W \to qq') = 0.6741$ and $\mathcal{BR}(W \to \ell\nu_\ell) = 0.3259$, with values taken from [177] assuming $\mathcal{BR}(t \to Wb) = 1$. In order to check the data/MC agreement, a dedicated control region has been defined

(Section 4.8.3).

5.2.3 Drell-Yan: Z+Jets (DY)

Another dominant source of background comes from the Z+Jets Drell Yan (DY) events. This background contribution relies on MC simulation, modeled via the MAGRAPH5_AMC@NLO as ME generator and PYTHIA as general purpose one, at NLO precision. Three sets of MC simulation have been considered to account for this background: the inclusive sample, samples binned in the transverse momentum of the Z boson, transverse momentum component (p_T) (Z), and in the number of jets (N_{jets}). The reason of this choice is to have a more precise modelling of this background and to have higher statistics to reduce statistical uncertainties in the tails of distributions. To properly combine the different MC sets, the simplified stitching technique described in Section 5.3.1 is adopted. The cross sections for each MC set are shown in Tab. 5.1. As for the $t\bar{t}$ processes, also for the DY background a dedicated control region has been defined to check the the agreement between data and the MC combined samples.

5.2.4 $W^{\pm}$+Jets (W)

The associated production of a leptonically decaying W boson ($W \rightarrow \ell\nu_\ell$ with $\ell = e, \mu, \tau$) together with hadronic jets is simulated via MAGRAPH5 using the MLM technique for jet merging, at LO precision [178, 179].
In this analysis the contribution of W + jets background is highly suppressed since, at the most, only one real τ comes from W decay and in this case the other three jets in the event must be misidentified as a τ plus two b-jets, but this happens with a low probability since the object identification criteria for both leptons and jets are highly performing. Three sets of samples are considered to include this process: the *W+Jets* inclusive production, the *W+N jets* (N=1,2,3,4) binned in the number of jets together with the W boson, and the samples binned in hadronic transverse energy (HT). As for the DY background, the different MC sets are combined via the simplified stitching technique described in Section 5.3.1. The cross sections for each MC set are shown in Tab. 5.1.

5.2.5 Electroweak $W^{\pm}$/Z + 2 Jets (EWK)

Electroweak production of W or Z bosons associated with two jets relies on the MAGRAPH5 and PYTHIA generators combination. The LO cross sections are obtained from the MC generator and their values used for the normalization in this analyses are: $\sigma(W^+ \rightarrow \ell^-\bar{\nu}_\ell$ + 2 jets) = 39.05 pb, $\sigma(W^+ \rightarrow \ell^+\nu_\ell$ + 2 jets) = 32.05 pb, $\sigma(Z \rightarrow \ell^\pm\ell^\mp$ + 2 jets) = 10.65 pb and $\sigma(Z \rightarrow \ell\nu\bar{\nu}$ + 2 jets) = 6.215 pb.

5.2.6 Single Higgs (H)

The background sources involving single Higgs considered for this analysis are: gluon gluon and vector boson fusion single Higgs production with in both cases the Higgs boson decays are splitted in $H(ZZ^*)$, $H(WW^*)$ and $H(\tau\tau)$. The associated production of single Higgs with a vector boson (VH) or a pair of top quarks (ttH) can emulate the final state of interest, therefore also these samples are considered. The ttH production is splitted in two subsamples having as final states either a pair of tau leptons or a pair of b quarks. The VH processes are splitted in $W^{\pm}H$ and ZH with the Higgs boson decaying in two tau leptons or b quarks as in the ttH case, and also considering the $W^{\pm}W^{\mp}$ decays. These samples are generated considering $m_H = 125\,\text{GeV}$. In all the cases, the MC samples are produced at NLO via POWHEG and PYTHIA . The cross section adopted to normalize these samples are reported in Tab. 5.2.

5.2.7 Single Top (ST)

In this analysis, the background contribution coming from single top quark events is mainly due to the the *t-channel* single top production and to associated production of a single top and a W boson. The MC samples are produced at NLO via POWHEG and PYTHIA and are normalized to the Next-to-Next Leading Order (NNLO) theoretical cross sections: $\sigma(t, t-channel) = 136.02\,\text{pb}$, $\sigma(\bar{t}, t-channel) = 80.95\,\text{pb}$, $\sigma(t+\bar{t}, tW) = 71.7\,\text{pb}$. They are obtained from [180].

5.2.8 Double Vector bosons (VV)

To account for vector bosons pair production, WW, WZ and ZZ samples are considered. In all the cases, the inclusive decays are taken in account. The samples are modeled via PYTHIA. The cross section considered for those processes are: $\sigma(W^{\pm}W^{\mp}) = 118.7\,\text{pb}$ at NNLOQCD [181], $\sigma(W^{\pm}Z) = 46.74\,\text{pb}$ and $\sigma(ZZ) = 15.99\,\text{pb}$ at NLO [182].

5.2.9 Triple Vector bosons (VVV)

Triple vector bosons production is included considering inclusive WWW, WWZ, WZZ and ZZZ samples. As for the VV samples, also in this case the modelling is performed via AMC@NLO and PYTHIA. The cross section considered for those processes are: $\sigma(ZZZ) = 0.01476\,\text{pb}$, $\sigma(WZZ) = 0.05709\,\text{pb}$, $\sigma(WWZ) = 0.1707\,\text{pb}$ and $\sigma(WWW) = 0.2158\,\text{pb}$. These are all taken from the CMS MC generator tool, the *CMS Cross Section DataBase (XSDB).*

5.2.10 $t\bar{t}$ + X (TTX)

The associated production of top quark pairs togheter with other particles can emulate the signal final states, e.g. if the two top quark decay hadronically in two b quarks and the other particles have tau leptons in the final state. However, they represent a very small contribution due to their small cross sections. The sources of background included are: the top quark pair production associated to:

- double vector boson ($t\bar{t}$ VV);
- Higgs plus vector boson ($t\bar{t}$ VH);
- double Higgs production ($t\bar{t}$ HH);
- two gluons ($t\bar{t}$ gg);
- one gluon and jets ($t\bar{t}$ g + jets);
- a single top and jets ($t\bar{t}$ t + jets);
- associated tW production ($t\bar{t}$ tW);
- another top quark pair ($t\bar{t}$ $t\bar{t}$).

All the processes taken into account are simulated with MAGRAPH5_AMC@NLO and PYTHIA. The cross sections of these samples, as well as their accuracy and the reference used for the analsis are reported in Tab. 5.3.

5.3. MC simulated events

To correctly account for data composition contributions from processes relying on MC simulations, for each proces (proc) the following weight is applied to the j-th generated event:

$$w_j^{MC,proc} = \frac{L \cdot s^{proc} \cdot w_j^{gen,proc}}{\sum_{k=0}^{N^{gen,proc}} (w_k^{pu} \cdot w_k^{gen,proc})} \tag{5.1}$$

Where:

- L is the integrated luminosity.
- s^{proc} is the cross section of the process. When a simulated process targets a specific final state, the cross section is multiplied by the final state branching fraction. Moreover, when processes are taken into account combining multiple MC (Section 5.3.1), the final expression of s^{proc} is reported in Eq. 5.2.
- $N^{gen,proc}$ is the number of generated events of the simulated process, before applying any selection.
- $w^{gen,proc}$ is the weight assigned to the event by the MC generator.
- The denominator contains the sum over the generated events of the products of w^{pu} and $w^{gen,proc}$ weights applied to the original sample without any preselection, where w^{pu}, ensures that the pile-up (PU) distributions in simulation matches the one in data[1]

5.3.1 The simplified stitching technique

When accounting for the DY ad $W^{\pm}$+Jets backgrounds, different MC production sets have been used. When considering different MC production schemes, the overlap of

[1]These weights are calculated comparing the PU profile in data and simulation. These are centrally provided by the CMS Collaboration.

Process	σ (pb)	Precision	Reference
W+Jets (inclusive)	61526.7	NNLO	[183]
W+Jets (inclusive)	66680	NLO	XSDB
W+Jets (inclusive)	53870	LO	XSDB
W+ 1 Jets	8927	LO	XSDB
W+ 2 Jets	2809	LO	XSDB
W+ 3 Jets	826.3	LO	XSDB
W+ 4 Jets	544.3	LO	XSDB
W+Jets HT 70-100 GeV	1264	LO	XSDB
W+Jets HT 100-200 GeV	1256	LO	XSDB
W+Jets HT 200-400 GeV	335.5	LO	XSDB
W+Jets HT 400-600 GeV	45.25	LO	XSDB
W+Jets HT 600-800 GeV	10.97	LO	XSDB
W+Jets HT 800-1200 GeV	4.933	LO	XSDB
W+Jets HT 1200-2500 GeV	1.16	LO	XSDB
W+Jets HT > 2500	0.008	LO	XSDB
Z + Jets (inclusive)	6077.22	NNLO	[183]
Z + Jets (inclusive)	6424	NLO	XSDB
Z + 0 Jets	5090	NLO	XSDB
Z + 1 Jets	983.5	NLO	XSDB
Z + 2 Jets	353.6	NLO	XSDB
Z + Jets p_T (Z) 0-50 GeV	1485	NLO	XSDB
Z + Jets p_T (Z) 50-100 GeV	397.4	NLO	XSDB
Z + Jets p_T (Z) 100-250 GeV	97.2	NLO	XSDB
Z + Jets p_T (Z) 250-400 GeV	3.7	NLO	XSDB
Z + Jets p_T (Z) 400-650 GeV	0.51	NLO	XSDB
Z + Jets p_T (Z) > 650 GeV	0.0047	NLO	XSDB

Table 5.1: Cross sections used for Z/W + Jets backgrounds. The XSDB is CMS MC generator tool.

Process	σ (pb)	Precision	Reference
gg H $\rightarrow \tau^+\tau^-$	3.046	NNLO	[184]
VBF H $\rightarrow \tau^+\tau^-$	0.237	NNLO	[184]
ggH, H(ZZ*), $2\ell 2$q	0.193	NNLO	[184]
ggH, H(WW*), $2\ell 2\nu$	0.102	NNLO	[184]
VBF, H(WW*), $2\ell 2\nu$	0.086	NNLO	[184]
W^+ H $\rightarrow \tau^+\tau^-$	0.053	NNLO	[184]
W^- H $\rightarrow \tau^+\tau^-$	0.033	NNLO	[184]
Z H $\rightarrow \tau^+\tau^-$	0.055	NNLO	[184]
ZH, H $\rightarrow$ bb , Z($\ell\ell$)	0.056	NNLO	[184]
ZH, H $\rightarrow$ bb , Z(qq)	0.356	NNLO	[184]
W^+H+Jets, H(WW)	0.872	NLO	XSDB
W^-H+Jets, H(WW)	0.5445	NLO	XSDB
ZH+Jets, H(WW)	0.9595	NLO	XSDB
ggZH, H(WW), Z($\ell\ell$)	0.003	NNLO	[184]

Table 5.2: Cross sections used for Higgs processes. The XSDB is CMS MC generator tool.

Process	σ (pb)	Precision	Reference
$t\bar{t}$ WW	0.092	LO	XSDB
$t\bar{t}$ WZ	0.004	LO	XSDB
$t\bar{t}$ ZZ	0.002	LO	XSDB
$t\bar{t}$ WH	0.002	LO	XSDB
$t\bar{t}$ ZH	0.001	LO	XSDB
$t\bar{t}$ HH	0.001	LO	XSDB
$t\bar{t}$ + t + Jets	0.0004	LO	XSDB
$t\bar{t}$ + tW	0.0007	LO	XSDB
$t\bar{t}$ $t\bar{t}$	0.008	LO	XSDB
$t\bar{t}$ + g + Jets	3.757	LO	XSDB
$t\bar{t}$ gg	0.017	LO	XSDB

Table 5.3: Cross sections used for $t\bar{t}$ +X backgrounds. The XSDB is CMS MC generator tool.

the phase space has to be taken into account properly weighting the cross section of each separated contribution, and this is done with the stitching method [185]. This analysis uses a "simplified" version of this method, defining the stitched cross section for each set of MC samples as follows:

$$s_i(proc) = \sigma_i^{NLO/LO}(proc) \cdot \frac{1}{K} \cdot \frac{\sigma_{incl}^{NNLO}}{\sigma_{incl}^{NLO}} \tag{5.2}$$

Where:

- $\sigma_i(proc)$ is the cross section of the specific process (e.g. DY+1Jet).
- K is the number of simulation sets, e.g. in the DY it is either 2 or 3 as shown in the next paragraph.
- all samples are normalized by applying the ratio of inclusive cross sections at NNLO and NLO (e.g. the inclusive DY cross section) $\sigma_{incl}^{NNLO}/\sigma_{incl}^{NLO}$. The NNLO cross sections are derived from theoretical predictions, while the NLO/LO are taken from the CMS cross section database.

DY stitching

The DY contribution is modelled combining the inclusive sample, the samples binned in the transverse momentum of the Z boson, p_T (Z), and the samples binned in the number of jets (N_{jets}). The inclusive cross section for the DY is available at both NNLO and NLO precision, and the binned ones are available at NLO precision. The modified stitched cross section for DY MC sets is:

$$s_i = \sigma_i^{NLO} \cdot \frac{1}{K} \cdot \frac{\sigma_{incl}^{NNLO}}{\sigma_{incl}^{NLO}}; \tag{5.3}$$

Where

$$K = \begin{cases} 2 & \text{if } p_T^{gen}(Z) = 0\,\text{GeV} \\ 3 & \text{otherwise} \end{cases} \tag{5.4}$$

The value of K shold be 3, but the lowest p_T (Z) binned sliced samples were found out to have a bug in the production, requiring $0 < p_T(Z) \leq 50\,\text{GeV}$, hence the events with Z boson produced at rest are excluded.

W+Jets stitching

The W+Jets contribution modelling relies on the combination of the inclusive sample, the samples binned in the transverse hadronic energy, HT, and the samples binned in the number of jets (N_{jets}). The inclusive cross section for the W+Jets is available at both NNLO and NLO precision, while the cross sections of the binned samples are available only at LO precision. When combining the W+Jets MC sets, the stitching cross section is defined as:

$$s_i = \sigma_i^{LO} \cdot \frac{1}{K} \cdot \frac{\sigma_{incl}^{NNLO}}{\sigma_{incl}^{NLO}}; \tag{5.5}$$

Where

$$K = \begin{cases} 1 & \text{if } N_{jets} = 0 \\ 2 & \text{if } HT < 70\,\text{GeV} \\ 3 & \text{otherwise} \end{cases} \tag{5.6}$$

As for the DY case, also for W+Jets combination, K can have different values based on the phase space. The reason is that the binned sample with 0 jets was not available, and that the samples binned in HT are provided with the lowest bin of HT = 70 GeV.

5.4. The QCD background estimation

Another source of background in this analysis arises when the non-prompt, *fake* leptons are selected, coming from the QCD multi-jet events; although this is a reducible background, its cross section is several order of magnitude greater than the signal one ($\sim$ µb/mb) therefore it still remains one of the most relevant background contributions. The most affected channels and categories are the ones with two hadronic taus ($\tau_h\tau_h$) and resolved jets (Resolved 1b category (res1b), Resolved 2b category (res2b)).

This background does not rely on the simulations as it is difficult to model due to the complex and poorly understood nature of QCD processes, moreover MC simulations are impractical due to the large sample size required. Non-perturbative effects such as hadronization and parton showering are difficult to simulate precisely, and generating sufficient statistics for the tails of QCD multijet production is computationally prohibitive. The high multiplicity of particles in QCD events, dominated by gluon exchanges, creates a vast and intricate phase space that is computationally expensive to model accurately. Moreover, QCD background simulations are limited by uncertainties in theoretical models, MC generators, and detector effects like resolution, calibration, and PU, making them unable to fully replicate experimental data.

To address this limitations and properly estimate this source of background, a data-driven technique has been adopted: the *ABCD method*. It consists in the measurement of the transfer factor in dedicated control regions after removing all well simulated processes and creating shape templates on which match the mismodelled background, assuming similar behavior for QCD processes in signal and control regions. Therefore, the phase space is splitted in four different regions (signal region + 3 sideband regions), defined by requiring the failure/passing of two selection criteria, which for this analysis are the lepton isolations and charges. When applying isolation cuts on leptons during the event selection, described in Chapter 4, leptons will be classified as Isolated (Iso) and Not Isolated (Non-Iso). These definitions are reported in Tab. 5.4. On the other hand, when separating the phase space according to lepton charges, if the leptons have the same charges, the region will be labeled with Same Sign (SS), and conversely when the leptons will have different charge signs, the label will be Opposite Sign (OS).

- The **region A** is the Signal Region (SR), where both leptons are isolated and have opposite charge signs, therefore this can be also defined as **OS-Iso**.
- The **region B** is populated by events with both leptons isolated, but with the same charge sign, **SS-Iso**.
- **Regions C and D** contain events with the non-isolated second lepton and respectively opposite and same charge signs between the lepton pair, so they will be referred as **OS-Non-Iso** and **SS-Non-Iso**.

A schematic view of the regions definition is shown in Fig. 5.2.

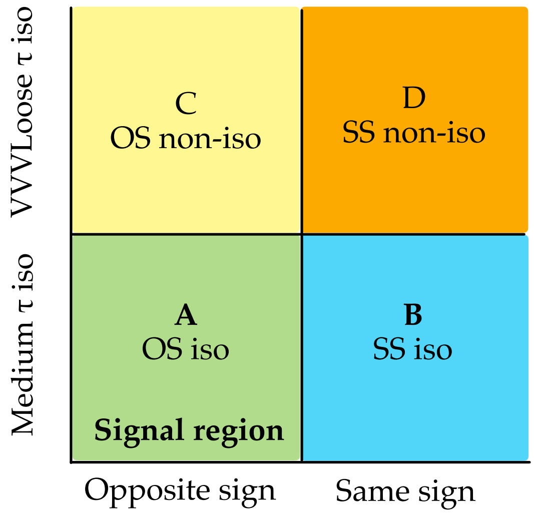


Figure 5.2

Figure 5.3: Schematic representation of signal and sidebands regions for QCD data-driven estimation.

The MC simulated background contributions are subtracted to the data in every region separately in order to obtain the contribution of the remaining QCD multi-jet background. The choice of the isolation range for the Non-Iso regions is determined by the fact that biases affecting the shape, when isolation is too far from the signal distribution, or a too low statistics sample, affecting the uncertainties on the background (yield and shape), when the range is too narrow, must be avoided. The differential distribution of the multijet background (*shape*), can be taken arbitrarily from either region B or region C without compromising the quality of data modeling with MC. However, this choice significantly impacts the treatment of systematic uncertainties. Incorporating shape uncertainties into the analysis requires defining an alternative template for the QCD multijet differential distribution. The QCD model in the signal region is obtained by extrapolating from the OS Non-Iso control region (region C), where the primary background is QCD. The QCD yield in the signal region is calculated using an extrapolation factor for each b-tag category in the B and D regions, following the relation:

$$N_A = N_C \cdot \frac{N_B}{N_D} \tag{5.7}$$

The shape and yield ratios are calculated after the subtraction of the other simulated backgrounds. In this analysis, the yield is taken from the region C and the shape is taken from the transfer factor in the other two regions. The upper and lower templates for the QCD shape uncertainty, discussed in Chapter 6, are chosen considering the transfer factor N_C/N_D and multiplying by the yield in region B.

Lepton	Iso	Non Iso
Electron	`PF RelIso03` < 0.15	$0.15 \leq$ `PF RelIso03` < 0.3
Muon	`PF RelIso03` < 0.15	$0.15 \leq$ `PF RelIso03` < 0.3
Tau	*DeepTauVSJet* WP $\geq$ Medium	VVVLoose $\leq$ *DeepTauVSJet* WP < Medium

Table 5.4: Iso and Non-Iso leptons definitions. Iso and Non-Iso leptons definitions. Electrons and muons conditions are based on the Particle Flow (PF) relative isolation, defined in Eq. 3.2, whereas hadronically decaying tau isolations rely on the *DeepTauVSJet* algorithm, described the τ_h Identification (ID) section.

Chapter 6

MC Corrections and systematic Uncertainties

While they are powerful tools, Monte Carlo (MC) simulations have several limitations, which arise mainly from modeling of hard scatter physics, e.g. leading order simulations may fail to accurately describe complex processes, and detector response. These limitations result in differences in performance related to object identification, reconstruction, and trigger efficiency between simulated and real data, affecting the sensitivity of the signal extraction. To mitigate these discrepancies, a comprehensive set of corrections is applied to the MC simulations to improve the agreement with observed distributions. The application of these MC corrections is discussed in the first part of this chapter. Despite these corrections, residual disagreements may remain, particularly in observables that are not sensitive to the signal. Such discrepancies are treated as potential sources of systematic uncertainty. Additional sources of systematic uncertainty include theoretical approximations (e.g., missing higher-order effects), experimental factors (e.g., detector calibration), and statistical uncertainties arising from the data-driven methods used to model the QCD plus jets background. The systematic uncertainties considered in this analysis for the signal extraction are described in detail in the second part of this chapter.

6.1. MC corrections

These corrections are applied on an event-by-event basis or as multiplicative factor to the total yield of a specific process. The MC corrections re-weight the simulated events and can take the form of **Scale Factors (SF)**, detailed in Appendix A.3, or **weights**.
Scale factors are determined by comparing selection efficiency (especially for triggers and analysis related selection) in data and MC simulations. They address discrepancies in detector modeling, reconstruction, or other experimental effects, ensuring simulations more accurately represent real experimental conditions. The majority of SFs used in this analysis is provided centrally by the CMS subgroups which provide

also triggers and particle Identification (ID) taggers, and are valid in certain phase space regions, which define the most inclusive requirements to the objects applied in this analysis, as described in Chapter 4.5. In addition to SFs, different types of **weights** are applied to correct for mismatches in effects such as pile-up (PU) profile differences between data and simulation, or L1 trigger prefiring probabilities. These weights are centrally provided. Thus after the event selection (Chapter 4.5) and the application of MC corrections, the number of events, $N^{\text{proc, weighted}}$ is given by:

$$N^{\text{proc, weighted}} = \sum_{j=1}^{N_{\text{tot}}} w_j^{MC,proc} \cdot \prod_k w_j^k \tag{6.1}$$

Where the j index runs from 0 to the number of selected events (N_{tot}), w_j^{MC} is the weight applied to the j-th MC event of process before any selection (defined in Eq. 5.1) and w_j^k is the weight applied to the j-th event to account the k-th MC correction. The MC corrections considered in this analysis are described in the following sections.

6.1.1 Pileup corrections

The PU modeling involves reweighting simulated events to match the PU distribution observed in data. Rather than a reweight based on the number of reconstructed primary vertices, the simulation uses the true number of PU interactions. This approach helps avoid biases that could arise from inefficiencies in reconstruction or event selection.

6.1.2 L1 prefiring

In 2016 and 2017, an issue with the Electromagnetic Calorimeter (ECAL) timing adjustment was not properly communicated to the Level 1 Trigger (L1T) Trigger Primitive (TP), described in Section 2.4.1. As a result, a significant portion of high η TPs were incorrectly associated with the previous Bunch Crossing (BX). Since two consecutive BXs cannot simultaneously trigger the L1T, this led to missing TPs in BX=0. Consequently, events effectively acts as self-vetoes when a substantial amount of ECAL energy is found in the pseudorapidity region $2. < |\eta| < 3$. The impact of this effect strongly depends on η and transverse momentum component (p_T), with a pre-firing rate exceeding 50% for jets with p_T of several hundred GeV in the region $2.75 < |\eta| < 3$.

For the muon system, the BX assignment of a trigger muon candidates can be inaccurate due to the limited time resolution of the muon system detectors. This issue was particularly pronounced in 2016 but also affected the 2017 and 2018 data to a lesser extent. The resulting pre-firing rate remains approximately uniform and impacts nearly the entire range, with a weight between 0% and 3%.

This prefiring effect is not modeled in simulations, the event weights accounting for these effects are centrally provided.

6.1.3 Pile up Jet ID

To address potential discrepancies between data and simulations in the performance of the PUJetID tagger, designed to distinguish prompt jets from those originating from PU interactions (as described in Section 3.3.4), a set of SFs is centrally provided by the CMS collaboration. Scale factors are determined by comparing the tagger's performance in data and simulations samples of Z plus jets events, and are provided as a function of jet η and p_T.

6.1.4 Leptons reconstruction and ID efficiencies

When working with multivariate discriminators which use many observables, the reconstruction/identification efficiencies in data and simulations will be different, mostly due to a not perfect modelling of the reconstruction and/or taggers observables distributions in simulations. In the following, MC corrections for electron and muon reconstruction and ID are described.

Electrons: reco, ID and Iso

For electron selection and veto the `MVAIsoWP80`, `MVAIsoWP90` and `MVANoIsoWP80` taggers, described in Section 3.3.3, are applied. The SFs that account for difference in efficiencies in data and MC samples are provided centrally by the CMS collaboration and depend on the specific Working Point (WP) chosen. They are provided as a function of electron η and p_T. The SFs measurement is performed on $Z \to ee$ + jets samples with the Tag and Probe (TnP) method (see Appendix A.2).

Muons: reco, ID, Iso

In Section 3.3.2 the muon efficiency definition has been introduced. The muon efficiencies are computed with the TnP method (see Appendix A.2) on $Z \to \mu\mu$ events and, if not otherwise specified, the tag is a muon selected with the Single Muon trigger (as defined in Tab. 4.1, 4.2 and 4.3), satisfying requirements on p_T, isolation and ID (the same described in Tables 4.5 and 4.7) and matching with the online trigger muon. Three different sets of SFs are applied to muons:

- **Reconstruction SFs**, computed considering as probe a tracker muon with requirements on p_T and longitudinal and transeverse impact parameters. The SFs are provided for candidate muons with $10 < p_T < 200\,\text{GeV}$ as a function of muon η and with $p_T \geq 200\,\text{GeV}$ in function of muon η and $|\vec{p}|$.
- **Muon ID SF**: in this case the probe muon is a global muon satisfying $|\text{dxy}| < 0.2\,\text{cm}$ and $|\text{dz}| < 0.5\,\text{cm}$. To pass, the probe must additionally satisfy the specified WP, which in this case is the tight muon ID. Scale factors are provided as a function of muon η and p_T.
- **Isolation SF**: since the isolation efficiency is computed relatively to a fixed ID working point (Section 3.3.2), and in this analysis Tight ID and Iso requirements

are applied, the SFs are computed on the probe muon passing the same ID and Iso criteria. Also in this case the SFs are centrally provided in bins of muon η and p_T.

6.1.5 DeepTau MC corrections

Different WPs of the *DeepTau* algorithm (Section 3.3.5) are employed for the $H \to \tau^+\tau^-$ candidate selection, as described in Section 4.6. To account for data/MC differences of the *DeepTau* algorithm performances, SFs are provided by the CMS subgroup that studies the tau objects, the *Tau Physics Object Group (POG)*.

- **DeepTauVSJet SFs**: Provided for multiple WPs (e.g., Loose, Medium, Tight), these SFs are computed with a fixed *DeepTauVSEle* WP and depend on τ_h observables such as p_T, Decay Mode (DM), and matching to generated particles. In this analysis, for each DM the SFs p_T dependence is modeled using linear functions.
- **DeepTauVSEle SFs**: Address discrepancies in data/MC for genuine electrons misidentified as hadronic taus. These SFs, derived from $Z \to ee$ plus jets events, are provided separately for barrel and endcap regions with binning based on ECAL geometry.
- **DeepTauVSMu SFs**: Account for data/MC differences in genuine muons misidentified as hadronic taus. Derived from $Z \to \mu\mu$ plus jets samples, these SFs are binned by η, reflecting the CMS muon and tracker systems geometries.

6.1.6 b-jet tagging MC corrections

The statistically correct treatment of b-jet taggers application is essential to enhance the signal sensitivity while minimizing biases in the final fit, addressing performance differences between data and simulations. The CMS B Tagging and Vertexing (B-Tagging and Vertexing (BTV)) subgroup provides dedicated taggers (*DeepFlavor* and *ParticleNet*) and their corrections.

b-jet tag Shape Scale Factors

The b-jet SF are evaluated using the IterativeFit method, which determines b-tagging SFs by calibrating the output of the DeepJet b-tagger to correct discrepancies between data and simulations. The method relies on flavor-enriched regions: $t\bar{t}$-enriched regions are used for heavy flavor jets (b- and c-jets), while Z-plus-jets enriched regions are used for light flavor jets (udsg-jets). The SFs are provided in bins based on the jet p_T, η and the b-tagger output.
In each region, MC efficiencies are normalized to match data, ensuring that the SF only impacts the shape of the b-tagger output distributions. After subtracting the contamination from other jet flavors in data, the data/MC ratio is computed to extract the SFs. Since the contamination itself depends on the SFs, this procedure is repeated iteratively until convergence is reached. The final SFs depend on p_T, η, and the b-tagger

output and are applied to correct tagging efficiencies and mis-tag rates in simulations. Since the SFs modify only the shape of the b-tagger distributions, the total event yield must remain preserved before and after the application b-tag selection criteria, as discussed in the following section.

b-jet tag Shape Calibration

Jet SFs depend on p_{T}, η, and the b-tagging discriminant D. The b-tag event weight ω_{event} is calculated as:

$$\omega_{\text{event}} = \prod_{i}^{N_{\text{jets}}} SF(D_i, p_{\mathrm{T},i}, \eta_i), \tag{6.2}$$

where i indexes the selected jets. To maintain consistency, the sum of event weights before and after b-tag selection must remain identical. A correction factor r is applied:

$$r = \frac{\sum \omega_{\text{before}}}{\sum \omega_{\text{after}}} \tag{6.3}$$

This correction ensures consistent event yields and is computed separately for each jet multiplicity bin in high-multiplicity analyses.

ParticleNet Corrections

The boosted category is defined by applying a cut on the AK8 jet ParticleNet mass decorrelated score, D_{bb} (Section 4.7). This selection is essential for identifying boosted objects but requires corrections to ensure an accurate data/MC agreement. The BTV POG provides SFs for "signal-like" jets, applied to jets matching a Higgs or Z boson decaying to a bb pair [145]. Thus no SFs are available for jets coming from other processes, as in some background samples the jets may not correspond to signal-like structures (e.g., X→bb). Therefore, independent corrections for these background processes are needed. These corrections have been estimated in the context of this analysis for the dominant backgrounds processes, Drell Yan (DY) and $\mathrm{t\bar{t}}$, as a function of the AK8 jet p_T.

6.1.7 Trigger corrections

Legacy triggers

The "Legacy triggers", summarized in Tables 4.1, 4.2 and 4.3, correspond to the ones previously used in the non-resonant analysis [156].

For the $e\tau_h$ and $\mu\tau_h$ channels, single-lepton triggers are combined with cross-lepton triggers ($e/\mu + \tau_h$) using a logical OR. When multiple triggers are active in a phase-space region, the overall trigger efficiency must be calculated accounting for the logical OR, avoiding overcounting events firing multiple triggers.

Considering the logical combination of N triggers, the probability of at least one trigger firing for the j-th event is given by [160]:

$$P_j = 1 - \prod_{i=1}^{N} \left[1 - x_{ij}\epsilon_i(\vec{q}_j)\right], \tag{6.4}$$

where ϵ_i is the i-th trigger efficiency depending on variables $\vec{q}_j$, and x_{ij} is the trigger bit (described in Section 2.4.2).
For $\mu\tau_h$ and $e\tau_h$ channels, SFs are calculated for the logical OR of the single-lepton and cross-lepton triggers. Assuming independence between the two legs[1] of the cross trigger, the efficiency is expressed as:

$$\text{Eff} = \varepsilon_L + \varepsilon_\ell\varepsilon_\tau - \varepsilon_\ell\varepsilon_\tau\varepsilon_{L|\ell}, \tag{6.5}$$

where ε_L is the single-lepton trigger efficiency, ε_ℓ represents the cross-lepton trigger efficiency for the e or μ leg, and ε_τ denotes the cross-lepton trigger efficiency for the τ_h leg. The term $\varepsilon_{L|\ell}$ represents the single-lepton efficiency provided that the same lepton leg fired the cross trigger.
The terms in Equations 6.5 can be visualized in terms of Venn diagrams as shown in Fig. 6.1. Unlike the others, the term $\varepsilon_{L|\ell}$ in Eq. 6.5 is not centrally provided and

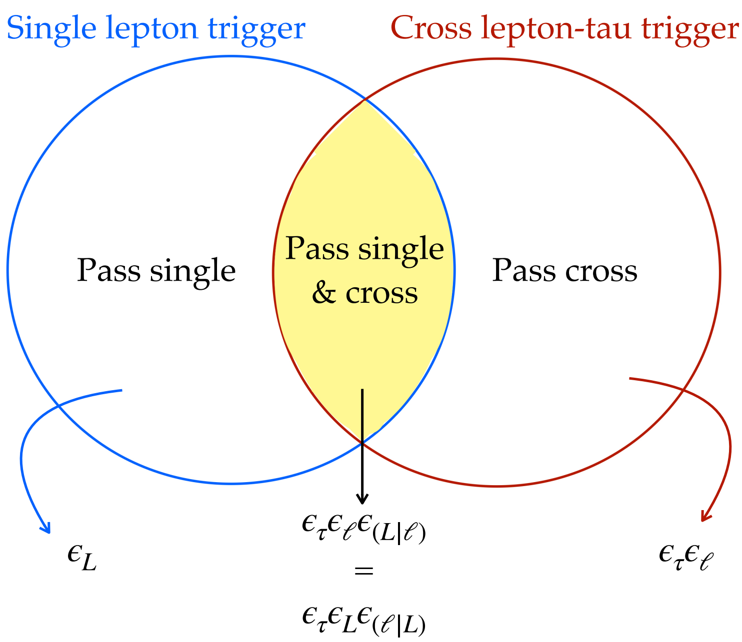


Figure 6.1: Venn diagram illustrating the single- and cross-trigger phase-spaces together with their intersection.

estimating it would require prior knowledge of the triggers used by each analysis and the storage of all possible combinations, which is impractical. However, a reasonable approximation of $\varepsilon_{L|\ell} \approx 1$ can be done, because the leptons requirements for the single lepton trigger and for the lepton leg of the cross trigger are very similar. The final term

[1] As already mentioned in Section 4.6.2 first leg refers to a specific lepton in the pair, which is chosen based on the channel. For $e\tau_h$and $e\mu$ channels, the first leg is the electron. For $\mu\tau_h$it is the muon, and for ee, $\mu\mu$ and $\tau_h\tau_h$ one either of the two e/μ/τ_h respectively can serve as the first leg.

of Eq. 6.5 is thus simplified to $\varepsilon_\ell \varepsilon_\tau$. Rare inefficiencies may result in $\varepsilon_L > \varepsilon_\ell$, resulting in $\varepsilon_{\ell|L} = 1$, but $\varepsilon_{L|\ell} \neq 1$.
Using Bayes' theorem, $\varepsilon_\ell \varepsilon_{L|\ell}$ can be rewritten as $\varepsilon_L \varepsilon_{\ell|L}$, which reduces to ε_L.
To account for both scenarios and to rely only on centrally provided SFs, the efficiency of the logical OR of the single and cross lepton triggers can be rewritten as:

$$\text{Eff} = \varepsilon_L + \varepsilon_\ell \varepsilon_\tau - \min(\varepsilon_L, \varepsilon_\ell)\varepsilon_\tau. \tag{6.6}$$

This formulation avoids the introduction of additional terms and reduces potential uncertainties in scale factors.
For the $\tau_h \tau_h$ channel, di-tau trigger efficiencies and SFs are derived from $Z \to \tau\tau \to \mu\nu_\mu\tau\nu_\tau$ events using a TnP method (see Appendix A.2). These efficiencies depend on the η and p_T of τ_h objects. SFs are applied event-by-event to MC events in the Legacy region (Section 4.8.2).

Single Tau Trigger
Flat SFs for the singleTau trigger are provided by the Tau POG for regions where trigger efficiency plateaus, defined as 10 GeV above the τ_h p_T threshold (130 GeV for 2016 and 190 GeV for 2017/2018). These SFs are applied to MC events in the Single Tau region (Section 4.8.2).

MET trigger
The Missing Transverse Energy (MET) trigger, optimized for high-mass resonance studies, is included in this analysis. Its efficiency is measured using orthogonal single-muon triggers and validated with the $\mu\mu$ channel. Efficiency curves are modeled with sigmoid functions, and SFs are derived as the ratio of data to MC fits. These are applied event-by-event to MC events in the MET region (Section 4.8.2) to account for trigger variations across years.

6.2. Systematic uncertainties

Systematic uncertainties are essential in this analysis, as they quantify potential biases in modeling, reconstruction, and event interpretation after the application of all MC corrections.
These uncertainties are divided into two categories: **normalization uncertainties** and **shape uncertainties**. They are incorporated as nuisance parameters in the final fit for the signal extraction, ensuring a robust interpretation of the results.
Normalization uncertainties affect the total yield, while shape uncertainties require event-specific weights or modified templates to account for their impact on the distribution of the variable used for the signal extraction. In the following, the systematic uncertainties included in the analysis are described.

6.2.1 Normalization Uncertainties

Normalization uncertainties affect the overall yield of a given process, whether signal or background, scaling the total event yield uniformly across the phase space and are simpler to model compared to shape uncertainties. They are treated by applying a multiplicative factor to the event yield, which accounts for all normalization effects. These factors are typically modeled using a **log-normal prior (lnN)** in the statistical framework. The log-normal distribution ensures that the scaling factors remain positive, a requirement for yield-based uncertainties.

Luminosity

The luminosity measurements with their uncertainties are centrally provided by the Compact Muon Solenoid (CMS) subgroup focused on luminosity studies (LUM-POG). Two groups of uncertainties are included: the uncertainties uncorrelated, one per era, for a total of 3 uncertainties and the 2016-2017-2018 and 2017-2018 correlated uncertainties, therefore resulting in 2 uncertainties. Their values span from 0.2% to 2%.

Theoretical uncertainties

Theoretical uncertainties arise from the choice of the PDF set according to the recommendations of [186], the uncertainty on α_s, the renormalization and the Quantum Chromodynamics (QCD) factorization scale, and the electroweark corrections. These are related to theoretical estimation of cross sections and branching fractions for signal and background processes and are reported in Tab. 6.1 with the references.

Final state branching fraction

Two different normalization systematics are obtained propagating the theoretical uncertainties on the Higgs decays [184] assuming $m_H = 125$ GeV:

- $\mathcal{B}(H \to bb) = 0.5824\ \pm 0.65\%(\text{theory})\ ^{+0.72\%}_{-0.74\%}(m_q)\ ^{+0.78\%}_{-0.80\%}(\alpha_S)$
- $\mathcal{B}(H \to \tau\tau) = 0.06272\ ^{+1.16\%}_{-1.17\%}(\text{theory})\ ^{+0.98\%}_{-0.99\%}(m_q)\ \pm 0.62\%(\alpha_S)$

The total uncertainties are $(+1.25\%, -1.27\%)$ for the $H \to bb$ decay, and $\pm 1.65\%$ for the $H \to \tau\tau$ decay.

Cross sections

Both signal and background simulated samples are affected by uncertainties related to the imperfect knowledge of their cross sections, impacting the simulation modelling. The theoretical uncertainties on the cross sections for the background processes are reported in Tab. 6.1.

Leptons reco, ID and Iso efficiencies

The uncertainties on electron and muon identification and isolation efficiencies, hence on SFs, are obtained by shifting their values by their statistical uncertainties, obtaining

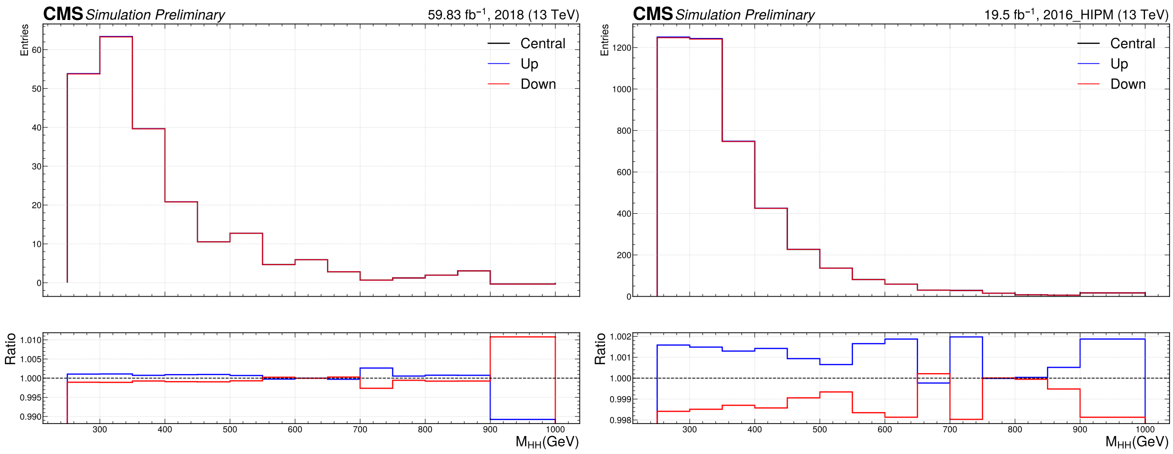
Process	PDF (%)	α_s (%)	QCD scale (%)	Ref.
ggH	$\pm$ 1.9	$\pm$ 2.6	$\pm$ 3.9	[187]
qqH	$\pm$ 2.1	$\pm$ 0.5	$^{+0.4}_{-0.3}$	[187]
ttH	$\pm$ 3	$\pm$ 2	$^{+5.8}_{-9.2}$	[176, 187]
ZH	$\pm$ 1.3	$\pm$ 0.9	$^{+3.8}_{-3.1}$	[187]
WH	$\pm$ 1.7	$\pm$ 0.9	$^{+0.5}_{-0.7}$	[187]
W + Jets	$\pm$3.7		$^{+0.8}_{-0.4}$	[183]
DY	$\pm$ 0.2		$\pm$2	[183]
$t\bar{t}$	$\pm$2.5		$^{+2.4}_{-3.6}$	[183]
ST, t channel, t	$^{+1.6}_{-1.0}$		$^{+1.12}_{-0.82}$	[180]
ST, t channel, $\bar{t}$	$^{+2}_{-1.5}$		$\pm$1	[180]
ST, tW	$\pm$2.8		$^{+2.4}_{-2.3}$	[180]
HW + Jets	$\pm$2			XSDB
HZ + Jets	$\pm$0.8			XSDB
EWK W/Z + Jets	$\pm$0.8			XSDB
WW	$^{+2.5}_{-2.4}$			[183]
ZZ	$^{+3.7}_{-2.6}$			[182]
WZ	$^{+4.1}_{-3.3}$			[182]
VVV	$\pm$0.3			XSDB

Table 6.1: Uncertainties on the PDF theoretical calculation, α_s, QCD scale theoretical calculation. From the 6th row of the table, uncertainties from PDFs and α_s are combined and from the 12th row all the theoretical uncertainties are combined. The XSDB is the CMS MC generator tool.

normalization uncertainties spanning between 0.1% and 3 % for muons, and between 1% and 10 % for electrons.
However, residual discrepancies have been observed during data/MC checks in the DYCR (Section 4.8.3) at high p_T for both electrons and muons. For muons, the residual discrepancies in spectra are more pronounced than those for electrons, which remain at acceptable levels. As suggested from the CMS Muon POG an additional 5% uncertainty is included.

PileUp Jet ID efficiencies

Uncertainties on the PU jet identification efficiencies and SFs are centrally provided by the CMS subgroup Jet and MEt (JME) POG, taking care of jets and MET studies (providing taggers and algorithms). The SFs and their uncertainties depend on the jet p_T and η, and are applied on an event-by-event base. The SFs uncertainties are obtained by varying their values by their statistical uncertainty, resulting in a normalization uncertainty between 0.1% and 0.2%. Examples of the PUJetID systematics (weights $\pm 1\sigma$, where σ is the standard deviation in each bin) on the distribution of M_{HH}^{KinFit} (Section 7.1), the observable used for the signal extraction in the resolved categories, are shown in Fig. 6.2, for the 2018 era, in the $\tau_h\tau_h$ channel for the Resolved 2b category (res2b) in the left panel, and for the 2016HIPM era, in the $e\tau_h$ channel for the Resolved 1b category (res1b) in the right panel.



(a) 2018 - res2b - $\tau_h\tau_h$ channel.

(b) 2016H - res1b - $e\tau_h$ channel.

Figure 6.2: M_{HH}^{KinFit} (M_{HH} in the figure) distribution varying the PUJetID normalization of $\pm 1\sigma$ around the central value. In each panel the era, category and channel are reported.

DeepTau corrections uncertainties

For the DeepTau tagger, all the SF and their uncertainties are centrally provided by the Tau POG.
The SFs for the *DeepTauVSEle* and *DeepTauVSMu* taggers are separated into different η regions according to the ECAL and muon systems geometries respectively. Each region has an associated uncertainty. These uncertainties are dominated by statistical sources, with values ranging from 0.1% to 1%. The uncertainties in the SFs are treated as uncorrelated across years and η bins, i.e. considering one uncorrelated nuisance parameter per year and per η bin.

Two groups of uncertainties are associated with the SFs for the *DeepTauVSJet* discriminator, depending on p_T (τ_h).

- For $20 < p_T(\tau_h) < 140\,\text{GeV}$, the SFs are derived using a simultaneous fit of $Z \to \tau_\mu\tau_h$ and $Z \to \mu\mu$ events. The associated uncertainties include: statistical uncertainties on the two linear fit parameters, resulting in 32 uncertainties in total (considering 4 DMs and 4 eras); systematic uncertainties, categorized as fully correlated across eras and DMs (1 uncertainty), correlated between DMs but

uncorrelated across eras (1 uncertainty per era, for a total of 4), and uncorrelated across both DMs and eras (16 uncertainties, calculated as 4 DMs $\times$ 4 eras).
- For $p_T(\tau_h) > 140\,\text{GeV}$, the SFs are measured using off-shell $W^* \to \tau\nu_\tau$ events across two p_T bins: $140 - 200\,\text{GeV}$ and $> 200\,\text{GeV}$. The associated uncertainties include two statistical uncertainties per era (one for each bin), totalizing 8 uncertainties, systematic uncertainties which are fully correlated across bins and eras. An additional extrapolation uncertainty accounts for the extrapolation of the SFs to higher p_T regions.

Triggers efficiency uncertainties

The uncertainties arising from the application of trigger SFs do not affect individual objects but rather modify the overall event weight.

- For the di-tau trigger, the uncertainties are provided by the Tau POG and are binned for each hadronic tau DM. When combining single-lepton and cross-lepton triggers in the $e\tau_h$ and $\mu\tau_h$ channels, the uncertainties are modified by the assumptions made in the calculation of the trigger efficiencies: they are computed per each leg propagating the uncertainties on data and MC efficiencies evaluated considering the logical OR between triggers, as described in Section 6.1.7.
- Uncertainties from the MET trigger SFs are extracted from the fitted parameters of the sigmoid function using error propagation for the data/MCs ratio.
- For the singleTau trigger, the uncertainties are provided from the Tau POG toghether with the SFs.

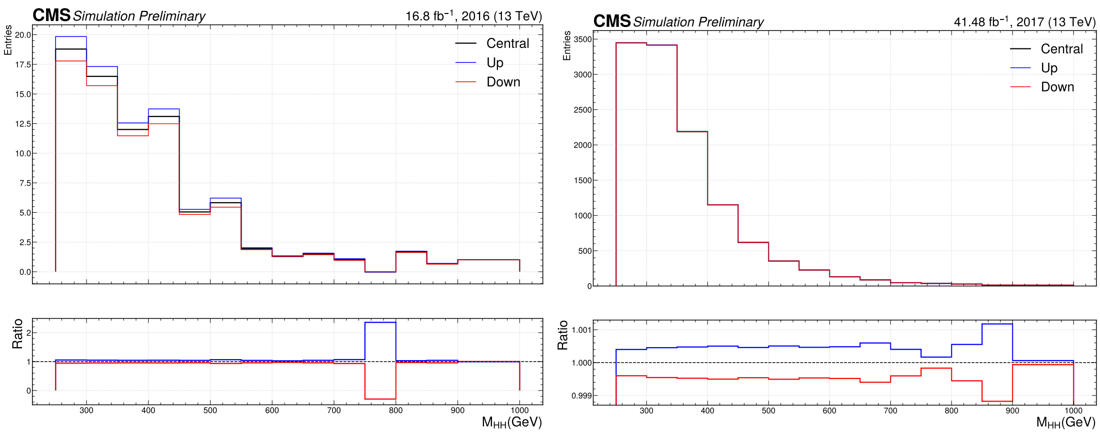
6.2.2 Shape uncertainties

Shape uncertainties impact the distribution of the variables used for the signal extraction in the final fit, rather than the total yield. These uncertainties are modeled using variations in templates or histograms, with **Up** and **Down** variations of $\pm 1\sigma_j$, with σ_j being the standard deviation with respect to the central value of the j-th bin of the variable distribution. Shape uncertainties are more complex to model as they require the creation of new histograms or templates to reflect the variations.

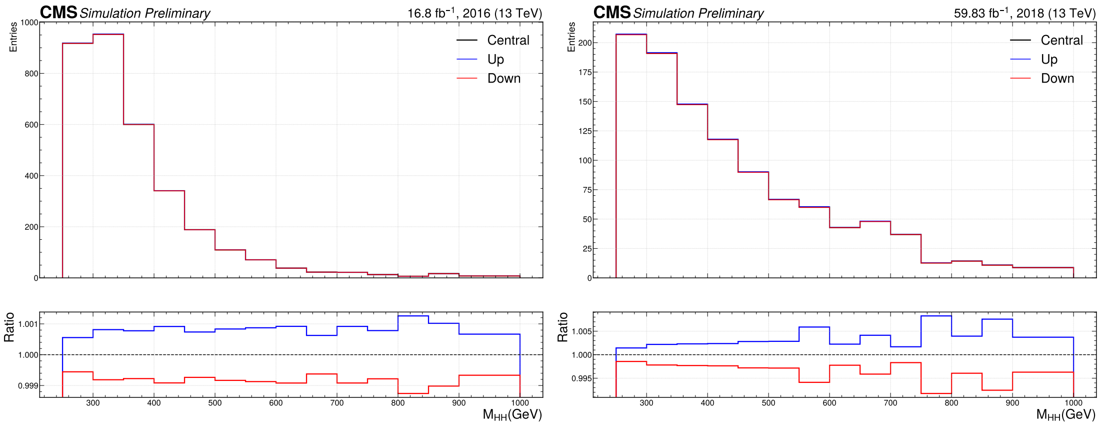
L1 Prefiring

The Level 1 (L1) prefiring uncertainties correspond to the upward and downward variations of the L1 prefiring weight. The uncertainty is taken as the square root of the quadratic sum of 20% of the object prefiring probability and the statistical uncertainty associated to the considered bin in the prefiring map for both electrons and muons. Correction factors were computed from data and applied to the acceptance evaluated by simulation.

In Fig. 6.4 two examples of the L1 prefiring systematics (up/down variated templates) on the distribution of M_{HH}^{KinFit} (Section 7.1) are shown, for the 2016 era, in the



(a) 2016 - res2b - $\tau_h\tau_h$ - τ_h triggers **(b)** 2017 - res1b - $\mu\tau_h$ - μ triggers



(c) 2016 - res1b - $e\tau_h$ - ele triggers **(d)** 2018 - res1b - $\tau_h\tau_h$ - MET triggers

Figure 6.3: M_{HH}^{KinFit} distributions (M_{HH} in the figure) varying the trigger normalization of $\pm 1\sigma$ around the central value. The uncertainties are grouped for each offline object selected by the trigger(s) in each region. In each panel the era, category, channel and object are reported.

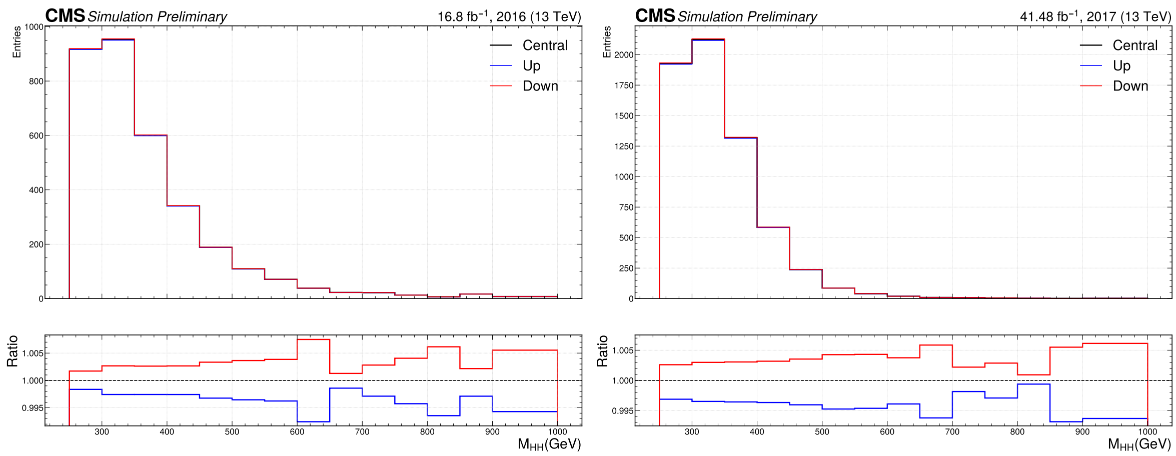
$e\tau_h$ channel and in the res1b category (Section 4.8.1) in the left panel, and for the 2017 era, in the $\mu\tau_h$ channel and in the res2b category (Section 4.8.1) in the right panel.

Pileup reweighting

The uncertainty on the pile-up reweighting technique is estimated varying the applied pile-up weights value by their uncertainty ($\pm 1\sigma$), reaching values up to 3%.

τ_h energy scale (TES)

τ_h Energy Scale (TES) is a correction applied to the energy measurements of τ_h to account for discrepancies between their measured and true energy. Uncertainties in the measurement of the τ_h energy can impact the distribution of variables used to search for a signal contribution.
The TES correction is determined using two distinct methods: for $p_T(\tau_h) \leq 34\,\text{GeV}$ it is derived from $Z \to \tau\tau \to \mu\tau_h$ events, while for $p_T(\tau_h) \geq 170\,\text{GeV}$ it is estimated from $W^* \to \tau_h\nu_\tau$ events, as the produced hadronic taus are more energetics. In the



(a) 2016 - res1b - $e\tau_h$ **(b)** 2017 - res2b - $\mu\tau_h$

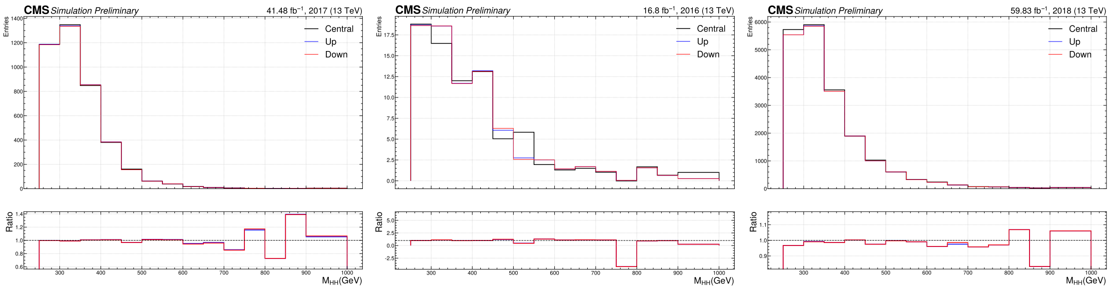
Figure 6.4: M_{HH}^{KinFit} distributions (M_{HH} in the figure) varying the L1 prefiring template of $\pm 1\sigma$ around the central value for each bin. In each panel the era, category and channel are reported.

intermediate range of $34 < p_T(\tau_h) < 170\,\text{GeV}$, it is linearly interpolated between the uncertainties of the low- and high-p_T measurements. TES corrections are provided per decay mode, with uncertainties decorrelated across data-taking periods and tau decay modes. Four uncertainties are provided by the Tau POG to account for the different tau decay modes considered in this analysis. When applying the uncertainty for a particular decay mode, the shift is applied exclusively to the reconstructed τ_h candidate matching with a generated MC true τ_h decaying in that mode.

Lepton faking τ_h energy scale

The energy scale for misidentified leptons, specifically electrons and muons faking hadronic taus, is separately calibrated. The misidentified electron energy scale is measured in $Z \to \tau\tau \to e\tau_h$ events, using the visible mass (m_{vis}) of the (e, τ_h) system as the observable. These corrections are τ_h decay-mode dependent and are treated as uncorrelated across different decay modes and data-taking periods.
The Hadron-Plus-Strips (HPS) algorithm [146, 147] identifies hadronically decaying taus, but occasionally misidentifies electrons and muons as hadronic taus. For electrons, this misidentification occurs primarily when one charged hadron is present, either alone or accompanied by one π^0.
As shown in Tab. 3.2, the branching fraction for τ hadronic decays involving only one charged hadron is $\sim 11.5\%$. In this case electrons can be misidentified as the charged hadron coming from the τ decay.
Additionally, $\sim 26.0\%$ of τ hadronic decays result in a charged hadron plus a π^0. In such cases, an electron radiating a bremsstrahlung photon can mimic the energy deposit of a π^0, especially when e^+e^- pair conversion occurs.
On the other hand, muons are less frequently misidentified as hadronic taus due to their lower bremsstrahlung emission compared to electrons. Muon misidentification is primarily observed in cases involving only one charged hadron. The uncertainty

associated with the muon-to-tau misidentification energy scale is $\pm 1\%$ across the entire p_T range, and it is uncorrelated across decay modes and data-taking periods.



(a) 2017 - res2b - $e\tau_h$ channel **(b)** 2016 - res2b - $e\tau_h$ channel **(c)** 2018 - res1b - $\mu\tau_h$ channel

Figure 6.5: M_{HH}^{KinFit} distributions (M_{HH} in the figure) varying the shape template of $\pm 1\sigma$ each bin for the TES (left), Muon-faking-tau energy scale (center) and Electron-faking-tau energy scale (right). In each panel the era, category and channel are reported.

Jet energy scale and resolution uncertainties

The uncertainties associated with both the Jet Energy Scale (JES) and the Jet Energy Resolution (JER) are accounted for by shifting the jet response by $\pm 1\sigma$ for each source [188]. Both JES and JER systematics introduce shape uncertainties in the response distributions of jet-related observables.

The JES corrections are applied to the measured jet energy in order to match the true particle-level energy [138, 139]. Due to the nonlinear response of the detector, JES ensures precise jet energy measurements by mapping the energy deposition to particle-level values. The JES uncertainties are characterized by various sources that describe fully correlated variations for a given p_T and η. Any JES variation, δ_{JES}, can be expressed as a linear combination of these sources, with coefficients α_i following Gaussian distributions:

$$\delta_{JES}(p_T, \eta; \alpha_i) = \sum_i \alpha_i \delta_{JES,i}(p_T, \eta)$$

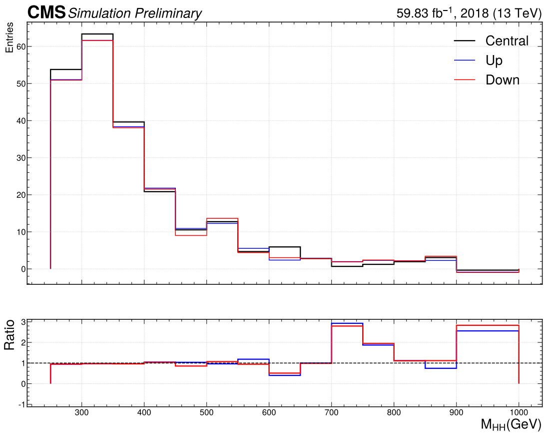
Here, the index i runs over the various JES sources, as detailed in Section 3.3.4. The total JES uncertainty at any point is computed as the quadratic sum of all individual uncertainty sources.

The JES uncertainties are centrally provided by the JME POG, and correspond to a set of 11 is uncertainties, equally applied to AK4 and AK8 jets, and propagated to the MET too.

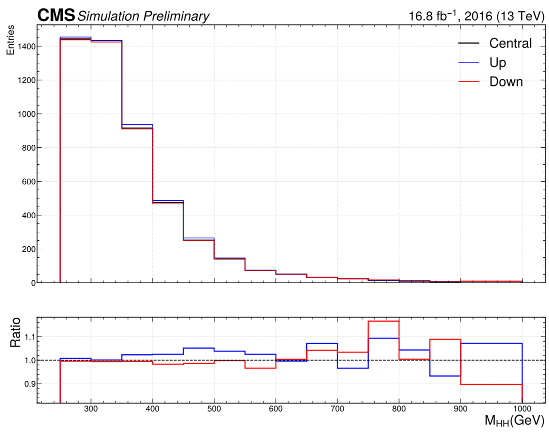
B-Tag shape SF uncertainties

Systematic uncertainties to the b-tag related SFs include:

- **Jet Energy Scale (JES)**: SFs are derived using JES-shifted jets, propagating both combined and individual JES source effects, ensuring that the scale factors reflect these calibration effects.



(a) 2018 - res2b - $\tau_h\tau_h$ channel



(b) 2016 - res1b - $\mu\tau_h$ channel

Figure 6.6: M_{HH}^{KinFit} distributions (M_{HH} in the figure) varying the shape template of $\pm 1\sigma$ each bin for the JES. In each panel the era, category and channel are reported.

- **Purity variations**: Light-flavor and charm jets contamination in heavy- and light-flavor regions is adjusted, varying its contribution by $\pm 20\%$. This is crucial to discriminates b-jets from light jets and avoids errors in the b-tagging SFs caused by misidentification.
- **Statistical uncertainties** are also considered, arising from fluctuations in data and simulations. These uncertainties are treated by applying linear and quadratic functions to the statistical variations in each b-tagging discriminant bin, ensuring accurate SF propagation.
- **Charm Jets uncertainties** are handled separately to prevent mixing with b- or light-flavor jet uncertainties; they are based on the total relative uncertainty for b-jets.

When combining data from different years (2016-2018) in Run 2, following the recommendations of JME POG, the JES, purity and charm systematics are correlated across the years, while statistical uncertainties are uncorrelated.

6.2.3 QCD estimation uncertainties

The QCD multi-jet background contribution is determined from data in jet-enriched regions using the ABCD method(Section 5.4). This background estimate is affected by various sources of systematic uncertainty. To account for statistical fluctuations in the jet-enriched regions, two normalization uncertainties and one shape uncertainty are considered. Assuming the QCD background contribution is constant across all four regions (A, B, C, and D), the following three uncertainties are defined:

- **Shape uncertainty:** The shape of the QCD background is estimated from region C and normalized using the ratio $C \times B/D$ in the main analysis. To estimate the shape uncertainty, a second histogram is constructed using the shape obtained from region B. This histogram is then used to define the uncertainty on the QCD shape.

- **Uncertainty on yield correction factor:** The statistical uncertainty on the correction factor B/D (Section 5.4) is treated as a normalization uncertainty. This uncertainty is calculated as the square root of the quadratic sum of the statistical uncertainties on the event yields in regions B and D.

6.2.4 Limited Size of MC Samples

The statistical uncertainty arising from the limited size of simulated event samples is assessed for each bin and process using the method proposed by Barlow and Beeston [189]. This method is designed to handle uncertainties in binned data, particularly those stemming from simulations or data-driven techniques.

The approach introduces a Nuisance Parameter (NP), β_{ji}, for each process j and bin i. This parameter scales the expected yield by multiplying it with the expected number of events in the i-th bin for the j-th sample. The nominal value of β_{ji} is set to 1, then in the fit procedure it is allowed to float following a prior distribution (e.g., Poisson, binomial, or log-normal). While Barlow and Beeston assumed a Poisson distribution, alternative choices such as a binomial or log-normal are also valid; the latter avoids the possibility of parameters approaching zero.

Although this method introduces a significant number of nuisance parameters, these can be efficiently managed using profile likelihood techniques, as the parameters for each bin can be optimized independently.

Assuming a Gaussian constraint on β_{ji}, with expectation value of 1 and standard deviation $\sigma_{\beta_{ji}}$, the contribution to the Negative Log-Likelihood (NLL) in the i-th bin for the j-th process can be expressed as:

$$-\ln(\mathcal{L})\Big|_{\text{Barlow-Beeston}} = -n_i \ln\left(\sum_j \beta_{ji}\mu_{ji}\right) + \sum_j \beta_{ji}\mu_{ji} + \sum_j \frac{(\beta_{ji}-1)^2}{2\sigma^2_{\beta_{ji}}}. \tag{6.7}$$

Where n_i is the measured yield in the i-th bin, μ_{ji} is the expected yield for the j-th process in the i-th bin .
During the minimization of the NLL (with other NPs fixed), a set of nonlinear equations for β_{ji} is obtained for each bin. These equations can be approximately solved using iterative Newton-type methods or more sophisticated techniques. When performing a profile likelihood using MINUIT [190] minimization, the Barlow-Beeston method can be implemented by solving for β_{ji} within the objective function, which computes NLL for given parameter values. This implementation includes the contribution of β_{ji} deviations from unity to NLL.

However, challenges arise with this approach. Any minimization algorithm approximates parameter values and the true function minimum. Near convergence, small discontinuous jumps in β_{ji} values or their contributions to NLL can confuse MINUIT's MIGRAD minimizer. This behavior may lead to a non-positive-definite covariance matrix of the β_{ji} parameters. In response, MINUIT may adjust the matrix, but this often destabilizes the fit, sending parameters to extreme values. A complete resolution would require rewriting MINUIT. The full Barlow-Beeston method is not strictly necessary to represent bin statistical uncertainties accurately. The critical factor is the overall statistical uncertainty in the predicted number of events per bin, which can be represented with a single nuisance parameter per bin. Independent uncertainties for each source in each bin, particularly Gaussian or Poisson uncertainties, can be combined into a single parameter, thus obtaining a simplified expression of the contribution to the NLL:

$$- \ln(\mathcal{L})\Big|_{\text{Barlow-Beeston}}^{\text{simplified}} = -n_i \ln(\beta_i \mu_i) + \beta_i \mu_i + \frac{(\beta_i - 1)^2}{2\sigma_{\beta_i}^2} \tag{6.8}$$

Since this is the only term in the NLL that depends on the β_i parameter, its value is determined by minimizing the NLL, which involves solving a quadratic equation:

$$\frac{\partial(-\ln(\mathcal{L}))}{\partial \beta_i} = 0 \implies \beta_i^2 + \left(\mu_i \sigma_{\beta_i}^2 - 1\right) \beta_i - n_i \sigma_{\beta_i}^2 = 0, \tag{6.9}$$

where μ_i represents the total expected yield in the i-th bin (given the values of all other parameters), σ_{β_i} is the relative statistical uncertainty on β_i and n_i is the yield in the i-th bin.

This simplified version, also known as “light” Barlow-Beeston approach is used in this analysis to treat the statistical uncertainties due to the limited size of MC samples.

All systematic uncertainties considered in this analysis for the signal extraction fit (Section 7.3) are summarized in Tab. 6.2.

[2]PU reweighting and cross section uncertainties are unified as they are included in the same step of the analysis.

	Systematic Uncertainty	CMS name	Affecting
Normalization	$\mathcal{B}(H \to \tau\tau)$	BR_htt	processes involving $H \to \tau\tau$
	$\mathcal{B}(H \to bb)$	BR_hbb	processes involving $H \to bb$
	Luminosity	lumi_13TeV_*	-
	PU reweighting + cross section[2]	CMS_pu_lumi_MC_*	-
	e reco/ID/Iso.	CMS_bbtt_eff_e_*	$e\tau_h$ channel
	μ reco/ID/Iso.	CMS_bbtt_eff_mu_*	$\mu\tau_h$ channel
	L1 prefiring	CMS_l1_prefiring_*	2016, 2016_HIPM, 2017 eras
	τ_h efficiency	CMS_eff_t_id_*	-
	QCD normalization	CMS_norm_qcd_*	QCD process
	α_s contribution	alpha_s	processes involving H and t
	PDF QCD contribution	QCD_scale	-
	PDF	pdf_*	pdf-specific dependent process
	pNet SFs	CMS_bbtt_pnet_*	boosted category
Shape	Trg. SFs	CMS_bbtt_trigSF_*	-
	TES + μ/e Fake Rate	CMS_scale_t_*	-
	b-tag shape	CMS_btag_*	resolved categories
	PU jet id.	CMS_eff_j_PUJET_id_*	resolved categories
	JER	CMS_res_j_*	-
	JES	CMS_scale_j_*	-
	QCD shape	CMS_scale_qcd_*	QCD process
	MC statistics	autoMCStats	low stat bins

Table 6.2: Summary of systematic uncertainties considered in the analysis. The first column reports the systematic uncertainty, the second column contaiнd the prefix of the uncertainty group name adopted by CMS. Finally, the last column reports the specific processes/years/eras affected by the uncertainty: unless otherwise specified, each uncertainty affects all processes/years/eras.

Chapter 7

Results

This chapter presents the results of the X $\rightarrow$ HH $\rightarrow$ $b\bar{b}\tau\tau$ search using the full Run 2 dataset collected at the CMS experiment. The discriminant observable and the reconstructed masses of the X resonance, calculated for each final state, along with the associated statistical and systematic uncertainties for the signal, background, and observed data, serve as primary inputs for the statistical analysis. These inputs are provided in the form of histograms, where the integrals of the distributions represent the expected yields for the signal and background, and are included in a fit designed to extract the signal contribution. The choice of the discriminant observable for the statistical inference, hence the signal extraction, is described in Section 7.1.
A multi-dimensional likelihood fit of the distribution of the final observable chosen for the signal extraction is done, with systematic uncertainties included in the likelihood function as Nuisance Parameter (NP) that are allowed to vary during the fit. The statistical procedure used is reported in Section 7.2.
Finally, the final results are shown in Section 7.3. The statistical procedure developed by the LHC Higgs Combination Group [191], has been adopted by both the ATLAS and CMS collaborations.

7.1. Discriminant observable

Resonances are characterized by their decay into specific final states, producing distinctive features in the invariant mass spectrum of the decay products. These features, often manifesting as localized peaks of the cross section in correspondance of the resonance mass, provide a direct and unambiguous signal of the existence of new particles.

The invariant mass of the decay products is a highly sensitive observable in resonance searches, as it encapsulates the energy and momentum information of the final-state particles in a Lorentz-invariant quantity. This property makes it a robust parameter for distinguishing signal events from complex background processes. The accurate reconstruction of the invariant mass is critical for identifying potential resonances and ensuring sensitivity to subtle signals that may otherwise be obscured by noise or experimental uncertainties.

Reconstruction of the invariant mass typically involves several techniques, ranging from simple calculations based on the visible decay products to more advanced methods that incorporate corrections for detector resolution and missing energy. The simplest approach uses only the visible decay products, while more sophisticated methods account for Missing Transverse Energy (MET) or employ kinematic fits to exploit known event topologies.

Three approaches for reconstructing the candidate resonance mass, M_X, have been investigated. They come from different ways to estimate the 4-momentum of the resonance (p_X) such that $M_X = \|p_X\|$.

Invariant Mass Using Visible Decay Products

The simplest method calculates the invariant mass using only the visible decay products, m_{HH}^{vis}. This approach involves summing the four-momenta of the $\text{H} \to \text{bb}$ and $\text{H} \to \tau^+\tau^-$ candidates to compute the mass:

$$p_{HH}^{vis} = p_{H(bb)} + p_{H(\tau\tau)} \qquad \text{where}$$
$$p_{H(bb)} + p_{H(\tau\tau)} = p_{b_1} + p_{b_2} + p_{\tau_1} + p_{\tau_1}$$

While straightforward, this method has limitations, as it neglects the contribution of undetected particles, such as neutrinos, and is affected by detector resolution. Consequently, m_{HH}^{vis} often deviates significantly from the true resonance mass and exhibits a broad distribution, making it less sensitive to the resonance signal.

Invariant Mass with Missing Transverse Energy

The second approach incorporates the missing transverse energy $\text{p}_\text{T}^\text{miss}$, to account for undetected particles. The reconstructed invariant mass, $m_{HH}^{vis+MET}$, is calculated by including the missing energy vector in the four-momentum sum:

$$p_{HH}^{vis+MET} = p_{H(bb)} + p_{H(\tau\tau)} + p_{miss}$$
$$m_{HH}^{vis+MET} = \left\| p_{HH}^{vis+MET} \right\|$$

This method improves upon $m_{HH}^{vis+MET}$ by providing a closer approximation of the resonance mass, particularly in events with significant missing energy. The inclusion of $\text{p}_\text{T}^\text{miss}$ reduces the width of the mass distribution and shifts the peak closer to the resonance mass. This approach is especially effective for boosted event categories, where decay products are highly collimated due to the resonance's high transverse momentum.

Invariant mass with kinematic Fit Procedure

The kinematic fit procedure represents the most advanced and precise method for reconstructing the resonance mass. This technique is based on some physically reasonable assumptions and exploits constraints from the event topology, such that in the kinematic parameter space that defines the resonance momentum p_X, only two independent parameters are left free to float in the fit [192]. The two independent free parameters chosen are the energy of the first τ and b-jet candidates, E_{τ_1} and E_{b_1}. In this algorithm, the two jets are assumed to be resolved, thus it is computed only for resolved categories. Possible extensions to boosted ones are under study. The following assumptions are made:

- The decay products of the two τ leptons, originating from a Higgs boson, follow the collinear approximation, since the τ leptons are highly boosted as they originate from an object that is heavy compared to their mass. This assumes that the τ leptons and their decay products are nearly collinear in the detector frame.
- The directions of all final-state objects are assumed to be accurately reconstructed. The primary uncertainties are attributed to the energy reconstruction, thus the errors on η and ϕ are negligible with respect to the energy reconstruction uncertainties.
- The transverse momentum of the resonance is balanced with the reconstructed transverse recoil, which is derived using the missing transverse energy, $\vec{E}_T^{\text{miss}}$, from the event.

The masses of the taus and jets are constrained to the Standard Model (SM) Higgs boson mass.

$$m(\tau_1, \tau_2) = m(b_1, b_2) = m_H = 125\,\text{GeV} \tag{7.1}$$

Then a global χ^2 fit is performed on the the total chi-squared function which inglobes all the kinematics of the event, defined as:

$$\chi^2 = \chi^2_{b1} + \chi^2_{b2} + \chi^2_{\text{recoil}} \tag{7.2}$$

The b-jet energies related chi-squared term is given by:

$$\chi^2_{b_{1,2}} = \left(\frac{E^{\text{fit}}_{b_{1,2}} - E^{\text{meas}}_{b_{1,2}}}{\sigma_{b_{1,2}}} \right)^2 \tag{7.3}$$

where $E^{\text{fit}}_{b_{1,2}}$ and $E^{\text{meas}}_{b_{1,2}}$ are the fitted and reconstructed energies, respectively, and $\sigma_{b_{1,2}}$ is the energy resolution.

The original τ lepton energies cannot be precisely measured due to the the presence of at least two neutrinos involved in the decay of the τ leptons. Therefore, the τ lepton energies are constrained by balancing the fitted $p_{T,X}$ transverse momentum and the reconstructed transverse recoil, $\vec{p}_{T,\text{recoil}}$.

The transverse momentum balance for the fitted resonance candidate is:

$$\vec{p}^{\,\text{fit}}_{T,X} = \vec{p}^{\,\text{fit}}_{T,b_1} + \vec{p}^{\,\text{fit}}_{T,b_2} + \vec{p}^{\,\text{fit}}_{T,\tau_1} + \vec{p}^{\,\text{fit}}_{T,\tau_2} \tag{7.4}$$

The measured transverse recoil is:

$$\vec{p}^{\,\text{meas}}_{T,\text{recoil}} = -\left(\vec{p}^{\,\text{meas}}_{T,\text{miss}} + \vec{p}^{\,\text{meas}}_{T,b_1} + \vec{p}^{\,\text{meas}}_{T,b_2} + \vec{p}^{\,\text{vis,meas}}_{T,\tau_1} + \vec{p}^{\,\text{vis, meas}}_{T,\tau_2} \right) \tag{7.5}$$

It is possible to define therefore the residual vector, which contributes to the last term of Eq. 7.2:

$$\chi^2_{\text{recoil}} = \vec{p}_T^{\,\text{res, recoil}} \cdot \mathbf{V}^{-1}_{\text{recoil}} \cdot \vec{p}_T^{\,\text{res, recoil}}, \tag{7.6}$$

where the residual vector is $\vec{p}_T^{\,\text{res, recoil}} = \vec{p}_T^{\,\text{fit, X}} + \vec{p}_T^{\,\text{meas, recoil}}$, and $\mathbf{V}_{\text{recoil}}$ is the covariance matrix.

These constraints reduce the number of fit parameters to two: the energy of the leading b-jet (E_{b_1}) and the energy of the leading τ-lepton (E_{τ_1}), and the minimisation of Eq. 7.3 is performed by varying these two parameters, performed iteratively using Newton's method.

In some cases, when the momenta combination of the signal objects is significantly incompatible with the $m_H = 125\,\text{GeV}$ hypothesis, the kinematic fit may not converge. However, for the signal events, such cases are very rare. Therefore, to ensure that the kinematic parameter space of the selected events are close to that for the signal, the events for which the kinematic fit does not converge are rejected for both the data and the simulations.

The fit provides the corrected 4-momenta the decay products within their uncertainties to satisfy these constraints, which can be summed determining the resonance mass. The mass of the X candidate estimated by the kinematic fit (m_H^{kinfit}) [192] is defined from the fitted signal 4-momenta ($p^{\text{fit}}_{\text{obj}}$) that minimizes χ^2_{kinfit}:

$$m_{HH}^{\text{kinfit}} = \left\| p^{\text{fit}}_{\tau_1} + p^{\text{fit}}_{\tau_2} + p^{\text{fit}}_{b_1} + p^{\text{fit}}_{b_2} \right\|. \tag{7.7}$$

Performance comparison

The m_H^{kinfit} has a better resolution than the $m_H^{bb\tau\tau}$, as shown in Figures 7.1-7.3 for resolved categories, and, therefore, has a better discriminating power between signal and background. For that reason, the distribution of m_H^{kinfit} in the data and in the modelled composition of the data sample have been chosen to be used in the final model fit procedure, described in the next chapter, for the signal extraction.

Among the three methods, the kinematic fit achieves the highest precision, producing a narrow and sharp peak at the resonance mass for signal events. In comparison, m_{HH}^{vis} yields the broadest distribution, with significant deviations from the true mass, while $m_{HH}^{vis+MET}$ provides intermediate performance by incorporating missing energy.

For boosted event categories, the kinematic fit has not been implemented due to computational complexity, leaving $m_{HH}^{vis+MET}$ as the standard approach. Despite its limitations, $m_{HH}^{vis+MET}$ effectively identifies resonances in such scenarios, though with reduced precision compared to the kinematic fit method as shown in Fig. 7.4.

Performance Comparison

The m_H^{kinfit} demonstrates superior resolution compared to $m_H^{bb\tau_h}$, as illustrated in Figures 7.1-7.3 for resolved event categories. This enhanced resolution translates into better discriminating power between signal and background. Consequently, the distribution of m_H^{kinfit}, both in the data and the modeled composition of the data sample, has been selected for use in the final model fit procedure, which is described in the

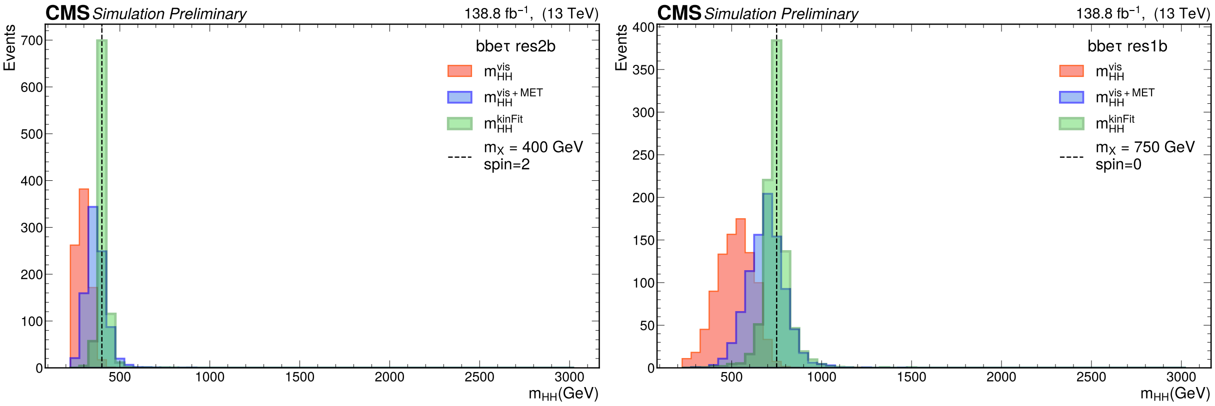


Figure 7.1: X mass distributions computed with KinFit algorithm (green), visible invariant HH mass (red) and visible HH mass + MET (blue) for the $e\tau_h$ channel and **resolved** categories. Left panel: Spin-2 hypothesis, $M_X = 400\,\text{GeV}$. Right panel: Spin-0 hypothesis, $M_X = 750\,\text{GeV}$.

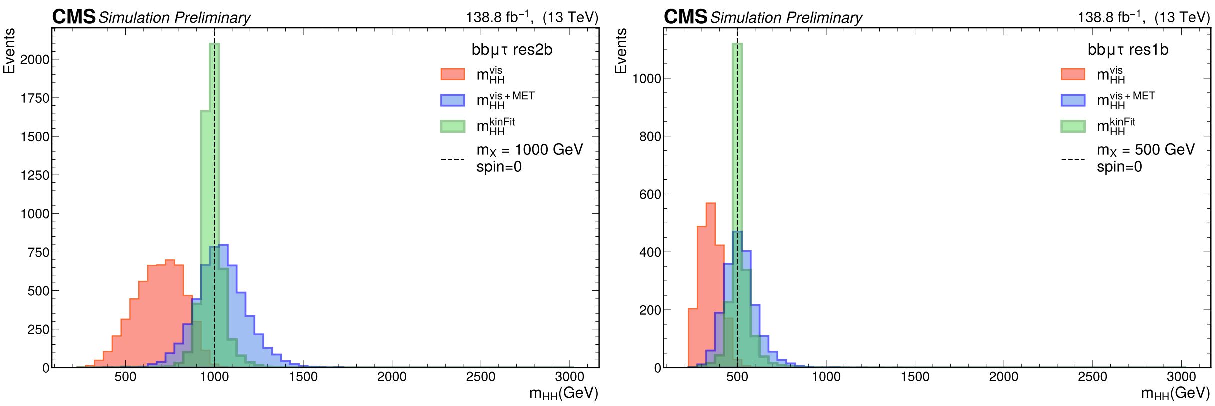


Figure 7.2: X mass distributions computed with visible invariant HH mass (red) and visible HH mass + MET (blue) for the $\mu\tau_h$ channel and **resolved** categories. Left panel: Spin-0 hypothesis, $M_X = 100\,\text{GeV}$. Right panel: Spin-0 hypothesis, $M_X = 500\,\text{GeV}$.

next chapter for the signal extraction.

Among the three reconstruction methods, the kinematic fit offers the highest precision, producing a narrow and well-defined peak at the resonance mass for signal events. In contrast, m_{HH}^{vis} results in the broadest distribution, exhibiting significant deviations from the true mass, while $m_{HH}^{vis+MET}$ provides intermediate performance by incorporating missing transverse energy into the calculation.

For boosted event categories, the kinematic fit has not been implemented due to computational constraints, making $m_{HH}^{vis+MET}$ the standard approach in such scenarios. Despite its limitations, $m_{HH}^{vis+MET}$ effectively identifies resonances, though with reduced precision compared to the kinematic fit method, as shown in Figure 7.4.

7.2. Statistical treatment

The statistical techniques employed to extract the signal or to put an upper limit to the cross section consist in the application of a modified frequentist approach for

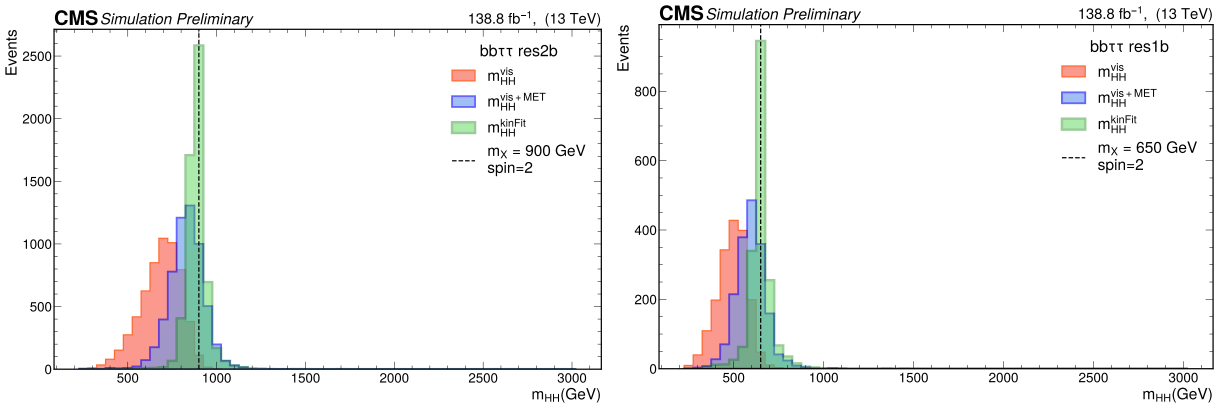


Figure 7.3: X mass distributions computed with KinFit algorithm (green), visible invariant HH mass (red) and visible HH mass + MET (blue) for the $\tau_h\tau_h$ channel and **resolved** categories. Left panel: Spin-2 hypothesis, $M_X = 900\,\text{GeV}$. Right panel: Spin-2 hypothesis, $M_X = 650\,\text{GeV}$.

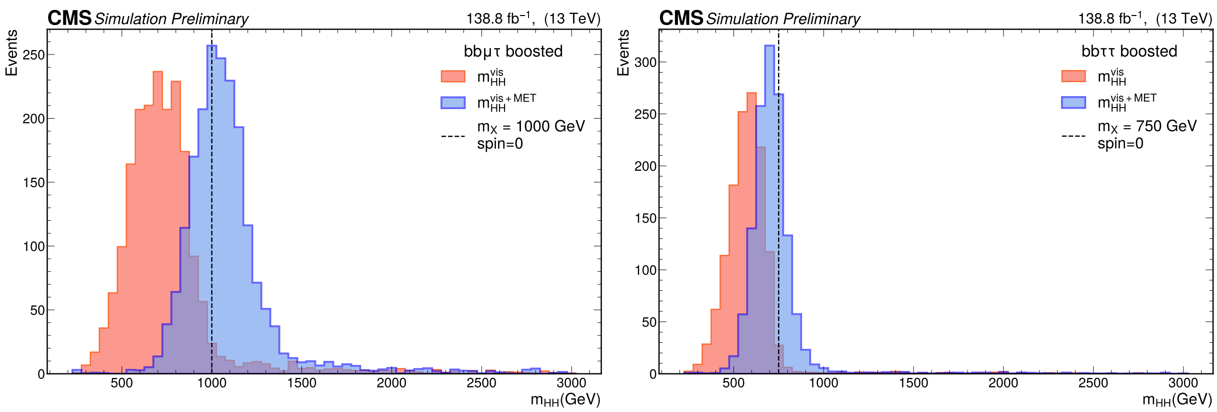


Figure 7.4: X mass distributions computed with KinFit algorithm (green), visible invariant HH mass (red) and visible HH mass + MET (blue) for the **boosted** category. Left panel: Spin-0 hypothesis, $M_X = 1000\,\text{GeV}$. Right panel: Spin-0 hypothesis, $M_X = 750\,\text{GeV}$.

hypothesis test, also defined CL_s *method*, initially developed at LEP and refined for Higgs boson discovery at the LHC [191].
The procedure computes exclusion limits, performing a maximum likelihood fit on the sensitive observables distribution (m_{HH}^{kinfit} for resolved categories and $M_{HH}^{\text{vis+miss}}$ for boosted category), and exploiting a test statistics based on the likelihood ratio under two different hypotheses:

- The **background-only hypothesis** (H_0 or H_b): There is no new particle (or not enough data) thus the data composition is made of background processes only.
- The **signal+background hypothesis** (H_1 or H_{s+b}): a signal is present indicating the presence of a new particle, so the data are composed of both signal and background processes.

The likelihood is built from the observables histograms, where the integrals of the distributions represent the expected yields for the signal and background, and are included in a fit designed to extract the signal contribution. The binned likelihood (see Appendix A.4.2) can be expressed as:

$$\mathcal{L}(\mu,\vec{\theta}) = \prod_{j=1}^{N_{bins}} \frac{(\mu s_j(\vec{\theta}) + b_j(\vec{\theta}))^{n_j}}{n_j} e^{-(\mu s_j(\vec{\theta})+b_j(\vec{\theta}))} \tag{7.8}$$

where N_{bins} is the number of bins in the histogram, n_j is the observed count in bin j, s_j and b_j are the signal and background yields.
The μ is the *signal strenghth*, which quantifies the deviation of the parameter from the SM one. From this expression it is clear that the two hypotheses are nested, i.e. H_b correspond to H_{s+b} where no signal is observed (when $\mu = 0$).
Uncertainties in predictions are managed using **Nuisance Parameters (NP)** $\vec{\theta}$[1], which influence signal and background expectations, thus s and b depend on θ.
To properly account these uncertainties contributions, since they are treated as NP, the *likelihood profiling* method is used, which integrates systematic errors into this framework, with auxiliary PDFs constraining the likelihood and constructing test statistic distribution.
Thus systematic uncertainties are represented by probability density functions (PDFs) $\rho(\theta|\tilde{\theta})$ which can be interpreted as posteriors derived from auxiliary measurements (e.g. calibrations, mathematical constraints, previous physics analyses...) via Bayes' theorem, such that $\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta)$, where $p(\tilde{\theta}|\theta)$ are the auxiliary measurements and $\pi_\theta(\theta)$ are the priors.
Typical pdfs for the systematics are the normal, log-normal, gamma distributions, keeping flat priors $\pi_\theta(\theta)$. In this way, s. Thus, when including also the systematic uncertainties, the resulting likelihood is defined as *profile likelihood* and its expression is:

$$\mathcal{L}(\mu,\theta) = \prod_{j=1}^{N_{bins}} \left[\frac{(\mu s_j(\theta) + b_j(\theta))^{n_j}}{n_j} e^{-(\mu s_j(\theta)+b_j(\theta))} \cdot \prod_{i=1}^{N_{syst}} p_i(\hat{\theta}_i|\theta_i) \right] \tag{7.9}$$

[1]From now on the vector symbol will be omitted for simplicity and the NPs will be labeled as θ but it is important to remind that they are a set of NP.

To compare the compatibility of the data with the background-only and signal+background hypotheses, the test statistic (see Appendix A.4.3) chosen is based on profile likelihood ratio:

$$\lambda(\mu) = \frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})} \tag{7.10}$$

where:

- the data are actual or toy MC data.
- $\hat{\theta}_\mu$ is the Maximum Likelihood Estimator (MLE), defined in Appendix A.4.1, of θ for a given μ, thus $\mathcal{L}(data|\mu, \hat{\theta}_\mu)$ is the maximum likelihood for a fixed signal strength μ and MLEs nuisance parameters $\hat{\theta}_\mu$.
- $(\hat{\mu}, \hat{\theta})$ are the MLEs such that $\mathcal{L}(data|\hat{\mu}, \hat{\theta})$ is the maximum likelihood across all values of both μ and $\hat{\theta}$.

$$\tilde{q}_\mu = -2ln(\lambda(\mu)) = -2\ln\frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}, \text{with} \qquad 0 \leq \hat{\mu} \leq \mu \tag{7.11}$$

The lower and upper constraints are applied to ensure that signal strength is zero or positive to have physical meaning and that the confidence interval remains one-sided and always includes the null hypothesis, respectively.
Upward fluctuations in the data, where $\hat{\mu} > \mu$, are not considered evidence against the signal hypothesis of strength μ. Instead, such fluctuations are treated as consistent with the hypothesis. By applying these constraints, the analysis avoids detached intervals that could include unphysical or contradictory results, focusing on excluding weaker-than-expected signals ($\hat{\mu} < \mu$). Upward fluctuations are thus interpreted conservatively, as they do not contradict the presence of a signal and may even support it.
The observed value of the test statistic, $\tilde{q}_\mu^{\text{obs}}$, is computed for a given hypothesis: in a first instance, the MLEs of the NPs are estimated via two separate maximum likelihood fits on data, in the background-only hypothesis ($\mu = 0$), yielding $\hat{\theta}_0$, and in the signal+background hypothesis ($\mu > 0$), yielding $\hat{\theta}_\mu$. Then an unconstrained maximum likelihood fit is done with both μ and θ free to vary, to find the overall MLEs $\hat{\mu}$ and $\hat{\theta}$ that maximize $\mathcal{L}(data|\mu, \theta)$, wit $\hat{\mu} \geq 0$ to ensure the signal strength remains physically meaningful. Finally, the test statistic $\tilde{q}_\mu^{\text{obs}}$ is evaluated in correspondance of the obtained values under the two different hypotheses.
This value is compared against test statistic distributions derived from toy Monte Carlo (MC) pseudo-data generated to construct pdfs under the two hypothesis $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0)$. For the pseudo-dataset generation, the nuisance parameters are fixed to their MLE ($\hat{\theta}_\mu^{\text{obs}}$ and $\hat{\theta}_0^{\text{obs}}$) for the two hypotheses respectively, but they are allowed to float in fits needed to evaluate the test statistic.
A large value of the test statistic $\tilde{q}_\mu^{\text{obs}}$ indicates that the observed data is less consistent with the signal hypothesis ($\mu > 0$) compared to the background-only hypothesis ($\mu = 0$). Conversely, a smaller value of $\tilde{q}_\mu^{\text{obs}}$ suggests that the observed data is more compatible with the signal hypothesis.

To obtain conservative results, two p-values (see Appendix A.4.6) are defined to be associated with the actual observation for the signal+background and background-only hypotheses:

$$p_{s+b} = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|H_{s+b}) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu$$
$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|H_b) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu$$

Where p_{s+b} is p-value for the signal+bacgkround hypothesis and p_b is the probability of obtaining a result less compatible with the background-only hypothesis than the observed result. The $CL_s(\mu)$ is the ratio of these two probabilities:

$$CL_s(\mu) = \frac{p_{s+b}}{1 - p_b}. \tag{7.12}$$

If the distributions for the signal hypothesis and the background-only hypothesis are well-separated, and H_{s+b} is true, then p_b will be very small. In this case, $1 - p_b \approx 1$, and the confidence level CL_s will approach the signal-plus-background hypothesis probability p_{s+b}. On the other hand, if the two distributions largely overlap, then p_b will be large, leading to a small value for $1 - p_b$, which prevents CL_s from becoming too small and ensures that the significance is not overestimated. This ensures that the null hypothesis is only rejected when there is strong evidence against it. The CL_s is then compared to the **significance level** (or simply *significance*, see Appendix A.4.7), denoted as α, which represents the probability of rejecting the alternative hypothesis (H_{s+b}) when the null hypothesis (H_b) is actually true, basically corresponding to the false positive. For a fixed significance level (α), the Neyman-Pearson lemma (see Appendix Section A.4.4) ensures that a selection criterion based on the likelihood ratio minimizes the misidentification probability (β). To avoid rejecting the null hypothesis in situations where the experiment has low sensitivity, the confidence level CL_s must satisfy:

$$1 - CL_s \leq CL \qquad \text{or, equivalently}$$
$$CL_s \geq \alpha$$

With $\alpha = 1 - CL$. When this condition is met, the signal hypothesis will be considered excluded at the confidence level CL. When choosing, for example $\alpha = 0.05$, the probability that the signal+background hypothesis is incorrectly taken to be true is 5% if $CL_s < \alpha$ while the signal+background hypothesis is excluded at a confidence level of 95% if $CL_s \geq \alpha$. This method method gives conservative limits, i.e. the actual confidence level is higher than $(1 - \alpha)$ [191].

Through this approach, model-independent 95% Confidence Level (CL) exclusion limits are derived for the production cross-section of a massive X resonance, under spin-0 and spin-2 hypotheses, multiplied by the branching fraction: $\sigma(gg \to X) \times \text{BR}(H \to b\bar{b}, H \to \tau^+\tau^-)$, as a function of the mass M_X. Statistical and systematic uncertainties are incorporated into the analysis using log-normal and shape uncertainties, which affect the shape of the discriminants. More details on the upper limit plots are provied in Appendix A.4.10.

7.3. Upper limits on cross section

The statistical method employed in this analysis follows the modified frequentist approach explained in Section 7.2, enabling the interpretation of p-values and CLs as probabilities. However, to ensure an unbiased methodology and avoid selection criteria based on signal-region data, the Compact Muon Solenoid (CMS) Collaboration enforces a rigorous blinding procedure. Thus the analysis developments must be conducted exclusively using simulated MC simulated events and sideband data, explicitly excluding the signal region. As a result, this search for the $HH \to b\bar{b}\tau^+\tau^-$ channel was performed in a blinded manner, preventing any influence of observed data in the sensitive observables and regions. This approach guarantees an impartial estimation of the expected limits and significance of the search while awaiting validation from the CMS internal review process.

To assess the analysis sensitivity, an *expected result* is defined, typically requiring the generation of pseudo-data (toys). Given the high computational cost of this process, the Asimov dataset [193] is exploited as a more efficient alternative, where statistical fluctuations are suppressed. Once the analysis is finalized and the greenlight for the unblinding will be given by the CMS collaboration, data will be included into the signal region and performing a maximum likelihood fit to extract the observed result. Model-independent upper limits at 95% CL are set on the production cross-section of a heavy resonance produced via gluon-gluon fusion ($gg \to X$), times the branching fraction $H \to hh \to b\bar{b}\tau^+\tau^-$. These limits are derived without assuming a specific theoretical framework, making them widely applicable to various Beyond Standard Model (BSM) scenarios. The limits are evaluated for resonance masses (m_X) ranging from 250 to 3000 GeV, utilizing all available simulated signal samples. To ensure full mass coverage, a linear interpolation technique is applied.

Despite being based solely on simulations, the obtained results demonstrate a significant sensitivity improvement over previous searches, surpassing the constraints set by earlier analyses [194]. This enhanced sensitivity highlights the robustness of the analysis and its potential impact on future searches for new physics.
The upper limits on the production cross-section times branching ratio, are presented in Figures 7.5-7.7, incorporating all channels, categories, and the full Run 2 dataset. In Figure 7.5, the inclusive (left) and channel-separated (right) limits are shown under the spin-0 (Radion) hypothesis, while Figure 7.7 presents the corresponding results for the spin-2 (Graviton) hypothesis.

The limits split by channel confirm that the $\tau\tau$ channel provides the highest sensitivity, as expected, reinforcing the motivation for comparing the performance of different tau-identification algorithms. Specifically, a dedicated study was conducted to compare *DeepTau2017v2p1* with the updated *DeepTau2018v2p5* algorithm. While both versions yield similar exclusion limits, *DeepTau2018v2p5* brings an important improvement in the data/MC agreement, as discussed in Section 4.8.4. This results underscores its reliability and suitability for future analyses, ensuring continued improvements in search sensitivity.

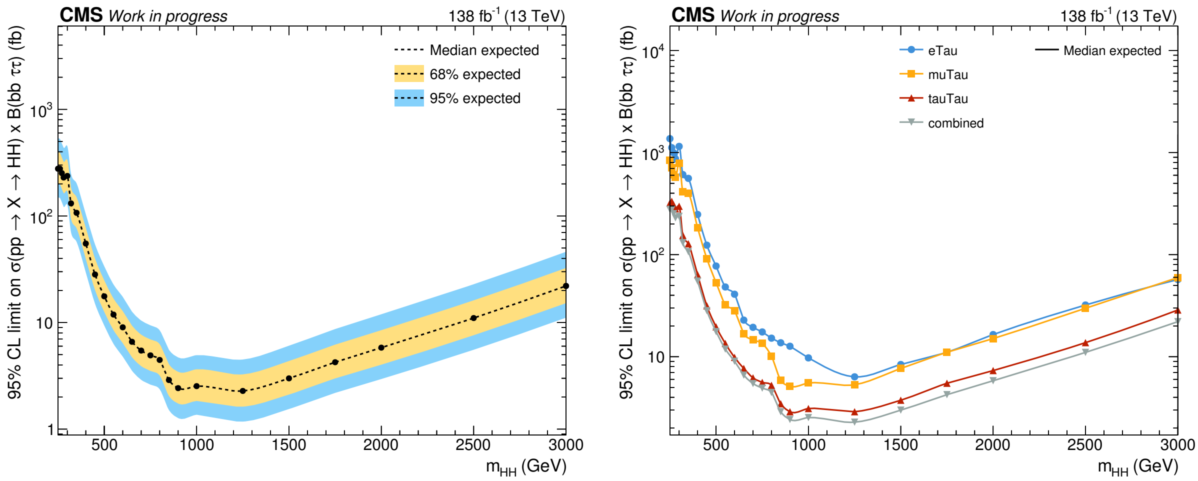


Figure 7.5: Left: Upper limits at 95% CL on the X $\to$ HH $\to$ $b\bar{b}\tau\tau$ cross section times the HH $\to$ $bb\tau\tau$ branching fraction for the $e\tau_h$+$\mu\tau_h$+$\tau_h\tau_h$ combined channels. Right: same upper limits, comparing the $e\tau_h$ $\mu\tau_h$ and$\tau_h\tau_h$ (blue, yellow, red) with the combined result (grey). These results are shown for the spin-2 (Radion-like) hypothesis.

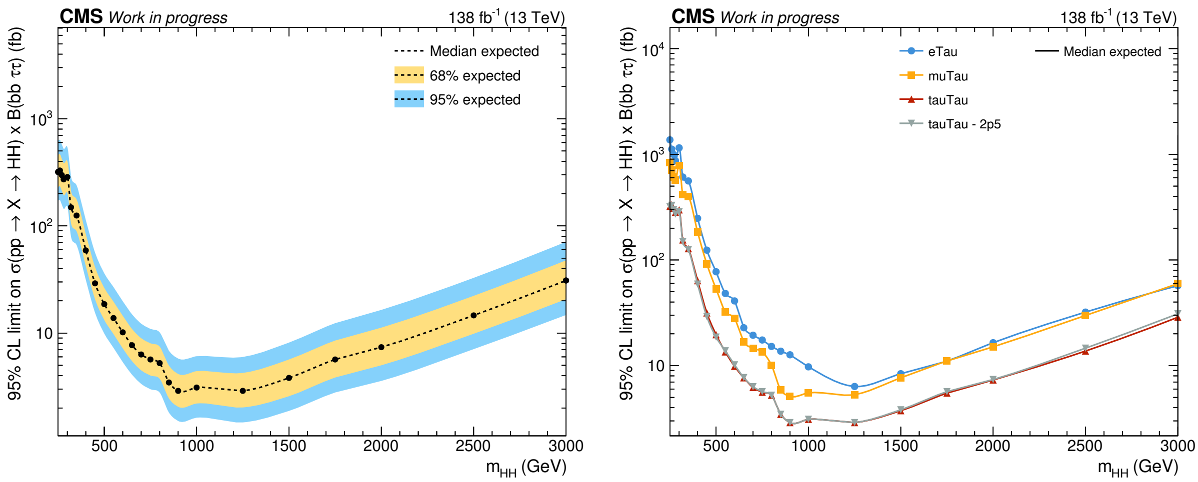


Figure 7.6: Left: Upper limits at 95% CL on the X $\to$ HH $\to$ $b\bar{b}\tau\tau$ cross section times the HH $\to$ $bb\tau\tau$ branching fraction for the $\tau_h\tau_h$ channel using the *DeepTau2018v2p5* for the analysis selection. Right: same upper limits, comparing the $e\tau_h$ $\mu\tau_h$ and$\tau_h\tau_h$ channels using the *DeepTau2017v2p1* selection (blue, yellow, red) with the $\tau_h\tau_h$ channel using the *DeepTau2018v2p5* selection (grey). These results are shown for the spin-2 (Radion-like) hypothesis.

These results mark an important milestone in the search for resonant Higgs boson pair production, setting new constraints on possible BSM scenarios and paving the way for future discoveries in high-energy physics.

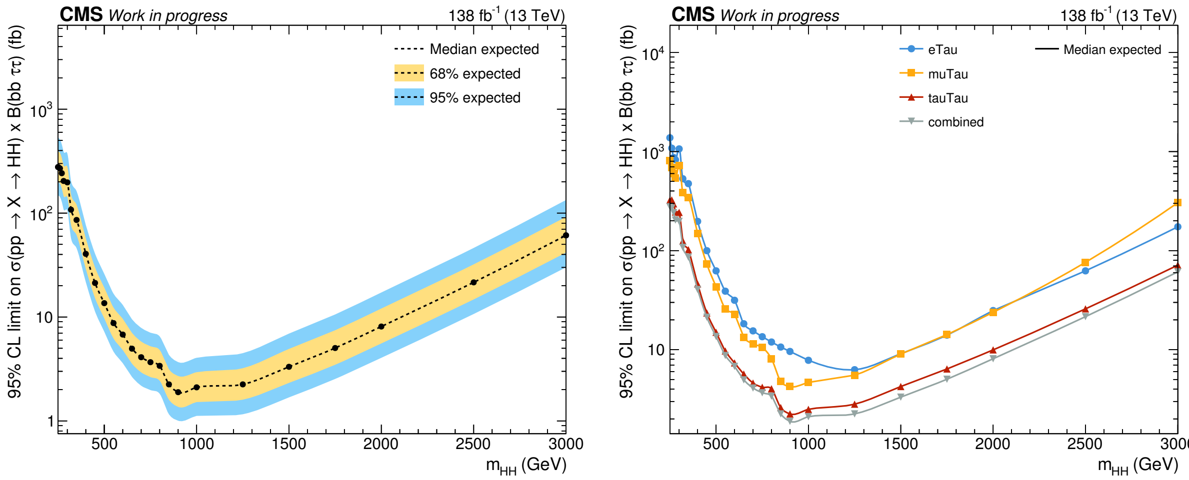


Figure 7.7: Left: Upper limits at 95% CL on the X $\to$ HH $\to$ b$\bar{\text{b}}\tau\tau$ cross section times the HH $\to bb\tau\tau$ branching fraction for the $e\tau_h$+$\mu\tau_h$+$\tau_h\tau_h$ combined channels. Right: same upper limits, comparing the $e\tau_h$ $\mu\tau_h$ and$\tau_h\tau_h$ (blue, yellow, red) with the combined result (grey). These results are shown for the spin-2 (Graviton-like) hypothesis.

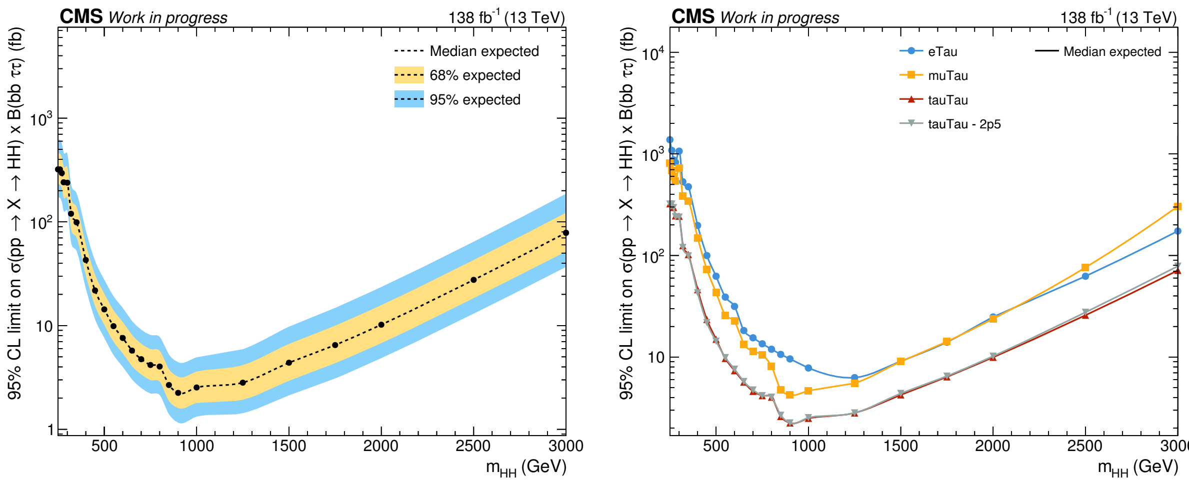


Figure 7.8: Left: Upper limits at 95% CL on the X $\to$ HH $\to$ b$\bar{\text{b}}\tau\tau$ cross section times the HH $\to bb\tau\tau$ branching fraction for the $\tau_h\tau_h$ channel using the *DeepTau2018v2p5* for the analysis selection. Right: same upper limits, comparing the $e\tau_h$ $\mu\tau_h$ and$\tau_h\tau_h$ channels using the *DeepTau2017v2p1* selection (blue, yellow, red) with the $\tau_h\tau_h$ channel using the *DeepTau2018v2p5* selection (grey). These results are shown for the spin-2 (Graviton-like) hypothesis.

7.4. Pulls and Impacts

A key step in validating the analysis is evaluating the impact of nuisance parameters through pulls and impacts (see Appendix A.4.9). Fig. 7.9 and Fig. 7.10 present the leading 25 nuisances, highlighting the dominant sources of systematic uncertainty, in the spin 0 and spin 2 hypotheses respectively. The systematic uncertainties names are defined in Tab. 6.2. The dominant systematic uncertainties, arise from the theoretical computation of the cross sections, the multijet background estimation, and jet and tau-lepton reconstruction.

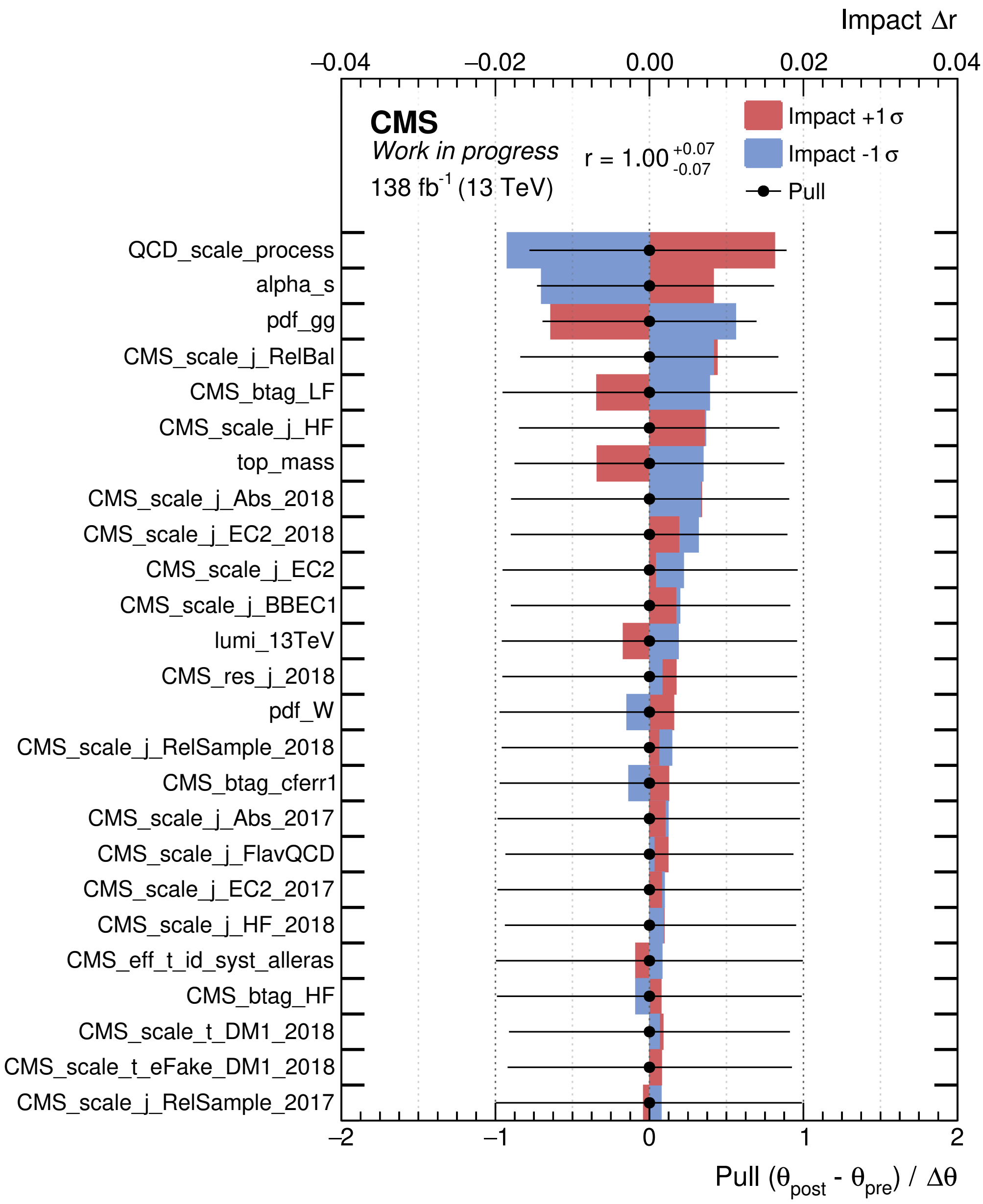


Figure 7.9: Some of the impact systematic uncertainties and relative impact on the signal strength, in the Spin-0 hypothesis, considering $m_X = 700\,\text{GeV}$. The black points and associated uncertainties report the deviation of the value of each nuisance θ from the maximum likelihood estimation with respect to their nominal value, divided by its uncertainty (pull). The blue and red bars report the impact of each nuisance parameter on the parameter of interest r (corresponding to μ), and they denote a positive and negative shift of the parameter by 1σ, respectively. The top panel reports the expected values obtained from an Asimov dataset, while the bottom panel shows the observed values after the maximum likelihood fit of the data.

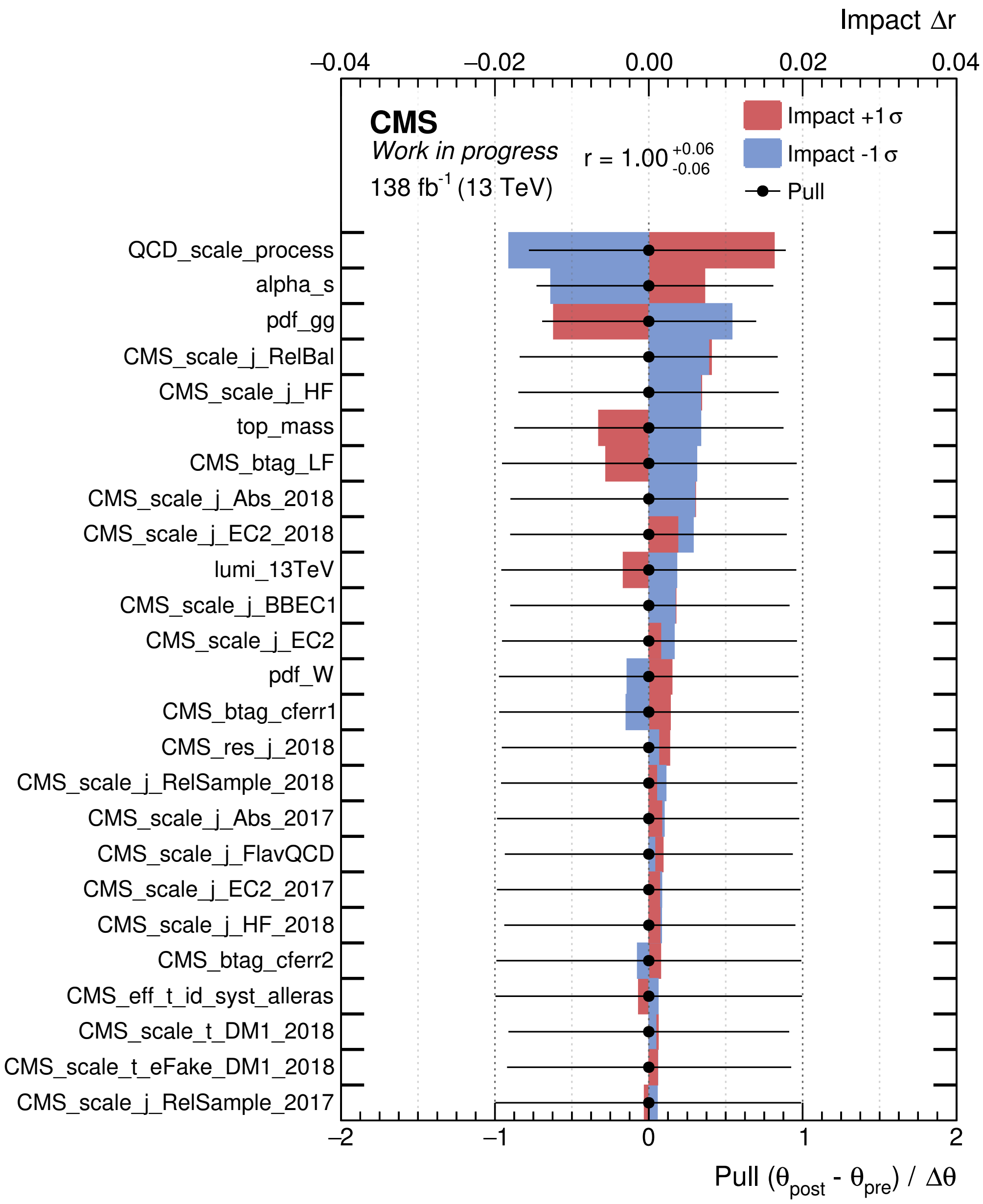


Figure 7.10: Some of the impact systematic uncertainties and relative impact on the signal strength in the Spin-2 hypothesis, considering $m_X = 700\,\text{GeV}$. The black points and associated uncertainties report the deviation of the value of each nuisance θ from the maximum likelihood estimation with respect to their nominal value, divided by its uncertainty (pull). The blue and red bars report the impact of each nuisance parameter on the parameter of interest r (corresponding to μ), and they denote a positive and negative shift of the parameter by 1σ, respectively. The top panel reports the expected values obtained from an Asimov dataset, while the bottom panel shows the observed values after the maximum likelihood fit of the data.

7.5. Final remarks

The derived upper limits show the great potential of the HH $\rightarrow bb\tau\tau$ channels in the context of the search for new physics and set the stage for future analyses with larger

datasets and improved analysis techniques.
Moreover, the integration of state-of-the-art machine learning techniques, such as *DeepTau*, gives an example of the importance of improving the analysis methodologies for maximizing the discovery potential of upcoming LHC runs.
As future data-taking periods will yield significantly larger datasets, ensuring optimal performance of advanced identification algorithms becomes increasingly critical. The demonstrated improvements in tau identification validate the importance of continuously refining analysis techniques to maintain high-precision measurements and enhance the sensitivity of searches for rare processes. This work underscores the necessity of ongoing algorithmic advancements and meticulous validation, ensuring that future analyses can fully exploit the amount of data anticipated in the next phases of the LHC program.

Chapter 8

Conclusions and future prospectives

In this thesis, a search for resonant Higgs pair production in the $bb\tau\tau$ final state has been presented. The analyzed data are taken from proton-proton (pp) collisions at $\sqrt{s} = 13\,\text{TeV}$ at the Large Hadron Collider (LHC) and collected by the Compact Muon Solenoid (CMS) experiment during Run 2 (2016–2018). This search is of particular interest due to its connection with Beyond Standard Model (BSM) theories, which could offer insights into the current limitations of the Standard Model (SM).

The analysis was initially performed using only 2016 data, already demonstrating a strong sensitivity [194]. This motivated the need to repeat the analysis with the full dataset, incorporating newly introduced advanced tools for particle reconstruction and identification. Several key improvements over previous analyses were implemented. Among these, new neural networks were developed to identify b-jets from Higgs boson decays (*HH-Btag*), classify objects such as *DeepTau* for tau leptons and *DeepJet* for hadronic jets, and to perform event categorization using the *ParticleNet* algorithm.

To obtain the final results, the *HH-kinFit* invariant mass observable was exploited to perform a fit on Monte Carlo (MC) simulations. Upper limits on the resonance production cross section were derived at the 95% Confidence Level (CL) across a mass range from 250 GeV to 3000 GeV, under both spin-0 (Radion-like) and spin-2 hypotheses (Graviton-like).

The results shown are based only on simulations, as the analysis is still under review within the CMS collaboration and is not yet public. However, these results already demonstrate the capability of this analysis to achieve high sensitivity when incorporating the full Run 2 statistics and exploiting state-of-the-art tools for particle reconstruction and identification. Although the sensitivity with the current dataset remains insufficient for SM measurements and precise BSM studies in the double Higgs boson system, the CMS detector's capabilities, along with the advanced reconstruction methods and statistical tools employed, lay the groundwork for highly precise measurements with additional data from future LHC runs.

A key aspect of the analysis was the study of the *DeepTau* algorithm for τ-lepton identification. While *DeepTau2017v2p1* was used for the primary results, an important validation study was conducted using *DeepTau2018v2p5*. The results obtained with *DeepTau2018v2p5* are consistent with those from *DeepTau2017v2p1* in terms of sensitivity, but a significant improvement was observed in the agreement between data and MC simulations. This enhanced agreement reduces the need for large data/MC corrections, leading to more reliable background estimations and systematic uncertainty reductions. Additionally, the validation of *DeepTau2018v2p5* on Run 2 data provided a crucial testbed for its implementation in Run 3 analyses, ensuring its robustness and effectiveness in high-luminosity conditions.

With the upcoming high-luminosity phase of the LHC (HL-LHC), which will collect approximately 3000 fb^{-1} over the next decade at $\sqrt{s} = 14$ TeV, there will be an unprecedented opportunity to study rare phenomena. HH production will be a top priority for the LHC research program, as the collected statistics will be crucial for probing the nature of double Higgs boson production, searching for deviations from the SM, and setting more stringent limits on potential BSM signals. Among the various decay channels, $\text{HH} \rightarrow \text{bb}\tau\tau$ is expected to play a crucial role in the challenging measurement of σ_{HH}, both within the SM and in the search for new physics.

The improvements in τ identification with *DeepTau2018v2p5*, demonstrated in this study, will be instrumental in these future analyses. Its ability to provide a better data/MC agreement and reduce systematic uncertainties ensures that upcoming searches, particularly those in Run 3 and HL-LHC, will benefit from a more precise and robust identification of hadronic τ decays.

Appendices

Appendix A

Useful definitions

A.1. Machine Learning Models in Particle Physics

A.1.1 Multivariate Analysis (MVA)

Multi Variate Analysis (MVA) refers to the process of analyzing datasets with multiple variables (features) to make predictions or decisions. In particle physics, MVA techniques are often employed for tasks such as classification and regression, aiming to predict a label or a continuous value based on various input features (e.g., kinematic properties of particles). Common MVA methods include Neural Network (NN), Decision Trees, and Boosted Decision Tree (BDT). The goal is to use these algorithms to extract the most significant patterns from the input data and make more accurate classification or regression.

MVA techniques are typically evaluated based on metrics like the accuracy of classification (signal vs. background) and regression performance, often using methods like the Receiver Operating Characteristic (ROC) curve to evaluate the quality of the model's predictions.

A.1.2 Working Point (WP)

A Working Point (WP) refers to a predefined threshold in classification algorithms that determines the balance between efficiency (correctly identifying signal events) and background rejection (correctly rejecting background or misidentified events). In particle physics, the WP defines the target efficiency and false positive rate to optimize the performance of a signal identification algorithm in a specific analysis.

WPs are crucial in tasks like jet flavor tagging or signal-event identification, where different cuts or thresholds are applied to determine whether an event is considered "signal" or "background." For example, a WP might define a target efficiency of 80% for signal identification, while keeping background rejection above a certain level (e.g., 90%).

The WP is associated with a **target efficiency**, a chosen threshold for a particular analysis, which may depend on parameters like transverse momentum (p_T) and pseudorapidity (η).

A good WP ensures that the signal identification performance remains consistent across different physics analyses with similar background conditions. This helps in making accurate comparisons between data and simulation and provides a reliable approach for optimizing signal identification.

A.1.3 Boosted Decision Trees (BDT)

BDT are an ensemble learning method where multiple weak learners (usually decision trees) are trained sequentially. Each new tree is trained to correct the errors of the previous trees by focusing on the misclassified data points.

The final output of the BDT is the weighted sum of the outputs from all individual decision trees:

$$y_{\text{BDT}} = \sum_{i=1}^{n} \alpha_i T_i(x), \tag{A.1}$$

where:

- α_i is the weight assigned to tree T_i,
- $T_i(x)$ is the output of the i-th decision tree for input x.

BDTs are particularly effective for classification tasks, such as distinguishing between signal and background in particle physics. In this case the output is a discriminant, i.e. a continuous score, for classification of signal and background; their separation is defined setting a WP.

A.1.4 Neural Networks (NN)

Artificial NNs are computational models inspired by the structure of the human brain, designed to recognize patterns and relationships in data. NNs consist of multiple layers of interconnected nodes, or neurons, each of which performs a mathematical operation. The primary goal of an NNs is to map an input x (which can have many dimensions) to an output y, often with fewer dimensions.

A simple NNs architecture consists of:

- An input layer, where the data enters.
- Hidden layers, where the input data is processed through weighted connections and activation functions.
- An output layer, where the prediction is made.

The output of a node j in a layer is given by:

$$y_j = f\left(\sum_i w_{ij} x_i + b_j\right), \tag{A.2}$$

where:

- x_i are the inputs to the node,
- w_{ij} are the weights of the connections between the nodes,
- b_j is the bias term for node j,
- f is the activation function, such as the sigmoid function, ReLU (Rectified Linear Unit), etc.

NNs can be used for both classification and regression tasks, and their performance improves as the amount of training data increases.

A.1.5 Neural Networks vs. Decision Trees

Both NNs and Decision Trees (including BDTs) are widely used in classification and regression problems in particle physics. While NNs excel at capturing complex relationships in high-dimensional data, decision trees have the advantage of being easier to interpret and implement, especially when physical insight is important.

A.1.6 Deep Learning and Deep Neural Networks (DNN)

Deep Learning refers to the use of Neural Networks with many hidden layers, known as Deep Neural Networks (DNN). These networks are capable of learning increasingly abstract features from data at each successive layer. The term "deep" refers to the large number of hidden layers between input and output layers.

In a DNNs, each additional layer allows the model to extract more complex and higher-level representations of the input data. This makes DNNss well-suited for tasks that involve large amounts of unstructured data, such as image recognition, natural language processing, and even in particle physics for classification tasks.

The training process in DNNss is similar to regular NNs but involves the optimization of more parameters due to the increased number of layers.

A.1.7 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are a specialized type of deep neural network particularly effective for image and signal processing. CNNs utilize convolutional layers that apply filters (kernels) to input data in order to capture spatial hierarchies in the data.

The output of a convolution operation is:

$$y_j = \sum_i w_{ij} \cdot x_i, \tag{A.3}$$

where w_{ij} are the filters applied to the input x_i. CNNs include pooling layers (such as max pooling) to down-sample the data, reducing the computational complexity while retaining important features.

CNNs are widely used in image recognition and have also found applications in high-energy physics to analyze detector data.

A.1.8 Graph Neural Networks (GNN)

Graph Neural Networks (GNN) are a class of neural networks designed for data that can be represented as graphs. In a GNN, nodes represent entities, and edges represent relationships between those entities. GNNs are particularly useful when data structures are inherently relational, such as in particle physics where interactions between particles or events can be represented as graphs.

A simple GNN layer aggregates information from neighboring nodes and updates the representation of each node based on this aggregation. Mathematically:

$$h_v^{(l+1)} = \text{Update}(h_v^{(l)}, \text{Aggregate}(\{h_u^{(l)} : u \in \mathcal{N}(v)\})), \quad (A.4)$$

where $h_v^{(l)}$ is the feature vector of node v at layer l, and $\mathcal{N}(v)$ is the set of neighbors of node v. This operation allows GNNs to capture relational information in complex datasets.

A.1.9 Receiver Operating Characteristic (ROC) Curve

The ROC curve is a graphical representation used to evaluate the performance of classification models. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) for different threshold values.

Mathematically, the TPR and FPR are given by:

$$\text{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}, \quad \text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}. \quad (A.5)$$

An ideal classifier would have a TPR of 1 and an FPR of 0, meaning it perfectly identifies all positive samples without misclassifying any negative samples.

The Area Under the ROC Curve (AUC) is often used as a single metric to summarize the performance of a classifier. A higher AUC indicates a better classifier.

A.2. The Tag-and-Probe Technique

The **Tag and Probe (TnP) method** is a data-driven technique widely used in particle physics to measure the efficiency of particle detection, identification, and reconstruction. This method uses events from well-known resonances, such as $Z \to \ell^+\ell^-$ (where ℓ represents a lepton, e.g., electron or muon), and evaluates detector performance by selecting clean and unbiased samples from data.

A.2.1 Concept

The T&P method leverages the decay of resonances into pairs of particles, where one particle is designated as the *tag* and the other as the *probe*:

- **Tag:** A particle that is selected using stringent criteria to ensure high confidence in its identification and reconstruction.
- **Probe:** The other particle in the event, selected with minimal or no bias, to evaluate the efficiency of the detector's criteria or response.

The invariant mass of the tag and probe system is required to match the mass of the resonance being studied. For instance, for $Z \to \ell^+\ell^-$, the invariant mass $m_{\ell\ell}$ must satisfy:

$$|m_{\ell\ell} - M_Z| < \Delta m, \tag{A.6}$$

where M_Z is the mass of the Z boson and Δm is a mass window chosen to include the signal region while minimizing background.

A.2.2 Procedure

1. Select a clean resonance decay sample, such as $Z \to \ell^+\ell^-$.
2. Identify one particle as the *tag*, which must pass strict selection criteria.
3. Select the other particle as the *probe*, with minimal requirements to avoid bias.
4. Evaluate whether the probe passes the criterion under study, e.g., particle identification or trigger selection.
5. Extract efficiencies by analyzing the invariant mass distribution and separating the signal from background.

A.2.3 Efficiency Measurement

The efficiency ϵ of a given criterion is calculated as:

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{total}}}, \tag{A.7}$$

where:

- N_{pass} is the number of probes passing the criterion under study.
- N_{total} is the total number of probes in the selected sample.

To account for background contamination, the invariant mass distribution is analyzed. The signal yield is extracted using techniques such as a fit to the invariant mass spectrum:

$$N_{\text{signal}}(m_{\ell\ell}) = N_{\text{total}}(m_{\ell\ell}) - N_{\text{background}}(m_{\ell\ell}). \tag{A.8}$$

The corrected efficiency can then be expressed as:

$$\epsilon = \frac{N_{\text{signal, pass}}}{N_{\text{signal, total}}}, \tag{A.9}$$

where $N_{\text{signal, pass}}$ and $N_{\text{signal, total}}$ correspond to the signal components of the passing and total probes, respectively.

A.2.4 Applications

The T&P method is commonly used for:

- **Reconstruction Efficiency:** Evaluating the efficiency of reconstructing a particle's trajectory in the detector.
- **Identification Efficiency:** Measuring the likelihood that a particle is correctly identified as a specific type.
- **Trigger Efficiency:** Estimating the probability that an event of interest is successfully recorded by the trigger system.

A.2.5 Advantages

- **Data-Driven:** The method uses real data, reducing reliance on simulations that may have inaccuracies.
- **Unbiased Probes:** Minimal selection criteria on probes ensure unbiased measurements.
- **Versatility:** Applicable across various particles and decay channels.

A.2.6 Challenges

- **Background Contamination:** Accurate background subtraction is essential to avoid bias in efficiency measurements.
- **Statistical Limitations:** Precision depends on the size and purity of the resonance sample.

A.3. Scale Factor in High-Energy Physics

In high-energy physics, **Scale Factors (SF)** are used to correct for differences in efficiencies between data and Monte Carlo (MC) simulations. These efficiencies are crucial for processes such as particle reconstruction, identification, and triggering. The scale factor is defined as the ratio of the efficiency observed in real data to that predicted by simulations:

$$\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}} \tag{A.10}$$

where:

- ϵ_{data} is the efficiency measured from experimental data.
- ϵ_{MC} is the efficiency obtained from MC simulations.

The efficiency ϵ is the fraction of events or objects that pass a certain selection criterion, relative to the total number of events or objects considered. This can be expressed as:

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{total}}} \tag{A.11}$$

where:

- N_{pass} is the number of events or objects passing the selection criteria.
- N_{total} is the total number of events or objects in the sample.

Once the scale factor is calculated, it is applied to the MC events to correct for the discrepancies between data and simulation. Each event that passes the selection criteria is assigned a weight equal to the scale factor:

$$w = \text{SF} \tag{A.12}$$

In cases where the efficiency varies with kinematic variables such as transverse momentum (p_T) or pseudorapidity (η), the scale factor is determined in different bins of these variables. The corrected number of events in a given bin can be calculated by summing the scale factor weights for all events:

$$N_{\text{corrected}} = \sum_{i=1}^{N_{\text{events}}} w_i = \sum_{i=1}^{N_{\text{events}}} \text{SF}(p_{T,i}, \eta_i) \tag{A.13}$$

This method ensures that the simulation reflects the efficiencies observed in real data, improving the precision of measurements and analyses in high-energy physics experiments.

A.4. Statistical Concepts in Particle Physics

A.4.1 The likelihood function

Applied to a dataset $x = \{x_1, x_2, \ldots, x_n\}$, it returns the probability distribution function (PDF) computed on data given the model parameters. For a dataset of n independent observations $x = \{x_1, x_2, \ldots, x_n\}$, the likelihood is the joint pdf:

$$\mathcal{L}(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta) \tag{A.14}$$

The likelihood represents a *good estimator* as it is consistent, efficient, robust and with a bias approaching 0 with the increase of measurements.

In maximum likelihood estimation, the best-fit parameters $\hat{\theta}$ are those that maximize the likelihood:

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta|x). \tag{A.15}$$

A.4.2 Binned vs. Unbinned Likelihood

Binned Likelihood: In binned analyses, data is grouped into m bins. The likelihood function assumes Poisson-distributed bin counts:

$$\mathcal{L}(\theta) = \prod_{i=1}^{m} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \tag{A.16}$$

where:

- n_i: observed count in bin i,
- $\mu_i = \int_{\text{bin } i} f(x|\theta)\, dx$: expected count in bin i given model $f(x|\theta)$.

Why the number of events in histograms bins are modeled as poissonian distributions? Histograms represent discrete counts of events in bins. These counts are modeled as Poissonian because:

- **Event Independence:** Each event is independent of others.
- **Fixed Mean Rate:** The mean number of events μ in each bin is determined by the underlying probability distribution.
- **Low Probabilities:** Individual events have a small probability of falling into a specific bin, but many trials occur.

The Poisson probability of observing n events in a bin with expected mean μ is:

$$P(n;\mu) = \frac{\mu^n e^{-\mu}}{n!}. \tag{A.17}$$

This distribution is appropriate for modeling rare and independent events, typical in physics experiments. For large μ, the Poisson distribution approaches a Gaussian, justifying why histograms with many entries appear normal-like. **Unbinned Likelihood:** In unbinned analyses, the likelihood directly uses the individual data points:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(x_i|\theta). \tag{A.18}$$

Unbinned likelihood retains all the data's information but is computationally more demanding.

A.4.3 Test Statistic

When comparing two hypotheses (e.g. to check if a signal of new physics is present in the data composition), hypothesis test is performed. Considering two hypotheses are defined as:

- **Null hypothesis (H_0)**: The data is consistent with the baseline or background model, with no signal or effect.
- **Alternative hypothesis (H_1)**: The data suggests the presence of a signal or effect, providing evidence for a model different from the null.

A **test statistic** quantifies the compatibility of data with a given hypothesis. A commonly used test statistics in HEP is based on the negative log-likelihood ratio:

$$q = -2 \ln \lambda = -2 \ln \left(\frac{\mathcal{L}(\vec{x}|H_0)}{\mathcal{L}(\vec{x}|H_1)} \right) \tag{A.19}$$

where $\vec{x} = x_1, x_2, \ldots x_N$ are N independent measurements, $\mathcal{L}(H_0)$ is the likelihood under the null hypothesis (H_0), and $\mathcal{L}(H_1)$ is the likelihood under the alternative hypothesis (H_1).

A.4.4 Neyman-Pearson Lemma

The Neyman–Pearson lemma [195] states that the likelihood ratio test is the most powerful test for distinguishing between two simple hypotheses H_0 and H_1 at a fixed significance level α.

$$\lambda(x) = \frac{\mathcal{L}(\vec{x}|H_0)}{\mathcal{L}(\vec{x}|H_1)}, \tag{A.20}$$

The test is optimal in the sense that, for a fixed background misidentification probability α, the selection that corresponds to the largest possible signal selection efficiency $1 - \beta$ is given by:

$$\lambda(x) = \frac{\mathcal{L}(\vec{x}|H_0)}{\mathcal{L}(\vec{x}|H_1)}, \geq k_\alpha. \tag{A.21}$$

where, by varying the value of the "cut" k_α, the required value of α may be achieved. This corresponds to choosing a point in the ROC curve (see Fig. 9.2) such that α corresponds to the required misidentification probability. [196, 197]

A.4.5 Wilks' Theorem

When a large number of measurements is available, *Wilks' theorem* [198] provides an asymptotic approximation for a test statistic based on the likelihood ratio.

Assuming Θ_0 to be the parameter space of the null hypothesis, H_0 and Θ_1 to be the parameter space for H_1, where H_0 and H_1 are nested, i.e., $\Theta_0 \subset \Theta_1$. Wilks' theorem states that the test statistic:

$$\chi^2 = -2 \log \frac{\sup\limits_{\vec{\theta} \in \Theta_0} \prod\limits_{i=1}^{N} L(x_i; \vec{\theta})}{\sup\limits_{\vec{\theta} \in \Theta_1} \prod\limits_{i=1}^{N} L(x_i; \vec{\theta})}, \tag{A.22}$$

follows a χ^2 distribution under H_0, with degrees of freedom equal to the difference in dimensionality between Θ_1 and Θ_0. Suppose μ is the parameter of interest (e.g., signal strength), and the remaining parameters $\vec{\theta}$ are nuisance parameters. The likelihood ratio test statistic becomes:

$$\chi^2(\mu) = -2\log \frac{\sup\limits_{\vec{\theta}} \prod\limits_{i=1}^{N} L(x_i; \mu, \vec{\theta})}{\sup\limits_{\mu,\vec{\theta}} \prod\limits_{i=1}^{N} L(x_i; \mu, \vec{\theta})}. \tag{A.23}$$

This statistic is asymptotically distributed as χ^2 with one degree of freedom under H_0 when $\mu = \mu_0$.

The numerator corresponds to fixing $\mu = \mu_0$ and maximizing the likelihood over the nuisance parameters $\vec{\theta}$, while the denominator involves maximizing the likelihood over all parameters $(\mu, \vec{\theta})$. This can be rewritten as:

$$\chi^2(\mu) = -2\log \frac{L(\hat{\hat{\vec{\theta}}}_{\mu_0}; \mu_0)}{L(\hat{\vec{\theta}}, \hat{\mu})}, \tag{A.24}$$

where $\hat{\hat{\vec{\theta}}}_{\mu_0}$ and $(\hat{\vec{\theta}}, \hat{\mu})$ represent the parameter values that maximize the likelihood under the respective conditions.

This test statistic is called the *profile likelihood* and is widely used for determining upper limits on parameters.

A.4.6 p-value

When performing an hypothesis test, comparing the null hypothesis (H_0) with the alternative hypothesis (H_1), to quantify if the data are inconsistent with H_0, the *p-value* is adopted. This is a function that under the null hypothesis is uniformly distributed between 0 and 1, whereas it tends to have very small values in the alternative hypothesis. In other words, the p-value is the probability of observing a result at least as extreme as the observed value under the null hypothesis H_0.

$$p = P(\text{Test Statistic } \geq \text{Observed Value } | H_0). \tag{A.25}$$

For a one-sided Gaussian distribution:

$$p = 1 - \Phi(Z), \tag{A.26}$$

where $\Phi(Z)$ is the cumulative Gaussian distribution.

A.4.7 Significance level

Another quantity related to the p-value which is the *significance level* (or simply *significance*), corresponding to the number of standard deviation in an area equal to the p-value integrating the rightmost tail of a Gaussian distribution. For a test statistic q_0:

$$\text{Significance } Z = \Phi^{-1}(1-p), \tag{A.27}$$

where Φ^{-1} is the inverse cumulative distribution function of a standard Gaussian, and p is the p-value associated with q_0.

For example, a $Z = 5\sigma$ corresponds to $p \approx 2.87 \times 10^{-7}$, the discovery threshold in physics.

A.4.8 Upper Limit

An **upper limit** is the maximum value of a parameter θ consistent with the data at a specified confidence level CL. For example, at 95% CL, the upper limit satisfies:

$$\int_0^{\theta_{\text{UL}}} \mathcal{L}(\theta|x)\, d\theta = 0.95 \int_0^{\infty} \mathcal{L}(\theta|x)\, d\theta. \tag{A.28}$$

In HEP, upper limits are often used to constrain parameters when no significant signal is observed.

A.4.9 Impacts

Impacts quantify the effect of systematic uncertainties (nuisance parameters ν) on the results. Each nuisance parameter's impact is assessed by varying ν and evaluating the change in the test statistic or the best-fit parameters. This is often visualized in an "impacts plot," ranking the most influential sources of uncertainty.

A.4.10 Upper-Limit and Confidence Bands for the Background-Only Hypothesis

When presenting an upper limit as the result of a data analysis, specially for searches or when defining the upper limit in a parameter scan, it is often useful to report also the expected value of the limit and the interval of excursions, quantified as percentiles corresponding to 1σ and 2σ, is typically provided. The direct approach to determine the expected median upper-limit ($\mu_{95\%CL}$) and its associated $\pm 1\sigma$ (68%) and $\pm 2\sigma$ (95%) confidence bands involves the following steps:

1. For each pseudo-dataset generated in the background only hypothesis, the CL_s and the upper-limit on the signal strength ($\mu_{95\%CL}$) are computed, as if the pseudo-data were real experimental data.

2. Subsequently, the cumulative probability distribution for $\mu_{95\%CL}$ is obtained by sorting the results and integrating from the side corresponding to low event yields.

3. Finally, the expected limits and contour bands are extracted, with the median expected limit, the value where the cumulative distribution crosses the 50% quantile and the $\pm 1\sigma$ (68%) and $\pm 2\sigma$ (95%) bands corresponding to the 16% and 84% quantiles and 2.5% and 97.5% quantiles respectively

Appendix B

The HH-bTag algorithm and performances

This appendix describes the `HH-BTag` machine learning algorithm, meant to improve the selection of b-jets in the resonant and non-resonant HH $\rightarrow b b \tau \tau$ searches, and its retraining carried out in order to refine the performance of the model. The version of the method upon which the new training is based is described in detail in AN-2019/283 [199]. The performance of the algorithm is evaluated and compared with respect to the previous version and to the standard algorithms (DeepFlavour and ParticleNet) used to select the b-jet pair of the HH $\rightarrow b b \tau \tau$ event candidates.

Simulated samples used for the training

For this analysis, the signal samples considered in the training of the model correspond to the nonresonant and resonant HH gluon-gluon fusion (ggF) production processes.

1. **Nonresonant ggF production samples:** These samples include the twelve benchmark modelling points described in Ref. [41] and listed in Ref. [199].

2. **Resonant ggF production samples:** Spin-0 radion production and spin-2 graviton production are included. For each considered spin, different mass points are taken within a range of 250 GeV to 3000 GeV.

Feature definition

The definition and selection of the variables (features) used as inputs for a neural network (NN) is a very important step in the development of trained discriminators. Several high-level variables were investigated, and the most discriminant ones were retained. The training used a total of 14 variables that can be subdivided in four groups:

- The score of the b-jet discriminator and kinematic variables, for the b-jet candidate.
- Kinematic variables, for the $H_{\tau\tau}$ candidate and Missing Transverse Energy (MET).
- Angular separation between the b-jet candidate and the $H_{\tau\tau}$ candidate.
- Global event variables (year and decay channel).

In order to cover the entire phase space of the signal processes (resonant and nonresonant, mass, spin, etc...), the kinematic variables were chosen in a way that they are independent of a specific scenario.

In the following, a short description of these variables is given.

b-jet score and kinematic variables

- The score of the b-jet candidates, given by the DeepFlavour b-tagger [142].
- p_T and η of each of the b-jet candidates from the Higgs boson decay.
- Relative mass and energy of the b-jet candidate defined as $\frac{M}{p_T}$ and $\frac{E}{p_T}$, where M is the invariant mass, E the energy, and p_T the transverse momentum of the b-jet candidate.
  The use of the relative variables allow to reduce correlations with p_T.

Kinematic variables of the $H_{\tau\tau}$ candidate

- Scalar sum of the p_T of the $H \to \tau\tau$ candidate, calculated as $p_{T,H_{\tau\tau}} = \sum \tau_{1p_T} + \tau_{2p_T}$.
- The visible p_T and η of the H $\to \tau\tau$ candidate, where the visible momentum of the $H_{\tau\tau}$ is calculated as the sum of the visible 4-momentum of the τ pair candidates.
- $\Delta\phi(MET, H_{\tau\tau}^{\text{vis}})$: for the signal events, the missing momentum usually points to the Higgs boson reconstructed from the τ pairs. In fact, since neutrinos are principally produced in the direction of the τ decays, signal events have a small angular separation, while for background events this variable is expected to be randomly distributed.
- Relative momentum of the missing energy, calculated as: $\frac{p_T(\text{MET})}{p_T(H_{\tau\tau})}$.

Angular separation between the b-jet candidate and the $H_{\tau\tau}$ candidate

The b-jet candidates are required to be out of a cone of radius $\Delta R = 0.5$ around the $H_{\tau\tau}$ candidate.

- $\Delta\eta(\text{jets}, H_{\tau\tau})$.
- $\Delta\phi(\text{jets}, H_{\tau\tau})$.

In order to help the convergence and improve the stability during the training, a scaling or normalization procedure is applied to some of the input variables. The features with a significantly non-Gaussian distribution on the finite validity domain, are confined to a range [-1,1]. A standardization procedure is instead applied, for features with a relatively narrow peak in their distributions, changing the x variable to $x \to (x - \mu)/\sigma$, where μ and σ are respectively the mean and the standard deviation of the variable distribution. It is necessary to scale the variables because many algorithms attempt to make trends in the inputs by comparing features of the input points, however there is an issue when the features have drastically different scales: the features with the larger scales completely dominates the others.

Network architecture and training

The `HH-BTag` discriminant is created though a Neural Networks (NN) implemented in Keras with TensorFlow backend. The new training of the model is based on its previous version. The architecture of the designed NNs and the process of the optimization of the hyperparameters are explained in detail in [199]. The training of the model uses:

- Five concatenated LSTM layers, with a sigmoid as activation function. After each layer, a batch normalization layer is applied.
- Ten TimeDistributed Dense layers, with sigmoid as activation function. Same as before, a batch normalization layer was applied between each layer.
- Binary cross-entropy as the loss function [200] and minimized with the Adam algorithm [201].
- A final TimeDistributed Dense layer with only one unit and with a sigmoid as activation function; the latter provides an output between 0 and 1, corresponding to the score of each b-jet candidates.

The total number of events is divided into two samples, depending on the odd/even number of the event. Then, two separate trainings and tests are performed. In one case, the odd sample is used in the training, while the even sample is used for the test. In the other case the opposite happens. In both trainings, the validation set is taken from the training dataset, and represents the 25% of the total training dataset. During the training, a weight is applied to each event according to its production cross-section (for resonant processes), production mechanism, decay channel and integrated luminosity. This procedure allows to renormalize the events in order to have equal importance for different processes.

B.0.1 The HH-bTag performances for $\mathrm{X} \rightarrow \mathrm{HH} \rightarrow \mathrm{b\bar{b}}\tau\tau$ resonant production

In order to evaluate the performance of the new discriminator, it is necessary to compute the predictions made by the trained model. The predictions were calculated using separately each set of weights determined by the training, evaluated on the corresponding validation set, using the sample with the opposite parity as input to be labeled. This procedure gives a score in the interval [0, 1] to each of the b-jet candidates, which depends on how confident the model is that the b-jet in question is one of the b-jets from the $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ decay. For the search of nonresonant $\mathrm{HH} \rightarrow bb\tau\tau$ and ggF resonant production, the performance of this new training of the already presented `HH-BTag` classifier for the selection of the pair of jets from the H_{bb} decay were compared with the previous version performance and with the ones of other two classifiers: DeepFlavour and ParticleNet.

After the $H \to \tau^+\tau^-$ candidates are reconstructed and the jet selection (in Tab. 4.8) is applied, the pair of b-jets candidates with the highest score of the classifier (`HH-BTag` v2, `HH-BTag` v1, DeepFlavour or ParticleNet) is selected. The variable used to evaluate the performance of the discriminators is the purity of the selection of the b-jet pair candidate, which is defined as:

$$\text{purity}^{\text{classifier}} = \frac{N_{\text{true}}^{\text{classifier}}}{N^{\text{classifier}}} \tag{B.1}$$

where $N_{\text{true}}^{\text{classifier}}$ is the number of events in which the selection of the b-jet pair candidate matched with the ground truth definition, taken from AN-2019/283 [199]. According to it, the two most energetic generated b-jets are used to perform the match, and the true b-jet label is assigned to the reconstructed b-jet if a matching generator level b-jet is found within a cone of size $\Delta R < 0.5$ around the direction of the reconstructed b-jet. $N^{\text{classifier}}$ is the total number of events where a candidate is reconstructed and the b-jet pair forming the $H \to b\bar{b}$ candidate is selected using the chosen classifier. The results shown in this section are produced using the combination of the purities (p_{even}, p_{odd}) which are computed with the predictions evaluated with the two sets of weights.

Figures B.1, B.2, B.3 and B.4 show the purity obtained in the selection of b-jet pair from the H_{bb} decay for the ggF resonant processes for spin-0 radion production and spin-2 graviton production as a function of the resonant mass, for 2016APV, 2016, 2017 and 2018 eras respectively.

The purity goes from 89.8% to 97.4% for `HH-BTag` v2, compared to 86.9% to 96.1% for `HH-BTag` v1. Comparing results with ParticleNet algorithm, the same purity ranges from 79.7% to 91.8%.

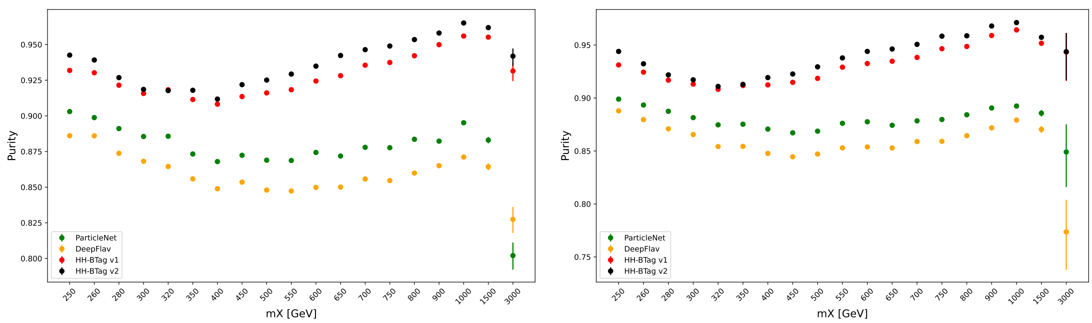


Figure B.1: Purity of the `HH-BTag` v2, `HH-BTag` v1, DeepFlavour and ParticleNet discriminators as a function of m_X, for the 2016 HIPM ggF spin-0 (left) and spin-2 (right) resonant process.

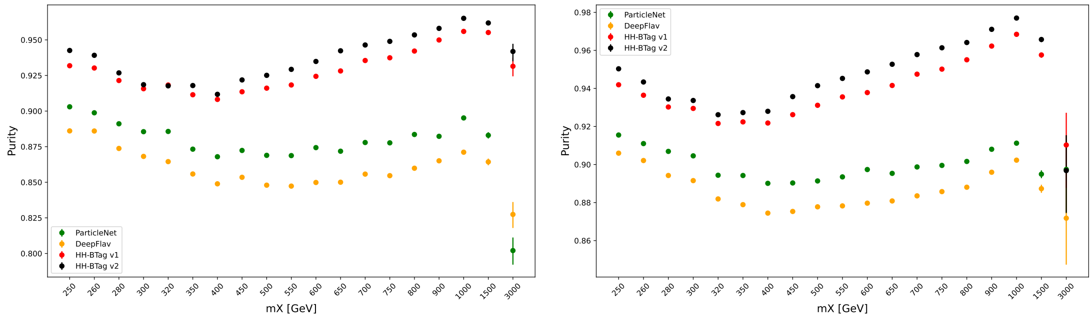


Figure B.2: Purity of the HH-BTag v2, HH-BTag v1, DeepFlavour and ParticleNet discriminators as a function of m_X, for the 2016 ggF spin-0 (left) and spin-2 (right) resonant process.

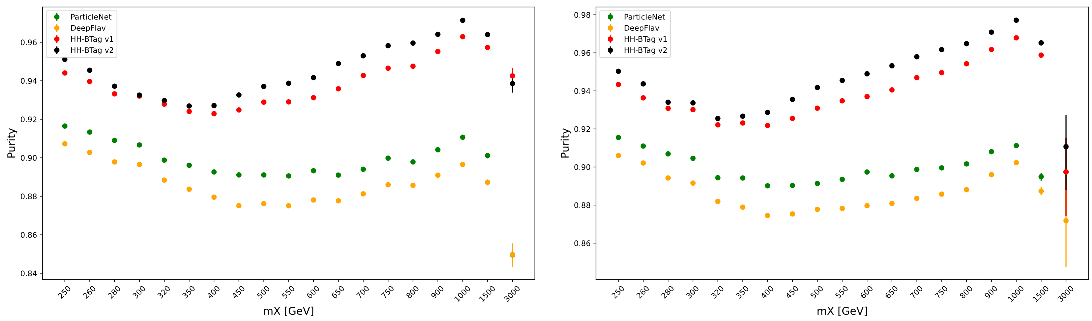


Figure B.3: Purity of the HH-BTag v2, HH-BTag v1, DeepFlavour and ParticleNet discriminators as a function of m_X, for the 2017 ggF spin-0 (left) and spin-2 (right) resonant process.

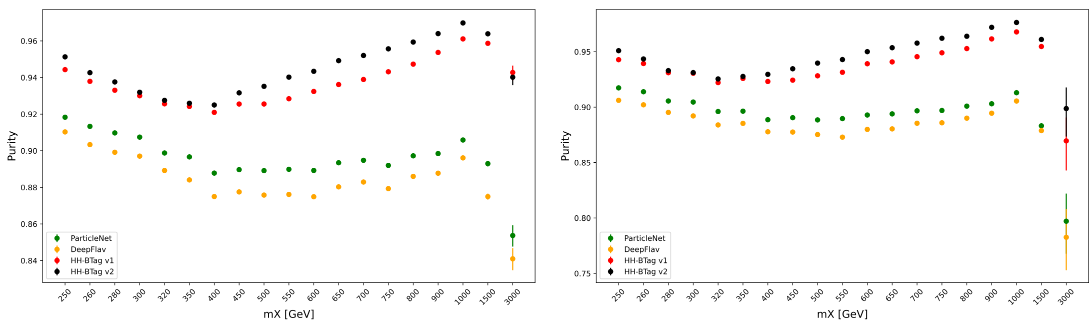


Figure B.4: Purity of the HH-BTag v2, HH-BTag v1, DeepFlavour and ParticleNet discriminators as a function of m_X, for the 2018 ggF spin-0 (left) and spin-2 (right) resonant process.

Appendix C

The Missing Transverse Energy (MET) and SingleTau triggers

In the analysis presented in this thesis, two trigger paths have been introduced together with the "Legacy" ones[1].
One key aspect of improving signal sensitivity is to increase trigger efficiency, avoiding to miss signal events while keeping non-interesting events rates low. To do this, it is important to verify how well additional triggers enhance signal identification.
In this work, the investigation of trigger efficiency increases in signal events becomes particularly important. By incorporating dedicated triggers, such as those based on MET and high-transverse momentum component (p_T) taus (SingleTau) candidates, into the baseline triggers, we can enhance the signal sensitivity, particularly in specialized regions of phase space where the baseline triggers might fall short.
In the following, the improvements made by adding MET and SingleTau triggers to the "Legacy" ones will be presented. These new triggers are designed to efficiently select events that might otherwise be lost in the baseline trigger configuration. The MET trigger allows for the identification of events with significant missing energy, which is often indicative of new physics signals, such as those from a Higgs boson decay. Similarly, the SingleTau trigger targets events involving high-p_T τ_h decays, which are crucial to save higher energetic events where one of the two tau candidates (either e, μ or τ_h) is missed.

The addition of these dedicated triggers in their respective regions of interest enhances the efficiency of signal event selection, thereby improving the overall sensitivity of the analysis. The following plots show how the gain of the inclusion of these additional triggers in the event selection, especially at high resonance mass, where other triggers may be less efficient, ultimately leading to a more complete and robust signal identification. These are computed by evaluating the triggers on signal samples,

[1] The term "Legacy" refers to the trigger paths and their associated application regions as defined in the non-resonant counterpart of this analysis [156].

and applying the triggers in their respective application regions as described in Section 4.8.2.
Figures C.3-**??** report the gain in signal acceptance by adding to the Legacy triggers: the SingleTau trigger (blue line), the MET trigger (green line) and both the SingleTau and MET triggers (red line). The results show that the inclusion of both MET and SingleTau triggers highly increases the signal acceptance across all the years, channels and resonance spin hypotheses.
However, it is interesting to note that in 2018 and 2017 eras, the impact of the adding only MET trigger or adding only SingleTau trigger with the Legacy ones is compatible, whereas for 2016 and 2016H eras the inclusion of just SingleTau trigger significantly improves the selection with respect to the addition of only the MET one. The reason could be related to the different implementation of the SingleTau triggers, as in 2016 and 2016H the offline τ_h p_T threshold for the matching is 130 GeV whereas in the other cases the offline τ_h p_T threshold is 190 GeV.

The reported plots are esemplificative and shown for the Spin-0 (Radion) hypothesis and in the Spin-2 (Graviton) hypothesis.

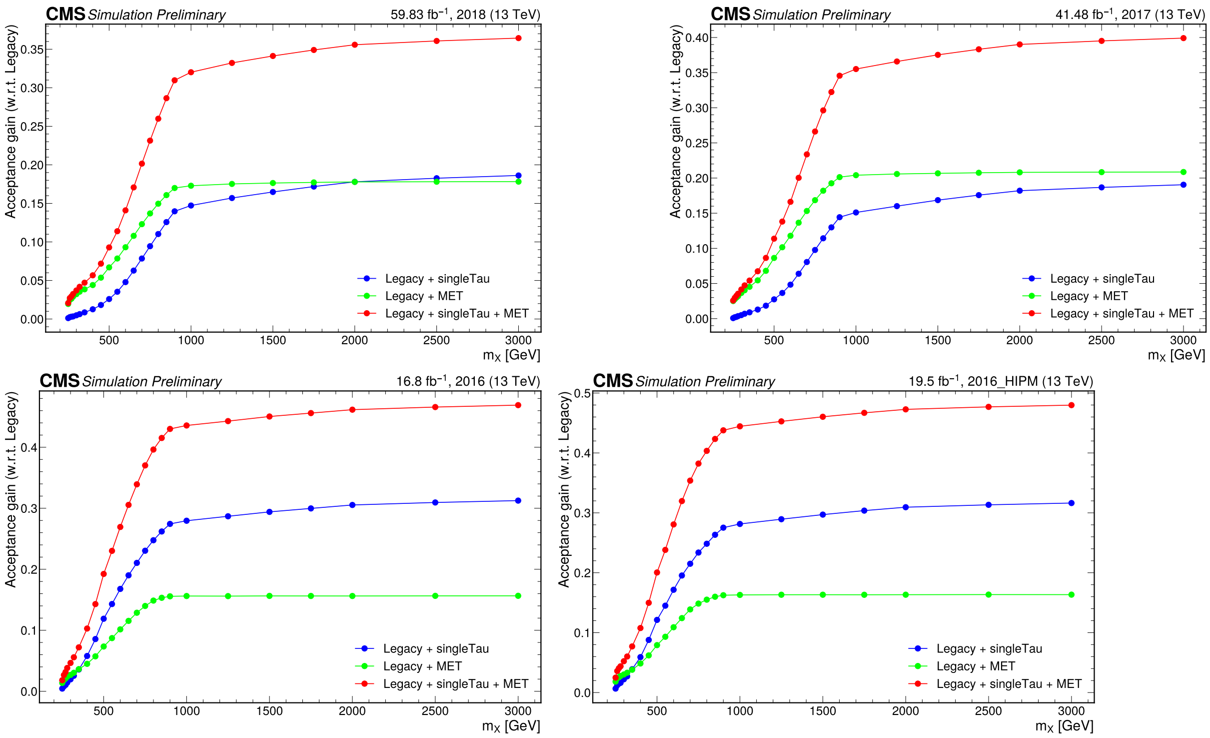


Figure C.1: Signal efficiency gain with the application of new triggers for the $e\tau_h$, channel, using 2018 (upper left), 2017 (upper right), 2016 (lower left), and 2016H (lower right) data. The study is performed on signal event samples featuring a Spin-0 (Radion) resonance. The blue line represents the efficiency when using only the Legacy trigger combined with the SingleTau trigger. The green line shows the efficiency for the OR combination of the Legacy and MET triggers. The red line corresponds to the OR combination of the Legacy, MET, and SingleTau triggers. In all cases, triggers are considered only if fired within their respective application regions.

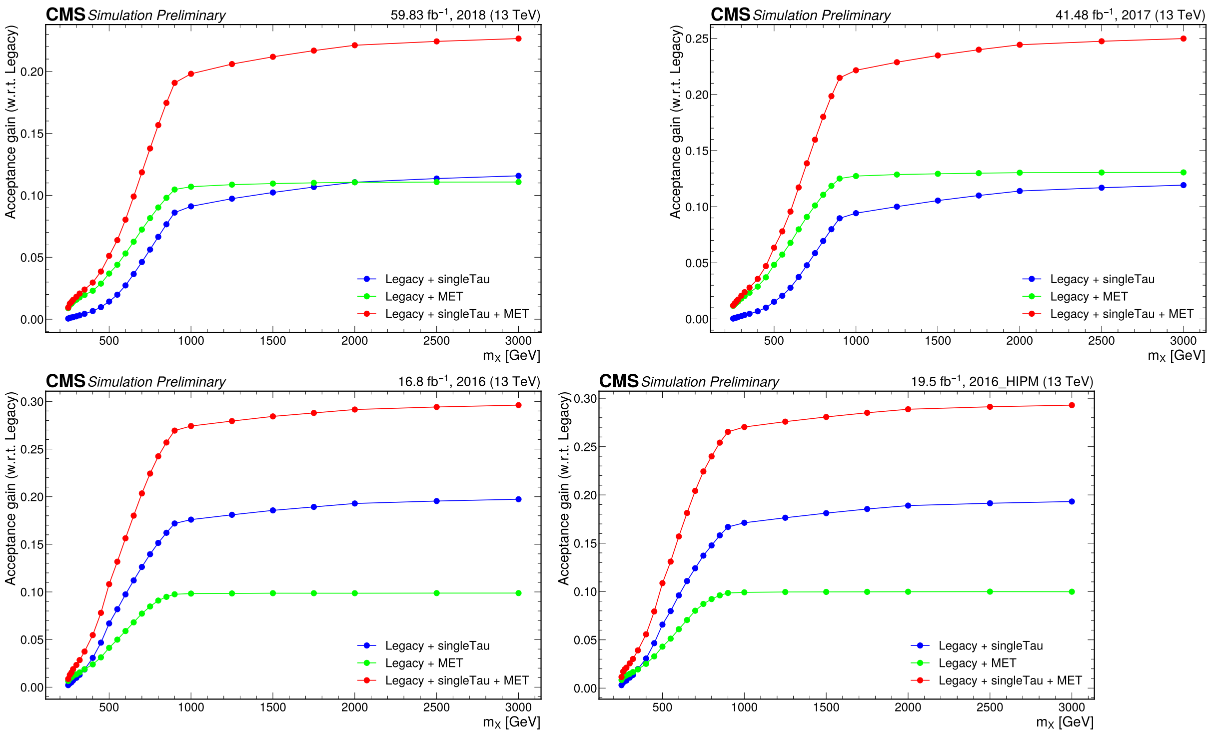


Figure C.2: Signal efficiency gain with the application of new triggers for the $\mu\tau_h$, channel, using 2018 (upper left), 2017 (upper right), 2016 (lower left), and 2016H (lower right) data. The study is performed on signal event samples featuring a Spin-0 (Radion) resonance. The blue line represents the efficiency when using only the Legacy trigger combined with the SingleTau trigger. The green line shows the efficiency for the OR combination of the Legacy and MET triggers. The red line corresponds to the OR combination of the Legacy, MET, and SingleTau triggers. In all cases, triggers are considered only if fired within their respective application regions.

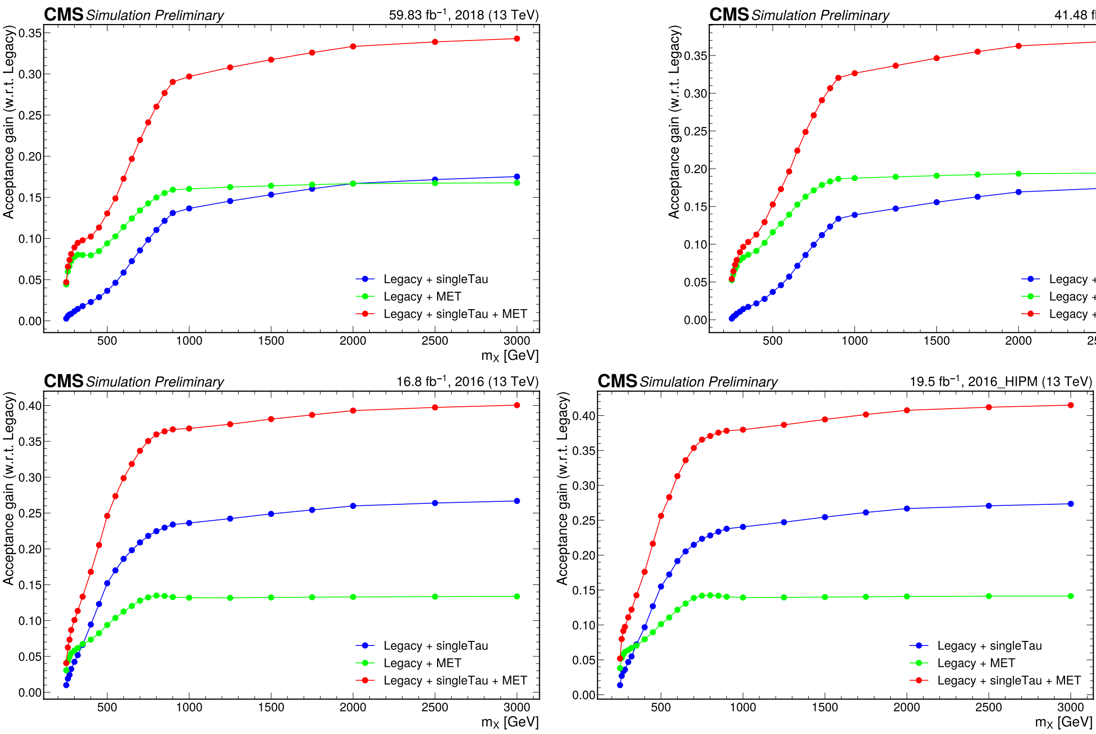


Figure C.3: Signal efficiency gain with the application of new triggers for the $\tau_h\tau_h$, channel, using 2018 (upper left), 2017 (upper right), 2016 (lower left), and 2016H (lower right) data. The study is performed on signal event samples featuring a Spin-0 (Radion) resonance. The blue line represents the efficiency when using only the Legacy trigger combined with the SingleTau trigger. The green line shows the efficiency for the OR combination of the Legacy and MET triggers. The red line corresponds to the OR combination of the Legacy, MET, and SingleTau triggers. In all cases, triggers are considered only if fired within their respective application regions.

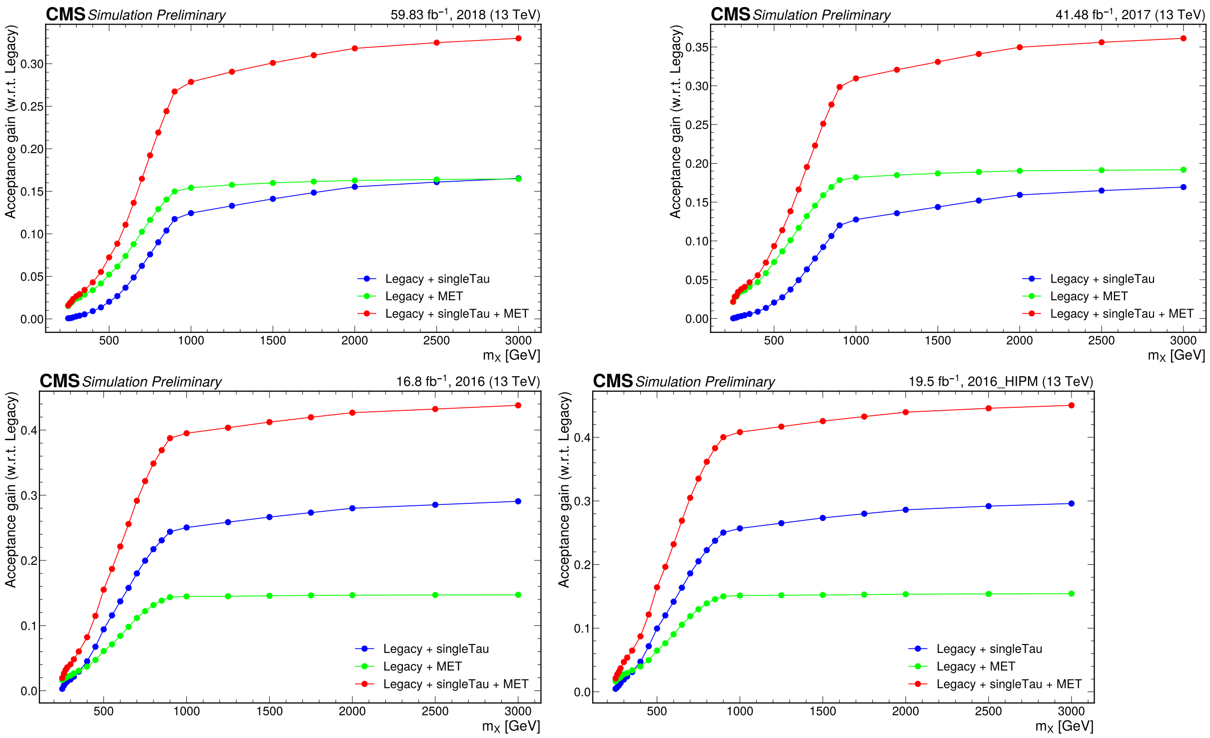


Figure C.4: Signal efficiency gain with the application of new triggers for the $e\tau_h$, channel, using 2018 (upper left), 2017 (upper right), 2016 (lower left), and 2016H (lower right) data. The study is performed on signal event samples featuring a Spin-0 (Radion) resonance. The blue line represents the efficiency when using only the Legacy trigger combined with the SingleTau trigger. The green line shows the efficiency for the OR combination of the Legacy and MET triggers. The red line corresponds to the OR combination of the Legacy, MET, and SingleTau triggers. In all cases, triggers are considered only if fired within their respective application regions.

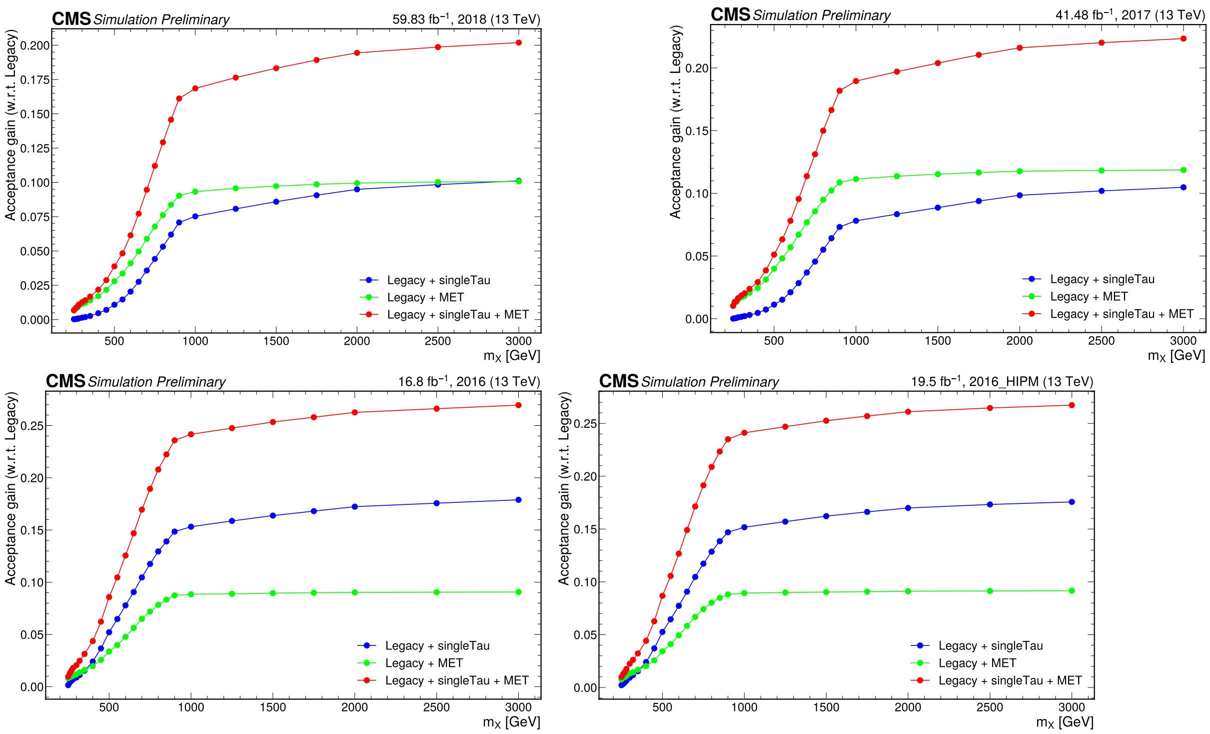


Figure C.5: Signal efficiency gain with the application of new triggers for the $\mu\tau_h$, channel, using 2018 (upper left), 2017 (upper right), 2016 (lower left), and 2016H (lower right) data. The study is performed on signal event samples featuring a Spin-0 (Radion) resonance. The blue line represents the efficiency when using only the Legacy trigger combined with the SingleTau trigger. The green line shows the efficiency for the OR combination of the Legacy and MET triggers. The red line corresponds to the OR combination of the Legacy, MET, and SingleTau triggers. In all cases, triggers are considered only if fired within their respective application regions.

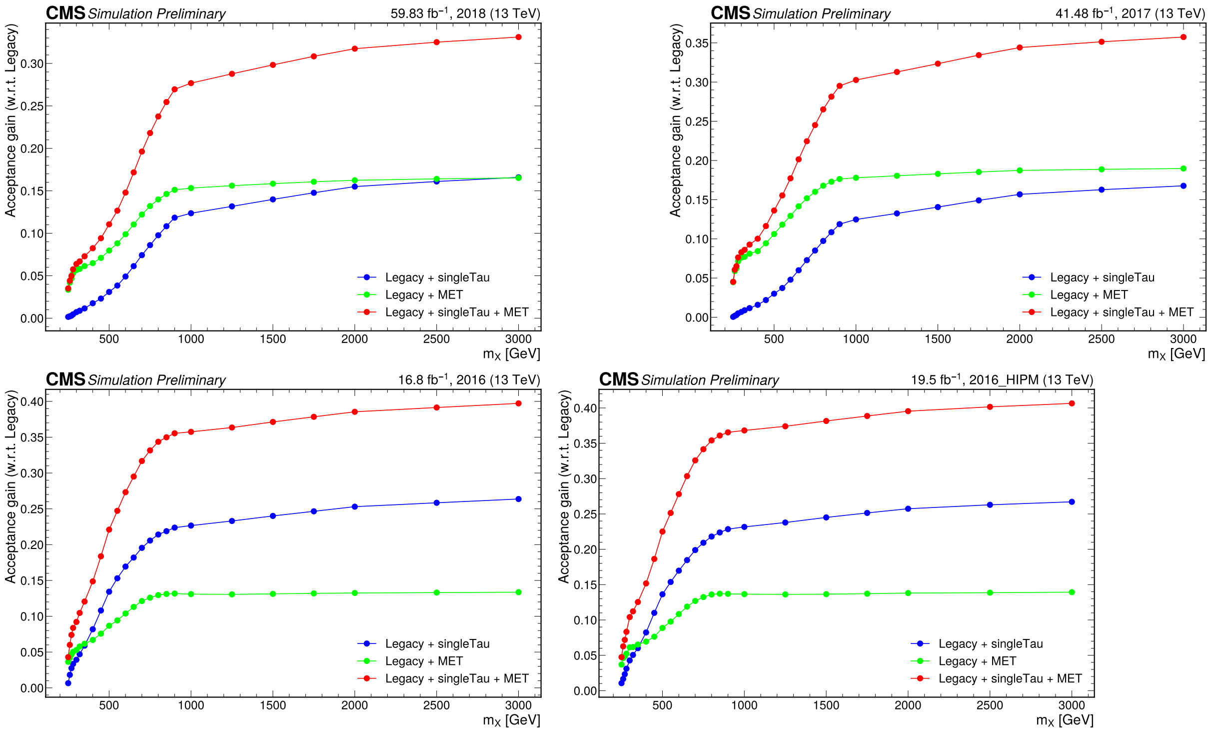


Figure C.6: Signal efficiency gain with the application of new triggers for the $\tau_h\tau_h$, channel, using 2018 (upper left), 2017 (upper right), 2016 (lower left), and 2016H (lower right) data. The study is performed on signal event samples featuring a Spin-0 (Radion) resonance. The blue line represents the efficiency when using only the Legacy trigger combined with the SingleTau trigger. The green line shows the efficiency for the OR combination of the Legacy and MET triggers. The red line corresponds to the OR combination of the Legacy, MET, and SingleTau triggers. In all cases, triggers are considered only if fired within their respective application regions.

Appendix D

Di-tau Mass Reconstruction – SVFit Algorithm

A key characteristic of the $H \to \tau\tau$ signal is the invariant mass of the $\mathrm{H} \to \tau^+\tau^-$ candidate ($\mathrm{m}_{\tau\tau}$). Thus, its accurate reconstruction is crucial for distinguishing the $H \to \tau\tau$ process from many other processes, especially the irreducible background $Z \to \tau\tau$ plus jets ("DY+jets").
However, due to the presence of neutrinos from tau decays in the final state, the visible mass of the tau pair ($\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$) has limited resolution. This makes it difficult to effectively separate the $H \to \tau\tau$ signal from background processes, particularly $Z \to \tau\tau$, since the Z boson mass ($m_Z \approx 91.2\,\mathrm{GeV}$) is close Higgs boson one ($m_H \approx 125\,\mathrm{GeV}$).

To improve mass reconstruction accuracy, kinematic parameters reconstructed from visible tau decay products are combined with missing transverse energy ($\mathrm{p}_\mathrm{T}^\mathrm{miss}$), used to account for neutrino contributions via the *SVFit* algorithm, based on the dynamic likelihood method [161]. *SVfit* has been previously exploited in Compact Muon Solenoid (CMS) analyses for SM and MSSM Higgs searches [202, 203], proving the effective benefits of using $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$ in place of $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$.

It provides event-by-event estimates of the invariant mass of $\mathrm{H} \to \tau^+\tau^-$ ($\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$). In addition to the observed momentum, the tau decay is parametrized in the detector coordinate system by considering the fraction of the tau energy carried by its visible decay products (X) and the angle between the original tau direction and the visible tau candidate direction. To fully describe the tau decay, the invariant mass of the two-neutrino system ($m_{\nu\nu}$) is also used.

The kinematics of tau pair decay varies with the tau decay mode. For hadronic decays, five parameters define the decay: the tau lepton's momentum, polar and azimuthal angles in the laboratory frame, and two decay angles in the tau rest frame. In leptonic decays, two neutrinos are produced, introducing an additional parameter: the invariant mass of the neutrino system ($m_{\nu\nu}$), which is absent in hadronic decays.

The unknown parameters of the $\tau\tau$ system are constrained by three observables:

the momentum, polar, and azimuthal angles of visible tau decay products measured in the laboratory frame. This results in two (three) unconstrained parameters for hadronic (leptonic) decays. An additional constraint is applied using the x and y components of missing transverse momentum, E_x^{miss} and E_y^{miss}. The remaining unconstrained parameters are:

- x: the fraction of tau lepton energy (in the lab. frame) carried by visible decay products,
- ϕ: the angle between the tau lepton momentum and the momentum of visible decay products,
- $m_{\nu\nu}$: the mass of the neutrino system (for leptonic tau decays).

Given x, ϕ, and $m_{\nu\nu}$, the tau lepton's energy and momentum in the laboratory frame are fully determined [161]. The values of $M_{\tau\tau}$ are reconstructed by combining the measured observables E_x^{miss} and E_y^{miss} with a probability model that accounts for tau decay kinematics and E_T^{miss} resolution. The probability density function for the observed missing transverse energy components, $p(\vec{x}|\vec{y},\vec{a})$, is expressed as:

$$P(M_i^{\tau\tau}) = \int \delta(M_i^{\tau\tau} - M_{\tau\tau}(\vec{y},\vec{a}))p(\vec{x}|\vec{y},\vec{a})d\vec{a}, \quad \text{(D.1)}$$

where $\vec{x} = (E_x^{\text{miss}}, E_y^{\text{miss}})$, $\vec{y} = (p_{\text{vis,1}}, p_{\text{vis,2}})$ represents the observed visible momenta, and $\vec{a} = (x_1, \phi_1, m_{\nu\nu,1}, x_2, \phi_2, m_{\nu\nu,2})$ are the unknown kinematic parameters. The best estimate $\hat{M}_{\tau\tau}$ is obtained by maximizing $P(M_i^{\tau\tau})$.

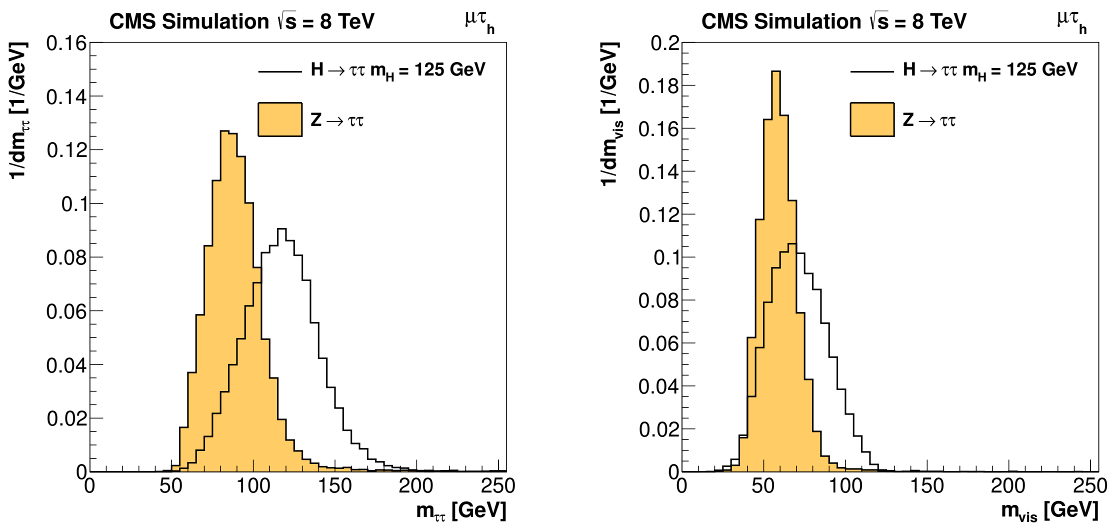


Figure D.1: Distributions of reconstructed $m_{\tau\tau}$by the SVFit algorithm (left) and of the mass of all visible tau decay products (right), for $H \to \tau^+\tau^-$ events, with mass $m_H =$ 125 GeV and for the irreducible $Z/\gamma^* \to \tau\tau$ background in the $\mu\tau_h$channel.

The global likelihood function consists of three components: two likelihood functions describing the decay parameters of each tau lepton and a third quantifying the compatibility of the tau decay hypothesis with the measured E_T^{miss}. This assumes that neutrinos from tau decays are the only source of missing energy. The likelihood functions used for tau decay kinematics differ for leptonic and hadronic decays. The likelihood function for leptonic decays is derived from matrix elements [204]:

$$L_{\text{lep}} \propto \frac{m_{\nu\nu}(m_\tau^2 + 2m_{\nu\nu}^2)(m_\tau^2 - m_{\nu\nu}^2)}{4m_\tau^2}, \tag{D.2}$$

within the constraints $0 \leq x \leq 1$ and $0 \leq m_{\nu\nu} \leq m_\tau\sqrt{1-x}$. For hadronic decays, a two-body phase-space model is assumed:

$$L_{\text{had}} \propto \frac{1}{2\pi}\left(1 - \frac{m_{\text{vis}}^2}{m_\tau^2}\right), \tag{D.3}$$

valid within $m_{\text{vis}}^2/m_\tau^2 \leq x \leq 1$.

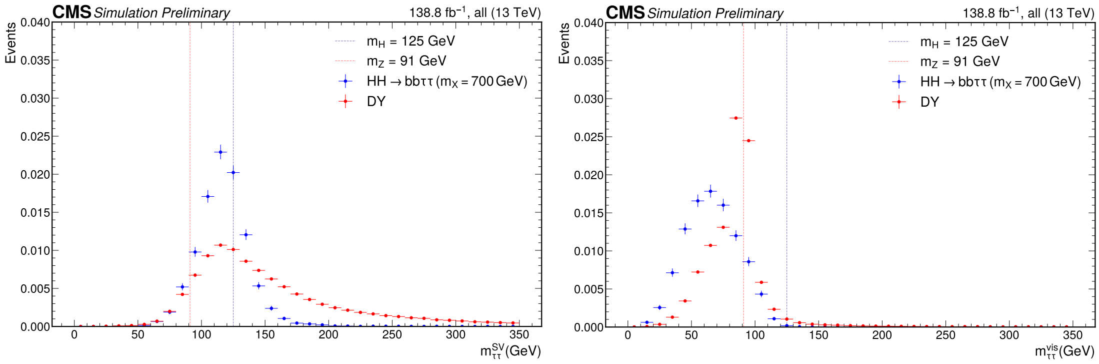
The likelihood function for E_T^{miss} is given by:

$$L_{E_T^{\text{miss}}} = \frac{1}{2\pi\sqrt{|V|}} \exp\left(-\frac{1}{2}\begin{bmatrix} E_x^{\text{miss}} - \sum p_{\nu x} \\ E_y^{\text{miss}} - \sum p_{\nu y}\end{bmatrix}^T V^{-1} \begin{bmatrix} E_x^{\text{miss}} - \sum p_{\nu x} \\ E_y^{\text{miss}} - \sum p_{\nu y}\end{bmatrix}\right), \tag{D.4}$$

where V is the covariance matrix accounting for E_T^{miss} resolution (Section 3.3.6), and $|V|$ is the determinant of V. Differences between summed neutrino momenta ($\sum p_{\nu x(y)}$) and E_T^{miss} components ($E_{x(y)}^{\text{miss}}$) could arise due to resolution effects and are included in the probability model assuming a Gaussian resolution. Figure D.1 shows that the di-tau lepton mass reconstructed with SVFit algorithm improves the shape separation. Moreover, in the context of the Standard Model (SM) H $\rightarrow \tau^+\tau^-$ studies, the sensitivity achieved by the CMS analysis using $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$ increased by about 30%, compared to performing the same analysis using $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ [202].

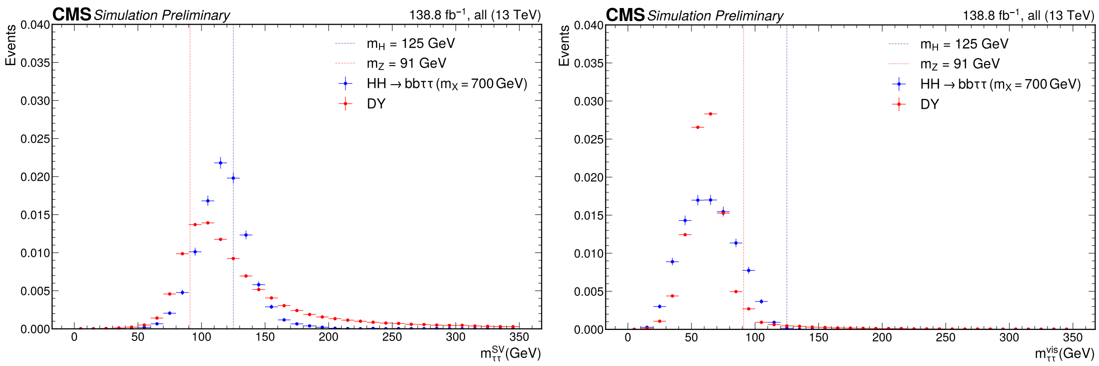
D.0.1 The SVfit algorithm in the X $\rightarrow$ HH $\rightarrow$ b$\bar{\text{b}}\tau\tau$ analysis

In this section, the H $\rightarrow \tau^+\tau^-$ candidate pair distributions of the invariant visible mass, $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ and the invariant mass reconstructed by *SVFit*, $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$, are shown. The distributions are presented for the $e\tau_h$ and $\mu\tau_h$ channels considering hadronic tau selection with the *DeepTau2017v2p1* discriminator. Figures D.1, D.2, D.3, and D.4 show that SVFit significantly improves the shape separation between signal and background distributions across all channels ($e\tau$, $\mu\tau$, and $\tau\tau$). The same distributions for $\tau_h\tau_h$ channel are displayed in Fig. D.4, but in the following, the $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ VS $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$ comparisons for $\tau_h\tau_h$ are reported considering the *DeepTau2018v2p5* version (see Appendix H). The application of SVFit in the X $\rightarrow$ HH $\rightarrow$ b$\bar{\text{b}}\tau\tau$ analysis is further illustrated with comparisons of $\mathrm{m}_{\tau\tau}^{(\mathrm{vis})}$ and $\mathrm{m}_{\tau\tau}^{(\mathrm{SVFit})}$ using the *DeepTau2017v2p1* and *DeepTau2018v2p5* discriminators. These results confirm the enhanced discrimination power of SVFit, reinforcing its importance in di-tau mass reconstruction for Higgs analyses.



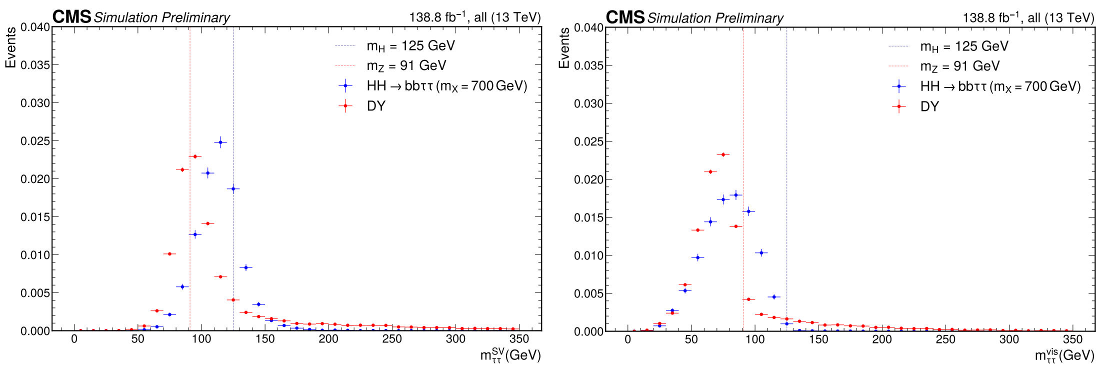
(a) **(b)**

Figure D.2: Distributions of *SVfit* invariant mass,$m_{\tau\tau}^{(\mathrm{SVFit})}$(Fig. D.2a) and visible invariant mass $m_{\tau\tau}^{(\mathrm{vis})}$(Fig. D.2b) of the τ pair candidates in the $e\tau_h$ channel, for $X \to HH \to b\bar{b}\tau\tau$ (for spin 0 and $m_X = 700\,\mathrm{GeV}$, in blue) and $Z \to \tau\tau$ + Jets ("DY", in red). Distributions are renormalized to the unit area.



(a) **(b)**

Figure D.3: Distributions of *SVfit* invariant mass,$m_{\tau\tau}^{(\mathrm{SVFit})}$(Fig. D.3a) and visible invariant mass $m_{\tau\tau}^{(\mathrm{vis})}$(Fig. D.3b) of the τ pair candidates in the $\mu\tau_h$ channel, for $X \to HH \to b\bar{b}\tau\tau$ (for spin 0 and $m_X = 700\,\mathrm{GeV}$, in blue) and $Z \to \tau\tau$ + Jets ("DY", in red). Distributions are renormalized to the unit area.



(a) **(b)**

Figure D.4: Distributions of *SVfit* invariant mass,$m_{\tau\tau}^{(\mathrm{SVFit})}$(Fig. D.4a) and visible invariant mass $m_{\tau\tau}^{(\mathrm{vis})}$(Fig. D.4b) of the τ pair candidates in the $\tau_h\tau_h$ channel, for $X \to HH \to b\bar{b}\tau\tau$ (for spin 0 and $m_X = 700\,\mathrm{GeV}$, in blue) and $Z \to \tau\tau$ + Jets ("DY", in red). Distributions are renormalized to the unit area.

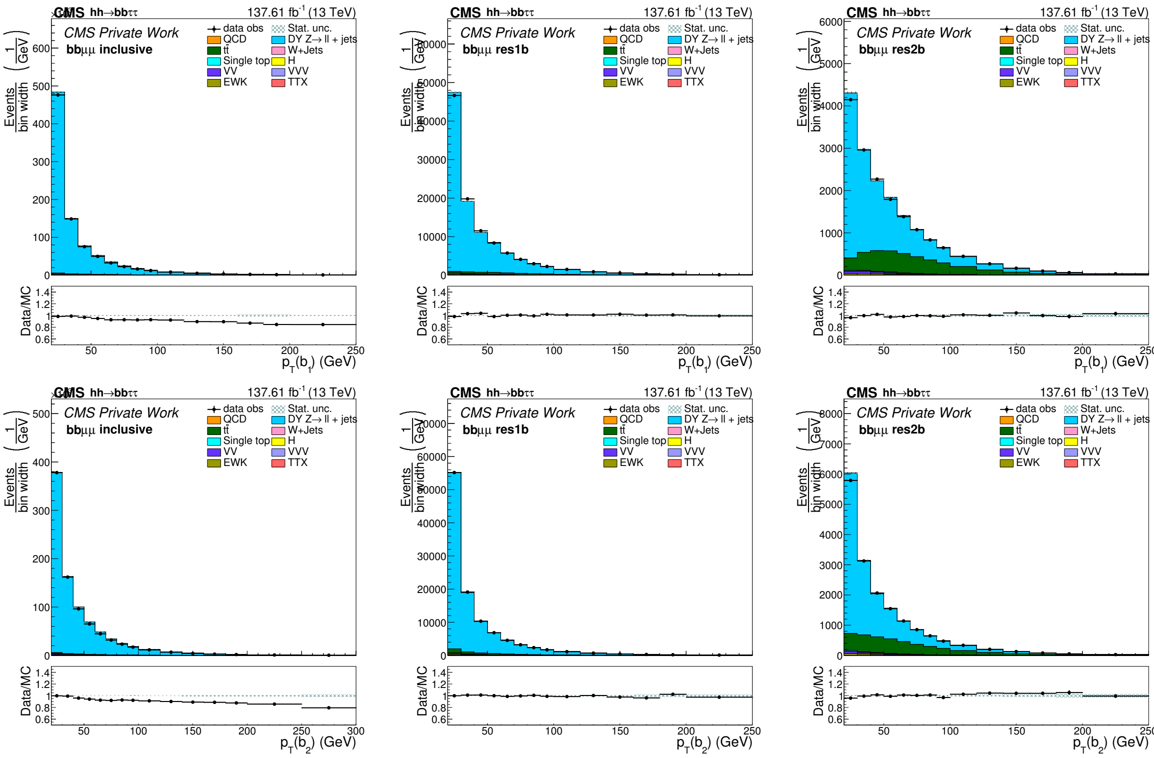
Appendix E

Control plots for DY-enriched region

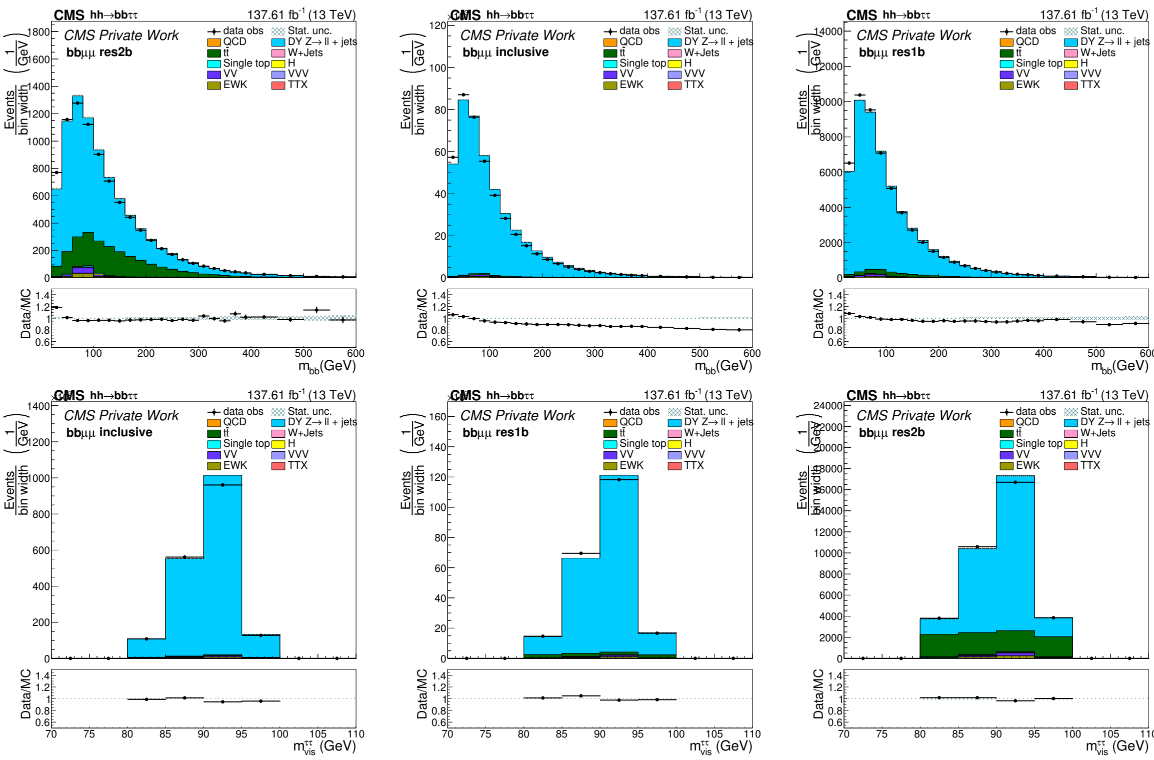
In this chapter the binned distributions obtained in the DY-enriched region are shown. The distributions are presented separately for the ee and $\mu\mu$ channels. The DYCR is defined in Section 4.8.3, considering the inclusive, Resolved 1b category (res1b) and Resolved 2b category (res2b) categories and requiring $m_{\ell\ell} \in [80, 100]$GeV , to be close to the Z peak. These are reported in Section **??**

E.1. Resolved categories

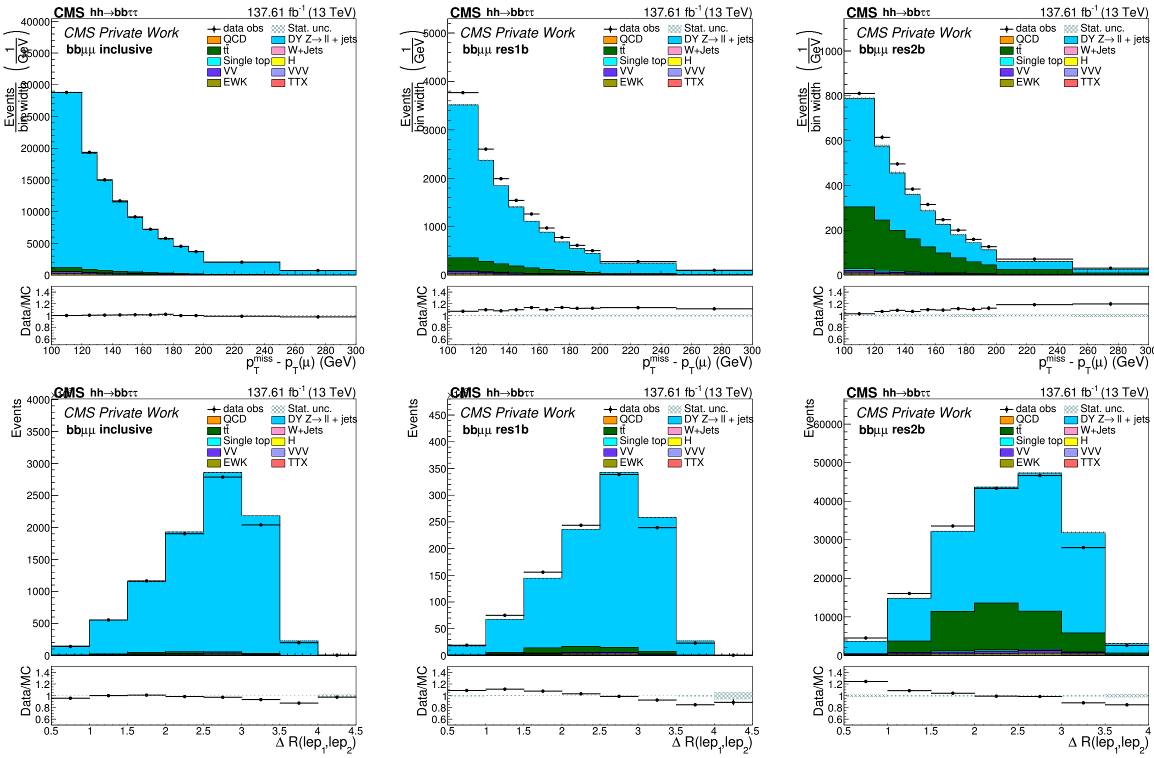
E.2. Boosted category



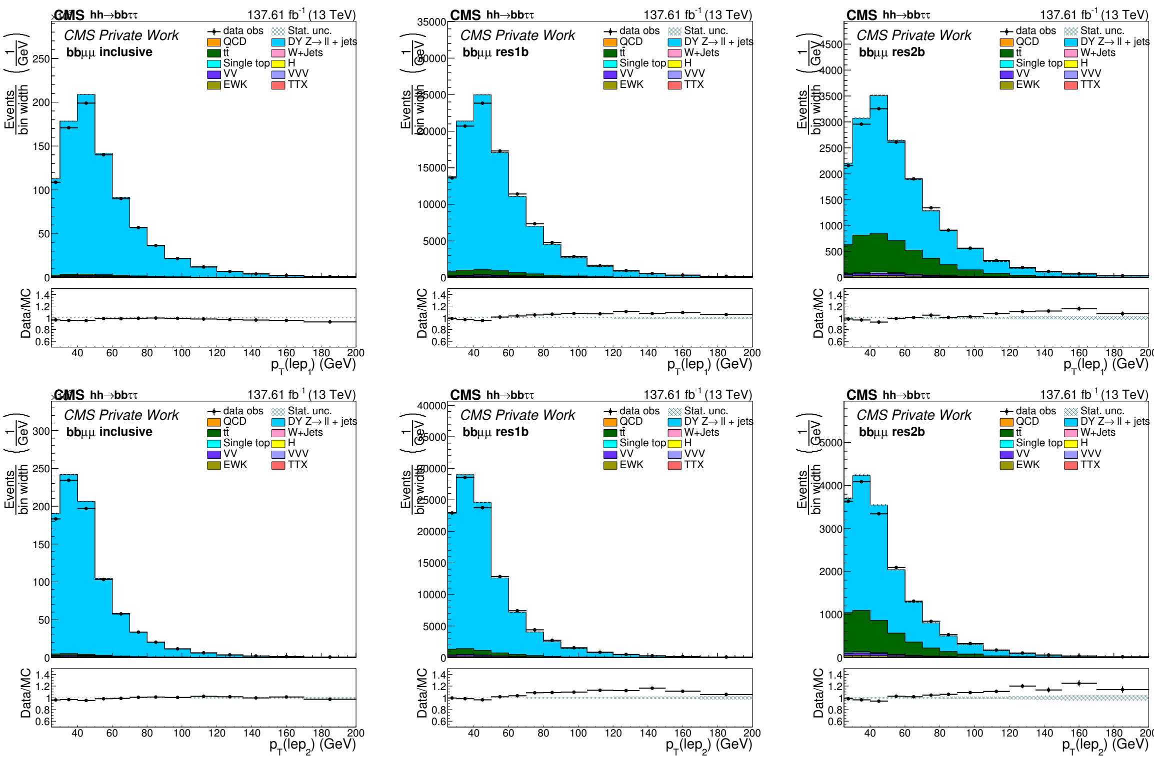
Distribution of the first and second b-jet transverse momentum component (p_T) in the inclusive, res1b and res2b categories in the Drell Yan Control Region (DYCR), for the $\mu\mu$, channel.



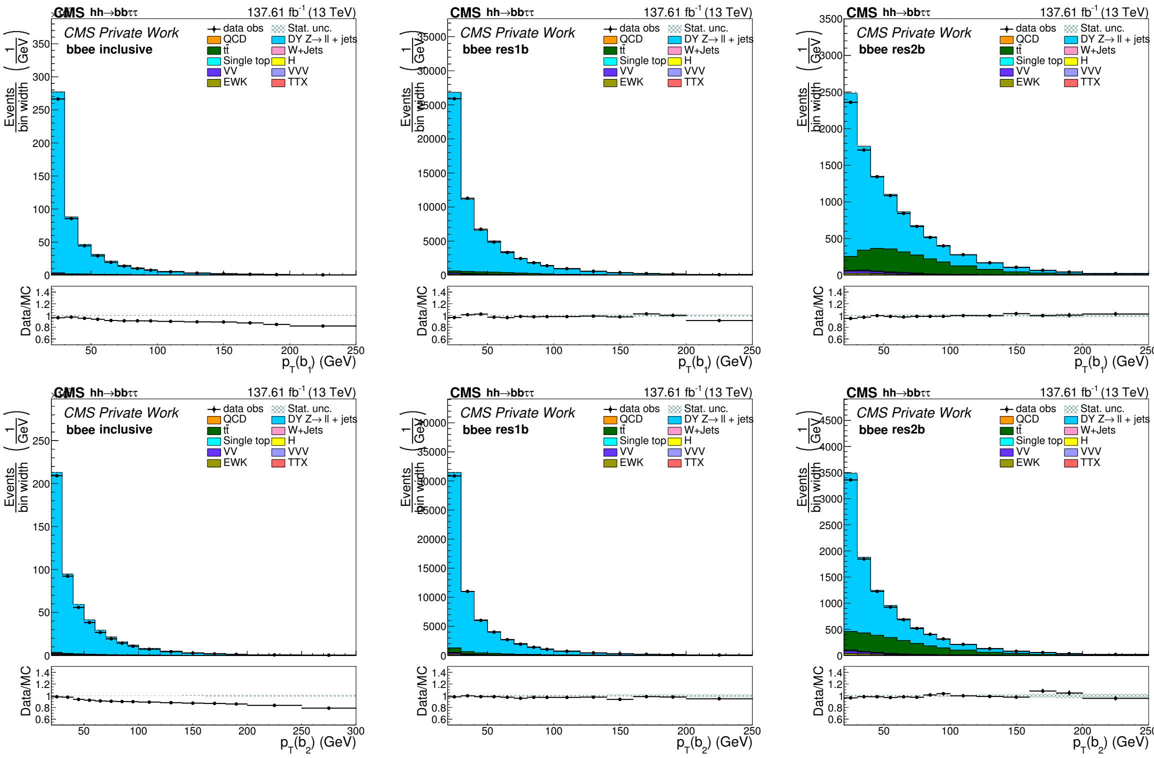
Distribution of the visible dilepton mass and invariant mass of the two b-jets in the inclusive, res1b and res2b categories in the DYCR, for the $\mu\mu$, channel.



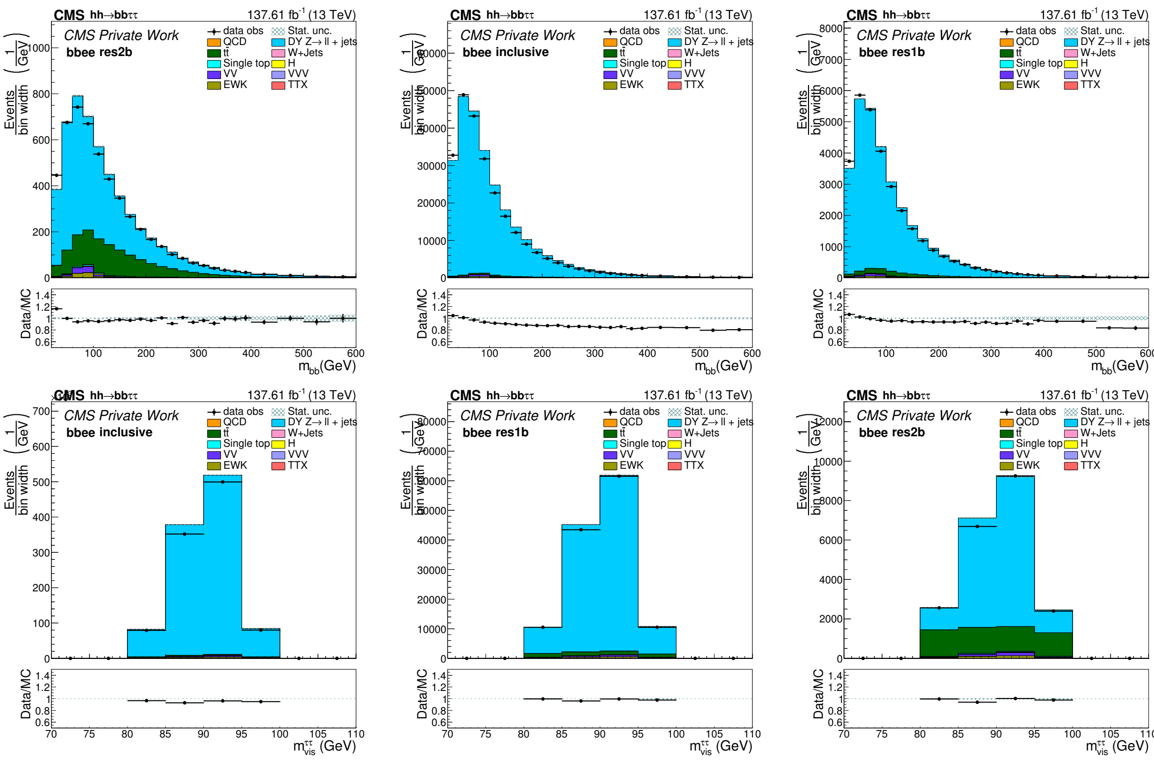
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive, res1b and res2b categories in the DYCR, for the $\mu\mu$, channel.



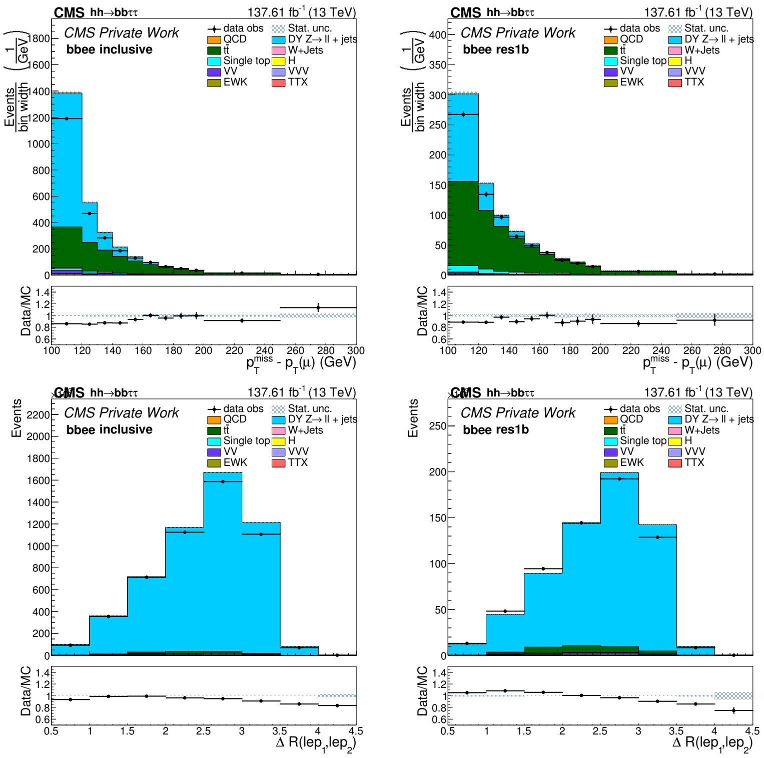
Distribution of the first and second lepton p_T, in the inclusive, res1b and res2b categories in the DYCR, for the $\mu\mu$, channel.



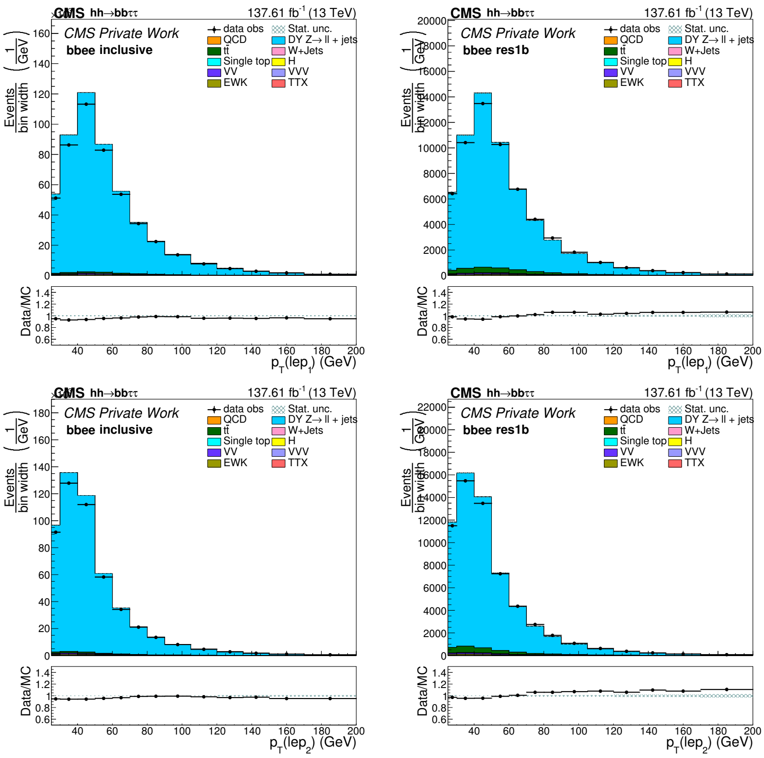
Distribution of the first and second b-jet p_T in the inclusive, res1b and res2b categories in the DYCR, for the *ee* , channel.



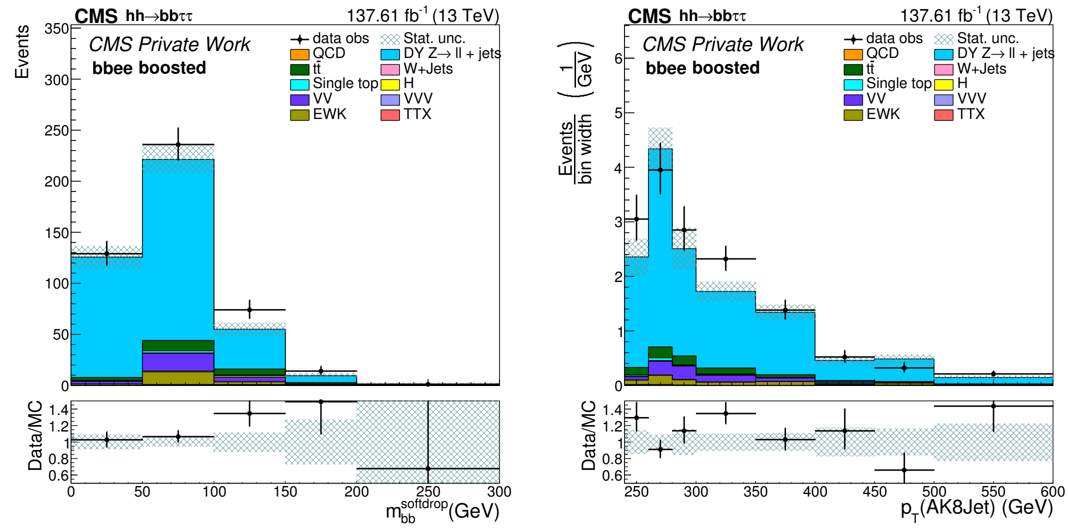
Distribution of the visible dilepton mass and invariant mass of the two b-jets in the inclusive, res1b and res2b categories in the DYCR, for the *ee* , channel.



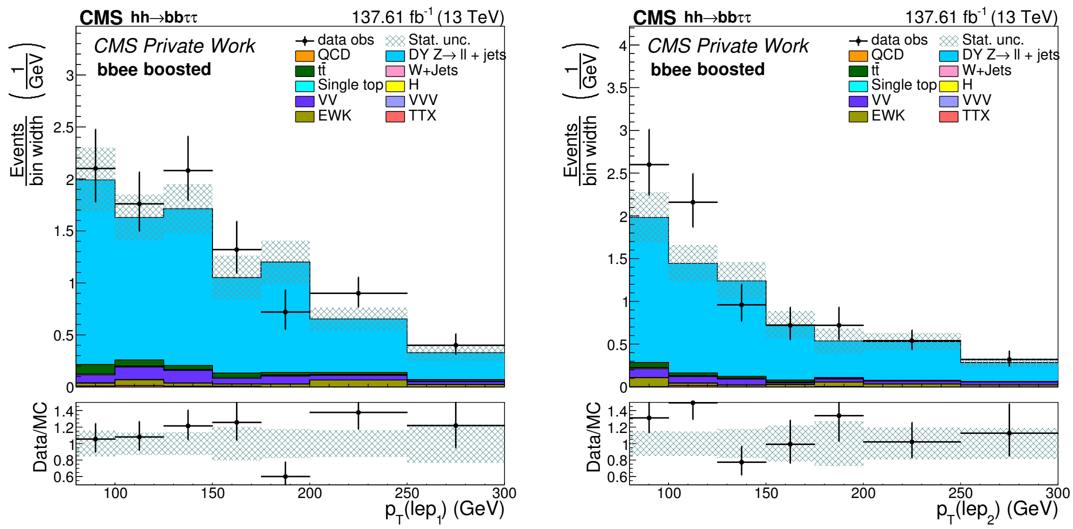
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive, res1b and res2b categories in the DYCR, for the *ee*, channel.



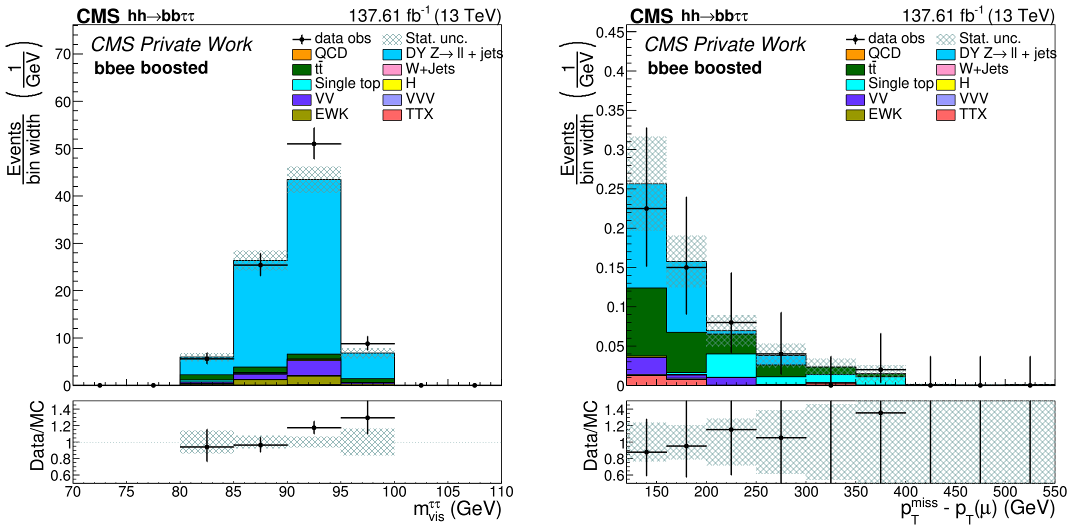
Distribution of the first and second lepton p_T, in the inclusive, res1b and res2b categories in the DYCR, for the *ee*, channel.



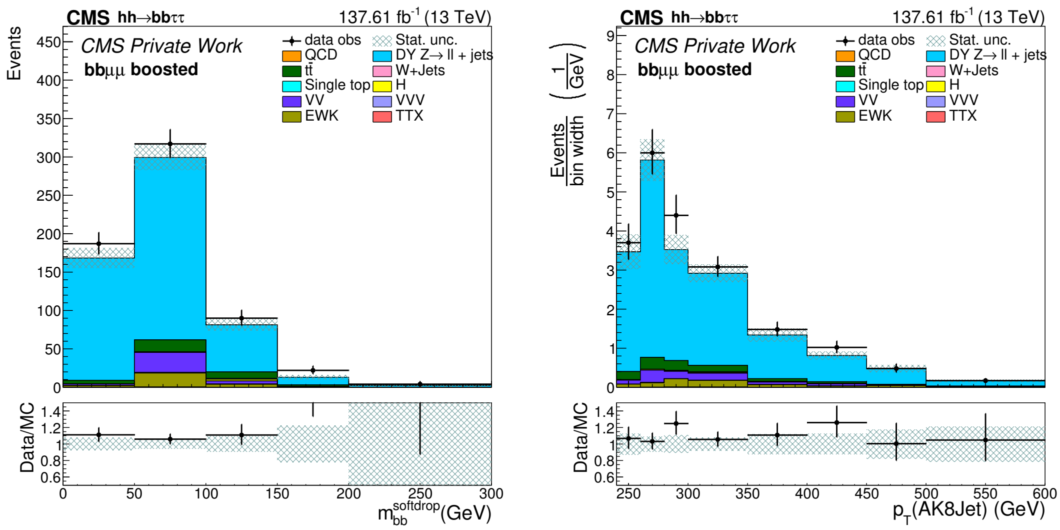
Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category in the DYCR, for the *ee* channel.



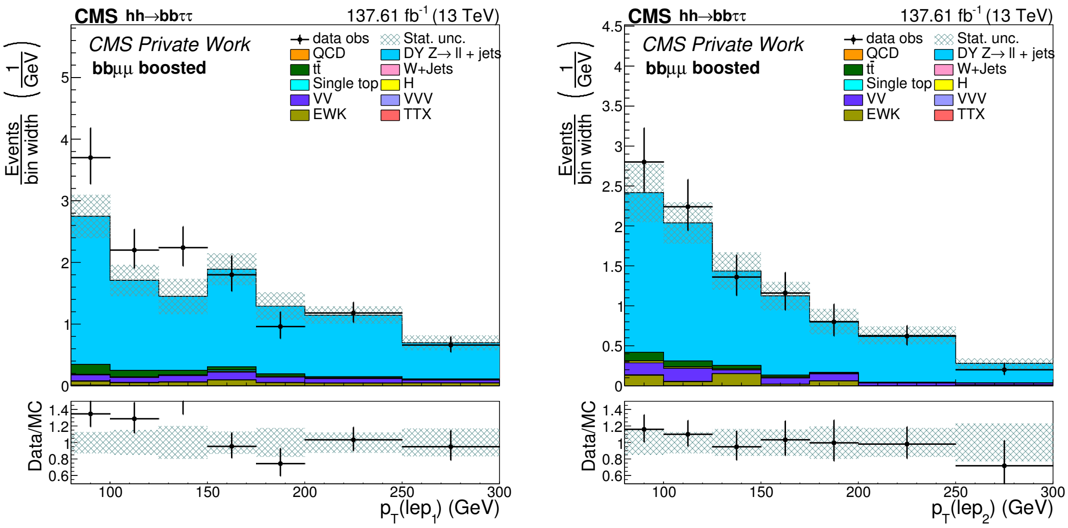
Distribution of the first and second lepton p_T in the boosted category in the DYCR, for the *ee* channel.



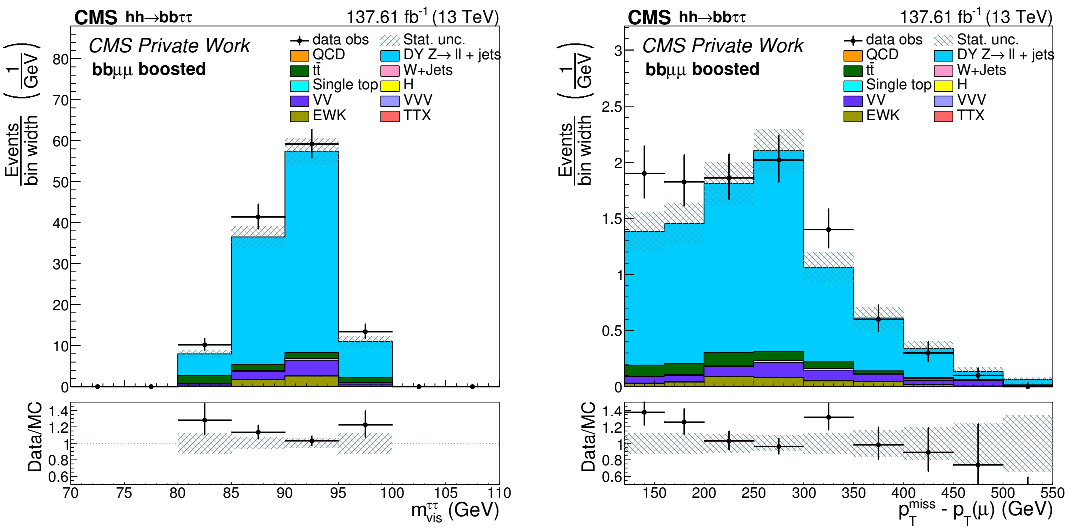
Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category in the DYCR, for the ee channel.



Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category in the DYCR, for the $\mu\mu$ channel.



Distribution of the first and second lepton p_T in the boosted category in the DYCR, for the $\mu\mu$ channel.



Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category in the DYCR, for the $\mu\mu$ channel.

Appendix F

Control plots for $t\bar{t}$-enriched region

In this chapter the binned distributions obtained in the $t\bar{t}$-enriched region are shown. The distributions are presented separately for the control channels, ee, $e\mu$ and $\mu\mu$ (Section F.1) and for the signal channels $e\tau_h$,$\mu\tau_h$ and $\tau_h\tau_h$, in Section F.2. The TTCR is defined in Section 4.8.3, considering the inclusive and Resolved 2b category (res2b) categories in Section F.1.1 and Section **??** for control and signal channels respectively. The lepton pair mass cuts are shown in Tab. 4.10. The boosted category is exploited to check the data/MC agreement with the application of the TT-like SFs for the ParticleNet-MD discriminator thus also the boosted category distributions are shown in Section F.1.2 and Section F.3.1 for control and signal channels respectively.

F.1. Control Channels

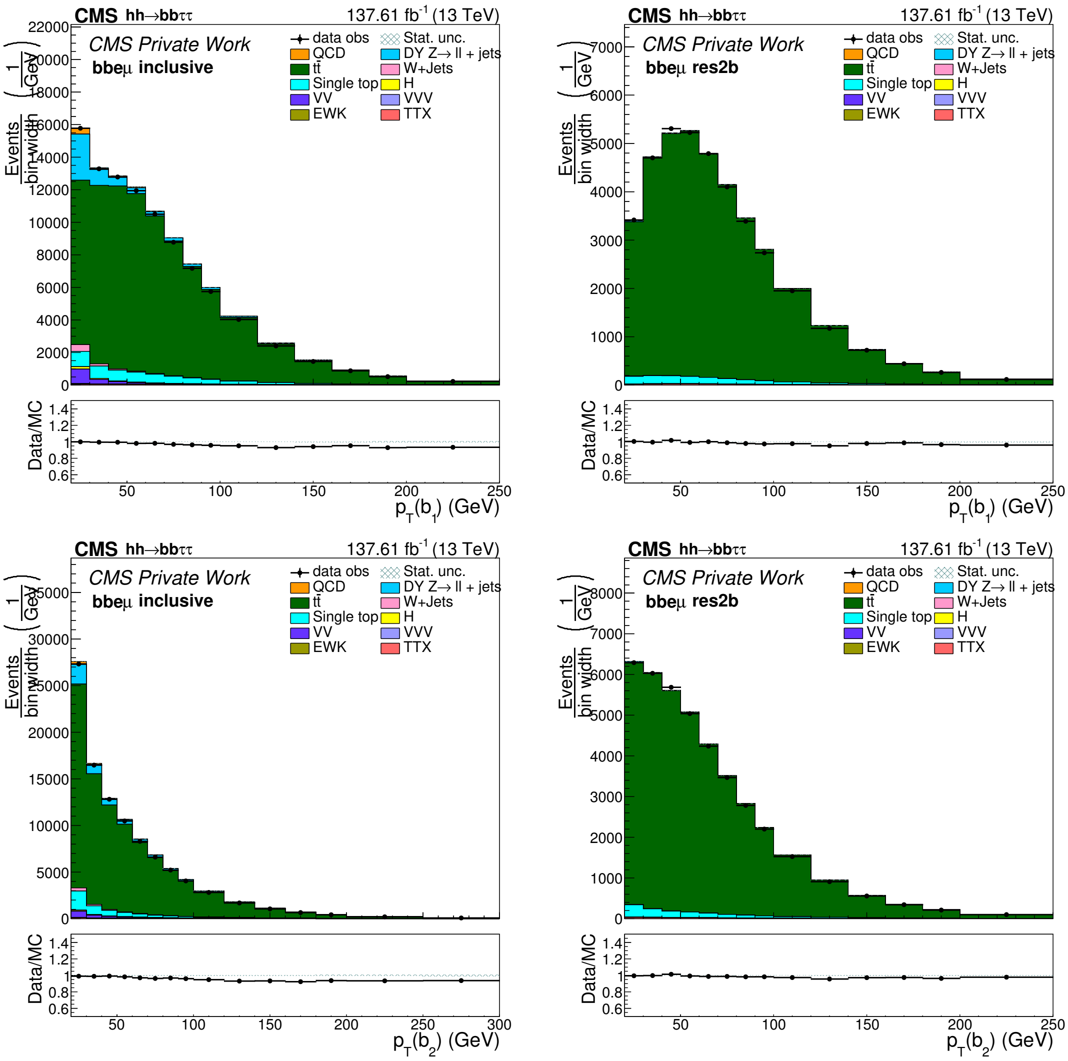
F.1.1 Resolved categories

F.1.2 Boosted category

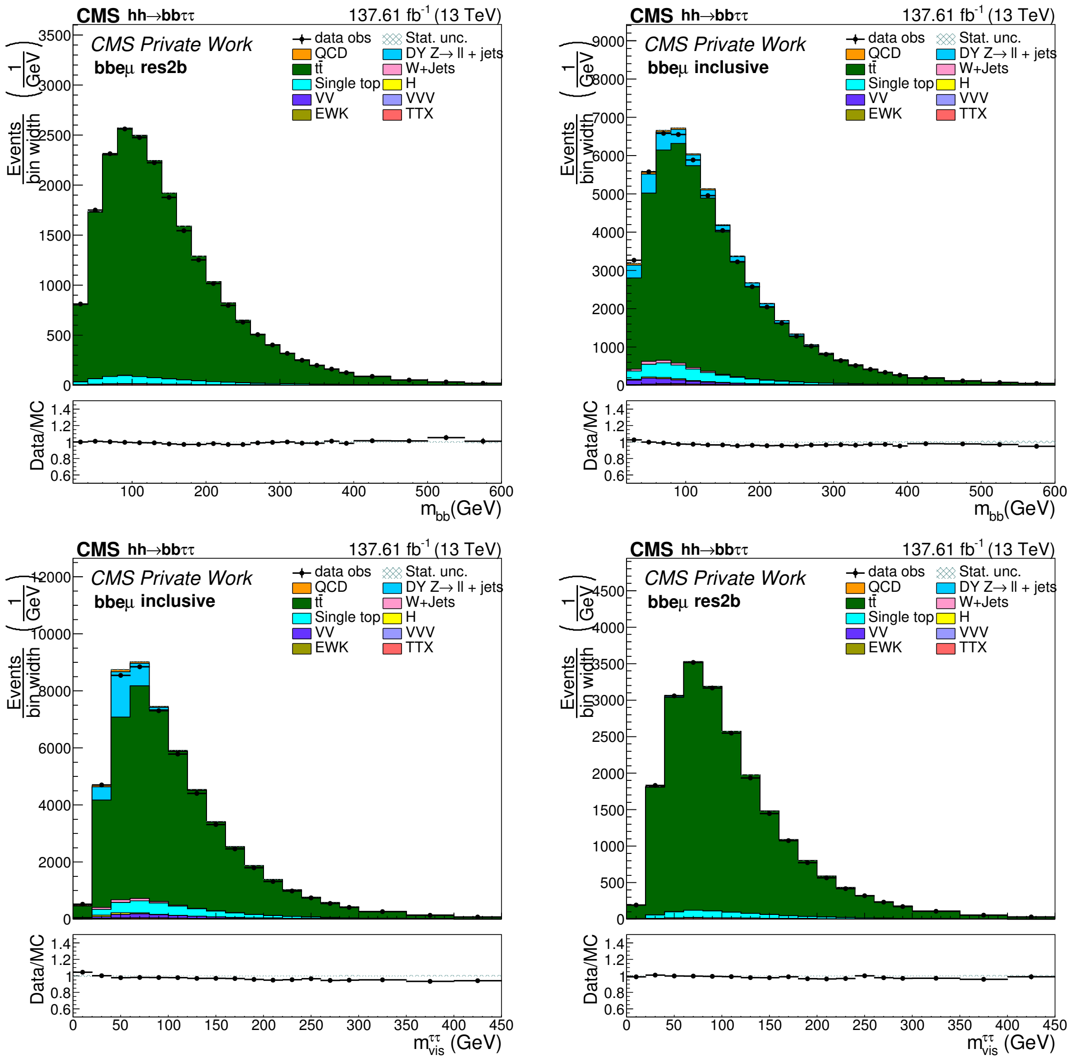
F.2. Signal channels

F.3. Resolved categories

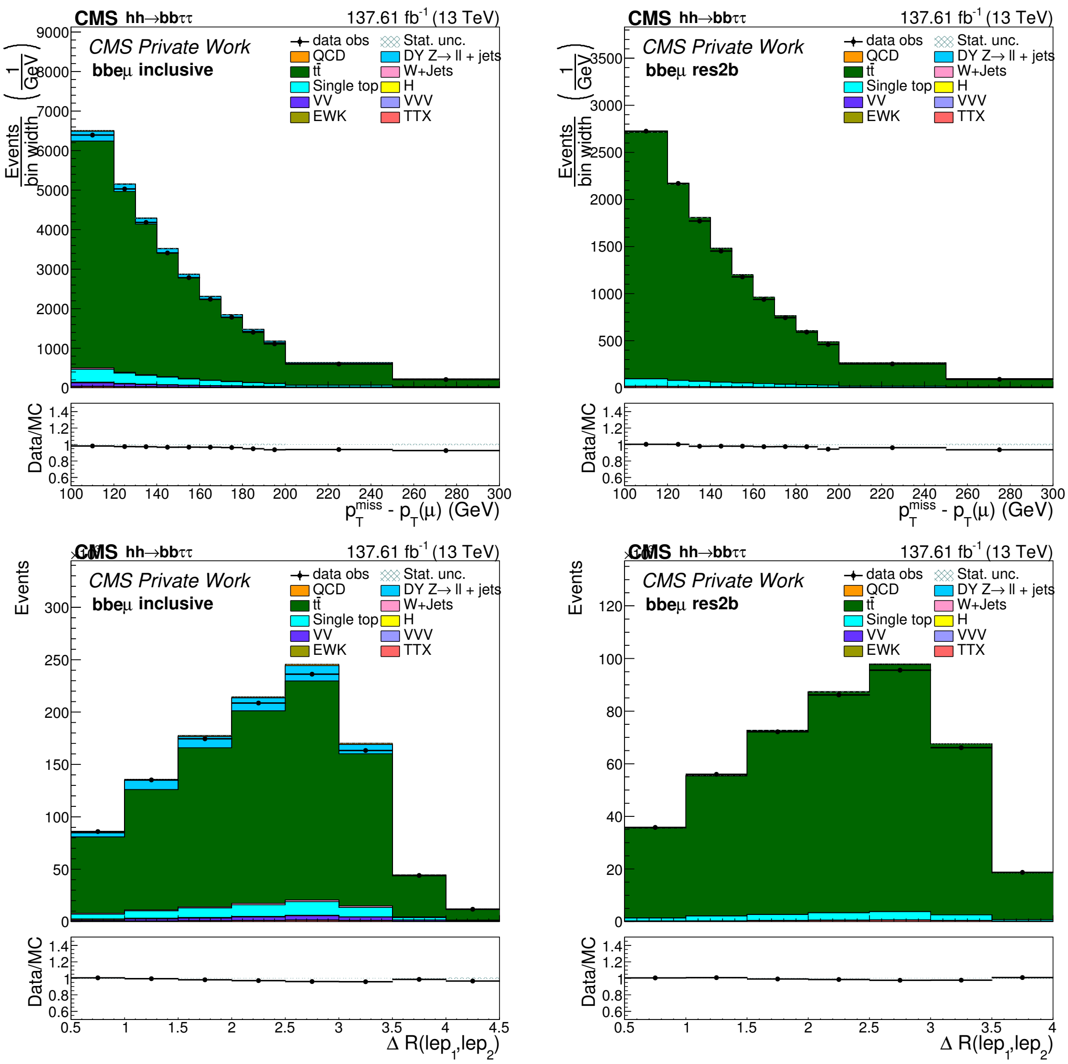
F.3.1 Boosted category



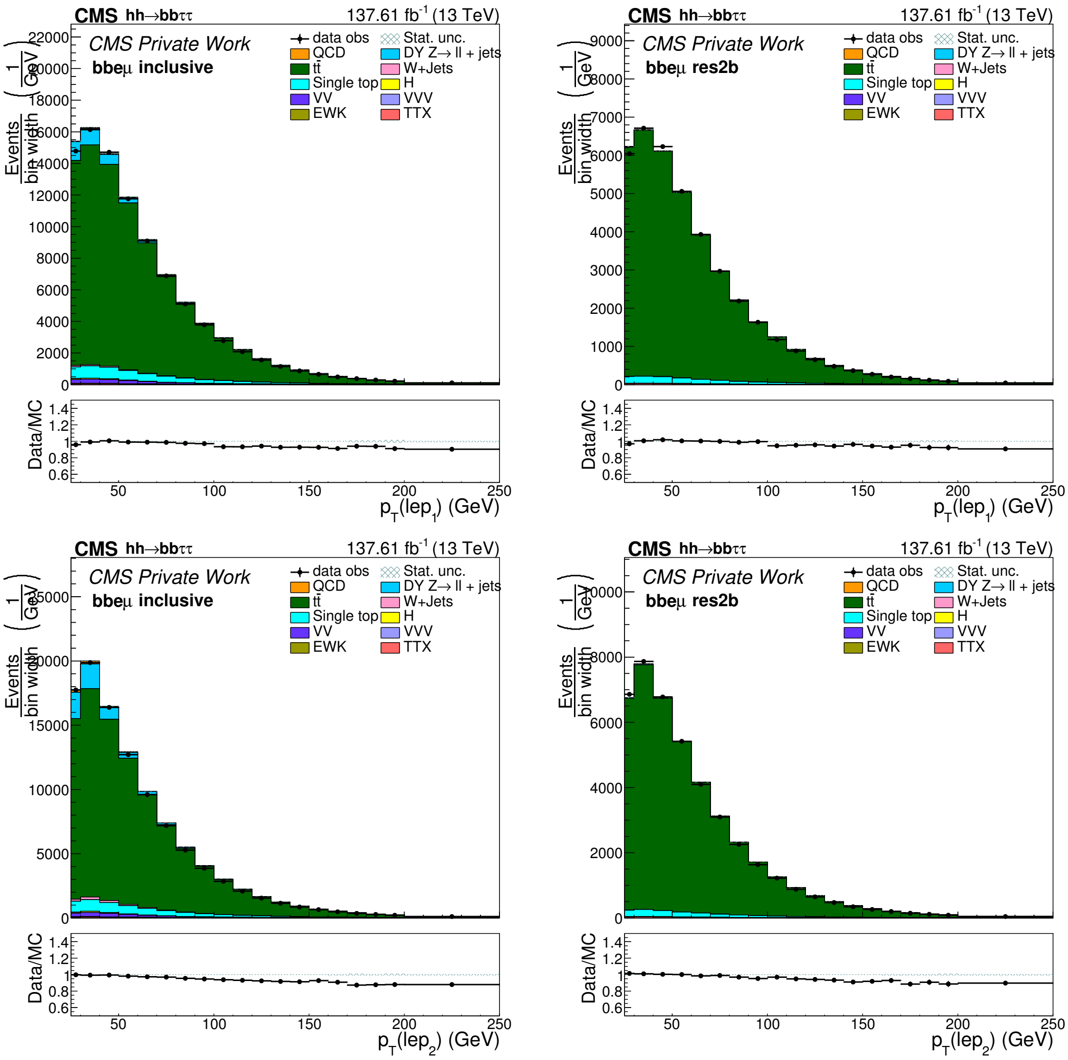
Distribution of the first and second b-jet transverse momentum component (p_T) in the inclusive and res2b categories in the $t\bar{t}$ Control Region (TTCR), for the $e\mu$ channel.



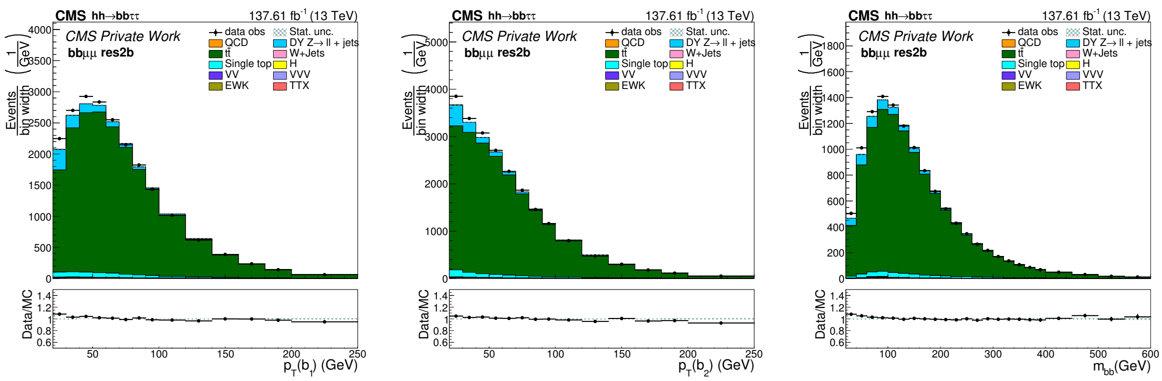
Distribution of the visible dilepton mass and invariant mass of the two b-jets in the inclusive and res2b categories in the TTCR, for the $e\mu$ channel.



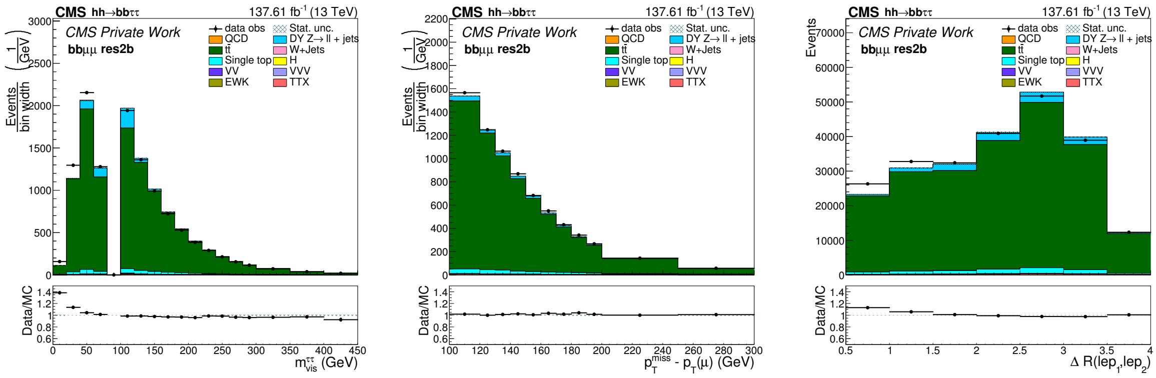
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive and res2b categories in the TTCR, for the $e\mu$ channel.



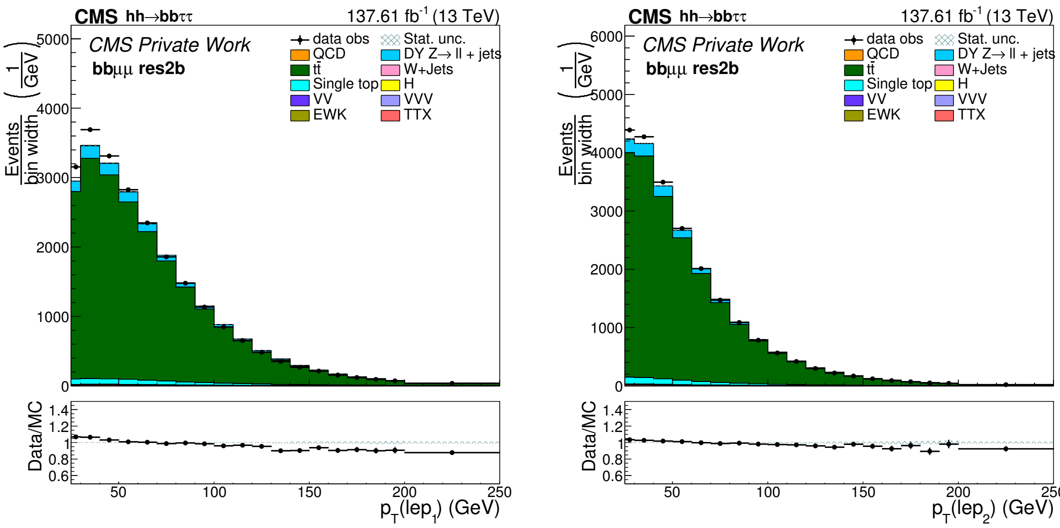
Distribution of the first and second lepton p_T , in the inclusive and res2b categories in the TTCR, for the $e\mu$ channel.



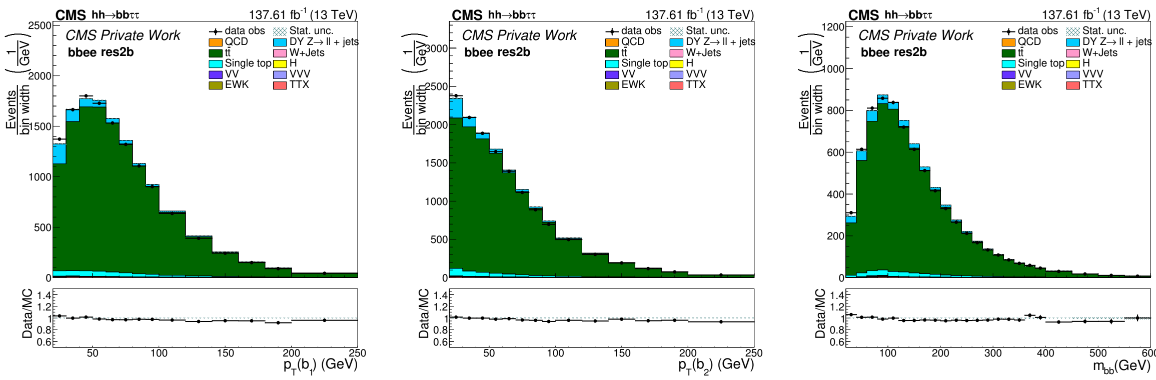
Distribution of the first and second b-jet p_T in the inclusive and res2b categories in the TTCR, for the $\mu\mu$, channel.



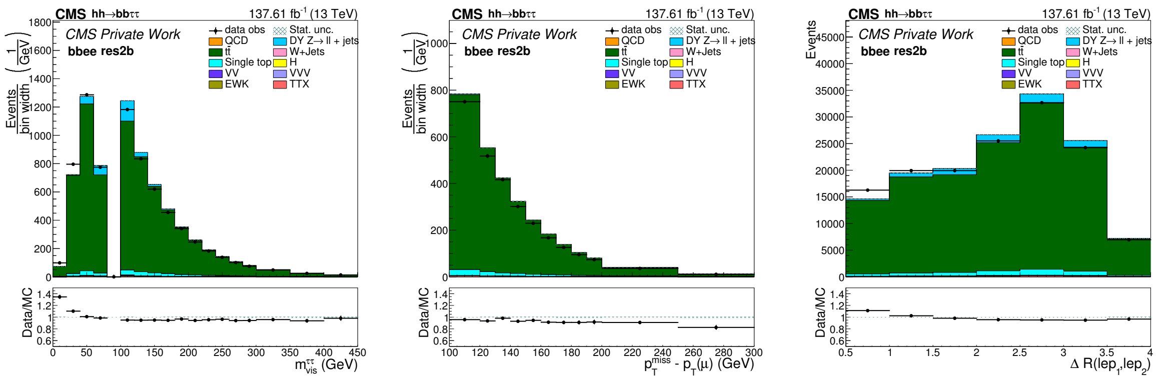
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive and res2b categories in the TTCR, for the $\mu\mu$, channel.



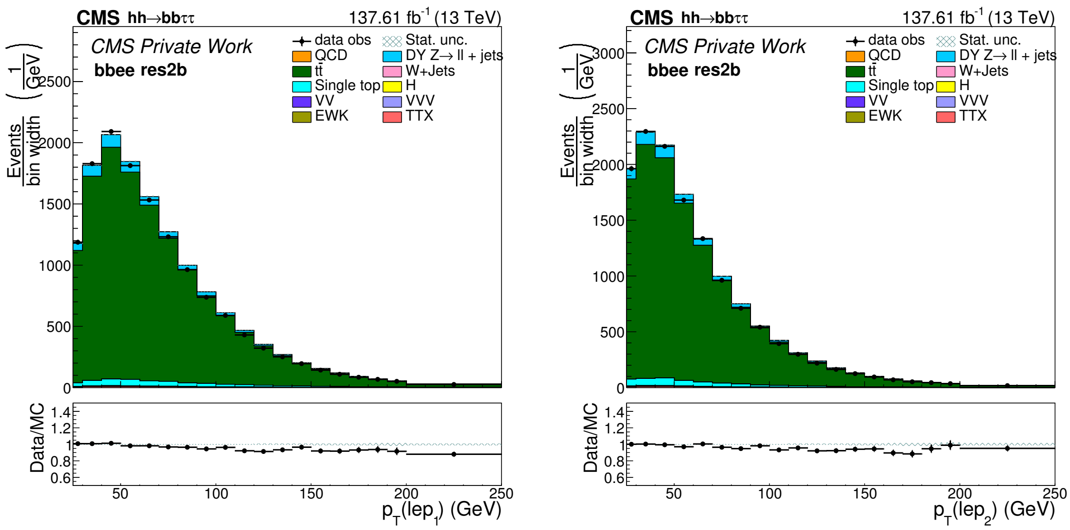
Distribution of the first and second lepton p_T, in the inclusive and res2b categories in the TTCR, for the $\mu\mu$, channel.



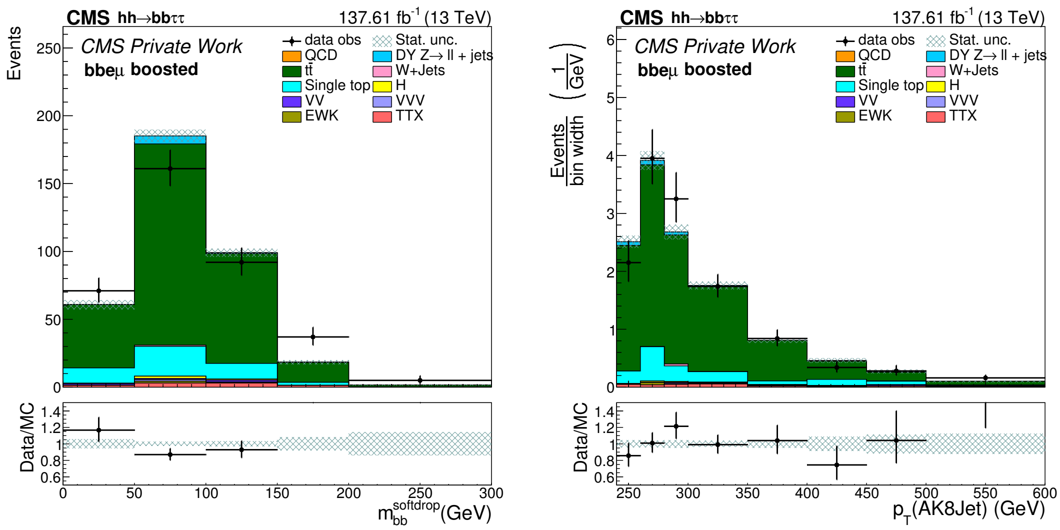
Distribution of the first and second b-jet p_T in the inclusive and res2b categories in the TTCR, for the ee channel.



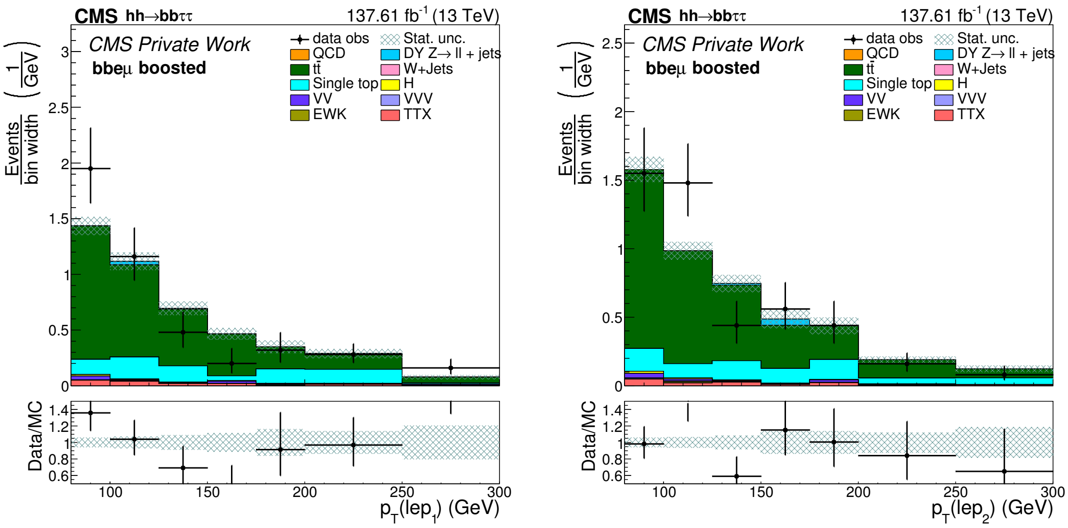
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive and res2b categories in the TTCR, for the ee channel.



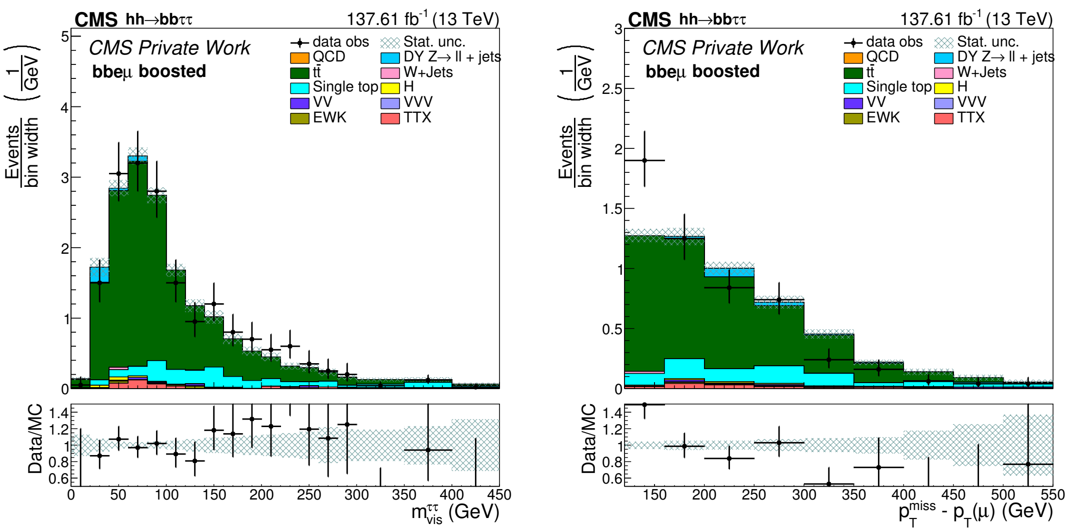
Distribution of the first and second lepton p_T, in the inclusive and res2b categories in the TTCR, for the ee channel.



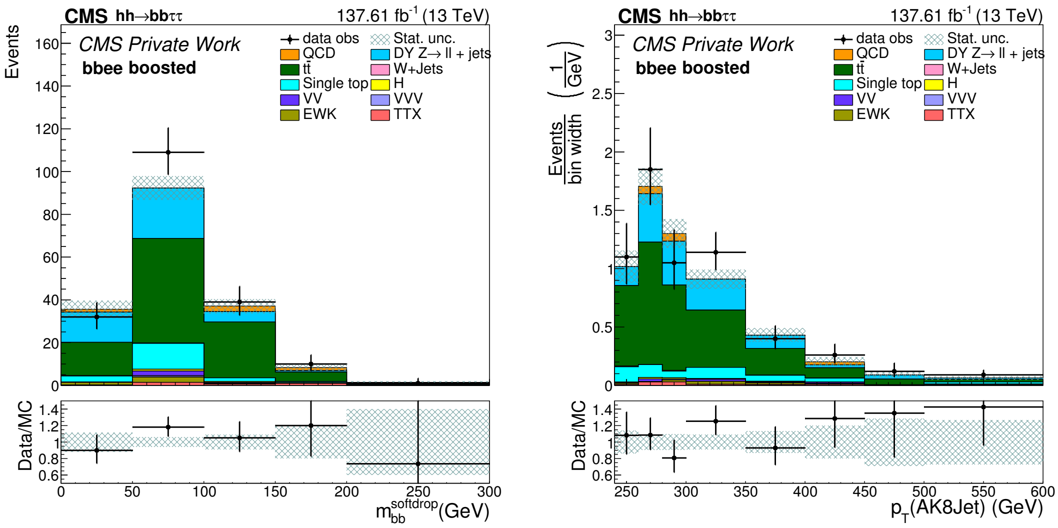
Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category for the TTCR, for the $e\mu$ channel.



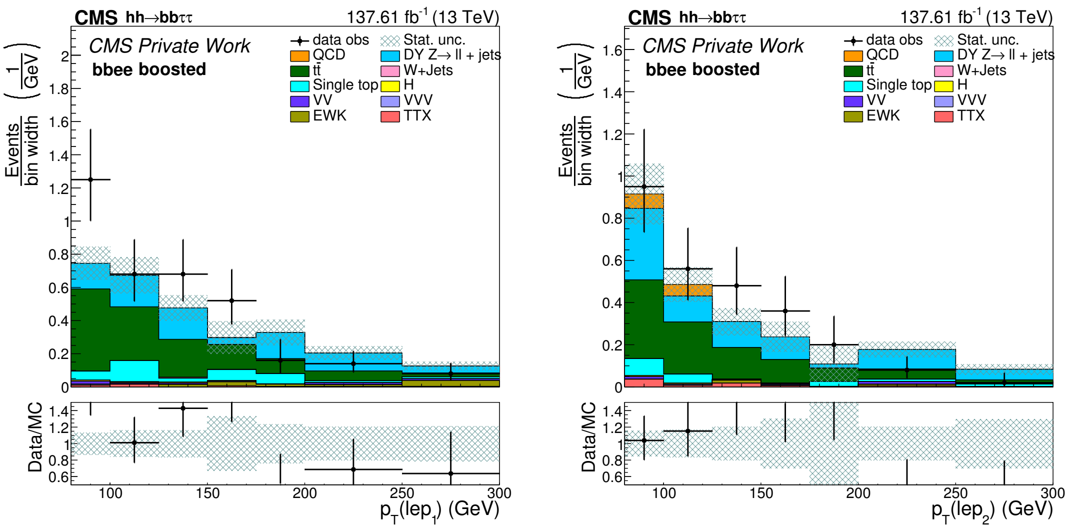
Distribution of the first and second lepton p_T in the boosted category for the TTCR for the $e\mu$ channel.



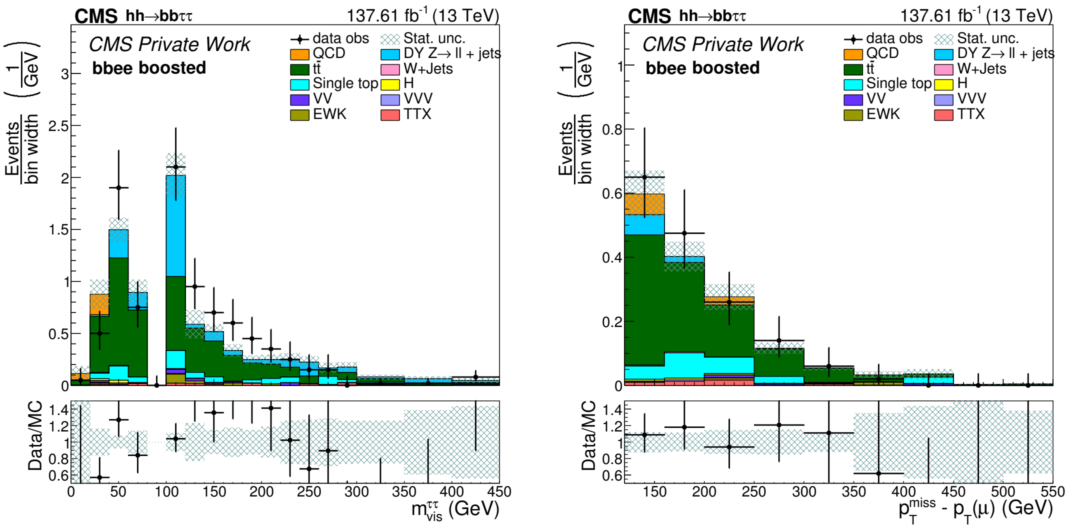
Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category for the TTCR for the $e\mu$ channel.



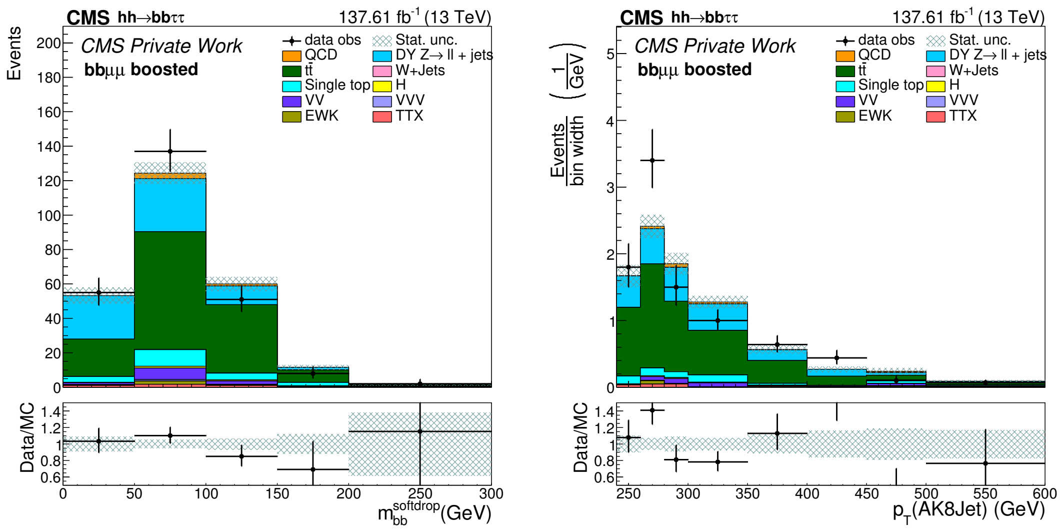
Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category for the TTCR, for the *ee* channel.



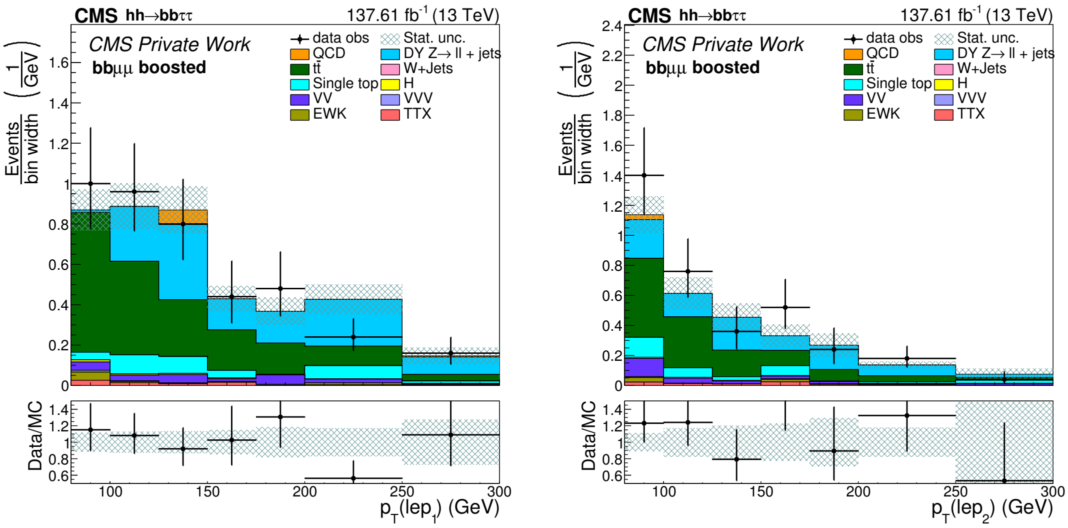
Distribution of the first and second lepton p_T in the boosted category for the TTCR for the *ee* channel.



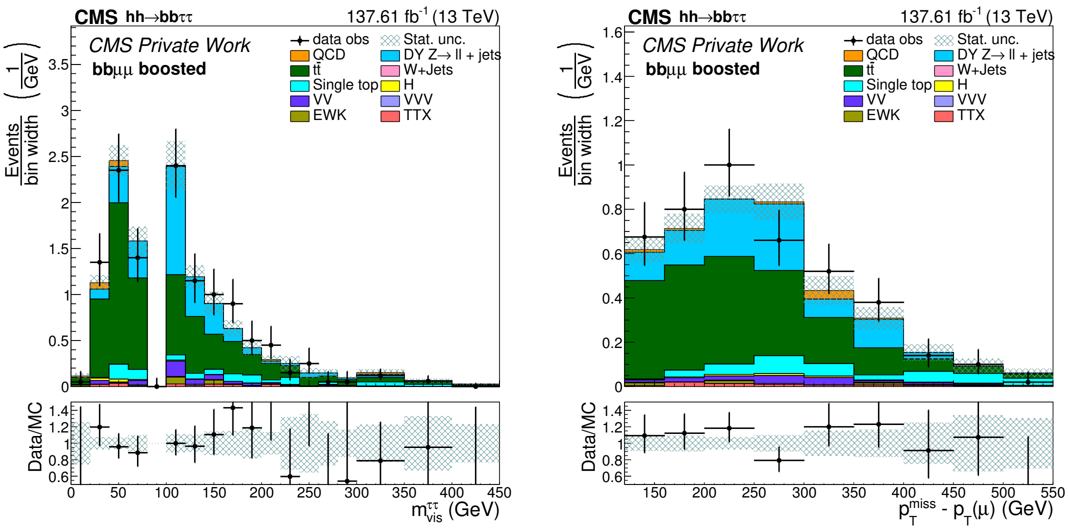
Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category for the TTCR for the ee channel.



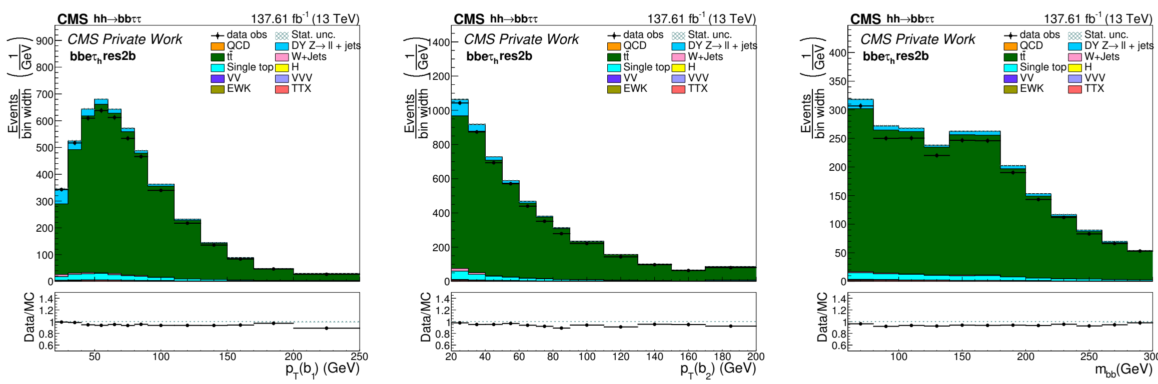
Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category for the TTCR, for the $\mu\mu$ channel.



Distribution of the first and second lepton p_T in the boosted category for the TTCR for the $\mu\mu$ channel.

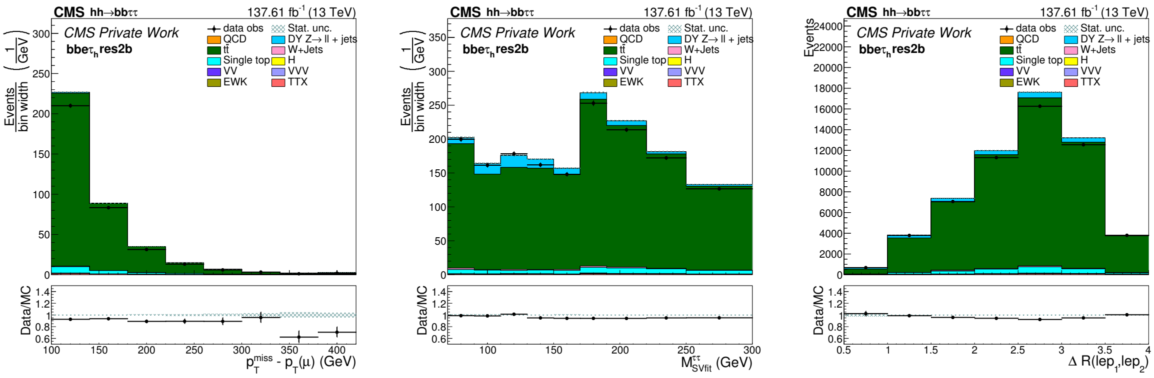


Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category for the TTCR for the $\mu\mu$ channel.

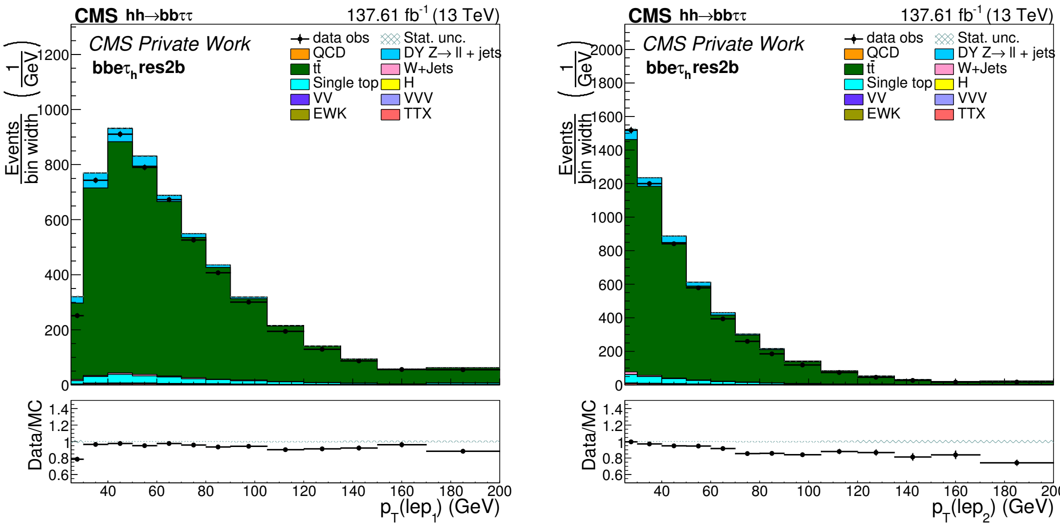


Distribution of the first and second b-jet p_T in the inclusive and res2b categories in the TTCR, for the $e\tau_h$ channel.

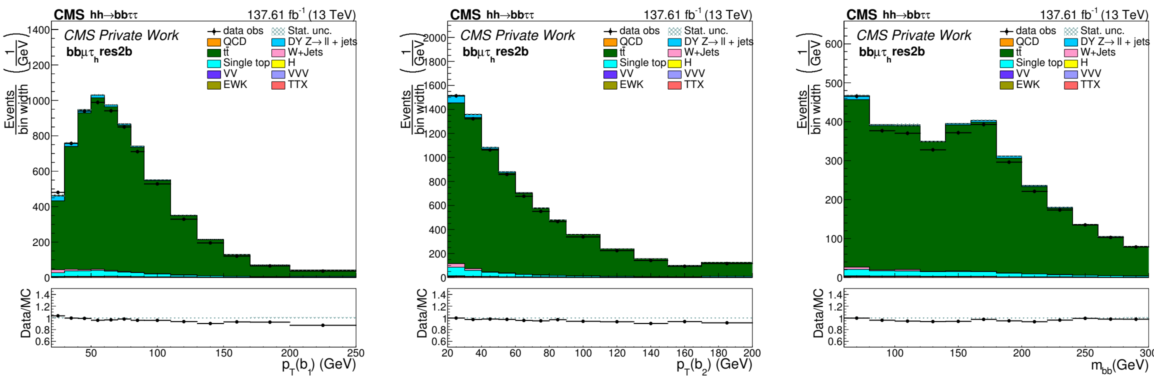
Distribution of the visible dilepton mass and invariant mass of the two b-jets in the inclusive and res2b categories in the TTCR, for the $e\tau_h$ channel.



Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive and res2b categories in the TTCR, for the $e\tau_h$ channel.

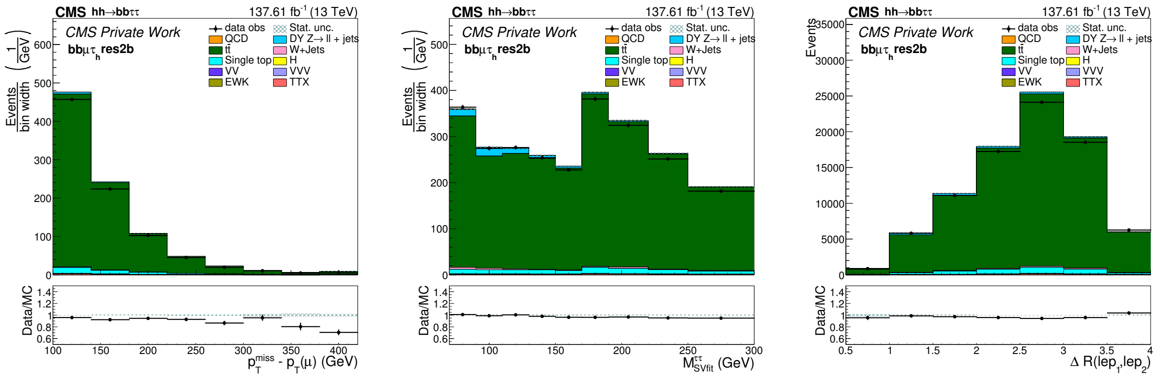


Distribution of the first and second lepton p_T , in the inclusive and res2b categories in the TTCR, for the $e\tau_h$ channel.

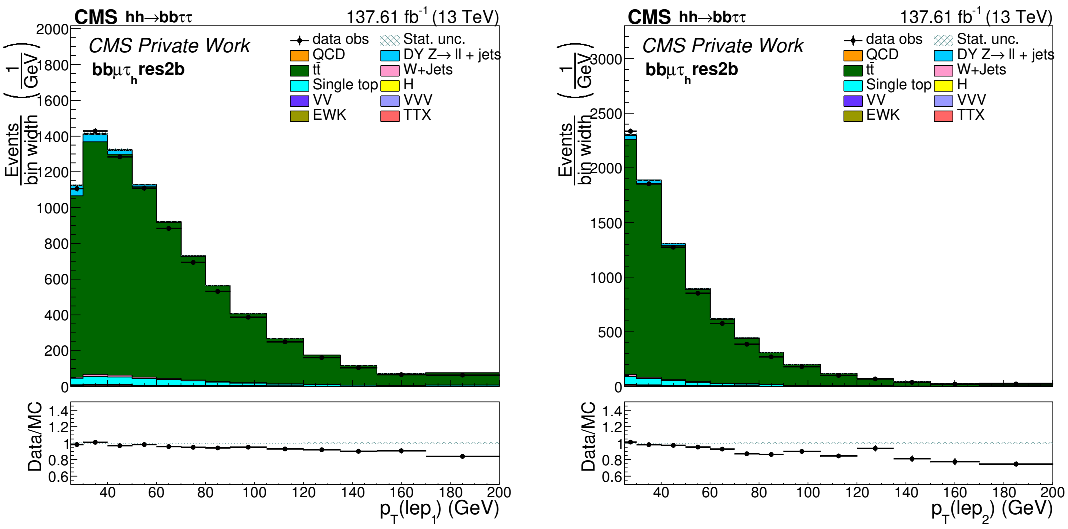


Distribution of the first and second b-jet p_T in the inclusive and res2b categories in the TTCR, for the $\mu\tau_h$ channel.

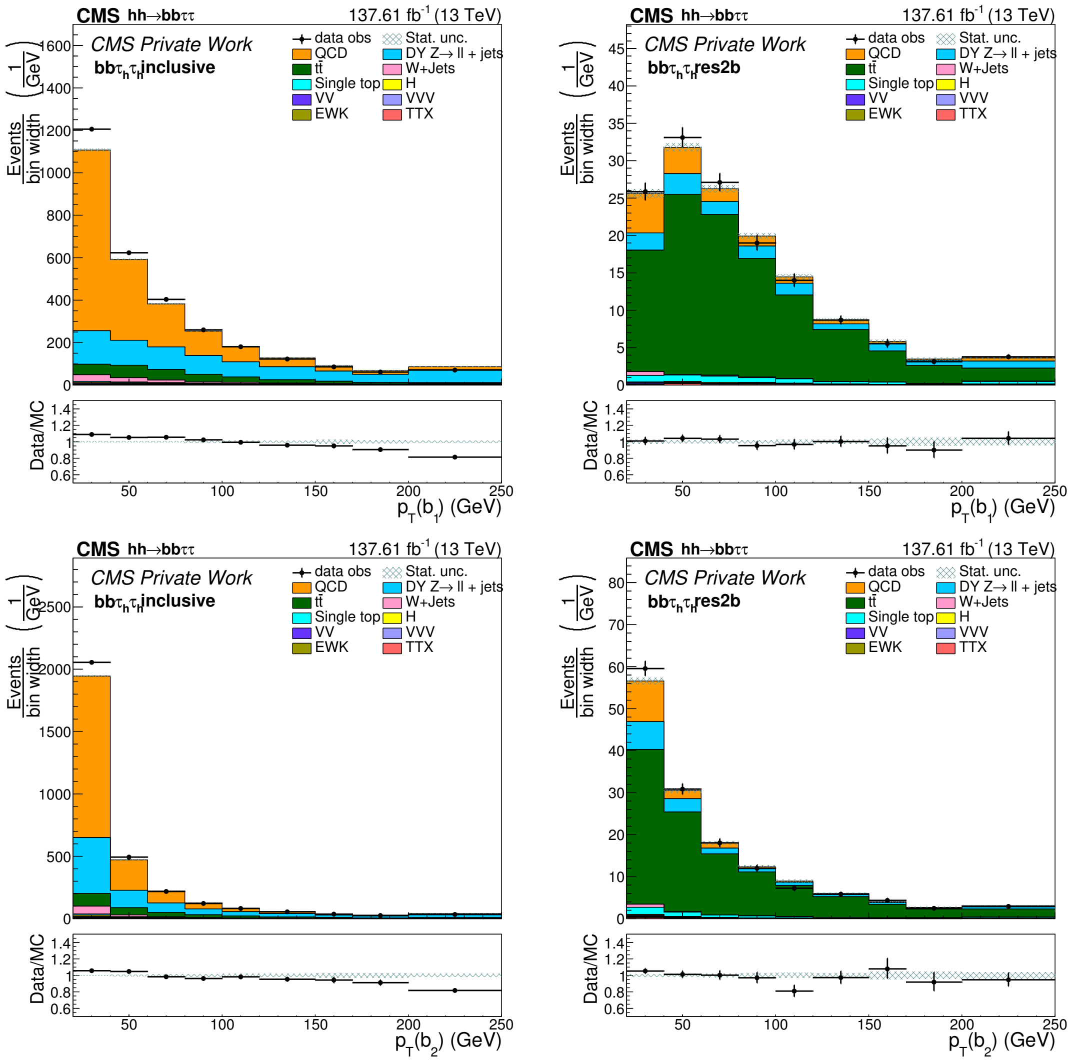
Distribution of the visible dilepton mass and invariant mass of the two b-jets in the inclusive and res2b categories in the TTCR, for the $\mu\tau_h$ channel.



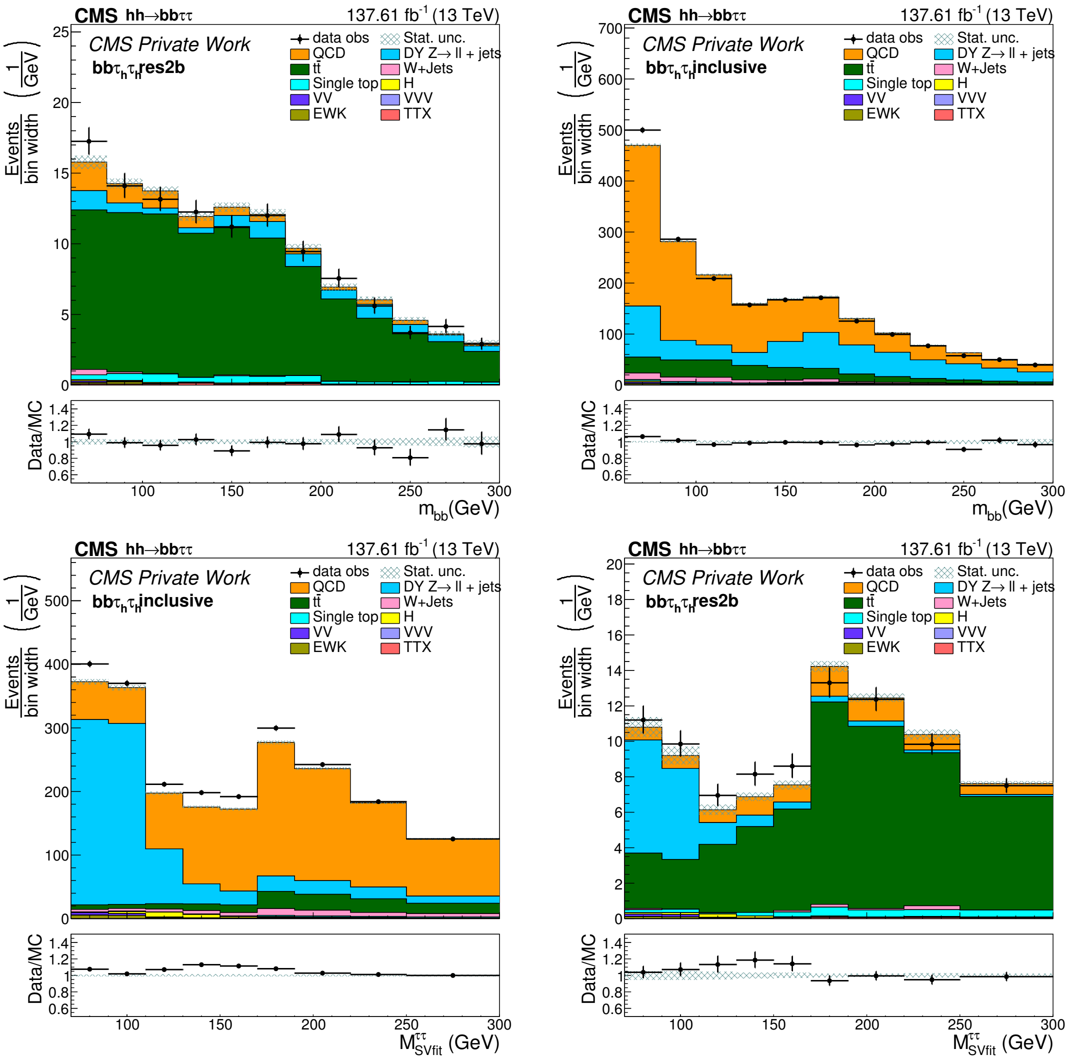
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive and res2b categories in the TTCR, for the $\mu\tau_h$ channel.



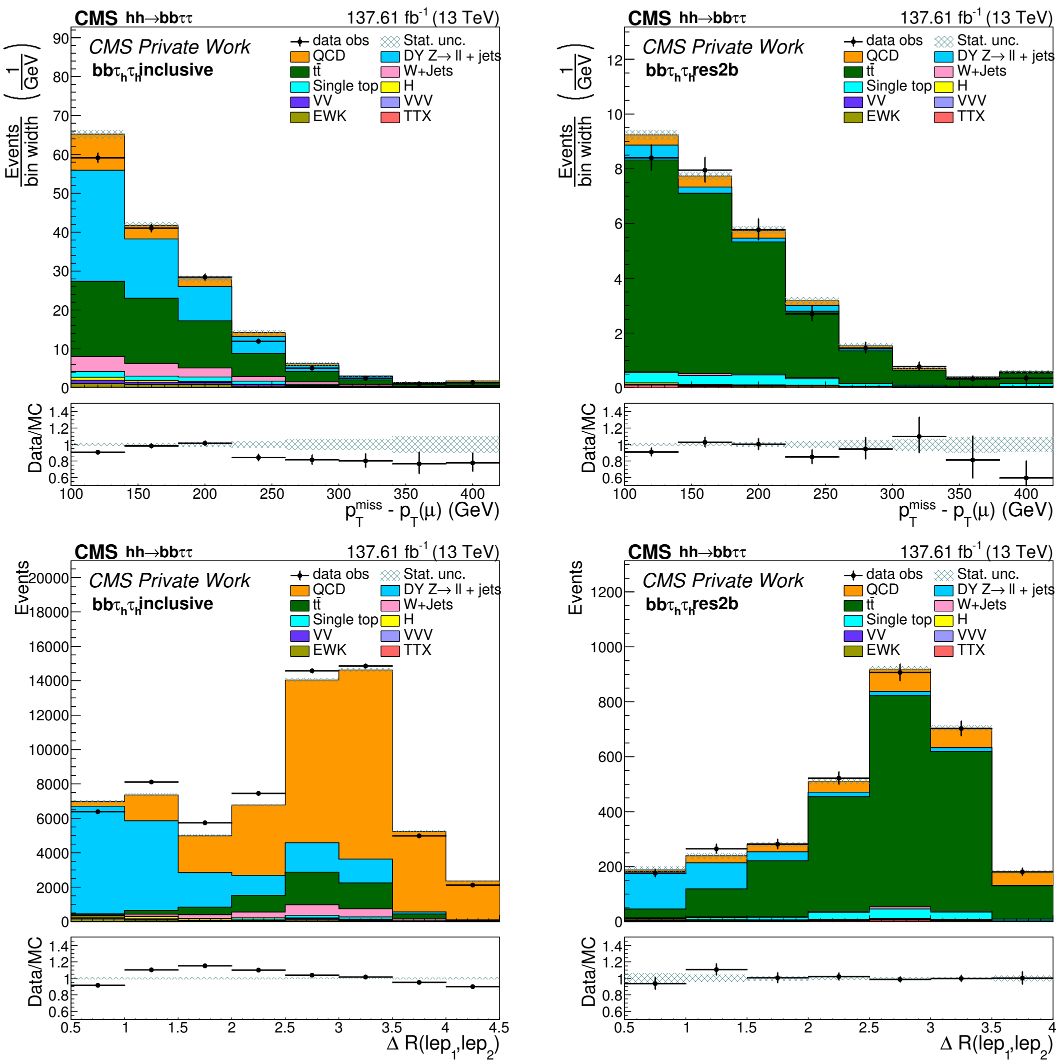
Distribution of the first and second lepton p_T, in the inclusive and res2b categories in the TTCR, for the $\mu\tau_h$ channel.



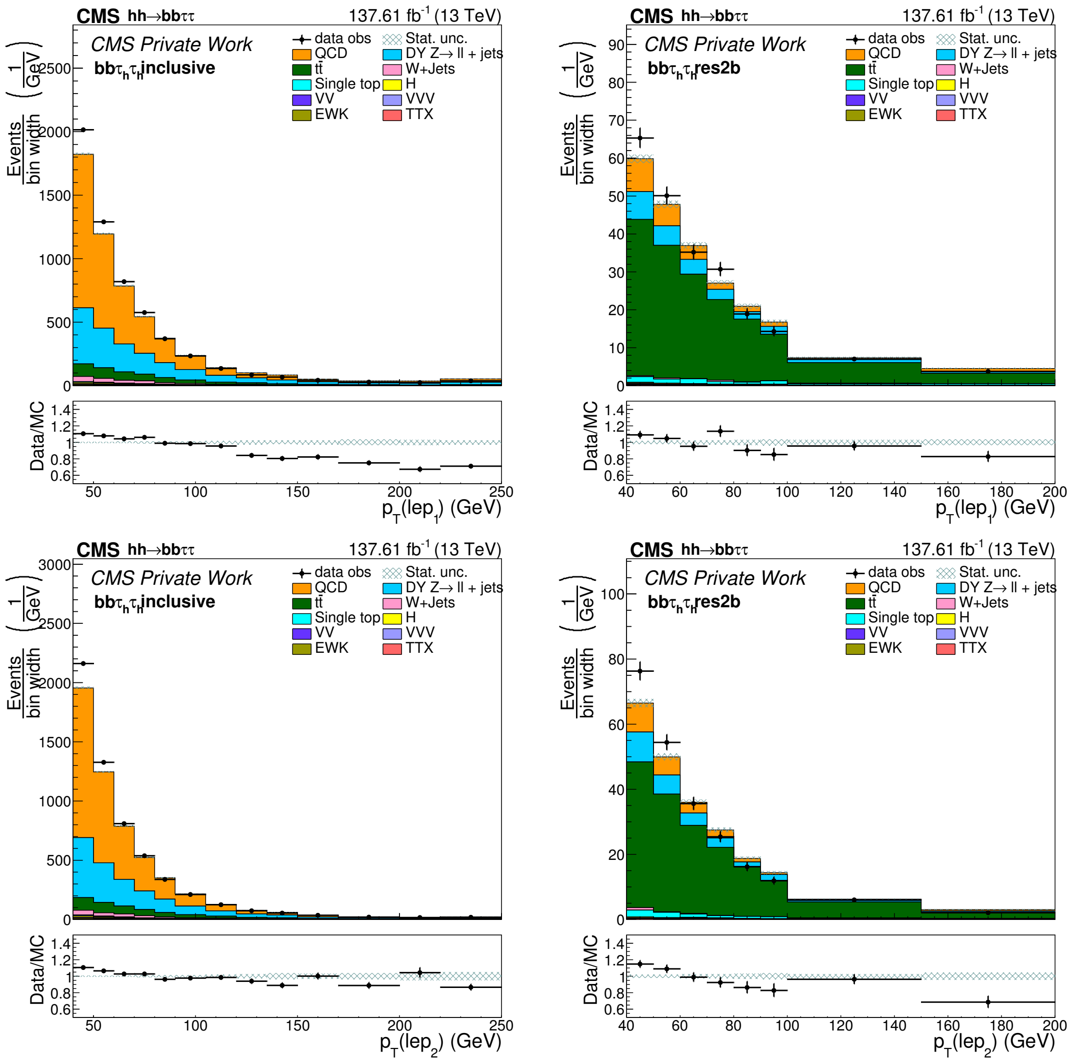
Distribution of the first and second b-jet p_T in the inclusive and res2b categories in the TTCR, for the $\tau_h\tau_h$, channel.



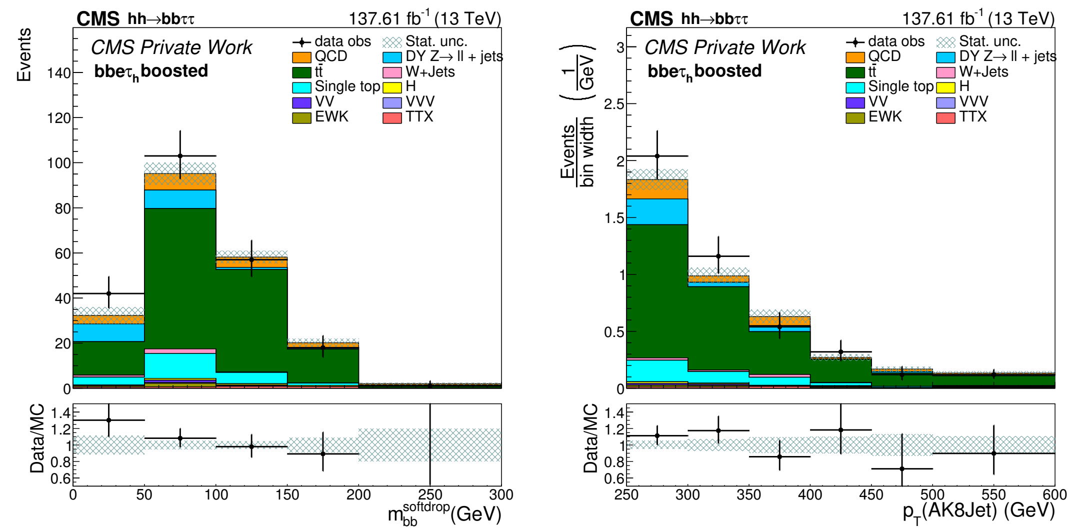
Distribution of the visible dilepton mass and invariant mass of the two b-jets in the inclusive and res2b categories in the TTCR, for the $\tau_h\tau_h$, channel.



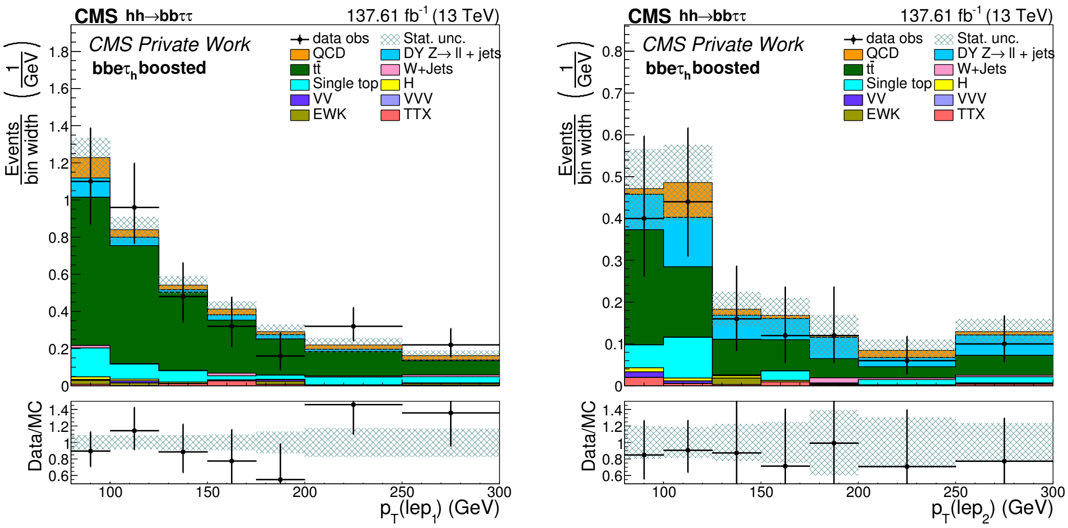
Distribution of the visible dilepton mass, of the MET-noμ and of angular separation of the two leptons, in the inclusive and res2b categories in the TTCR, for the $\tau_h\tau_h$, channel.



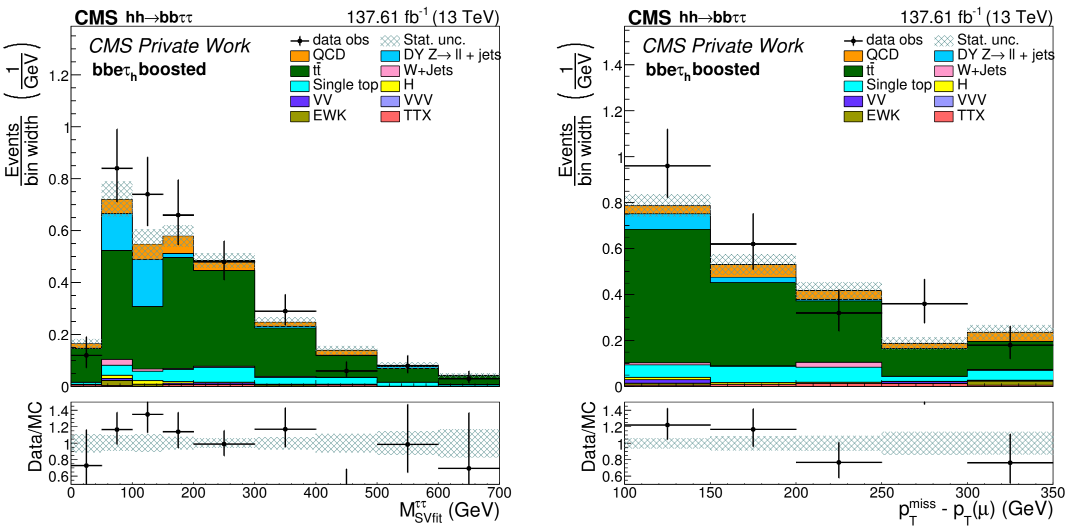
Distribution of the first and second lepton p_T , in the inclusive and res2b categories in the TTCR, for the $\tau_h\tau_h$, channel.



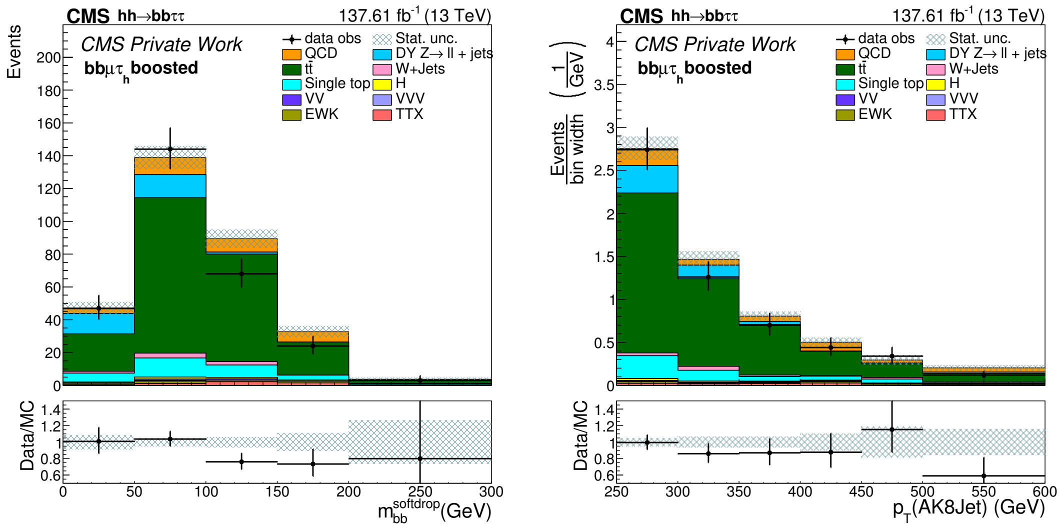
Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category for the TTCR, for the $e\tau_h$ channel.



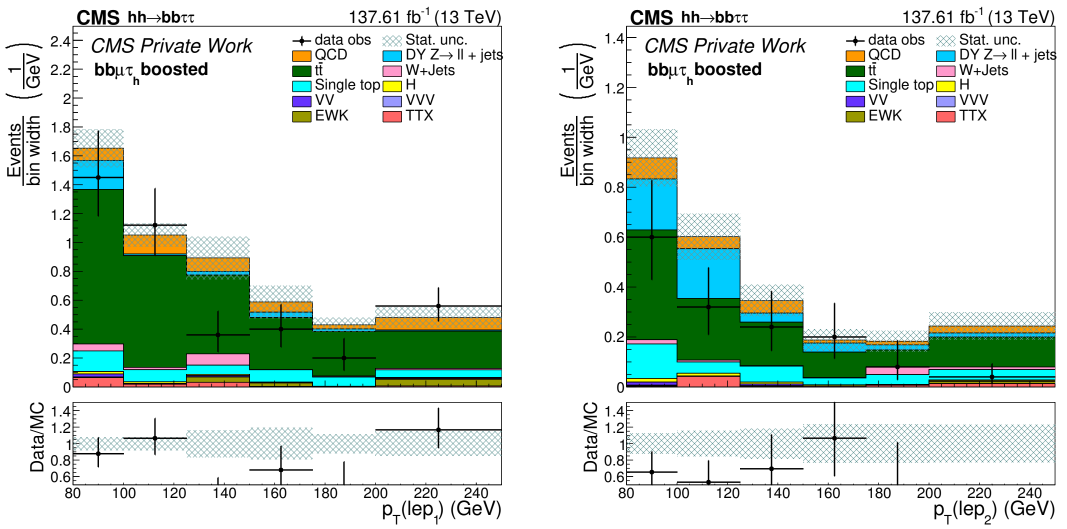
Distribution of the first and second lepton p_T in the boosted category for the TTCR for the $e\tau_h$ channel.



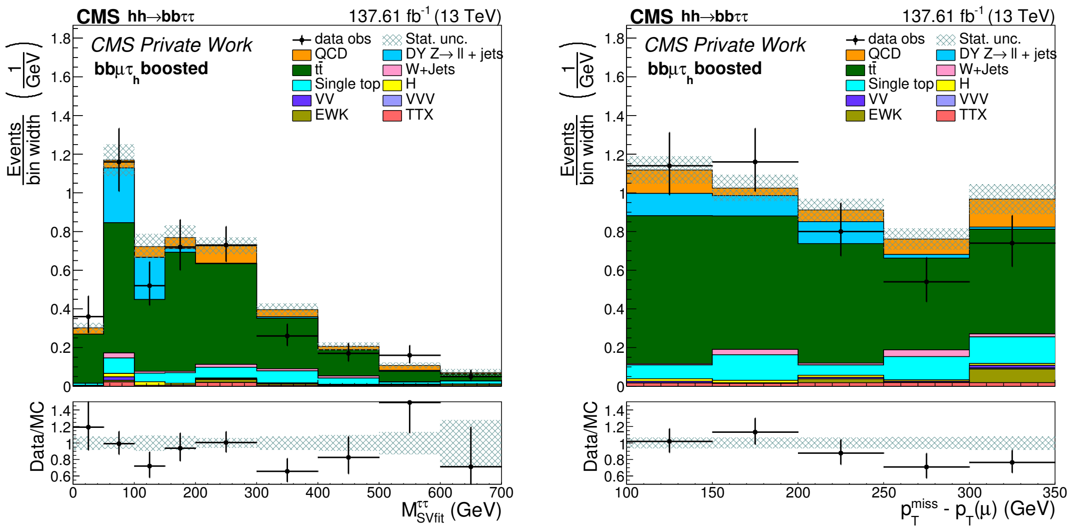
Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category for the TTCR for the $e\tau_h$ channel.



Distributions of the selected AK8 jet softdrop mass and p_T in the boosted category for the TTCR, for the $\mu\tau_h$ channel.



Distribution of the first and second lepton p_T in the boosted category for the TTCR for the $\mu\tau_h$ channel.



Distribution of the the dilepton invariant mass and the MET-noμ in the boosted category for the TTCR for the $\mu\tau_h$ channel.

Appendix G

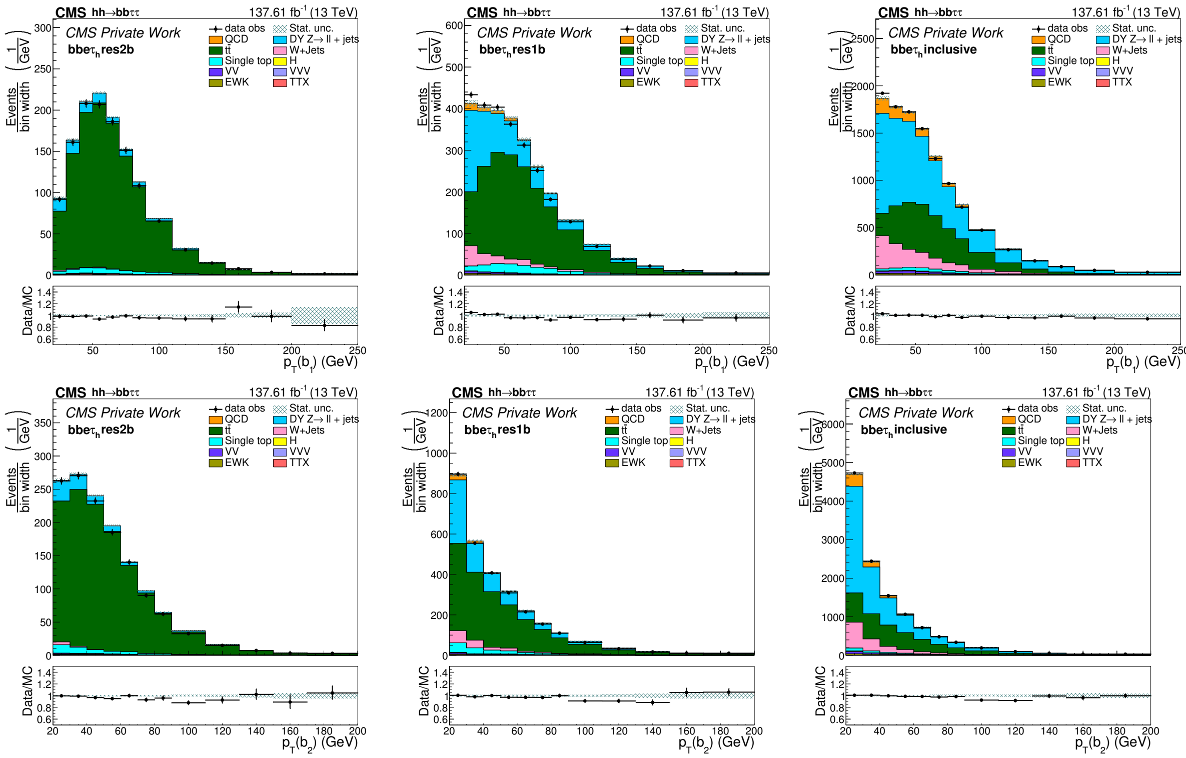
Control plots for signal region

In this section the binned distributions obtained in the Signal Region (SR) are shown, for lepton and b-jet observables not sensitive to the signal. The distributions are presented separately for the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels. The SR is defined in Section 4.8.3, considering the inclusive, Resolved 1b category (res1b) , Resolved 2b category (res2b) and boosted categories. The resolved categories plots are reported in Section G.1 whereas the boosted ones are reported in Section G.1.2.

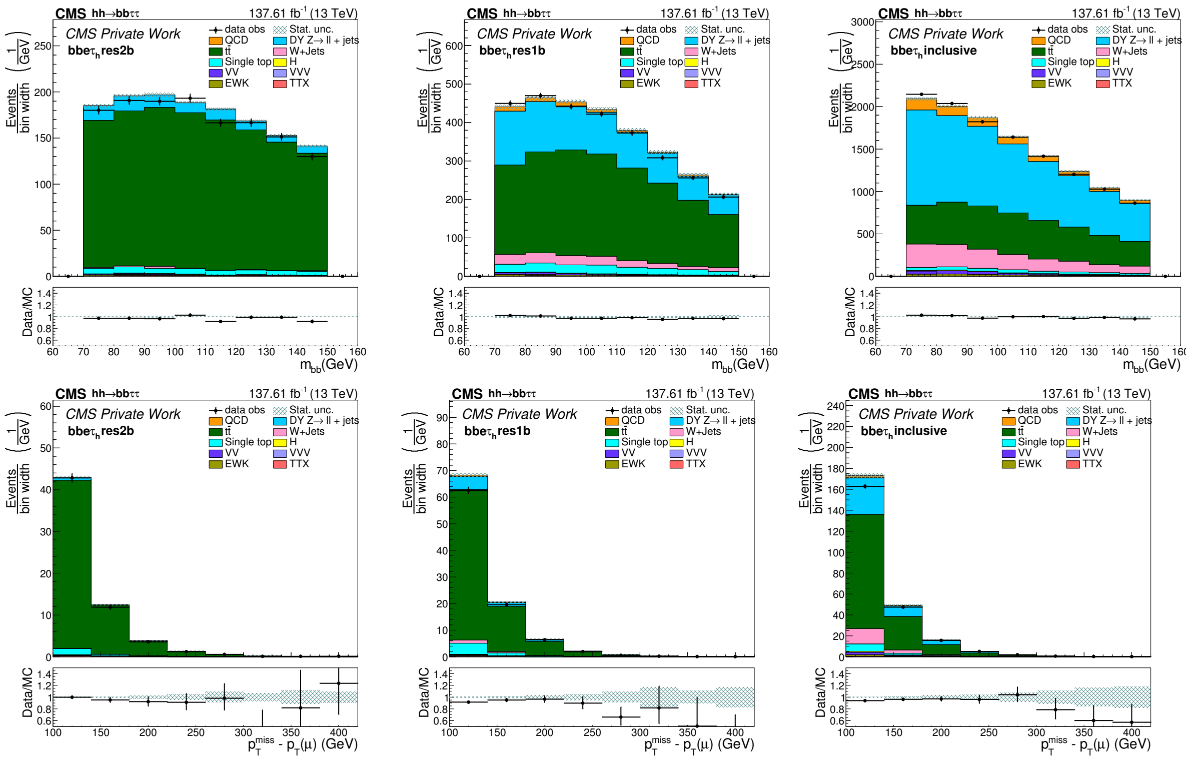
G.1. Resolved categories

G.1.1 Resolved categories

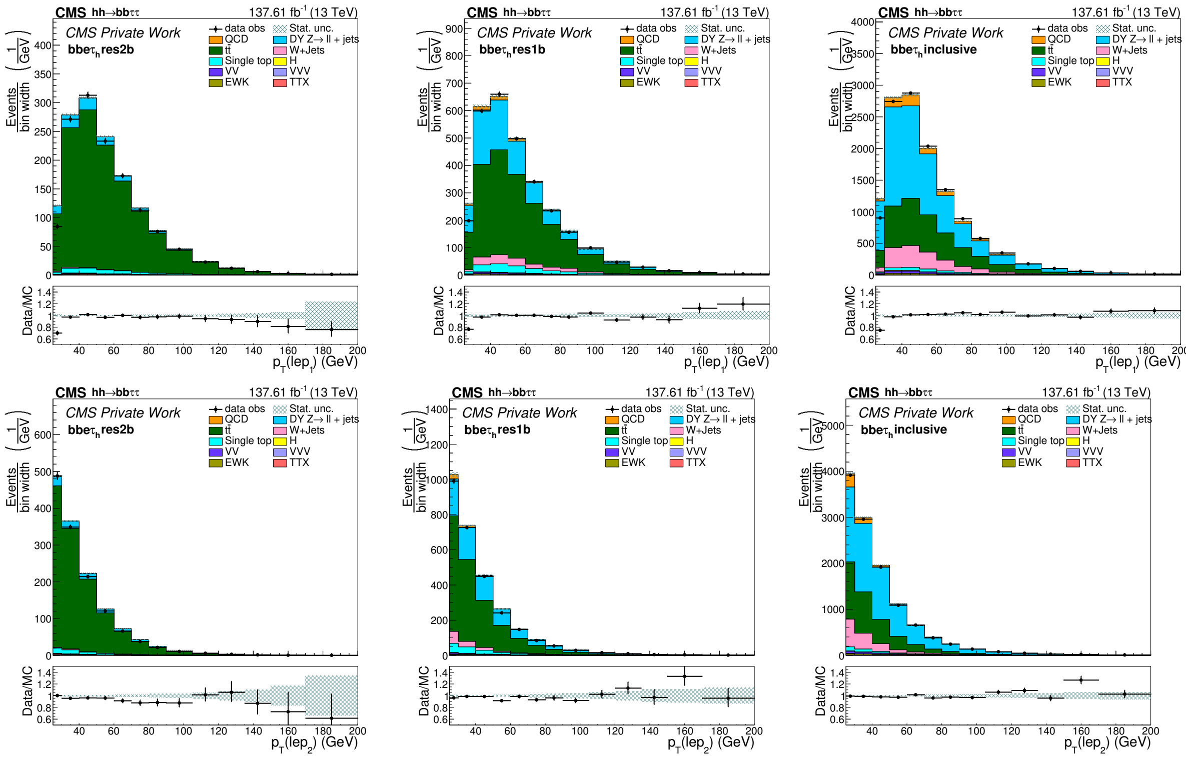
G.1.2 Boosted category



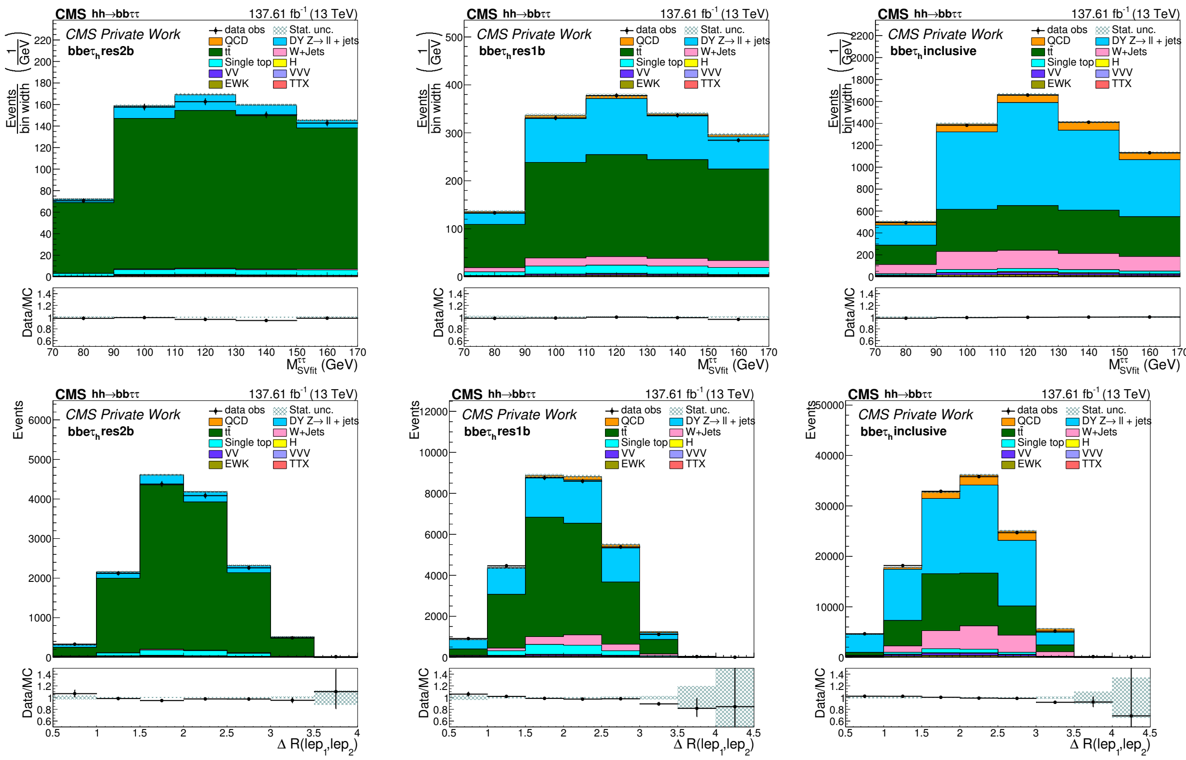
Distribution of the first (upper panel) and second (lower panel) b-jet transverse momentum component (p_T) , in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $e\tau_h$ channel.



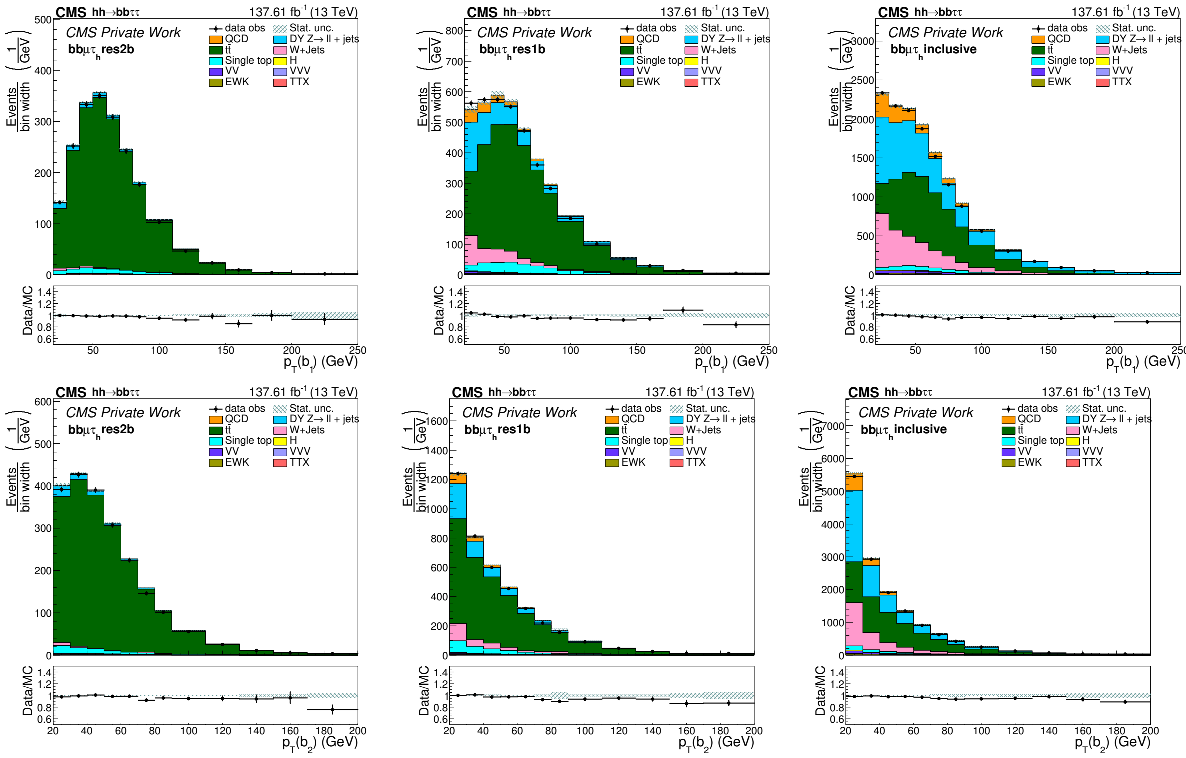
Top row: invariant mass of the two b-jets in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Bottom row: MET-noμ distributions in the same categories/channel in the SR. Plots are shown for the $e\tau_h$ channel.



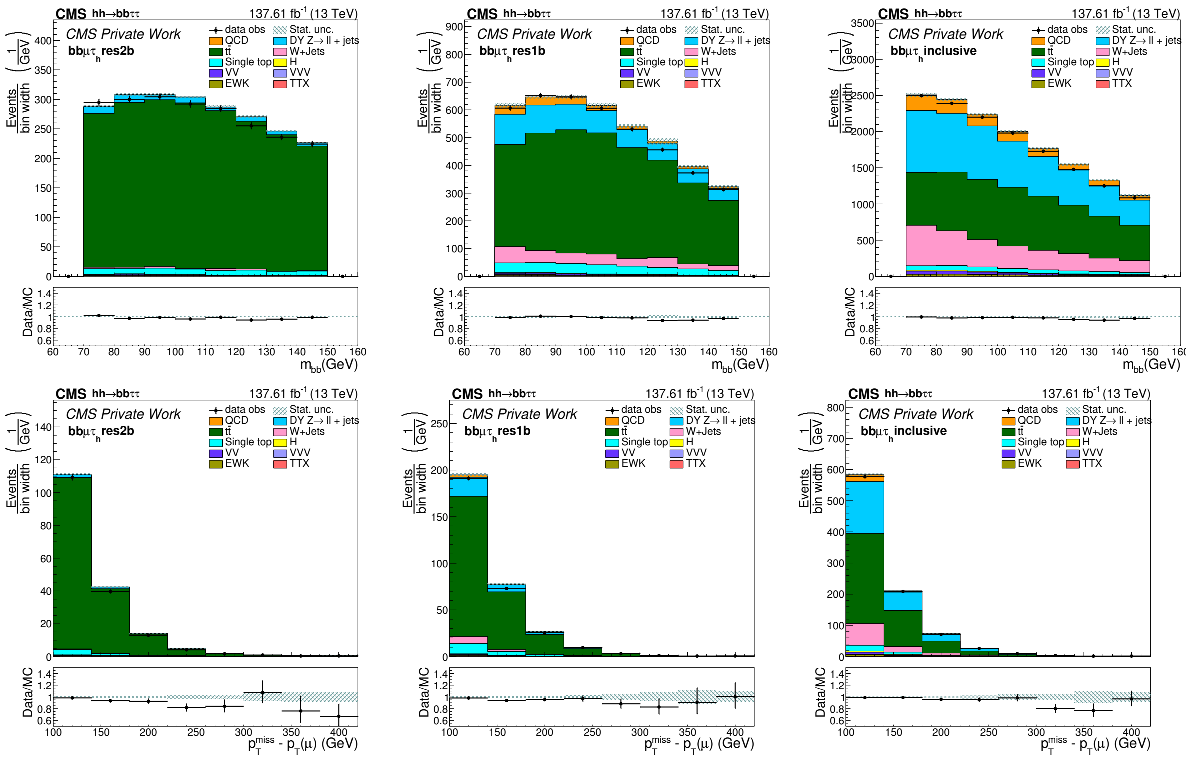
Distribution of the first (upper row) and second (lower row) lepton p_T , in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $e\tau_h$ channel.



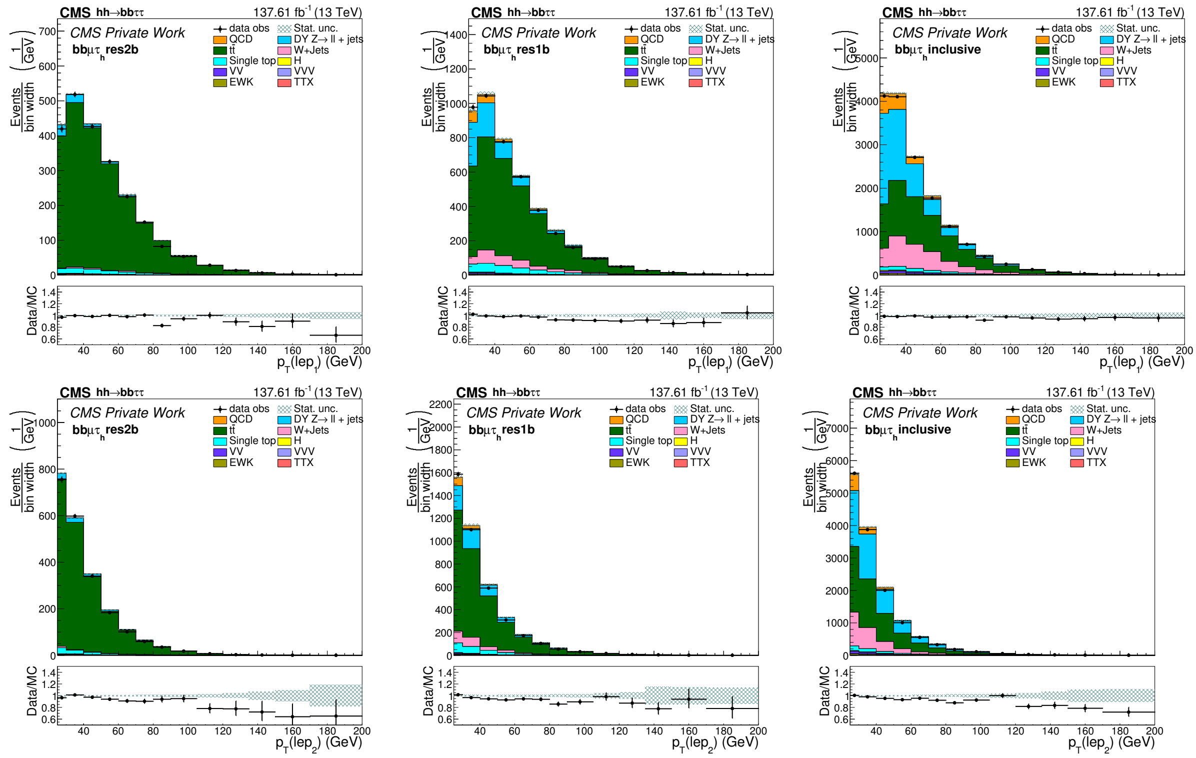
Di-lepton mass (top line) and angular separation between leptons (bottom line) in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $e\tau_h$ channel.



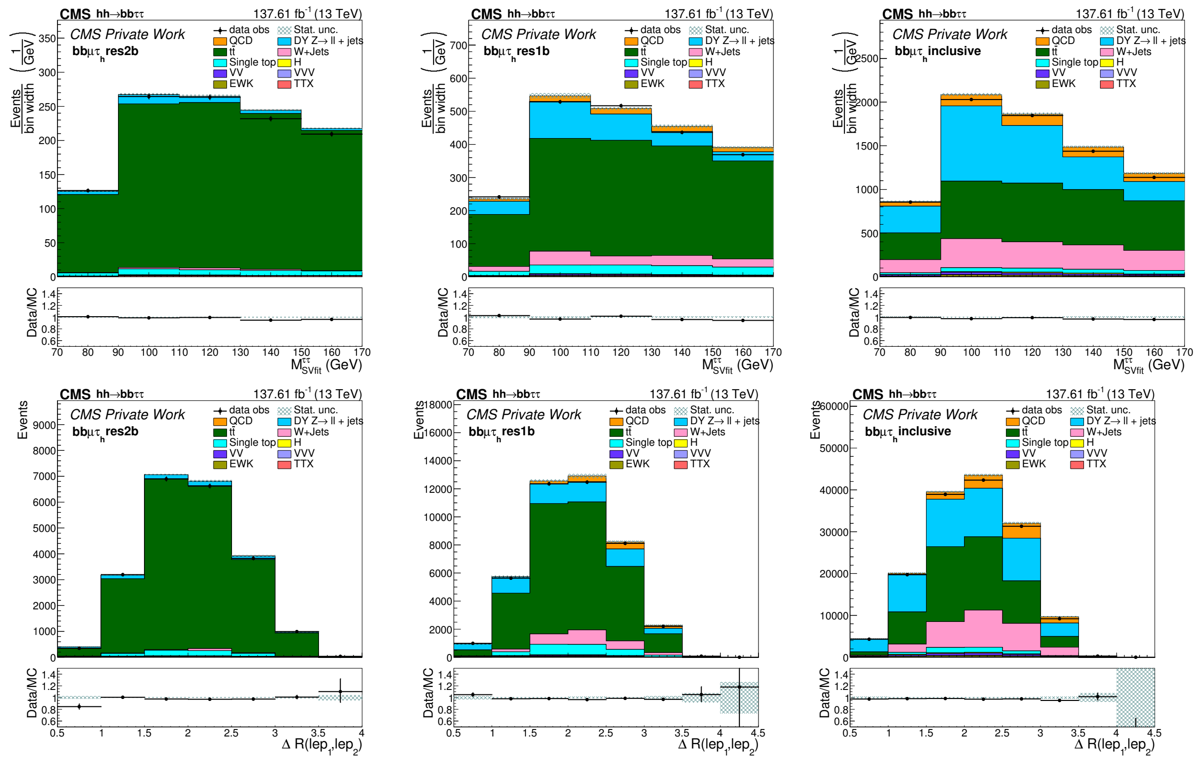
Distribution of the first (upper panel) and second (lower panel) b-jet p_T , in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $\mu\tau_h$ channel.



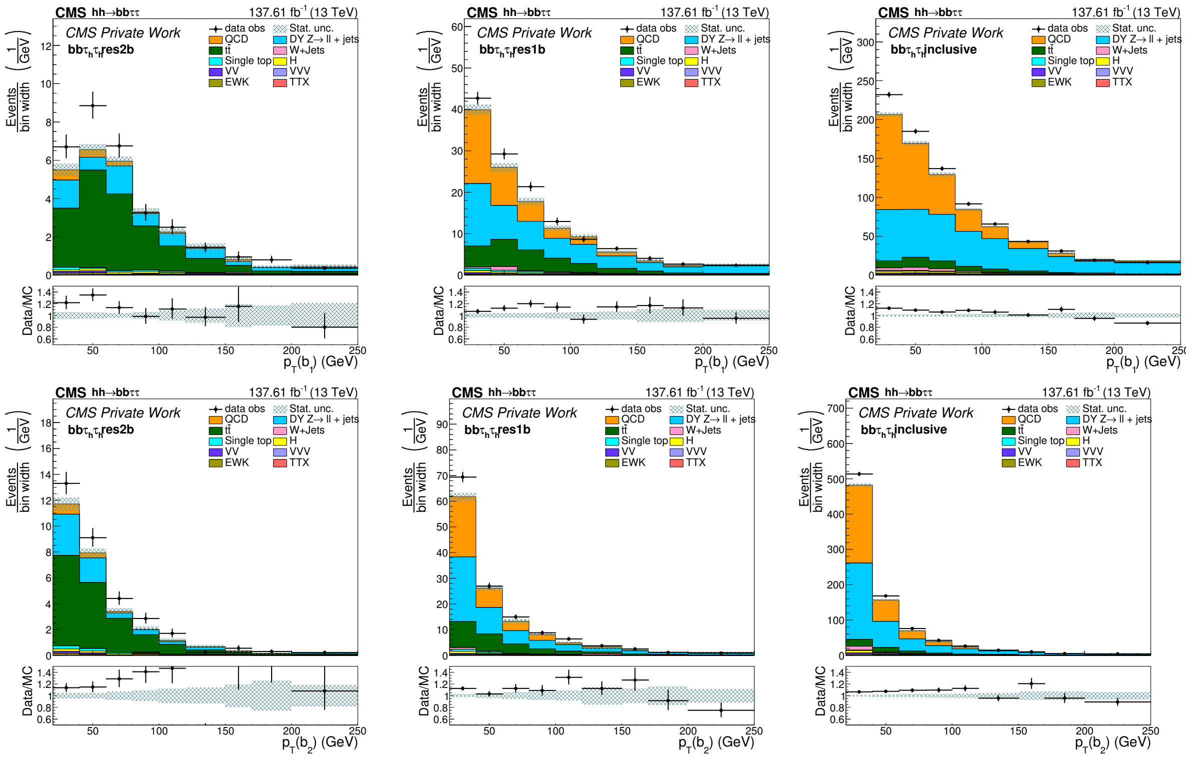
Top row: invariant mass of the two b-jets in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Bottom row: MET-noμ distributions in the same categories/channel in the SR. Plots are shown for the $\mu\tau_h$ channel.



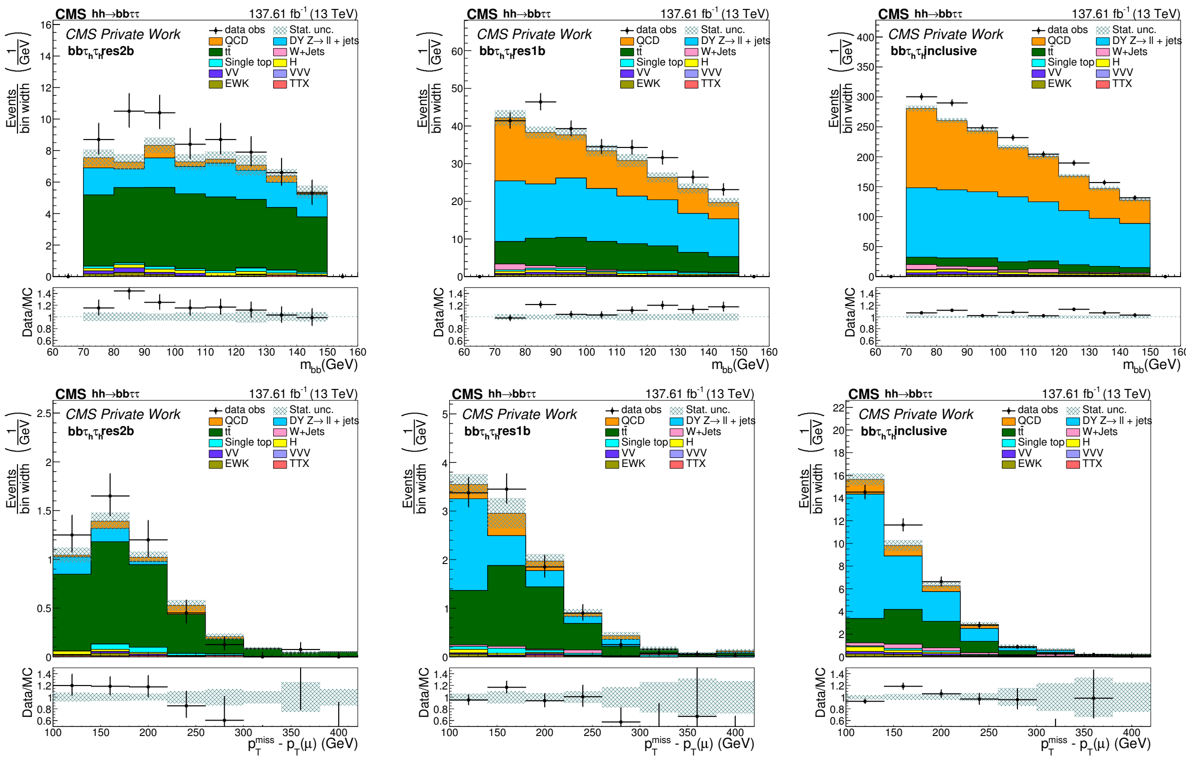
Distribution of the first (upper row) and second (lower row) lepton p_T , in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $\mu\tau_h$ channel.



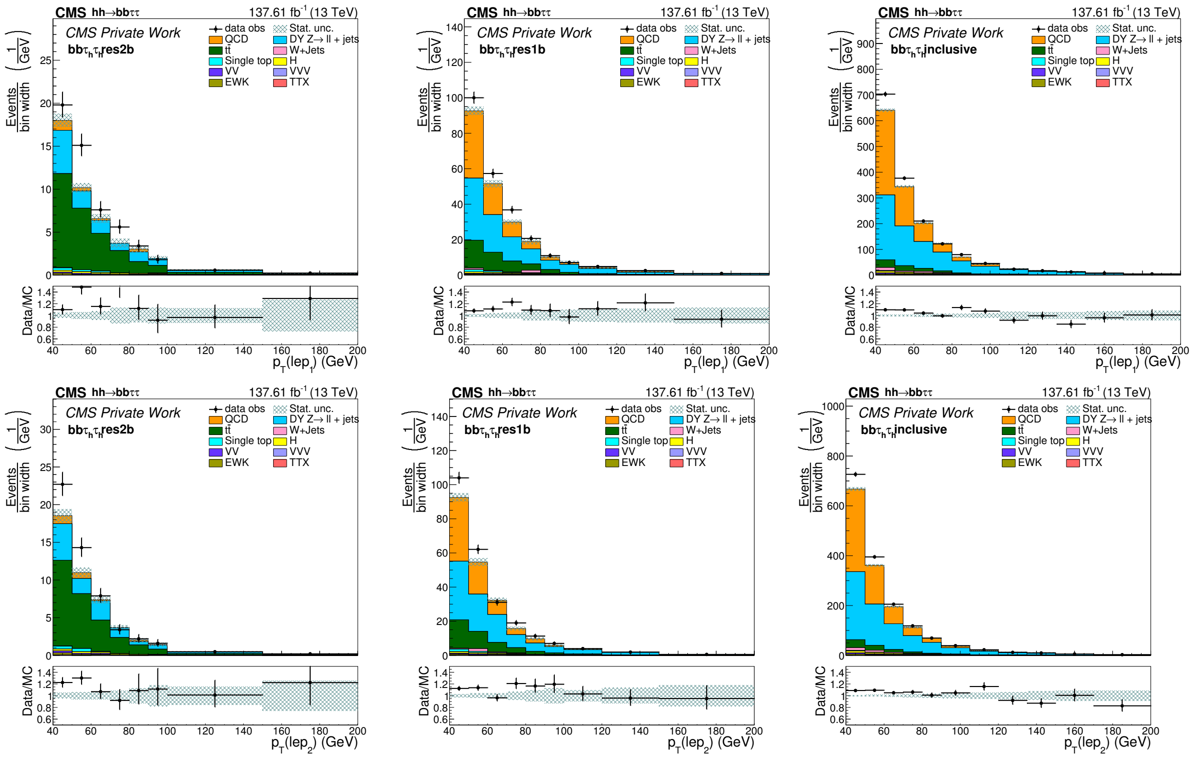
Di-lepton mass (top line) and angular separation between leptons (bottom line) in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $\mu\tau_h$ channel.



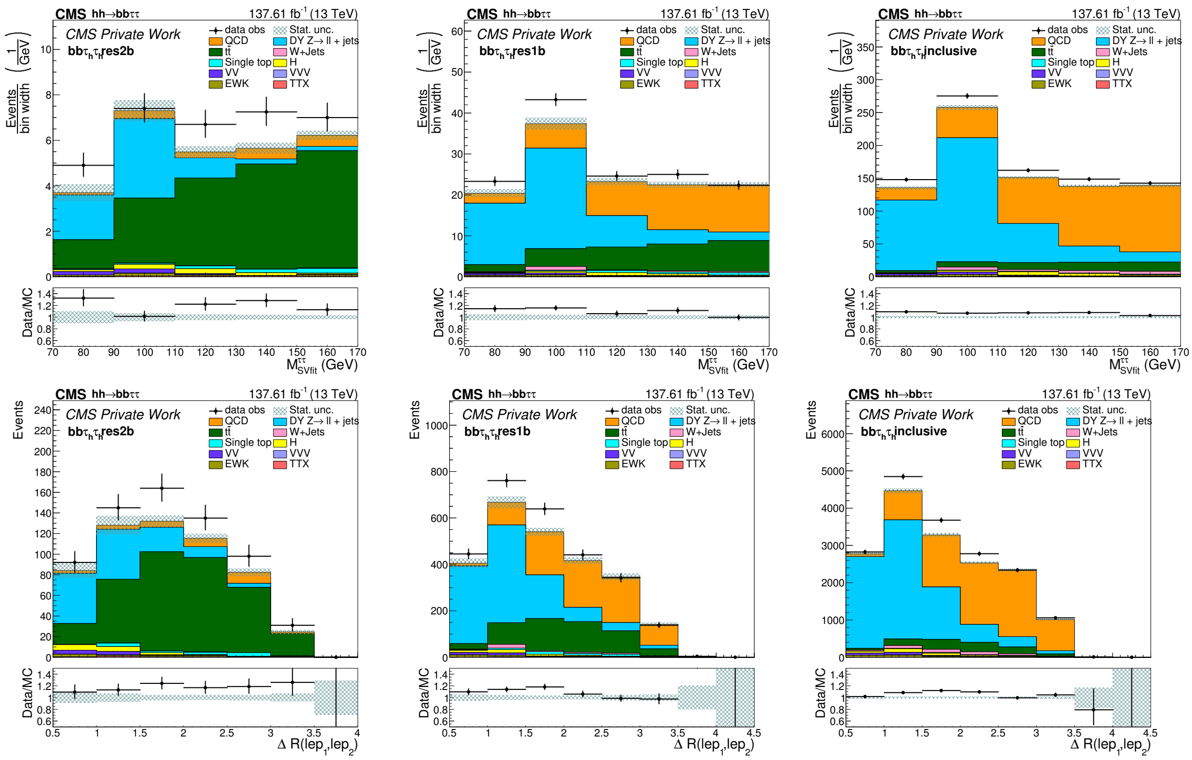
Distribution of the first (upper panel) and second (lower panel) b-jet p_T , in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $\tau_h \tau_h$ channel.



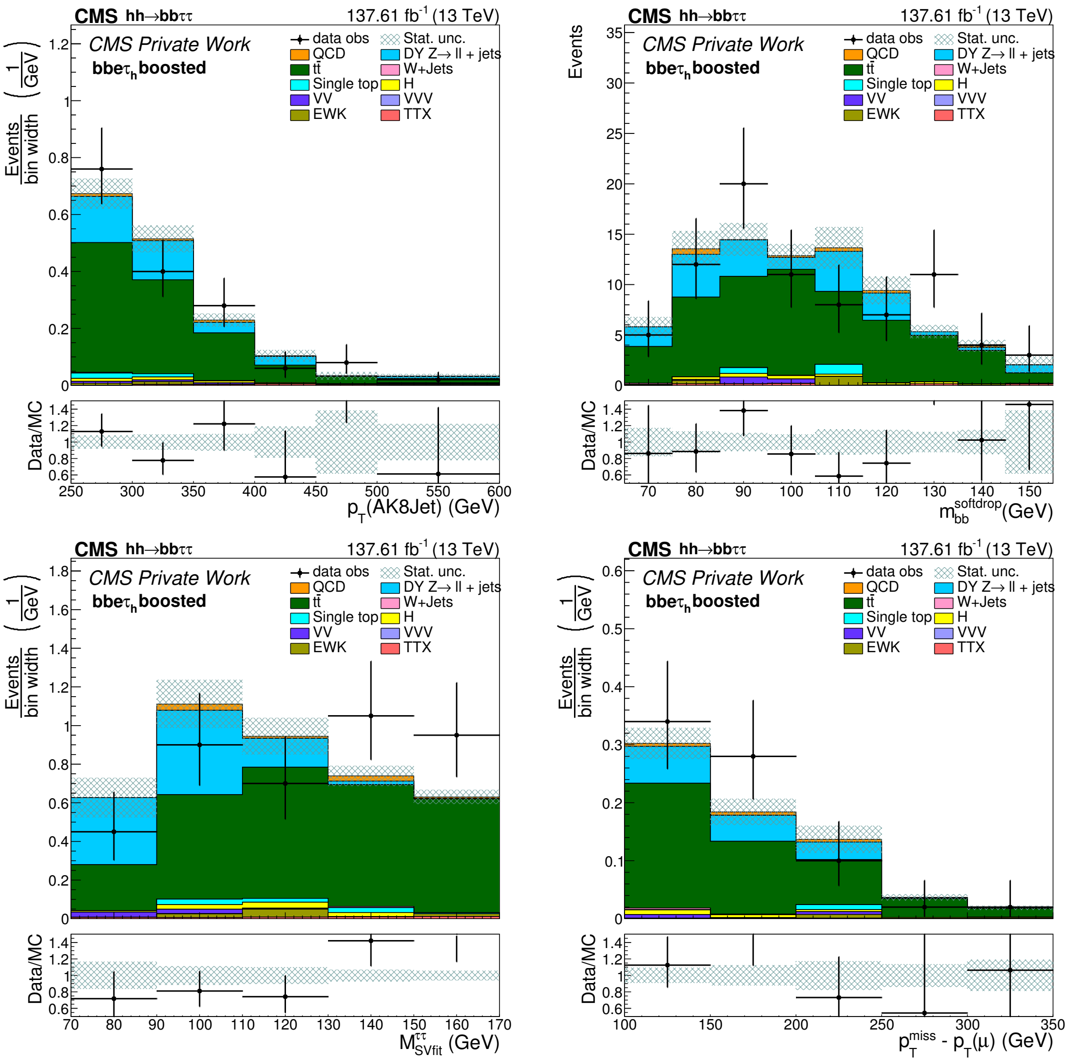
Top row: invariant mass of the two b-jets in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Bottom row: MET-noμ distributions in the same categories/channel in the SR. Plots are shown for the $\tau_h \tau_h$ channel.



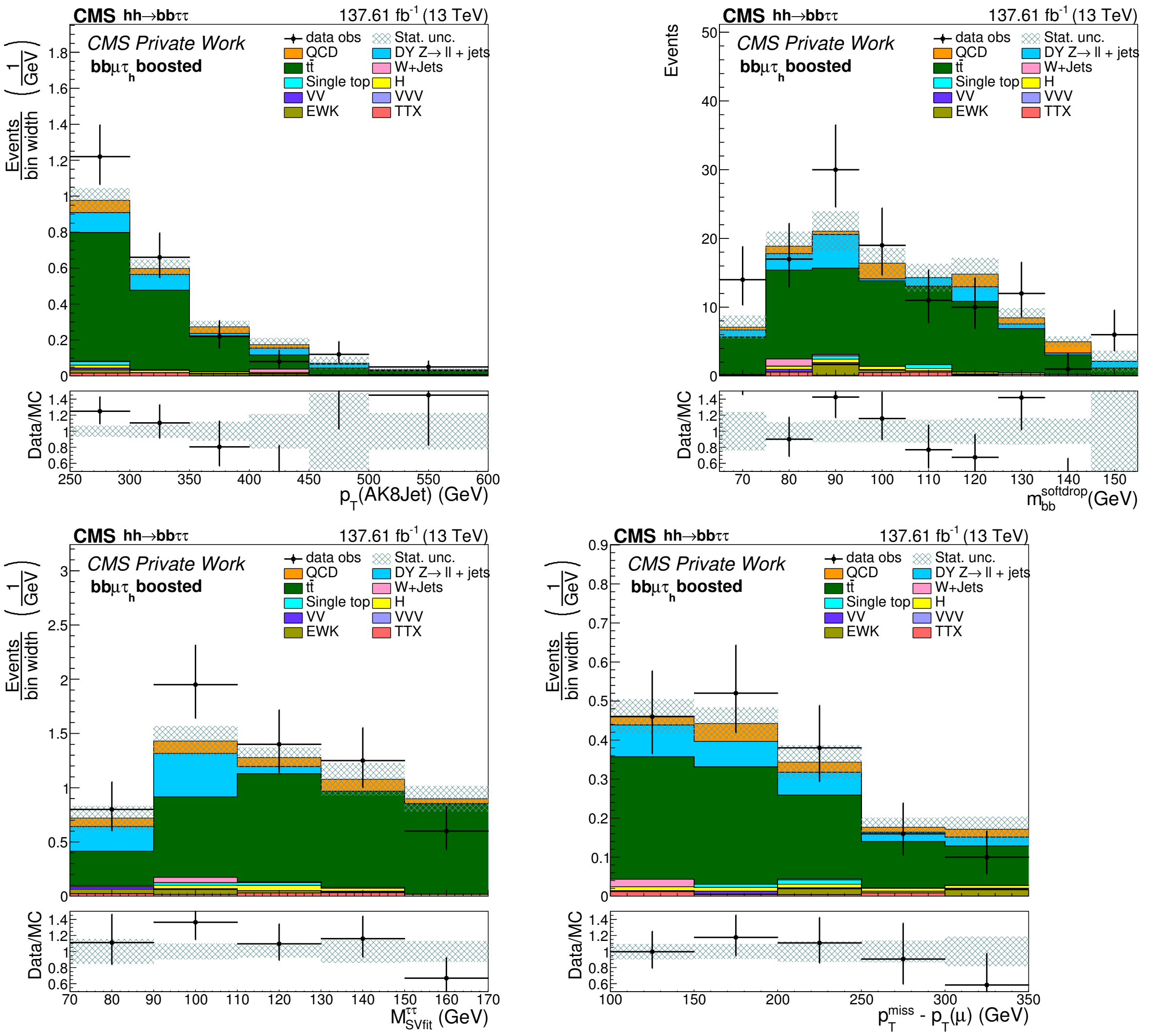
Distribution of the first (upper row) and second (lower row) lepton p_T , in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $\tau_h\tau_h$ channel.



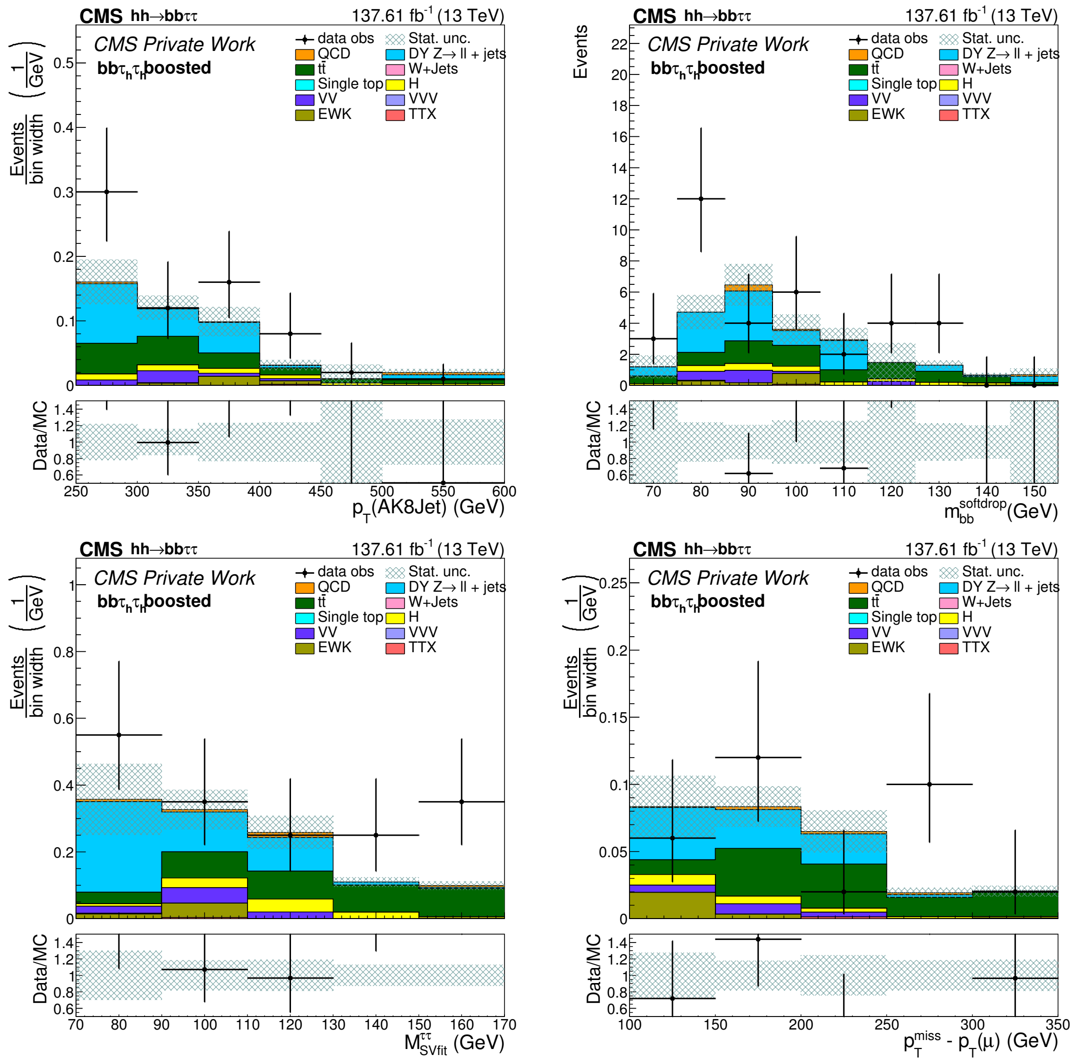
Di-lepton mass (top line) and angular separation between leptons (bottom line) in the res2b (left), res1b (center) and inclusive (right) categories in the SR. Plots are shown for the $\tau_h\tau_h$ channel.



Top row: Selected AK8 jet p_T (left) and softdrop mass (right) distribution in the boosted category. Bottom row: $m_{\tau\tau}^{(SVFit)}$ (left) and MET-noμ (right) distributions in the boosted category. Plots are shown for the $e\tau_h$ channel.



Top row: Selected AK8 jet p_T (left) and softdrop mass (right) distribution in the boosted category. Bottom row: $m_{\tau\tau}^{(\mathrm{SVFit})}$ (left) and MET-noμ (right) distributions in the boosted category. Plots are shown for the $\mu\tau_h$ channel.



Top row: Selected AK8 jet p_T (left) and softdrop mass (right) distribution in the boosted category. Bottom row: $m_{\tau\tau}^{(SVFit)}$ (left) and MET-noμ (right) distributions in the boosted category. Plots are shown for the $\tau_h\tau_h$ channel.

Appendix H

The *DeepTau2018v2p5* algorithm

This appendix describes the updated version of the *DeepTau* algorithm, *DeepTau2018v2p5* and compares the preformances with its predecessor, *DeepTau2017v2p1*. At the end, the data/Monte Carlo (MC) distributions obtained in the context of the search presented in this thesis, using *DeepTau2018v2p5* for tau selection, identification and discrimination against jets, muons and electrons, are shown.

H.1. From *DeepTau2017v2p1* to *DeepTau2018v2p5*

The hadronic tau reconstruction has been introduced in Section 3.3.5, highlighting the importance of identifying genuine τ_h and distinguishing them from jets, electrons, and muons for precision measurements and new physics searches, as well as the challenges it presents. Compact Muon Solenoid (CMS) employs the *DeepTau* algorithm (described in Section 3.3.5) for hadronic tau identification.
The first deployed version, used in CMS Run 2 analyses, was *DeepTau2017v2p1*. It was trained exclusively on MC simulated events. Although MC simulations generally provide a good representation of real collision data, some input features are not perfectly modeled, leading to discrepancies between data and MC.
These discrepancies are particularly significant at high DeepTau VSjet discriminator scores, where the genuine τ_h purity is highest.
Previously, these mismodeling effects were mitigated using Scale Factor (SF) extracted from dedicated calibration measurements, fitting simulations to data. However, this approach introduces larger uncertainties and prevents the raw *DeepTau* score from being reliably used in analyses. Additionally, training *DeepTau2017v2p1* entirely on MC events may result in suboptimal classification performance due to the presence of these discrepancies during training.

DeepTau2018v2p5 [205] is an enhanced version, built upon *DeepTau2017v2p1* and designed for CMS analyses with Run 3 data. The major improvement is the integration of adversarial machine learning techniques to mitigate mismodeling effects directly at

the training stage, reducing discrepancies and improving agreement between data and simulation and thus improving tau identification and making the raw *DeepTau* score a more reliable input for analyses. .
A domain classification subnetwork was implemented alongside the main tau classification network. This subnetwork attempts to distinguish MC events from data, forcing the primary network to use only features that generalize well across both domains. Unlike simply removing mismodeled variables, this approach retains discriminative power while minimizing domain-specific artifacts. Similar adversarial techniques have been successfully applied in previous CMS analyses [206].

The main features of *DeepTau2018v2p5* are:

- The training dataset includes both MC and real collision data.
- Improved data handling through feature standardization and data balancing across different kinematic regions.
- Hyperparameter optimization for better classification performance.
- Adversarial machine learning techniques to reduce data-MC mismodeling effects.
- Inclusion of additional high-energy tau samples to enhance performance at high p_T.
- More precise reconstruction of hadronic tau decays using information from multiple sub-detectors within CMS.

Domain adaptation is a technique that encourages a model to learn features that are useful for classification but invariant to the differences between data and MC. In DeepTau v2.5, this is implemented using adversarial machine learning, where two competing subnetworks are trained simultaneously:

- A primary classifier that distinguishes between tau types (τ_h, electrons, muons, and jets).
- A secondary discriminator that attempts to distinguish between MC and real collision data.

In the context of *DeepTau2018v2p5* developement, the goal is to maximize tau classification performance while minimizing the ability to differentiate between data and MC. This forces the classifier to rely on features that are robust across both domains. The advantages of this approach include:

- Reduced mismodeling effects at the training level, rather than correcting them post-hoc with scale factors.
- Smaller uncertainties in tau identification, improving sensitivity in analyses.
- Potential for using the raw DeepTau VSjet score directly in analyses, without needing additional corrections.
- The ability to maintain high tau identification efficiency while reducing misidentifications from jets, electrons, and muons.

The algorithm utilizes a comprehensive set of input variables:

- 42 high-level variables sensitive to tau kinematics, isolation, and lifetime characteristics.

- Detailed information from particle flow candidates, electrons, and muons within the tau isolation cone, encompassing 7 to 37 variables per candidate type.
- Candidates within the tau signal ($\Delta R < 0.1$) and isolation ($\Delta R < 0.5$) cones are organized into $\eta \times \phi$ grids, facilitating efficient processing by the CNN.

The network architecture includes the pre-processing of high-level variables and input cells through fully connected dense layers, the application of 2D convolutional layers to the processed cell data, extracting relevant features and finally it combines the processed features and additional dense layers to estimate the probabilities of a tau candidate being an electron, muon, quark or gluon jet, or a hadronic tau.

The final *DeepTau2018v2p5* model significantly improved agreement between data and MC, reducing discrepancies in the DeepTau discriminator score against jets from 13.3% to 0.80% in regions with high hadronic tau purity (>96%). However, domain adaptation significantly reduces discrepancies, it does not solve the fundamental issue of MC mismodeling, and in particular a slight degradation was observed in discrimination against electrons and muons. In any case, *DeepTau2018v2p5* outperforms above its predecessor in interesting kinematic regions for CMS benchmark analyses. A comprehensive description of *DeepTau2018v2p5* performances compared to *DeepTau2017v2p1* is reported in Refs. [205, 207].

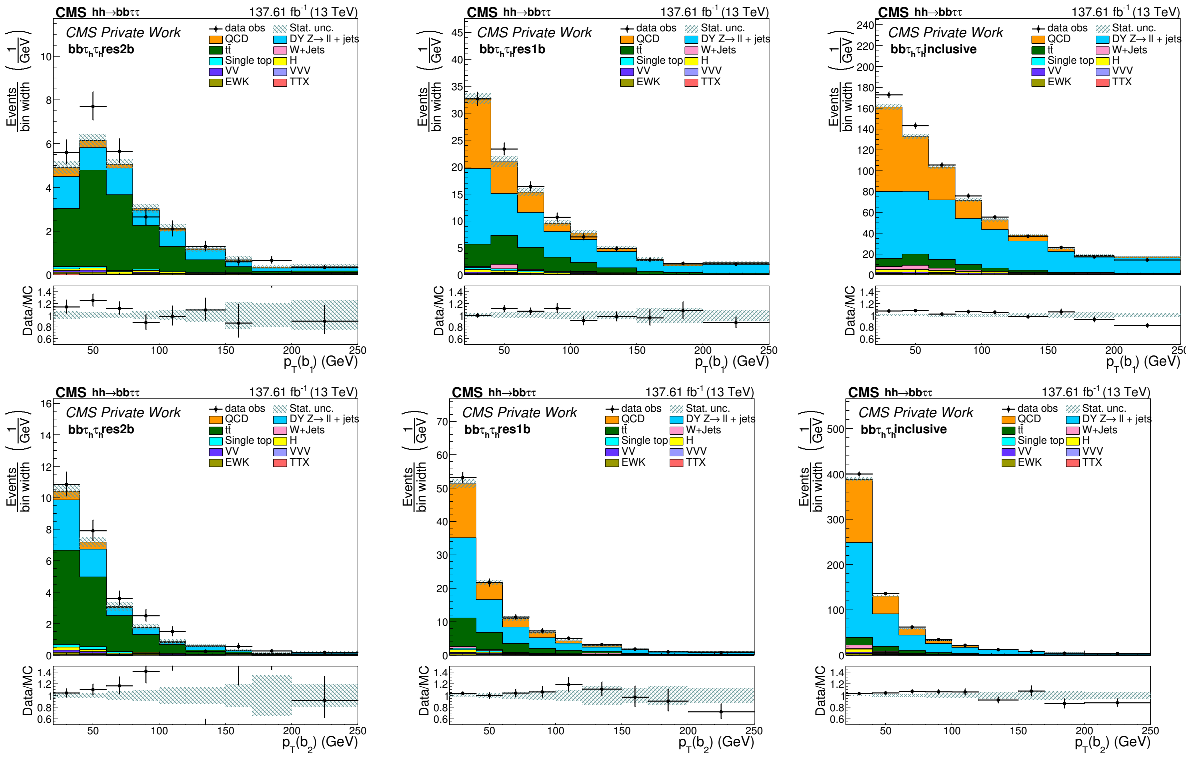
Although *DeepTau2018v2p5* has been deployed for use during the Run 3 data-taking period, testing it on Run 2 data is crucial to ensure backward compatibility, validate performance improvements in well-understood datasets, and facilitate seamless integration into combined Run 2 and early Run 3 analyses.

H.2. *DeepTau2018v2p5* in the $X \rightarrow HH \rightarrow b\bar{b}\tau\tau$ search

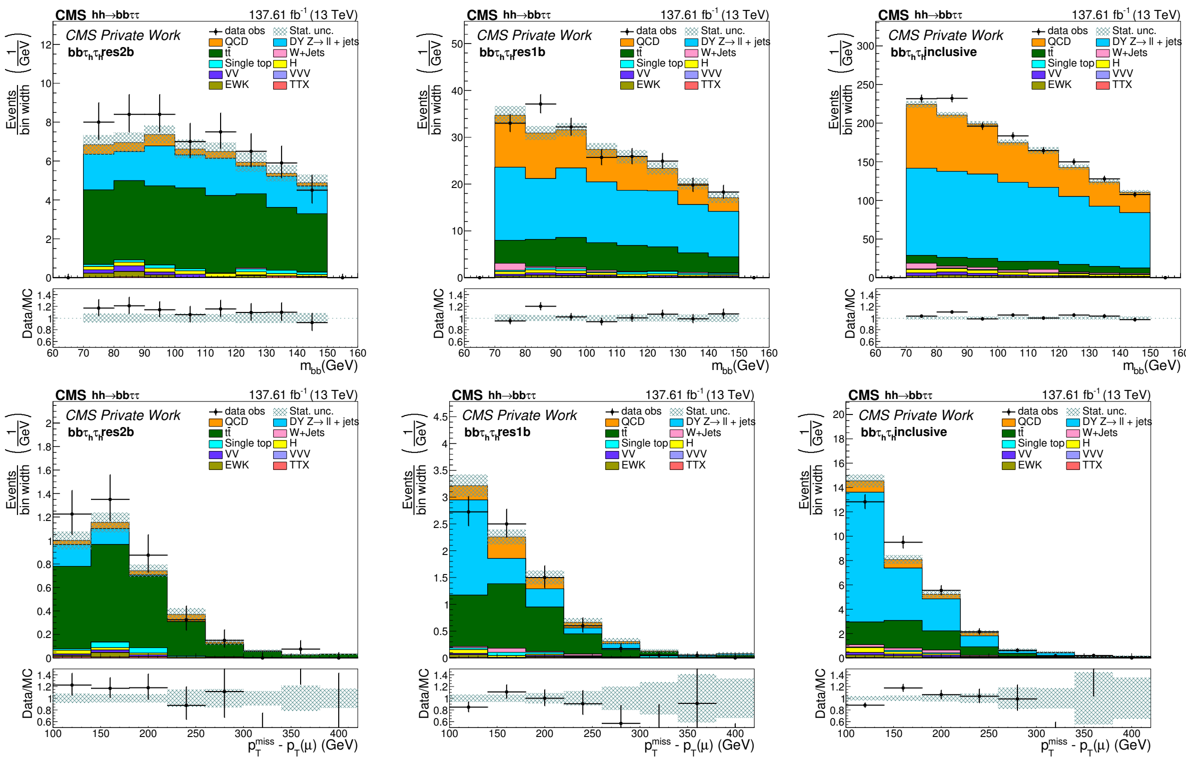
In this section the binned distributions obtained in the Signal Region (SR) are shown, for lepton and b-jet observables not sensitive to the signal. The distributions are presented for the and $\tau_h\tau_h$ channels. The SR is defined in Section 4.8.3, considering the inclusive, Resolved 1b category (res1b), Resolved 2b category (res2b) and boosted categories.

H.2.1 Resolved categories

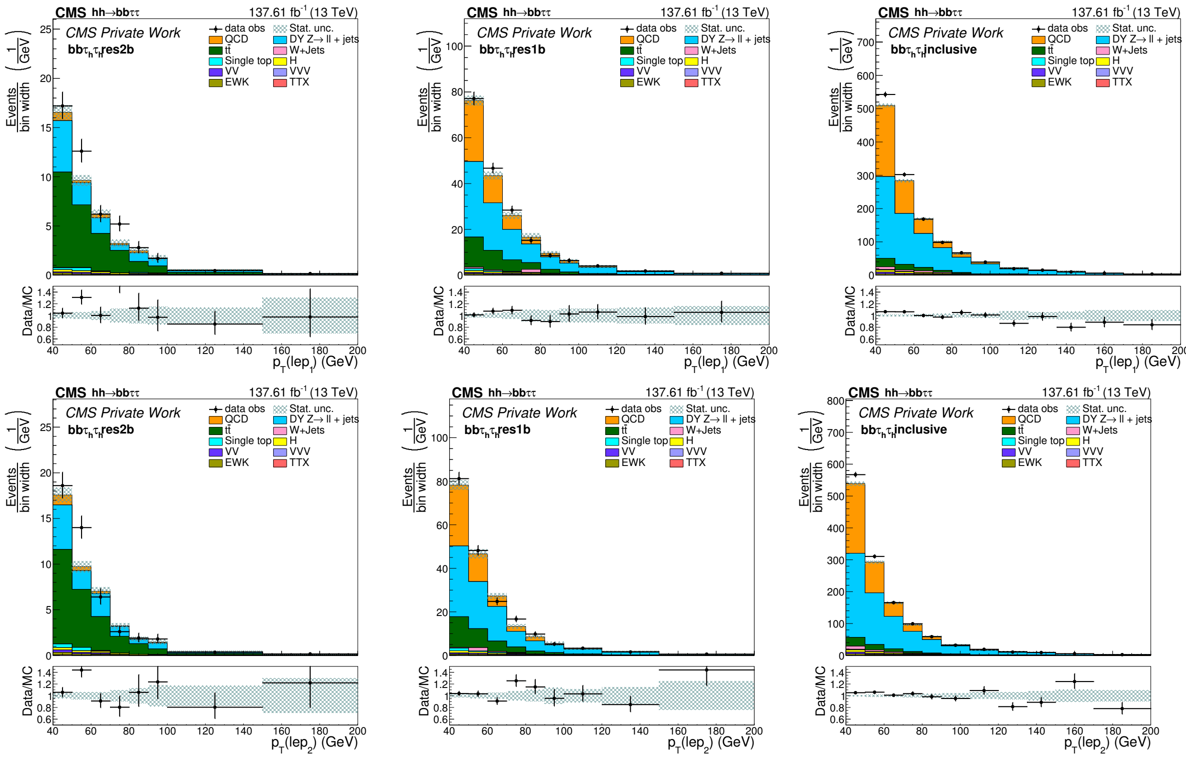
H.2.2 Boosted category



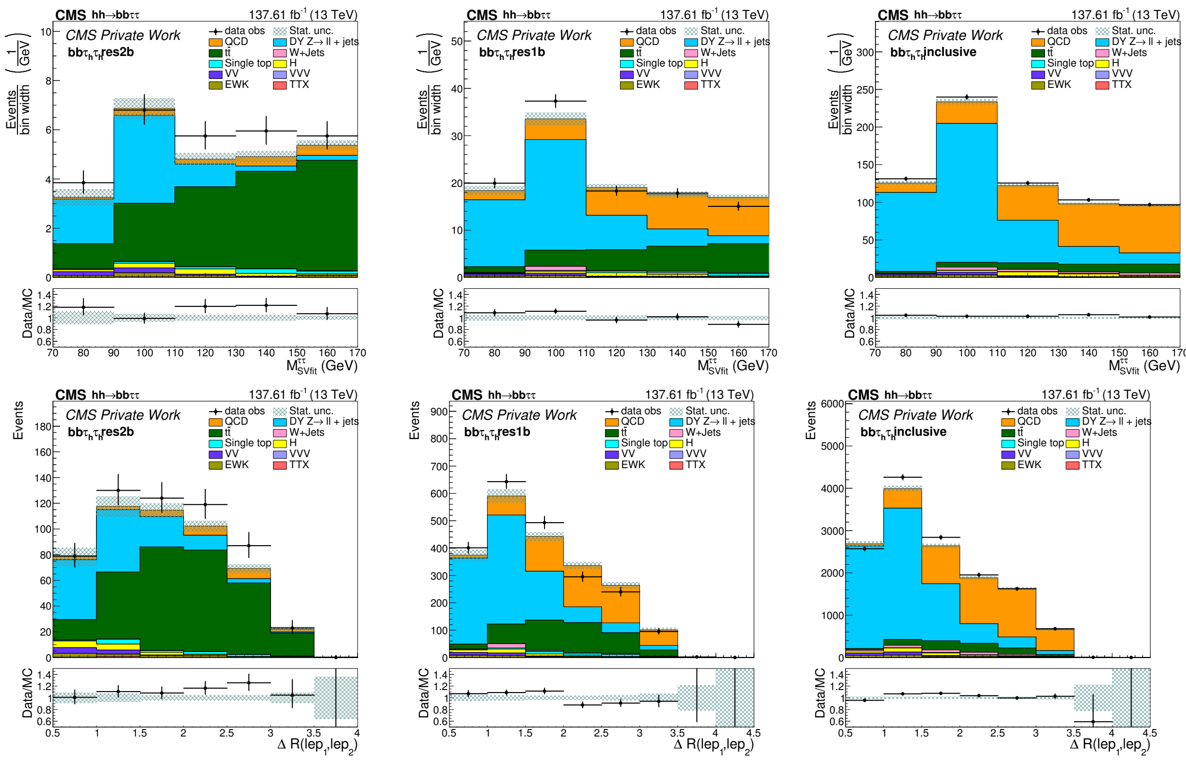
Distribution of the first (upper panel) and second (lower panel) b-jet transverse momentum component (p_T) , in the res2b (left), res1b (center) and inclusive (right) categories.



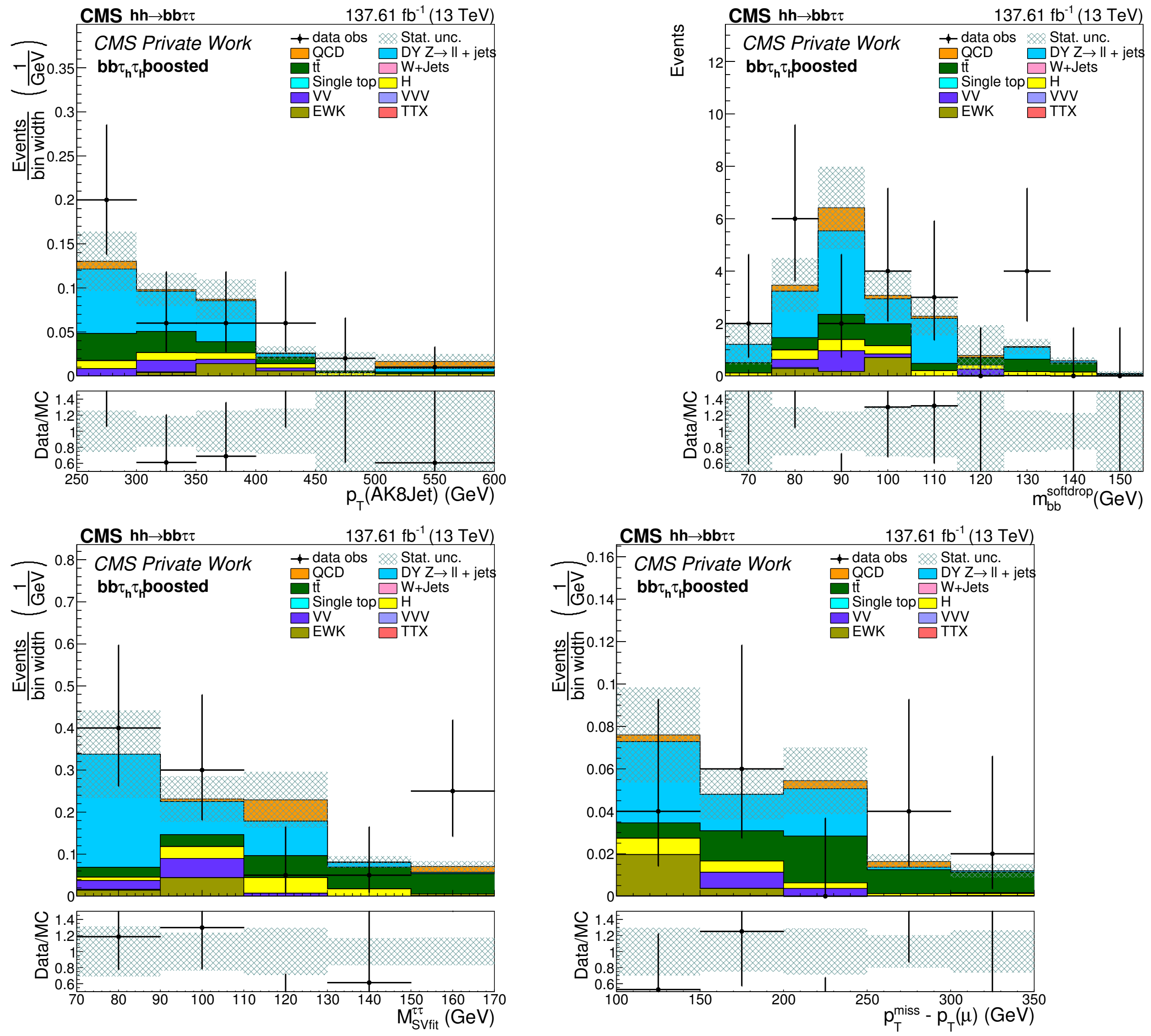
Top row: invariant mass of the two b-jets in the res2b (left), res1b (center) and inclusive (right) categories. Bottom row: p_T^{miss} -p_T (μ) distributions in the same categores/channel.



Distribution of the first (upper row) and second (lower row) lepton p_T , in the res2b (left), res1b (center) and inclusive (right) categories.



Di-lepton mass (top line) and angular separation between leptons (bottom line) in the res2b (left), res1b (center) and inclusive (right) categories.



Top row: Selected AK8 jet p_T (left) and softdrop mass (right) distribution in the boosted category. Bottom row: $m_{\tau\tau}^{(SVFit)}$(left) and p_T^{miss}-p_T (μ) (right) distributions in the boosted category.

Acronyms

AK Anti-k_T.
ATLAS A Toroidal LHC ApparatuS.
AUC Area Under the ROC Curve.

BDT Boosted Decision Tree.
BEH Brout-Englert-Higgs.
BPIX Pixel Barrel.
BR Branching Ratio.
BS Beam Spot.
BSM Beyond Standard Model.
BTV B-Tagging and Vertexing.
BX Bunch Crossing.

CERN Conseil Eurpoeen pour la Recherche Nucleare / European Center for the Nuclear Research.
CHS Charged Hadron Subtraction.
CKM Cabibbo-Kobayashi-Maskawa.
CL Confidence Level.
CMS Compact Muon Solenoid.
CNN Convolutional Neural Network.
CR Control Region.
CSC Cathode Strip Chamber.
CTF Combinatorial Track Finder.

DAQ Data Acquisition system.
DM Decay Mode.
DNN Deep Neural Network.
DT Drift Tube Chamber.
DY Drell Yan.
DYCR Drell Yan Control Region.

EB ECAL Barrel.
ECAL Electromagnetic Calorimeter.
EE ECAL Endcap.
EFT Effective Field Theory.
ES ECAL preshower.
EVF Event Filter Farm.
EWSB Electroweak symmetry breaking.

FPIX Pixel Forward.
FPR False Positive Rate.

GEM Gas Electron Multiplier.
ggF Gluon-gluon Fusion.
GNN Graph Neural Network.
GSF Gaussian Sum Filter.

HB HCAL Barrel.
HCAL Hadronic Calorimeter.
HE HCAL Endcap.
HEP experimental High Energy Physics.
HF Hadron Forward.
HH Higgs bosons pair.
HL-LHC HiLumi-LHC.
HLT High Level Trigger.
HO Hadron Outer.
HPS Hadron-Plus-Strips.

ID Identification.
IP Interaction Point.
Iso Isolated.
Iso Isolation.

JEC Jet Energy Corrections.
JER Jet Energy Resolution.
JES Jet Energy Scale.
JME Jet and MEt.

KF Kalman Filter.
KK Kaluza Klein.

L1 Level 1.
L1T Level 1 Trigger.
LEP Large Electron Positron Collider.
LHC Large Hadron Collider.
LO Leading Order.

MC Monte Carlo.
ME Matrix Element.
MET Missing Transverse Energy.
ML Machine Learning.
MLE Maximum Likelihood Estimator.
MSSM Minimal Supersymmetric Standard Model.
MVA Multi Variate Analysis.

NLL Negative Log-Likelihood.
NLO Next to Leading Order.
NN Neural Network.
NNLO Next-to-Next Leading Order.
Non-Iso Not Isolated.
NP Nuisance Parameter.

OS Opposite Sign.

$\mathbf{p}_T$ transverse momentum component.
$\mathbf{p}_Z$ longitudinal momentum component.
PF Particle Flow.
PMNS Pontecorvo-Maki-Nakagawa-Sakata.
POG Physics Object Group.
pp proton-proton.
PU pile-up.
PUPPI PileUp Per Particle Identification.
PV Primary Vertex.

QCD Quantum Chromodynamics.
QED Quantum Electrodynamics.
QFT Quantum Field Theory.

res1b Resolved 1b category.
res2b Resolved 2b category.
ROC Receiver Operating Characteristic.
RPC Resistive Plate Chamber.

SC SuperCluster.
SF Scale Factor.
SM Standard Model.
SR Signal Region.
SS Same Sign.
SSB Spontaneous symmetry breaking.
SUSY Super symmetry.

TEC Tracker Endcap.
TES τ_h Energy Scale.
TIB Tracker Inner Barrel.
TID Tracker Innter Disks.
TnP Tag and Probe.
TOB Tracker Outer Barrel.
TP Trigger Primitive.
TPR True Positive Rate.
TTCR $t\bar{t}$ Control Region.

VBF Vector Boson Fusion.
VdM Van der Meer.
VEV Vacuum Expectation Value.

WED Warped Extra Dimensions theory.
WP Working Point.

XSDB CMS Cross Section DataBase.

Bibliography

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

[2] CMS Collaboration, S. Chatrchyan *et al.*, "*Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*", Phys. Lett. B **716**, 30 (2012), https://doi.org/10.1016/j.physletb.2012.08.021.

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

[4] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, Reading, USA, 1995), https://doi.org/10.1017/CBO9780511622618.

[5] M. D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, 2014), https://doi.org/10.1017/9781139540940.

[6] T. Kajita, "*Nobel Lecture: Discovery of atmospheric neutrino oscillations*", Rev. Mod. Phys. **88**, 030501 (2016), https://link.aps.org/doi/10.1103/RevModPhys.88.030501.

[7] Particle Data Group, R. L. Workman *et al.*, "*Review of Particle Physics*", PTEP **2022**, 083C01 (2022), https://doi.org/10.1093/ptep/ptac097.

[8] D. J. Gross and F. Wilczek, "*Ultraviolet Behavior of Non-Aadbelian Gauge Theories*", Phys. Rev. Lett. **30**, 1343 (1973), https://doi.org/10.1103/PhysRevLett.30.1343.

[9] H. D. Politzer, "*Reliable Perturbative Results for Strong Interactions?*", Phys. Rev. Lett. **30**, 1346 (1973), https://doi.org/10.1103/PhysRevLett.30.1346.

[10] K. G. Wilson, "*Confinement of Quarks*", Phys. Rev. D **10**, 2445 (1974), https://doi.org/10.1103/PhysRevD.10.2445.

[11] Gell−Mann, Murray, *"The Eightfold Way: A Theory of strong interaction symmetry"*, (1961), https://doi.org/10.2172/4008239.

[12] R. P. Feynman, *"Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction"*, Phys. Rev. **80**, 440 (1950), https://link.aps.org/doi/10.1103/PhysRev.80.440.

[13] S. L. Glashow, *"Partial Symmetries of Weak Interactions"*, Nucl. Phys. **22**, 579 (1961), https://doi.org/10.1016/0029-5582(61)90469-2.

[14] A. Salam, *"Weak and Electromagnetic Interactions"*, Conf. Proc. C **680519**, 367 (1968), https://doi.org/10.1007/BF02726525.

[15] M. Thomson, *Modern particle physics* (Cambridge University Press, New York, 2013), https://doi.org/110.1017/CBO9781139525367.

[16] Z. Maki, M. Nakagawa, and S. Sakata, *"Remarks on the unified model of elementary particles"*, Prog. Theor. Phys. **28**, 870 (1962), https://doi.org/10.1143/PTP.28.870.

[17] Particle Data Group Collaboration, S. Navas *et al.*, *"Review of Particle Physics"*, Phys. Rev. D **110**, 030001 (2024), https://link.aps.org/doi/10.1103/PhysRevD.110.030001.

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

[19] P. W. Higgs, *"Spontaneous Symmetry Breakdown without Massless Bosons"*, Phys. Rev. **145**, 1156 (1966), https://link.aps.org/doi/10.1103/PhysRev.145.1156.

[20] F. Englert and R. Brout, *"Broken Symmetry and the Mass of Gauge Vector Mesons"*, Phys. Rev. Lett. **13**, 321 (1964), https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[21] P. W. Higgs, *"Broken Symmetries and the Masses of Gauge Bosons"*, Phys. Rev. Lett. **13**, 508 (1964), https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

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

[23] T. W. B. Kibble, *"Symmetry Breaking in Non-Abelian Gauge Theories"*, Phys. Rev. **155**, 1554 (1967), https://link.aps.org/doi/10.1103/PhysRev.155.1554.

[24] J. Goldstone, *"Field theories with «Superconductor »solutions"*, Il Nuovo Cimento (1955-1965) **19**, 154 (1961), https://doi.org/10.1007/BF02812722.

[25] J. Riebesell, Higgs potential, https://tikz.net/higgs-potential/.

[26] MuLan, D. M. Webber *et al.*, "*Measurement of the Positive Muon Lifetime and Determination of the Fermi Constant to Part-per-Million Precision*", Phys. Rev. Lett. **106**, 041803 (2011), https://doi.org/10.1103/PhysRevLett.106.079901.

[27] "*Precision electroweak measurements on the Z resonance*", Physics Reports **427**, 257 (2006), https://www.sciencedirect.com/science/article/pii/S0370157305005119.

[28] S. P. MARTIN, A supersymmetry primer, 1998, http://dx.doi.org/10.1142/9789812839657_0001.

[29] A. Carvalho, Gravity particles from warped extra dimensions, predictions for lhc, 2018, https://arxiv.org/abs/1404.0102, 1404.0102.

[30] The ATLAS and CMS Collaborations, G. Aad *et al.*, "*Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*", Physical Review Letters **114** (2015), http://dx.doi.org/10.1103/PhysRevLett.114.191803.

[31] CMS Collaboration, V. Khachatryan *et al.*, "*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** (2015), http://dx.doi.org/10.1103/PhysRevD.92.012004.

[32] ATLAS Collaboration, G. Aad *et al.*, "*Evidence for the spin-0 nature of the Higgs boson using ATLAS data*", Physics Letters B **726**, 120–144 (2013), http://dx.doi.org/10.1016/j.physletb.2013.08.026.

[33] V. Khachatryan *et al.*, "*Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8*

TeV

TeV", The European Physical Journal C **75** (2015), http://dx.doi.org/10.1140/epjc/s10052-015-3351-7.

[34] V. Khachatryan *et al.*, "*Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs*", Physics Letters B **736**, 64–85 (2014), http://dx.doi.org/10.1016/j.physletb.2014.06.077.

[35] CERN, Cern yellow reports: Monographs, vol 2 (2017): Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector, 2017, https://e-publishing.cern.ch/index.php/CYRM/issue/view/32.

[36] LHC Higgs Working Group, (LHCHWG), Lhc higgs cross section wg picture gallery, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGCrossSectionsFigures.

[37] CMS Collaboration, A. Tumasyan *et al.*, "*A portrait of the Higgs boson by the CMS experiment ten years after the discovery*", Nature **607**, 60 (2022), https://doi.org/10.1038/s41586-022-04892-x.

[38] J. Baglio *et al.*, "*The measurement of the Higgs self-coupling at the LHC: theoretical status*", Journal of High Energy Physics **2013** (2013), http://dx.doi.org/10.1007/JHEP04(2013)151.

[39] B. D. Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, "*Higgs boson potential at colliders: Status and perspectives*", Reviews in Physics **5**, 100045 (2020), http://dx.doi.org/10.1016/j.revip.2020.100045.

[40] F. Goertz, A. Papaefstathiou, L. L. Yang, and J. Zurita, "*Higgs boson pair production in the D = 6 extension of the SM*", Journal of High Energy Physics **2015** (2015), http://dx.doi.org/10.1007/JHEP04(2015)167.

[41] A. Carvalho *et al.*, "*Higgs pair production: choosing benchmarks with cluster analysis*", Journal of High Energy Physics **2016**, 1–28 (2016), http://dx.doi.org/10.1007/JHEP04(2016)126.

[42] R. Frederix *et al.*, "*Higgs pair production at the LHC with NLO and parton-shower effects*", Physics Letters B **732**, 142–149 (2014), http://dx.doi.org/10.1016/j.physletb.2014.03.026.

[43] T. Gherghetta, Tasi lectures on a holographic view of beyond the standard model physics, 2010, https://arxiv.org/abs/1008.2570, 1008.2570.

[44] A. Anglés-Castillo and A. Pérez, "*A quantum walk simulation of extra dimensions with warped geometry*", Scientific Reports **12**, 1926 (2022), https://doi.org/10.1038/s41598-022-05673-2.

[45] E. PONTÓN, *Four Lectures on TeV Scale Extra Dimensions* (WORLD SCIENTIFIC, 2013), p. 283–374, http://dx.doi.org/10.1142/9789814390163_0007.

[46] L. Randall and R. Sundrum, "*Large Mass Hierarchy from a Small Extra Dimension*", Physical Review Letters **83**, 3370–3373 (1999), http://dx.doi.org/10.1103/PhysRevLett.83.3370.

[47] H. Collins and B. Holdom, "*Randall-Sundrum scenario with an extra warped dimension*", Physical Review D **64** (2001), http://dx.doi.org/10.1103/PhysRevD.64.064003.

[48] T. Gherghetta and A. Pomarol, "*Bulk fields and supersymmetry in a slice of AdS*", Nuclear Physics B **586**, 141–162 (2000), http://dx.doi.org/10.1016/S0550-3213(00)00392-8.

[49] W. D. Goldberger and M. B. Wise, "*Modulus Stabilization with Bulk Fields*", Physical Review Letters **83**, 4922–4925 (1999), http://dx.doi.org/10.1103/PhysRevLett.83.4922.

[50] CMS Collaboration, Searches for higgs boson production through decays of heavy resonances, 2024, https://arxiv.org/abs/2403.16926, 2403.16926.

[51] H. Davoudiasl, S. Gopalakrishna, E. Pontón, and J. Santiago, *"Warped five-dimensional models: phenomenological status and experimental prospects"*, New Journal of Physics **12**, 075011 (2010), http://dx.doi.org/10.1088/1367-2630/12/7/075011.

[52] K. S. Agashe *et al.*, *"LHC signals from cascade decays of warped vector resonances"*, Journal of High Energy Physics **2017** (2017), http://dx.doi.org/10.1007/JHEP05(2017)078.

[53] A. Oliveira and R. Rosenfeld, *"Hidden sector effects on double Higgs production near threshold at the LHC"*, Physics Letters B **702**, 201–204 (2011), http://dx.doi.org/10.1016/j.physletb.2011.06.086.

[54] CMS Collaboration, A. M. Sirunyan *et al.*, *"Search for high-mass resonances in dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV"*, Journal of High Energy Physics **2018** (2018), http://dx.doi.org/10.1007/JHEP06(2018)120.

[55] G. F. Giudice, R. Rattazzi, and J. D. Wells, *"Graviscalars from higher-dimensional metrics and curvature-Higgs mixing"*, Nuclear Physics B **595**, 250–276 (2001), http://dx.doi.org/10.1016/S0550-3213(00)00686-6.

[56] H. de Sandes and R. Rosenfeld, *"Radion-Higgs mixing effects on bounds from LHC Higgs boson searches"*, Phys. Rev. D **85**, 053003 (2012), https://link.aps.org/doi/10.1103/PhysRevD.85.053003.

[57] C. Csáki, J. Hubisz, and S. J. Lee, *"Radion phenomenology in realistic warped space models"*, Physical Review D **76** (2007), http://dx.doi.org/10.1103/PhysRevD.76.125015.

[58] C. Csáki, M. L. Graesser, and G. D. Kribs, *"Radion dynamics and electroweak physics"*, Physical Review D **63** (2001), http://dx.doi.org/10.1103/PhysRevD.63.065002.

[59] G. Aad *et al.*, *"A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery"*, Nature **607**, 52 (2022), https://doi.org/10.1038/s41586-022-04893-w.

[60] G. Aad *et al.*, *"Evidence for the Higgs Boson Decay to a Z Boson and a Photon at the LHC"*, Physical Review Letters **132** (2024), http://dx.doi.org/10.1103/PhysRevLett.132.021803.

[61] M. Gouzevitch and A. Carvalho, *"A review of Higgs boson pair production"*, Reviews in Physics **5**, 100039 (2020), https://www.sciencedirect.com/science/article/pii/S2405428320300022.

[62] M. J. Dolan, C. Englert, and M. Spannowsky, *"Higgs self-coupling measurements at the LHC"*, JHEP **10**, 112 (2012), http://dx.doi.org/10.1007/JHEP10(2012)112.

[63] ATLAS, G. Aad *et al.*, *"Combination of Searches for Higgs Boson Pair Production in pp Collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS Detector"*, Phys. Rev. Lett. **133**, 101801 (2024), https://link.aps.org/doi/10.1103/PhysRevLett.133.101801.

[64] CMS Collaboration, *"Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at sqrt(s) = 13 TeV"*, (2024), https://cds.cern.ch/record/2917252.

[65] ATLAS Collaboration, G. Aad *et al.*, *"Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at* $\sqrt{s}$ *= 13 TeV"*, Physics Letters B **843**, 137745 (2023), http://dx.doi.org/10.1016/j.physletb.2023.137745.

[66] A. Hayrapetyan *et al.*, *"Constraints on the Higgs boson self-coupling from the combination of single and double Higgs boson production in proton-proton collisions at s=13TeV"*, Physics Letters B **861**, 139210 (2025), https://www.sciencedirect.com/science/article/pii/S0370269324007688.

[67] ATLAS Collaboration, G. Aad *et al.*, *"Combination of Searches for Resonant Higgs Boson Pair Production Using pp Collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS Detector"*, Physical Review Letters **132** (2024), http://dx.doi.org/10.1103/PhysRevLett.132.231801.

[68] E. Gross and O. Vitells, *"Trial factors for the look elsewhere effect in high energy physics"*, The European Physical Journal C **70**, 525 (2010), https://doi.org/10.1140/epjc/s10052-010-1470-8.

[69] E. Lopienska, *"The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022"*, (2022), https://cds.cern.ch/record/2800984, General Photo.

[70] CMS Collaboration, S. Chatrchyan *et al.*, *"The CMS Experiment at the CERN LHC"*, JINST **3**, S08004 (2008), https://cds.cern.ch/record/1129810.

[71] ATLAS Collaboration, G. Aad *et al.*, *"The ATLAS Experiment at the CERN Large Hadron Collider"*, JINST **3**, S08003 (2008), https://cds.cern.ch/record/1129811.

[72] LHCb Collaboration, A. A. Alves, Jr. *et al.*, *"The LHCb Detector at the LHC"*, JINST **3**, S08005 (2008), https://cds.cern.ch/record/1129809.

[73] ALICE Collaboration, K. Aamodt *et al.*, *"The ALICE experiment at the CERN LHC"*, JINST **3**, S08002 (2008), https://cds.cern.ch/record/1129812.

[74] TOTEM Collaboration, G. Anelli *et al.*, *"The TOTEM experiment at the CERN Large Hadron Collider"*, JINST **3**, S08007 (2008), https://cds.cern.ch/record/1129807.

[75] LHCf Collaboration, O. Adriani *et al.*, *"The LHCf experiment at LHC"*, Czech. J. Phys. **56**, A107 (2006), https://doi.org/10.1007/s10582-006-0147-x.

[76] MoEDAL Collaboration, J. Pinfold, *"The MoEDAL experiment at the LHC"*, EPJ Web Conf. **145**, 12002 (2017), http://dx.doi.org/10.1051/epjconf/201614512002.

[77] FASER Collaboration, J. Boyd, *"FASER Experiment"*, (2023), https://doi.org/10.1142/9789811280184_0019.

[78] FASER Collaboration, H. Abreu *et al.*, *"Technical Proposal: FASERnu"*, (2020), https://arxiv.org/abs/2001.03073.

[79] SND@LHC Collaboration, E. Zaffaroni, *"SND@LHC - A new Scattering and Neutrino Detector at the LHC"*, PoS **PANIC2021**, 280 (2022), https://doi.org/10.22323/1.414.1236.

[80] CMS Collaboration, Public CMS Luminosity Information, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[81] The HiLumi-LHC Project, https://hilumilhc.web.cern.ch/article/ls3-schedule-change.

[82] CMS Collaboration, G. L. Bayatian *et al.*, *"CMS Physics: Technical Design Report Volume 1: Detector Performance and Software"*, (2006), https://cds.cern.ch/record/922757.

[83] CMS Collaboration, S. Chatrchyan *et al.*, *"Aligning the CMS muon chambers with the muon alignment system during an extended cosmic ray run"*, JINST **5**, T03019 (2010), https://cds.cern.ch/record/1223872.

[84] CMS Collaboration, V. Karimäki *et al.*, The CMS tracker system project: Technical Design Report, 1997, https://cds.cern.ch/record/368412.

[85] CMS Collaboration, The CMS tracker: addendum to the Technical Design Report, 2000, https://cds.cern.ch/record/490194.

[86] CMS Collaboration, *"Commissioning and performance of the CMS silicon strip tracker with cosmic ray muons"*, Journal of Instrumentation **5**, T03008 (2010), https://dx.doi.org/10.1088/1748-0221/5/03/T03008.

[87] P. Azzurri, *"The CMS Silicon Strip Tracker"*, J. Phys.: Conf. Ser. **41**, 127 (2006), https://cds.cern.ch/record/914891, 8 pages, 8 figures, talk given at XIX EPS NPDC Conference on New Trends in Nuclear Physics Applications and Technology, September 5-9, 2005 Pavia, Italy Subj-class: Instrumentation and Detectors.

[88] L. Borrello, A. Messineo, E. Focardi, and A. Macchiolo, *"Sensor Design for the CMS Silicon Strip Tracker"*, (2003), https://cds.cern.ch/record/687861.

[89] C. Collaboration, *"Description and performance of track and primary-vertex reconstruction with the CMS tracker"*, Journal of Instrumentation **9**, P10009 (2014), https://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[90] CMS Collaboration, W. Adam *et al.*, *"The CMS Phase-1 pixel detector upgrade"*, Journal of Instrumentation **16**, P02027 (2021), https://dx.doi.org/10.1088/1748-0221/16/02/P02027.

[91] L. Viliani, *"CMS tracker performance and readiness for LHC Run II"*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **824**, 67 (2016), https://www.sciencedirect.com/science/article/pii/S0168900215011018, Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[92] Tracker Material Budget plots Twiki, https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots.

[93] CMS Collaboration, "the cms electromagnetic calorimeter project: Technical design report ", 1997, https://cds.cern.ch/record/349375.

[94] CMS Collaboration, The cms hadron calorimeter project: Technical design report, 1997, https://cds.cern.ch/record/357153.

[95] CMS Collaboration, S. Chatrchyan *et al.*, *"Performance and Operation of the CMS Electromagnetic Calorimeter"*, JINST **5**, T03010 (2010), https://cds.cern.ch/record/1213246.

[96] CMS Collaboration, A. Benaglia, *"The CMS ECAL performance with examples"*, Journal of Instrumentation **9**, C02008 (2014), https://dx.doi.org/10.1088/1748-0221/9/02/C02008.

[97] CMS Collaboration, M. Chadeeva and N. Lychkovskaya, *"Calibration of the CMS hadron calorimeter in Run 2"*, Journal of Instrumentation **13**, C03025 (2018), https://dx.doi.org/10.1088/1748-0221/13/03/C03025.

[98] CMS Collaboration, The CMS hadron calorimeter, 2008, https://jinst.sissa.it/LHC/CMS/2008_JINST_3_S08004.pdf.

[99] CMS Collaboration, C. Isik, *"Phase 1 upgrade of the CMS Hadron Barrel Calorimeter"*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1042**, 167389 (2022), https://www.sciencedirect.com/science/article/pii/S0168900222006933.

[100] CMS Collaboration, S. Chatrchyan *et al.*, *"Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data. Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data"*, JINST **5**, T03012 (2010), https://cds.cern.ch/record/1223950.

[101] CMS Collaboration, *The CMS magnet project: Technical Design Report*Technical design report. CMS (CERN, Geneva, 1997), https://cds.cern.ch/record/331056.

[102] CMS Collaboration, S. Chatrchyan *et al.*, "*The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV at the LHC*", JINST **8**, P11002 (2013), https://cds.cern.ch/record/1558674, Comments: Submitted to JINST.

[103] CMS Collaboration, J. G. Layter, *The CMS muon project: Technical Design Report*Technical design report. CMS (CERN, Geneva, 1997), https://cds.cern.ch/record/343814.

[104] CMS Collaboration, A. M. Sirunyan *et al.*, "*Performance of the CMS muon detector and muon reconstruction with pp collisions at $\sqrt{s} = 13$ TeV*", JINST **13**, P06015 (2018), http://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[105] D. Abbaneo *et al.*, A gem detector system for an upgrade of the high-eta muon endcap stations ge1/1 + me1/1 in cms, 2012, https://arxiv.org/abs/1211.1494, 1211.1494.

[106] CMS Collaboration, G. L. Bayatyan *et al.*, "*CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*", (2002), https://cds.cern.ch/record/706847.

[107] CMS Collaboration, S. Cittolin, A. Rácz, and P. Sphicas, "*CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*", (2002), https://cds.cern.ch/record/578006.

[108] G. Bauer *et al.*, "*The data-acquisition system of the CMS experiment at the LHC*", Journal of Physics: Conference Series **331**, 022021 (2011), https://dx.doi.org/10.1088/1742-6596/331/2/022021.

[109] CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV, 2019, https://cds.cern.ch/record/2676164.

[110] S. van der Meer, Calibration of the effective beam height in the ISR, 1968, https://cds.cern.ch/record/296752/.

[111] CMS Collaboration, E. Ayala *et al.*, "*The Pixel Luminosity Telescope: a detector for luminosity measurement at CMS using silicon pixel sensors*", The European Physical Journal C **83** (2023), http://dx.doi.org/10.1140/epjc/s10052-023-11713-6.

[112] CMS Collaboration, M. Guthoff, The new fast beam condition monitor using poly-crystalline diamond sensors for luminosity measurement at cms. the new fast beam condition monitor using diamond and silicon sensors for luminosity measurement at cms, 2019, https://cds.cern.ch/record/2799585.

[113] D. Forkel-Wirth, D. Perrin, L. Scibile, G. Segura, and P. Vojtyla, "*Radiation monitoring system for the environment and safety project*", (2002), https://cds.cern.ch/record/687619.

[114] CMS Collaboration, Cms luminosity based on pixel cluster counting - summer 2012 update, 2012, https://cds.cern.ch/record/1482193.

[115] CMS Collaboration, A. M. Sirunyan *et al.*, *"Particle−flow reconstruction and global event description with the CMS detector"*, JINST **12**, P10003 (2017), https://doi.org/10.1088/1748-0221/12/10/P10003.

[116] N. Dawe, https://github.com/ndawe/tikz-track.

[117] R. Frühwirth, *"Application of Kalman filtering to track and vertex fitting"*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262**, 444 (1987), https://www.sciencedirect.com/science/article/pii/0168900287908874.

[118] W. Adam, B. Mangano, T. Speer, and T. Todorov, Track Reconstruction in the CMS tracker, 2006, https://cds.cern.ch/record/934067.

[119] CMS Collaboration, *"2017 tracking performance plots"*, (2017), https://cds.cern.ch/record/2290524.

[120] CMS Tracking POG Performance Plots For 2017 with PhaseI pixel detector, https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPerformance2017MC.

[121] CMS Collaboration, *"Tracking Performance in the CMS High Level Trigger - June 2018"*, (2018), https://cds.cern.ch/record/2629370.

[122] CMS Collaboration, Tracking performance in the CMS High Level trigger, https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTplots2018DataTracking.

[123] K. Rose, *"Deterministic annealing for clustering, compression, classification, regression, and related optimization problems"*, Proceedings of the IEEE **86**, 2210 (1998), https://doi.org/10.1109/5.726788.

[124] W. Waltenberger, R. Frühwirth, and P. Vanlaer, *"Adaptive vertex fitting"*, Journal of Physics G: Nuclear and Particle Physics **34**, N343 (2007), https://dx.doi.org/10.1088/0954-3899/34/12/N01.

[125] ALEPH, DELPHI, L3, OPAL, Collaborations and LEP Electroweak Working Group, *"Electroweak measurements in electron–positron collisions at W-boson-pair energies at LEP"*, Physics Reports **532**, 119 (2013), https://www.sciencedirect.com/science/article/pii/S0370157313002706, Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP.

[126] CMS Collaboration, *"Muon identification and isolation efficiencies with 2017 and 2018 data"*, (2018), https://cds.cern.ch/record/2629364.

[127] CMS Collaboration, *"Muon Reconstruction and Identification Performance with Run-2 data"*, (2020), https://cds.cern.ch/record/2727091.

[128] CMS Collaboration, *"Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-probe technique"*, (2020), https://cds.cern.ch/record/2724492.

[129] CMS Collaboration, A. Sirunyan *et al.*, *"Performance of the CMS muon trigger system in pp collisions at $\sqrt{s} = 13$ TeV"*, Journal of Instrumentation **16**, P07001 (2021), http://dx.doi.org/10.1088/1748-0221/16/07/P07001.

[130] H. A. Bethe and L. C. Maximon, *"Theory of Bremsstrahlung and Pair Production. I. Differential Cross Section"*, Phys. Rev. **93**, 768 (1954), https://link.aps.org/doi/10.1103/PhysRev.93.768.

[131] CMS Collaboration, *"Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s}$ = 8 TeV"*, Journal of Instrumentation , P08010–P08010 (2015), http://dx.doi.org/10.1088/1748-0221/10/08/P08010.

[132] CMS Collaboration, A. Sirunyan *et al.*, *"Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC"*, Journal of Instrumentation **16**, P05014 (2021), https://dx.doi.org/10.1088/1748-0221/16/05/P05014.

[133] M. Cacciari, G. P. Salam, and G. Soyez, *"The anti-kt jet clustering algorithm"*, Journal of High Energy Physics **2008**, 063 (2008), https://dx.doi.org/10.1088/1126-6708/2008/04/063.

[134] M. Cacciari, G. P. Salam, and G. Soyez, *"FastJet user manual: (for version 3.0.2)"*, The European Physical Journal C **72** (2012), http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[135] CMS Collaboration, A. M. Sirunyan *et al.*, *"Pileup mitigation at CMS in 13 TeV data"*, JINST **15**, P09018 (2020), https://doi.org/10.1088/1748-0221/15/09/P09018.

[136] D. Bertolini, P. Harris, M. Low, and N. Tran, *"Pileup Per Particle Identification"*, JHEP **10**, 059 (2014).

[137] A. Malara, Reconstruction of jets and missing transverse momentum at the cms experiment: Run 2 and perspective for run 3, 2021, https://arxiv.org/abs/2012.06271, 2012.06271.

[138] CMS Collaboration, *"Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV"*, (2021), https://cds.cern.ch/record/2792322.

[139] V. Khachatryan *et al.*, *"Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV"*, Journal of Instrumentation **12**, P02014–P02014 (2017), http://dx.doi.org/10.1088/1748-0221/12/02/P02014.

[140] CMS Collaboration, Jet algorithms performance in 13 TeV data, 2017, https://cds.cern.ch/record/2256875.

[141] CMS Collaboration, *"Performance of the pile up jet identification in CMS for Run 2"*, (2020), https://cds.cern.ch/record/2715906.

[142] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, *"Jet Flavour Classification Using DeepJet"*, JINST **15**, P12012 (2020), https://cds.cern.ch/record/2730134, 14 pages, 9 figures, accepted for publication in JINST.

[143] CMS Collaboration, "*Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector*", (2018), https://cds.cern.ch/record/2627468.

[144] CMS Collaboration, "*Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques*", (2020), https://cds.cern.ch/record/2707946.

[145] CMS Collaboration, Performance of heavy-flavour jet identification in boosted topologies in pp collisions at $\sqrt{s} = 13$ TeV, 2023, https://cds.cern.ch/record/2866276.

[146] CMS Collaboration, "*Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS*", Journal of Instrumentation **11**, P01019 (2016), https://dx.doi.org/10.1088/1748-0221/11/01/P01019.

[147] CMS Collaboration, "*Performance of τ-lepton reconstruction and identification in CMS*", Journal of Instrumentation **7**, P01001 (2012), https://dx.doi.org/10.1088/1748-0221/7/01/P01001.

[148] CMS Collaboration, A. Tumasyan *et al.*, "*Identification of hadronic tau lepton decays using a deep neural network*", JINST **17**, P07023 (2022), https://doi.org/10.1088/1748-0221/17/07/P07023.

[149] PDG table for τ leptons, https://pdg.lbl.gov/2024/listings/rpp2024-list-tau.pdf.

[150] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016), http://www.deeplearningbook.org.

[151] CMS Collaboration, A. Sirunyan *et al.*, "*Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s}$=13 TeV*", Journal of Instrumentation **13**, P10005–P10005 (2018), http://dx.doi.org/10.1088/1748-0221/13/10/P10005.

[152] CMS Collaboration, A. Sirunyan *et al.*, "*Performance of missing transverse momentum reconstruction in pp collisions at $\sqrt{s}$ = 13 TeV using the CMS detector*", Journal of Instrumentation **14**, P07004–P07004 (2019), http://dx.doi.org/10.1088/1748-0221/14/07/P07004.

[153] CMS Collaboration, A. Sirunyan *et al.*, "*Search for heavy resonances decaying into two Higgs bosons or into a Higgs boson and a W or Z boson in pp collisions at 13 TeV*", JHEP **2019**, 051 (2019), https://doi.org/10.1007/JHEP01(2019)051.

[154] CMS, "*Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data*", (2020), https://cds.cern.ch/record/2740688.

[155] CMS, F. Cavallari and C. I. Rovelli, "*Calibration and Performance of the CMS Electromagnetic Calorimeter in LHC Run2*", (2020), https://cds.cern.ch/record/2798128.

[156] CMS Collaboration, A. Tumasyan *et al.*, "*Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two τ leptons in ppcollisions at $\sqrt{s}$=13 TeV*", Physics Letters B **842**, 137531 (2023), http://dx.doi.org/10.1016/j.physletb.2022.137531.

[157] The CMS Collaboration, MET Analysis.

[158] A. Tumasyan *et al.*, "*Measurements of Higgs boson production in the decay channel with a pair of τ leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV*", The European Physical Journal C **83** (2023), http://dx.doi.org/10.1140/epjc/s10052-023-11452-8.

[159] S. L. n. S. G. Marzani S., "*The jet mass distribution after Soft Drop*", Eur Phys J C Part Fields **78** (2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6560847/.

[160] V. Lendermann *et al.*, "*Combining Triggers in HEP data analysis*", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **604**, 707 (2009), https://www.sciencedirect.com/science/article/pii/S0168900209006627.

[161] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, "*Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques*", Journal of Physics: Conference Series **513**, 022035 (2014), https://dx.doi.org/10.1088/1742-6596/513/2/022035.

[162] L. Serkin, "*Treatment of top-quark backgrounds in extreme phase spaces: the "top p_T reweighting" and novel data-driven estimations in ATLAS and CMS*", (2021).

[163] N. Kidonakis, Nnll threshold resummation for top-pair and single-top production, 2012, https://arxiv.org/abs/1210.7813, 1210.7813.

[164] M. Czakon, D. Heymes, and A. Mitov, "*High-precision differential predictions for top-quark pairs at the LHC*", Phys. Rev. Lett. **116**, 082003 (2016), https://doi.org/10.1103/PhysRevLett.116.082003.

[165] M. Czakon *et al.*, "*Top-pair production at the LHC through NNLO QCD and NLO EW*", JHEP **10**, 186 (2017), https://doi.org/10.1007/JHEP10(2017)186.

[166] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, and J. Mazzitelli, "*Top-quark pair production at the LHC: Fully differential QCD predictions at NNLO*", JHEP **07**, 100 (2019), https://doi.org/10.1007/JHEP07(2019)100.

[167] C. Bierlich *et al.*, A comprehensive guide to the physics and usage of pythia 8.3, 2022, https://arxiv.org/abs/2203.11601, 2203.11601.

[168] J. Alwall *et al.*, "*The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*", Journal of High Energy Physics **2014** (2014), http://dx.doi.org/10.1007/JHEP07(2014)079.

[169] P. Nason, *"A New method for combining NLO QCD with shower Monte Carlo algorithms"*, JHEP **11**, 040 (2004), https://arxiv.org/abs/hep-ph/0409146.

[170] S. Frixione, P. Nason, and C. Oleari, *"Matching NLO QCD computations with Parton Shower simulations: the POWHEG method"*, JHEP **11**, 070 (2007), https://arxiv.org/abs/0709.2092.

[171] S. Alioli, P. Nason, C. Oleari, and E. Re, *"A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX"*, JHEP **06**, 043 (2010), https://arxiv.org/abs/1002.2581.

[172] J. Allison *et al.*, *"Geant4 developments and applications"*, IEEE Transactions on Nuclear Science **53**, 270 (2006), https://doi.org/10.1109/TNS.2006.869826.

[173] S. Agostinelli *et al.*, *"Geant4—a simulation toolkit"*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003), https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[174] CMS Collaboration, The cms nanoaod data tier, https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD.

[175] R. Brun and F. Rademakers, *"ROOT: An object oriented data analysis framework"*, Nucl. Instrum. Meth. A **389**, 81 (1997).

[176] CMS Collaboration, , Nnlo+nnll top-quark-pair cross sections twiki, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO.

[177] Top quark pdg review, https://pdg.lbl.gov/2024/reviews/rpp2024-rev-top-quark.pdf.

[178] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *"Matching matrix elements and shower evolution for top-pair production in hadronic collisions"*, Journal of High Energy Physics **2007**, 013–013 (2007), http://dx.doi.org/10.1088/1126-6708/2007/01/013.

[179] Mlm jet merging, https://pythia.org/latest-manual/JetMatching.html.

[180] CMS Collaboration, , Nlo single-top channel cross sections twiki, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopNNLORef.

[181] T. Gehrmann *et al.*, *"W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD"*, Physical Review Letters **113** (2014), http://dx.doi.org/10.1103/PhysRevLett.113.212001.

[182] J. M. Campbell, R. K. Ellis, and C. Williams, *"Vector boson pair production at the LHC"*, Journal of High Energy Physics **2011**, 18 (2011), https://doi.org/10.1007/JHEP07(2011)018.

[183] Y. Li and F. Petriello, *"Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code"*, Phys. Rev. D **86**, 094034 (2012), https://link.aps.org/doi/10.1103/PhysRevD.86.094034.

[184] LHC Higgs Cross Section Working Group, Sm higgs branching ratios and total decay widths (update in cern report4 2016), https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR.

[185] K. Ehatäht and C. Veelken, *"Stitching Monte Carlo samples"*, Eur. Phys. J. C **82**, 484 (2022), https://doi.org/10.1140/epjc/s10052-022-10407-9.

[186] J. Butterworth *et al.*, *"PDF4LHC recommendations for LHC Run II"*, Journal of Physics G: Nuclear and Particle Physics **43**, 023001 (2016), http://dx.doi.org/10.1088/0954-3899/43/2/023001.

[187] LHC Higgs Cross Section Working Group, (LHCHXSWG), Sm higgs production cross sections at $\sqrt{s}$ = 13 tev (update in cern report4 2016), https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV.

[188] T. C. collaboration, *"Determination of jet energy calibration and transverse momentum resolution in CMS"*, Journal of Instrumentation **6**, P11002–P11002 (2011), http://dx.doi.org/10.1088/1748-0221/6/11/P11002.

[189] R. Barlow and C. Beeston, *"Fitting using finite Monte Carlo samples"*, Comput. Phys. Commun. **77**, 219 (1993), https://cds.cern.ch/record/249779.

[190] F. James, *"MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1"*, (1994), https://cds.cern.ch/record/2296388/files/minuit.pdf.

[191] ATLAS, CMS, LHC Higgs Combination Group, *"Procedure for the LHC Higgs boson search combination in Summer 2011"*, (2011), http://cds.cern.ch/record/1379837.

[192] V. Khachatryan *et al.*, *"Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$"*, Physics Letters B **755**, 217 (2016), https://www.sciencedirect.com/science/article/pii/S0370269316000757.

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

[194] A. Sirunyan *et al.*, *"Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV"*, Physics Letters B **778**, 101–127 (2018), http://dx.doi.org/10.1016/j.physletb.2018.01.001.

[195] J. Neyman and E. S. Pearson, *"On the Problem of the Most Efficient Tests of Statistical Hypotheses"*, Phil. Trans. Roy. Soc. Lond. A **231**, 289 (1933), https://doi.org/10.1098/rsta.1933.0009.

[196] L. Lista, *"Statistical Methods for Data Analysis in Particle Physics"*, **909** (2016), https://doi.org/10.1007/978-3-319-20176-4.

[197] G. Cowan, *"Statistical Data Analysis"*, **909** (1998), https://www.pp.rhul.ac.uk/~cowan/sda/.

[198] S. S. Wilks, *"The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses"*, Annals Math. Statist. **9**, 60 (1938), https://doi.org/10.1214/aoms/1177732360.

[199] K. Androsov, M. A. Ciocci, and M. R. Di Domenico, *"HH-BTag a ML-based method to select b jets from $H \to b\bar{b}$ decay, in $HH \to b\bar{b}\tau\bar{\tau}$ searches"*, AN-2019/283 (2020).

[200] Z. Zhang and M. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels, 2018, https://proceedings.neurips.cc/paper/2018/file/f2925f97bc13ad2852a7a551802feea0-Paper.pdf.

[201] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun, 2015.

[202] S. Chatrchyan *et al.*, *"Evidence for the 125 GeV Higgs boson decaying to a pair of tau leptons"*, Journal of High Energy Physics **2014**, 104 (2014), https://doi.org/10.1007/JHEP05(2014)104.

[203] V. Khachatryan *et al.*, *"Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions"*, Journal of High Energy Physics **2014**, 160 (2014), https://doi.org/10.1007/JHEP10(2014)160.

[204] B. Bullock, K. Hagiwara, and A. Martin, *"Tau polarization and its correlations as a probe of new physics"*, Nuclear Physics B **395**, 499 (1993), https://www.sciencedirect.com/science/article/pii/055032139390045Q.

[205] L. Russell, Identification of Hadronic Tau Lepton Decays with Domain Adaptation using Adversarial Machine Learning at CMS, 2022, https://cds.cern.ch/record/2827366, Presented 28 Jun 2022.

[206] CMS Collaboration, *"A deep neural network to search for new long-lived particles decaying to jets"*, Machine Learning: Science and Technology **1**, 035012 (2020), https://dx.doi.org/10.1088/2632-2153/ab9023.

[207] CMS Collaboration, *"Performance of the CNN-based tau identification algorithm with Domain Adaptation using Adversarial Machine Learning for Run 3"*, (2024), https://cds.cern.ch/record/2904699.

Acknowledgements

Nel corso del dottorato mi sembra di aver vissuto tante vite tutte insieme, e sarà molto difficile condensarle in poche righe dato che, come i lettori e le lettrici sapranno, non ho il dono della sintesi.
In questi anni ho trovato la mia "famiglia scientifica", Agnese e Konstantin. Sono grata di aver trovato delle persone che si sono prese cura della mia formazione accademica, tutti i loro consigli ed insegnamenti sono stati preziosi, e tutte le strigliate lo sono state ancor di più. Come diciamo a Napoli, *mazz' e panell' fanne 'e figli bell*. Siete dei modelli a cui aspirare. Grazie per avermi supportata, sopportata e per avermi insegnato a nuotare nell'oceano della ricerca.
Durante il mio percorso ho trovato tantissime persone, e ne ho perse altrettante. Sono cresciuta e maturata, ho imparato a superare i miei limiti, ad affrontare con determinazione le mie paure, e grazie alle persone a me care ho trovato tanto coraggio nell'affrontare ogni tipo di sfida, dalle più banali alle più importanti.
Una delle esperienze più straordinarie della mia vita è stato l'anno al CERN, indimenticabile per tanti aspetti, ma soprattutto perchè mi ha portato a ritrovare un tesoro, un'amica *ballerina*, Silvia. Grazie per tutto l'affetto, per la stima e per esserci sempre, anche dopo che ti ho fatto dire addio al tuo telefono nel Rodano. Grazie per ogni momento, per tutti gli audio da 10 minuti che ti ascolti e per quelli che mi mandi che ascolterò sempre, a qualsiasi fuso orario mi troverò.
Ringrazio le mie amiche di sempre, che hanno sempre creduto in me e ci sono sempre state. Grazie Giorgia, perchè dici che io sono l'"amica geniale", perchè hai da sempre sopportato tutte le mie paure mediche (accompagnandomi anche a fare i prelievi), ma soprattutto per avermi insegnato la determinazione, dal primo giorno di scuola media in cui ancor prima del nome mi dicesti che saresti diventata un medico. Tu più di qualsiasi altra persona mi hai insegnato a non fermarsi davanti a nulla e a perseguire la propria passione sempre, e comunque. Grazie *Pisy*, perchè siamo così diverse, il giorno e la notte, e come tali non potremmo mai vederci l'una senza l'altra. Da quando ci siamo conosciute non abbiamo mai avuto dubbi sul fatto che il nostro rapporto di sorellanza non si sarebbe mai sciolto. E così sarà, per sempre. Grazie Alice, perchè trovo in te un porto sicuro. Perchè in te ho trovato sempre rifugio dalle mie paure, e mi hai insegnato ad affrontarle. Grazie Chiara, per ogni fiorellino che ci ricorda le bellezze della vita anche quando siamo a terra, e Maria, perchè a distanza di anni e di km non ci siamo perse, e non è scontato.
Non manca il grande *Maurito*, che anche se ci vediamo poco, quando ci rivediamo è

come se non fosse passato un attimo da quando vivevamo le giornate in $\hbar$, e quando eravamo alla fermata del bus in attesa della C33 (che forse è la C33 stessa). Tu mi hai insegnato la resilienza, e sono grata di avere un amico come te.
Al mio fianco in questi anni c'è sempre stata *la Zezi*, la mia prima coinquilina in un'esperienza in cui avevo paura di essere una "B", che grazie a te ho imparato che non esistono "A" o "B", ma solo persone che possono imparare da nuove esperienze. Grazie perchè ci sei sempre, nelle nostre infinite passeggiate alla torre (quella vera), in cui sembriamo un articolo "il" ambulante.
Ringrazio la mia amica *berliner goddess* Ilaria, perchè ho trovato in te un'amica dal cuore puro, che mi ha insegnato a non avere paura di essere me stessa, di affrontare a testa alta tutte le asperità della vita. Sei un'amica preziosa, ricordalo sempre.
Quando sono tornata a Pisa dal CERN ero totalmente spaesata, mi sentivo anche un po' persa, ma grazie a Peppe, il mio coinquilino del cuore, e Martina, la mia compagna di sushi, ho ritrovato la grinta e mi sono ri-innamorata di Pisa. Grazie per ogni abbraccio tra coinquilini, per ogni serata passata alla Torre, per ogni singolo momento in cui ci siete stati. A proposito di *Torre*, a Pisa ho trovato non una, ma ben due case. Se non sono nella dimora dove dormo, è certo che sarò alla *Torre del Luppolo*. Ringrazio i ragazzi della torre (Alessio, Andrea, Alessandro, Gianluca P., Guido, Gianluca C., Allegra, Angela, Daniele, Dario, Luca, Francesco, Giulio, Vera) per aver passato insieme tante serate, sessioni di gioco e braciate, regalandomi tanti momenti di gioia, specie quando ne avevo più bisogno. I want to thank Namitha, my indian-vomerese-princess alter-ego, for all the time spent together in front of Margaritas, for all the good chats and for cheering me up at *Tau2023*, where I had another water-technology fight experience. Another important mention goes to Foteinì (hoping I wrote correctly the name, but *Fofo* in doubt), thank you for all the beer and the good time together. Ringrazio la *donna* Fulvia, perchè la nostra amicizia mi ha insegnato che il tempo non misura l'intensità di un'amicizia, se è vera, e che non è mai tardi per trovare nuove amicizie inestimabili.
E come posso non menzionare, tra tutte le persone che mi hanno insegnato tanto in questo percorso, Francino. Ti voglio bene, sempre.
Grazie, Stefano, per essere entrato nella mia vita ed aver portato la luce, che è sempre stata lì, ma tu hai acceso l'interruttore. E te ne sarò per sempre grata. Grazie per essere rimasto, per tutte le stampe 3D, per condividere la vita con me, per i viaggi. Con te ho capito cosa davvero significa comunicare in modo costruttivo e sano, e questo me lo porterò per sempre.
Ringrazio inoltre la dolcissima Faye, che è stata sempre un toccasana per il mio cuore, una buffa creaturina piena di allegria e tenerezza che saltella e corre con le sue minute zampette.

Infine, ringrazio i miei genitori. Non riesco a quantificare quanto io sia grata di avervi come genitori. Non esistono altre parole, se non che vi amo con tutto il mio cuore, e anche più se potessi.