Search for Higgs Boson Pair Production in the $b\bar{b}\tau^+\tau^-$ Final State with the ATLAS Experiment at the LHC

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Christopher Deutsch
aus
Neuwied

Bonn, 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Jochen Dingfelder
2. Gutachter:	Prof. Dr. Ian Brock
Tag der Promotion:	14.11.2023
Erscheinungsjahr:	2023

Abstract

A search for Higgs boson pair (HH) production in final states with two b quarks and two τ leptons using 139 fb^{-1} of proton–proton collision data at a centre-of-mass energy of 13 TeV recorded by the ATLAS detector at the Large Hadron Collider (LHC) is presented. Searches for HH production are vital for solidifying our understanding of electroweak symmetry breaking in the Standard Model (SM) since they serve as direct probes of the Higgs boson self-coupling. In addition, they provide sensitivity to new phenomena introduced by theories beyond the SM.

The search presented in this thesis targets channels with one or more hadronic τ-lepton decays. It relies on effective tau identification algorithms, which are algorithms that distinguish signatures of hadronic τ-lepton decays in the ATLAS detector from other sources, such as quark- or gluon-initiated jets. A novel tau identification algorithm based on recurrent neural networks is introduced. It exploits information from reconstructed charged-particle tracks, topological clusters of calorimeter cell signals, and purposefully constructed discriminating variables for tau identification. The new technique outperforms the method previously employed in the ATLAS experiment.

Three different types of Higgs boson pair production are considered in this thesis: non-resonant HH production predicted by the SM (SM HH production); resonant HH production via scalar, narrow-width resonances; and non-resonant HH production with anomalous values of the Higgs boson self-coupling constant λ_{HHH}.

The search for SM HH production targets the gluon–gluon fusion and vector boson fusion production modes. No statistically significant deviation from the background-only hypothesis is observed; therefore, upper limits are set on the SM HH production cross section. The search yields an upper limit of 4.7 times the SM cross section prediction at 95 % confidence level (CL).

The search for resonant HH production considers intermediate resonances with masses ranging from 251 to 1 600 GeV. Over the considered mass range, the largest deviation from the background-only hypothesis is observed for a mass of 1 000 GeV with a local (global) significance of 3.1σ (2.0σ). Upper limits at 95 % CL are set on the resonant HH production cross section, which range from 20 to 900 fb depending on the mass of the resonance.

Finally, the search for SM HH production is reinterpreted in the context of anomalous values of the Higgs boson self-coupling constant. Upper limits at 95 % CL are set on the non-resonant HH production cross section as a function of the self-coupling modifier $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{\text{SM}}$, where $\lambda_{HHH}^{\text{SM}}$ is the Higgs boson self-coupling predicted by the SM. The upper limits on the non-resonant HH production cross section are compared to theory predictions, providing constraints on the self-coupling modifier of $-2.4 \leq \kappa_\lambda \leq 9.2$.

Contents

CHAPTER 1

Introduction

The turn of the 20$^{\text{th}}$ century marked the origin of modern physics, ushering in a century of remarkable advancements in fundamental physics. The key to this success was the development of the *theory of special relativity* and the rise of *quantum mechanics*. On the one hand, the theory of special relativity allowed the description of the mechanics of physical systems with energies much larger than the energy of the system at rest. On the other hand, the introduction of quantum mechanics was instrumental in describing phenomena occurring at length scales of atomic and sub-atomic physics. The field of experimental high-energy physics (HEP) is situated at the intersection of special relativity and quantum mechanics, probing the nature of elementary particles and their interactions at the largest, experimentally accessible energy scales.

The *Standard Model* (SM) is the theory at the centre of particle physics and combines special relativity and quantum mechanics in a paradigm known as a *relativistic quantum field theory*. It describes the currently known elementary particles and three (out of four) fundamental forces of nature. The formulation of the SM traces back to the 1970s, when the theory of the strong interaction [1–3] and the Glashow–Salam–Weinberg model [4–6] of the electroweak interaction were consolidated in a single theory. Since then, the SM had tremendous success in describing the properties of elementary particles and their interactions and notably in predicting the existence of several elementary particles before their experimental discovery. Among the predicted particles is the *Higgs boson* (H), which was discovered in 2012 by the ATLAS and CMS experiments [7, 8] at the Large Hadron Collider (LHC). The Higgs boson arises as part of the Brout–Englert–Higgs (BEH) mechanism [9, 10] that is employed in the SM to explain how elementary particles acquire mass in a process referred to as *electroweak symmetry breaking* (EWSB).

In the SM, the Higgs boson has the ability to interact with itself—a property referred to as the *Higgs boson self-coupling*. The strength of the self-coupling is determined by the shape of the Higgs potential, a central element of the BEH mechanism responsible for EWSB. Therefore, direct measurements of the Higgs boson self-coupling are instrumental to test our understanding of EWSB in the SM. These measurements can be performed using high-energy particle collision events, such as the ones occurring at the LHC, that result in the production of pairs of Higgs bosons. Higgs boson pair production is an exceedingly rare process, for which experimental evidence has yet to be established.

In the presence of new physics, the phenomenology of Higgs boson pair production might differ considerably from the SM. While the SM is a very successful theory, it is known to be incomplete as it leaves a number of phenomena (e.g. gravitation, dark matter, etc.) unexplained. Eventually,

these shortcomings have to be addressed, requiring a theory that goes beyond the SM (BSM). Some BSM theories predict Higgs boson pair production via new massive resonances or deviations of the (effective) Higgs boson self-coupling strength from the SM expectation. The presence of such signatures would be indicative of new physics, making searches for Higgs boson pair production an important probe of the BSM sector.

This thesis presents a search for Higgs boson pair production in final states with two b quarks and two τ leptons. In particular, final states with two hadronic τ-lepton decays ($b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$) or one hadronic and one leptonic τ-lepton decay ($b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$) are considered.[1] The search is performed using data of proton–proton (pp) collision events at a centre-of-mass energy of 13 TeV recorded by the ATLAS detector in 2015–2018. The size of the analysed dataset amounts to an integrated luminosity of 139 fb^{-1}. Three different modes of Higgs boson pair production are considered: non-resonant Higgs boson pair production as predicted by the SM (SM HH production); production of Higgs boson pairs via new, scalar resonances with masses of up to 1.6 TeV (resonant HH production); and non-resonant Higgs boson pair production with anomalous values of the Higgs boson self-coupling constant.

The structure of this thesis is as follows: Chapter 2 lays the foundation for searches for Higgs boson pair production. The SM is summarised, and the role of EWSB and the Higgs boson is illustrated. Moreover, the phenomenology of Higgs boson and Higgs boson pair production is discussed in the context of the SM and BSM theories. The chapter concludes with a summary of experimental results on searches for Higgs boson pair production prior to the work presented in this thesis. Chapter 3 describes the LHC, the ATLAS experiment, and the techniques used to reconstruct particle collision events with the ATLAS detector. Chapter 4 describes the statistical framework used to interpret the collision events recorded by the detector.

The bulk of this thesis is concerned with the search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state. Particular focus lies on the channel with two hadronic τ-lepton decays, referred to as the $\tau_{\text{had}}\tau_{\text{had}}$ channel. A crucial element of searches in this final state is the ability to reconstruct and identify hadronic τ-lepton decays based on their signature in the ATLAS detector. *Tau identification* refers to the process of identifying hadronic τ-lepton decays, aiming to distinguish detector signatures originating from τ_{had} from those of non-τ_{had} sources, such as quark- or gluon-initiated jets. A novel tau identification method using *recurrent neural networks* is introduced in Chapter 5 that exploits low-level tracking and calorimeter information as well as discriminating variables purposefully constructed for tau identification. This method has since become the default tau identification algorithm at the ATLAS experiment for analyses of data recorded during Run 2 (2015–2018) and the beginning of Run 3 (2022–) of the LHC [11, 12]. It is used in Chapter 6 for the search for SM HH production and resonant HH production in BSM models with an additional CP-even scalar particle. Moreover, the result of the search for SM HH production is reinterpreted in Chapter 7 to set constraints on models predicting anomalous values of the Higgs boson self-coupling constant. The thesis is concluded with a summary in Chapter 8.

[1] Decays of the form $\tau^- \to \text{hadrons} + \nu_\tau$ are referred to as a hadronic τ-lepton decays (τ_{had}), and $\tau^- \to \ell^-\bar{\nu}_\ell\nu_\tau$ for $\ell = e$ or μ as leptonic τ-lepton decays (τ_{lep}). The charge conjugate modes are named analogously.

CHAPTER 2

The Standard Model and Higgs Boson Pair Production

The Standard Model (SM) of particle physics is among the most precisely tested theories of physics, explaining numerous observed phenomena with high precision. It provides a description of the known fundamental particles and their interactions (excluding gravitation) in the framework of a relativistic quantum field theory. The current formulation of the SM stems from the early 1970s after the development of the theories of the electroweak and strong interaction [1–6, 9, 10, 13]. This chapter lays the theoretical foundation for the searches for Higgs boson pair production presented in this thesis.

Both SI and natural units ($\hbar = c = \varepsilon_0 = 1$) are used in the following, whichever is most suitable in a given context. In addition, this chapter adopts the Einstein summation convention, implying summation over repeated indices in a mathematical term. Greek indices represent the four dimensions of space-time, while the meaning of Latin indices is context-dependent. Finally, the metric tensor of special relativity is assumed to be $\eta = \mathrm{diag}(+1, -1, -1, -1)$.

This chapter is structured as follows: First, an overview of the SM and its particle content is given in Section 2.1. The fundamental interactions are described in Section 2.2 on the basis of symmetries of the theory. Subsequently, the SM phenomenology of the Higgs boson and Higgs boson pair production is presented in Section 2.3. The SM has several limitations in explaining certain experimental or theoretical phenomena, which suggests the existence of BSM physics. Section 2.4 lists some of these limitations as well as examples of BSM theories that can lead to an enhanced production of Higgs boson pairs at the LHC. This chapter concludes in Section 2.5 with the experimental status of searches for Higgs boson pair production prior to the work presented in this thesis. The initial parts of this chapter (Sections 2.1 and 2.2) are based on Refs. [14–16].

2.1 Particles of the Standard Model

The particles of the SM are illustrated in Figure 2.1. They can be broadly categorised into *fermions* and *bosons*, which are particles with half-integer and integer spin, respectively. With the discovery of the Higgs boson in 2012 [7, 8], experimental evidence for the existence of all SM particles is established. The following gives an overview of the fermions and bosons of the SM.

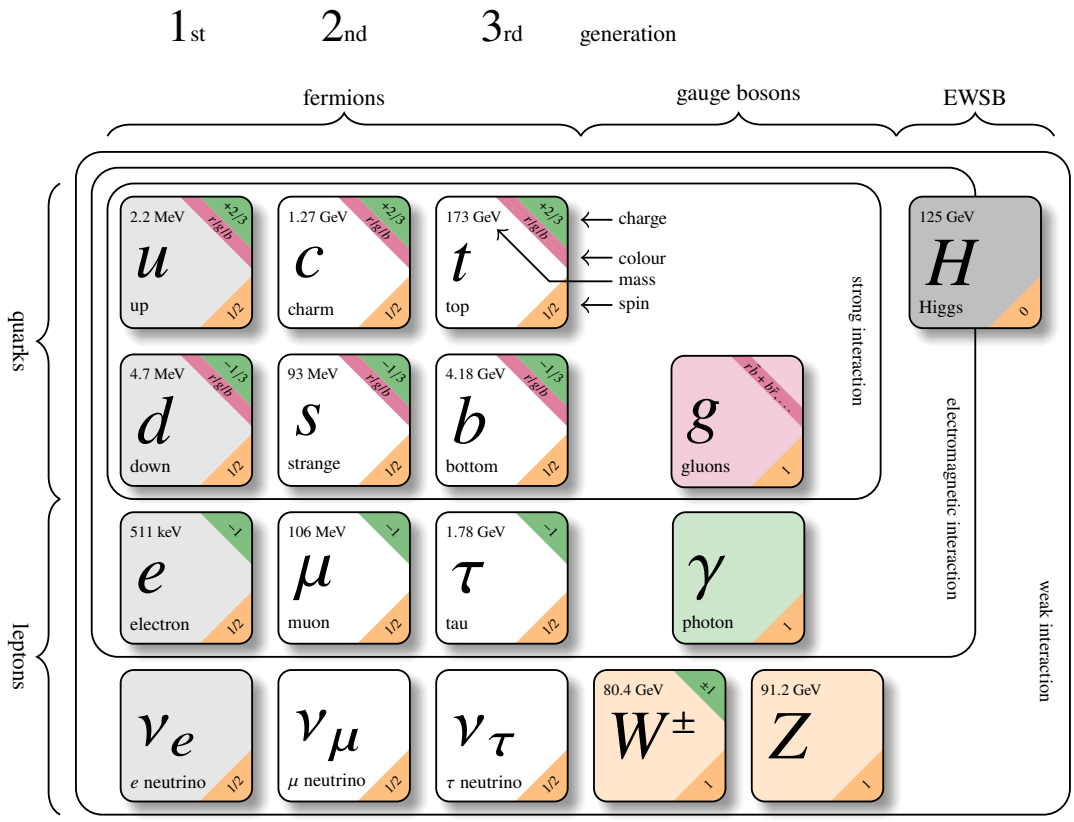


Figure 2.1: Particles of the SM. The diagram is adapted from Ref. [17] with particle masses from Ref. [18]. Antifermions are not shown explicitly. Not shown are the charges of the electroweak interaction, the weak isospin and weak hypercharge, which depend on the chirality of fermions. The gluons have eight different states of colour charge, which are combinations of colour and anticolour.

Fermions are massive[1] particles adhering to the Pauli exclusion principle, thus often referred to as *matter* particles. For every fermion there exists a corresponding antifermion that has the same properties but with additive quantum numbers of opposite sign. The fermions of the SM are divided into *quarks* that participate in the strong interaction, and *leptons* that do not. They are further divided into up-type quarks (u, c, t), down-type quarks (d, s, b), electrically charged leptons (e, μ, τ), and neutrinos (ν_e, ν_μ, ν_τ). The fermions come in three generations, the main difference between them being the mass of the fermions, which increases with every successive generation. All ordinary (stable) matter consists of fermions of the first generation: up quarks, down quarks, and electrons.

The fermions carry charge-like quantum numbers that dictate the fundamental interactions they participate in. Quarks carry *colour charge*, which comes in three discrete values of either *red*, *green*, or *blue*, and the corresponding anticolours for antiquarks. Particles carrying colour charge take part in the strong interaction. Quarks and (electrically) charged leptons carry *electric charge* and are therefore subject to the electromagnetic interaction. Fermions (antifermions) with left-handed (right-handed) chirality carry *weak isospin*; therefore, they take part in charged-current weak interactions. All fermions of the SM couple to the neutral-current weak interaction, the strength being determined by the weak isospin and the *weak hypercharge* (or analogously the weak isospin and the electric charge, cf. Section 2.2.2).

Gauge bosons are particles with spin-1 and are therefore also referred to as vector bosons. Gauge bosons are the quanta of fields arising in quantum field theories built on certain symmetry principles, referred to as gauge theories, which are discussed in Section 2.2. The gauge bosons mediate the strong, electromagnetic, and weak interaction through particle exchange.

The massless gluons are the mediators of the strong interaction between particles with colour charge. Gluons are carriers of colour charge themselves, allowing self-interactions between gluons. The massless photon mediates the electromagnetic interaction between electrically charged particles. The gauge bosons of the weak interaction, the $W^\pm$ and Z bosons, are the only massive gauge bosons in the SM. The $W^\pm$ bosons are electrically charged and mediate the charged-current weak interaction between particles carrying weak isospin. The neutral Z boson mediates the neutral-current weak interaction.

The Higgs boson is the only scalar particle in the SM. The Higgs boson arises as part of the BEH mechanism [9, 10] that is employed in the SM to generate the masses of fermions and massive gauge bosons ($W^\pm$ and Z) without violating the symmetry principles underlying the mathematical formulation of the SM. The role of the Higgs boson in the SM is discussed in Section 2.2.2.

2.2 Symmetries and Interactions

The SM is a relativistic quantum field theory, which describes particles and their interaction using space-time dependent fields $\phi_i(\boldsymbol{x})$. The field dynamics are determined by the *Lagrangian density* $\mathcal{L}$, which is a function of the fields ϕ_i and their space-time derivatives $\partial_\mu \phi_i = \partial \phi_i / \partial x^\mu$ ($\mu = 0, 1, 2, 3$).

[1] Neutrinos are considered as massless in the SM; however, the observation of neutrino oscillations [19, 20] is experimental evidence for non-zero neutrino masses.

The evolution of the fields in time follows from the principle of stationary action, i.e. by extremising the action $S = \int \mathrm{d}^4x\, \mathcal{L}(\phi_i, \partial_\mu \phi_i)$, which yields the Euler–Lagrange equations

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \,.$$

For a given Lagrangian density, the Euler–Lagrange equations provide the "equations of motion" of the fields. The Lagrangian density is hereafter simply referred to as the *Lagrangian*.

A continuous transformation of the fields that leaves the Lagrangian unchanged is referred to as a *gauge transformation*. The fields resulting from this transformation follow the same equations of motion and therefore describe the same physical system. This invariance is referred to as *gauge invariance* or *gauge symmetry*. Continuous symmetries of the Lagrangian are characterised by Lie groups, the most important ones for the description of the SM being the unitary group $U(1)$ and the special unitary groups $SU(2)$ and $SU(3)$. Any element of a unitary group can be written as

$$\hat{U} = \exp\left[i\theta_a G^a \right] ,$$

where θ_a are real parameters and G^a are the generators of the group. A Lagrangian that is invariant to a transformation $\hat{U}$ of the fields with parameters θ_a is said to possess *global* gauge invariance. The more restrictive case where the Lagrangian is invariant with respect to transformations of the fields with space-time dependent parameters $\theta_a(\boldsymbol{x})$ is referred to as *local* gauge invariance.

This is illustrated for the case of the Lagrangian of the Dirac field of mass m given by

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \,, \tag{2.1}$$

where ψ ($\bar{\psi} = \psi^\dagger \gamma^0$) are (adjoint) Dirac spinors and γ^μ are the Dirac matrices. Equation (2.1) possesses global gauge invariance with respect to $U(1)$ transformations given by $\psi \to \psi' = \exp[iq\theta]\psi$, where q is an arbitrary constant (for the moment). However, when performing a local transformation by letting the parameter be a function of space-time, i.e. $\theta \to \theta(\boldsymbol{x})$, the invariance of the Lagrangian is spoiled due to the derivative acting on the space-time dependent phase factor. One might impose $U(1)$ local gauge invariance on the Lagrangian by adding terms to Equation (2.1) that cancel the additional contributions. Conventionally, this is done by substituting the derivative ∂_μ by a gauge covariant derivative D_μ that transforms as $D_\mu \psi \to \exp[iq\theta(\boldsymbol{x})] D_\mu \psi$, thus recovering local gauge invariance. A definition of D_μ with these properties requires the introduction of a new massless vector field, referred to as a *gauge field*, with appropriate transformation properties:

$$D_\mu = \partial_\mu + iqA_\mu \qquad \text{with} \qquad A_\mu \xrightarrow{U(1)} A'_\mu = A_\mu - \partial_\mu \theta(\boldsymbol{x}) \,. \tag{2.2}$$

The additional term introduced by substituting $\partial_\mu \to D_\mu$ in Equation (2.1) is interpreted as an interaction between fermions and vector bosons of the gauge field.

The principle of local gauge invariance can be used to obtain the Lagrangian of quantum electrodynamics (QED), which describes electromagnetic interactions. The symmetry group of QED is $U(1)_Q$, the subscript Q indicating that the generator of the group is the electric charge operator. Imposing local gauge invariance by substituting Equation (2.2) into the Dirac Lagrangian yields the

interaction term

$$\mathcal{L}_{\text{int.}} = -q\bar{\psi}\gamma^{\mu}\psi A_{\mu}\,.$$

In the case of QED, the field A^{μ} is identified as the four-potential of the electromagnetic field and q as the electric charge of the fermion. Thus, the interaction term describes the coupling between photons and fermions with electric charge q. For a single type of fermion, the Lagrangian of QED is given by

$$\mathcal{L}_{\text{QED}} = \underbrace{\mathcal{L}_{\text{Dirac}}}_{\text{Free fermion field}} \quad \underbrace{-q\bar{\psi}\gamma^{\mu}\psi A_{\mu}}_{\text{Fermion–photon interaction}} \quad \underbrace{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{Photon field kinetic term}}\,,$$

which additionally includes the Lagrangian of the free photon field, referred to as the *kinetic term* of the field, defined by the electromagnetic tensor $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$. This additional term also fulfils the local gauge invariance with respect to $U(1)_Q$.

The principle of local gauge invariance is at the heart of the SM, where it is used to great success in describing the three fundamental interactions. The symmetry group of the SM is

$$SU(3)_{\text{colour}} \otimes SU(2)_{\text{L}} \otimes U(1)_Y\,,$$

where $SU(3)_{\text{colour}}$ is the symmetry of the strong interaction and $SU(2)_{\text{L}} \otimes U(1)_Y$ the symmetry of the unified description of the electromagnetic and weak interaction. These are introduced in Section 2.2.1 and Section 2.2.2, respectively.

2.2.1 Quantum Chromodynamics

Quantum chromodynamics (QCD) is the theory describing the interactions of quarks and gluons. The fundamental charge of QCD is colour charge, which comes in three distinct colours referred to as red, green, and blue (r, g, b). The quark fields are written in terms of the three component objects

$$\psi = \begin{pmatrix} q_{\text{r}} \\ q_{\text{g}} \\ q_{\text{b}} \end{pmatrix} \qquad \text{and} \qquad \bar{\psi} = \begin{pmatrix} \bar{q}_{\text{r}} & \bar{q}_{\text{g}} & \bar{q}_{\text{b}} \end{pmatrix}\,,$$

in which q_i ($\bar{q}_i$) represents the (adjoint) Dirac spinor describing the quark field with colour i. The symmetry group of the strong interaction is $SU(3)_{\text{colour}}$, the subscript indicating that elements of the group act in colour-space. The generators of $SU(3)_{\text{colour}}$ are taken to be

$$T_a = \frac{1}{2}\lambda_a \qquad \text{for} \qquad a = 1, \ldots, 8\,,$$

where λ_a are the Gell-Mann matrices.[2] The theory of QCD is referred to as a Yang–Mills gauge theory [21] since the generators of $SU(3)_{\text{colour}}$ do not commute in general, that is, the symmetry group is non-Abelian. The commutation relation between the generators is given by $[T_a, T_b] = if_{abc}T^c$, defining the structure constants f_{abc} of the group.

[2] Mathematical objects with an upper Latin index and those with a lower Latin index are assumed to be equivalent.

The principle of local gauge invariance with respect to $SU(3)_{\text{colour}}$ is used to obtain the Lagrangian of QCD. Let the local gauge transformation of the quark fields be

$$\psi \rightarrow \psi' = \exp[ig_s\theta_a(\boldsymbol{x})T^a]\psi\,,$$

where g_s is referred to as the strong coupling constant. The gauge covariant derivative is then

$$D_\mu = \partial_\mu + ig_s G^a_\mu T_a\,,$$

which introduces eight gauge fields G^a_μ corresponding to the eight gluons of QCD. To ensure local gauge invariance, the gluon fields have to transform according to

$$G^k_\mu \xrightarrow{SU(3)_{\text{colour}}} G^{k\,\prime}_\mu = G^k_\mu - \partial_\mu\theta^k(\boldsymbol{x}) - g_s f_{ij}{}^k\theta^i(\boldsymbol{x})G^j_\mu\,.$$

Furthermore, the Lagrangian of QCD has to account for the energy density of the gluon fields (and their interactions). This contribution is given by the kinetic term of the gluon fields and reads

$$\mathcal{L}_g = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a\,,$$

where $G^a_{\mu\nu}$ are the gluon field strength tensors defined as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^a{}_{bc}G^b_\mu G^c_\nu\,.$$

The Lagrangian of QCD for a single flavour of quark with mass m is consequently given by

$$\mathcal{L}_{\text{QCD}} = \underbrace{\bar\psi(i\gamma^\mu\partial_\mu - m)\psi}_{\text{Free quark field}} \quad \underbrace{-g_s(\bar\psi\gamma^\mu T_a\psi)G^a_\mu}_{\text{Quark–gluon interactions}} \quad \underbrace{-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{\text{Gluon field kinetic term}}\,.$$

This Lagrangian describes the free quark field, the interactions of quarks with the eight gluons, and the kinetic energy of the gluon fields. The non-Abelian nature of the $SU(3)_{\text{colour}}$ group, meaning a set of indices exists such that $f_{abc} \neq 0$, gives rise to the distinct structure of QCD through the kinetic term of the gluon fields. This term includes self-interactions between gluons, which correspond to, in the language of Feynman diagrams, triple and quartic interaction vertices between gluons. Gluons themselves are carriers of colour charge, which leads to this behaviour. This is unlike the photon of QED which does not carry electrical charge and thus does not couple to itself.

Two important features of the theory of QCD are highlighted in the following:

Colour confinement Due to the dynamics of the gluon self-interactions, free quarks or gluons cannot be observed in nature [22]. For example, separating the quarks of a quark–antiquark pair leads to the formation of *flux tubes* in the gluon field strength that result in a linear increase in field energy with separation of the quarks. Eventually, the energy stored in the gluon field is sufficiently large to create a quark–antiquark pair from the vacuum. This process repeats itself until only quarks or gluons bound into colourless composite particles (colour singlet states) remain. The most prevalent bound states of quarks are (anti-)baryons consisting of three (anti-)quarks and mesons consisting of a quark–antiquark pair. The $SU(3)_{\text{colour}}$ symmetry also

allows for colour singlet states of multiple gluons referred to as *glueballs* [23, 24] or other combinations of quarks such as tetra- ($q_1\bar{q}_2q_3\bar{q}_4$) or pentaquarks ($q_1q_2q_3q_4\bar{q}_5$) [25].

Running coupling & asymptotic freedom The strong coupling constant, frequently expressed as $\alpha_s = g_s^2/(4\pi)$ in analogy to the fine-structure constant $\alpha = e^2/(4\pi)$ of QED, is not constant but varies as a function of the momentum transfer Q of an interaction. A quark scattering process involving the exchange of a gluon is represented by an infinite number of Feynman diagrams with different higher-order corrections, for example diagrams with additional quark or gluon loops. A process referred to as *renormalisation* absorbs these corrections into an effective coupling constant, the coupling consequently becoming a function of Q^2. This effect occurs in both QED and QCD, however, with distinct signatures. While the coupling $\alpha(Q^2)$ of QED increases with Q^2, $\alpha_s(Q^2)$ of QCD decreases with Q^2 due to gluon self-interactions. The high-Q^2 behaviour of α_s is referred to as *asymptotic freedom* [2, 3].

2.2.2 Theory of the Electroweak Interaction

The principle of local gauge invariance was previously used to obtain the Lagrangians of QED and QCD. The characteristics of the weak interaction make this approach more difficult. First, the mediators of the interaction are massive with masses of [18]

$$m_W = (80.377 \pm 0.012)\,\text{GeV} \qquad \text{and} \qquad m_Z = (91.1876 \pm 0.0021)\,\text{GeV}.$$

Second, the symmetry with respect to parity (space-inversion) transformations is violated [26]. Lastly, the charged-current weak interaction couples fermions of different flavour that differ by one unit in electric charge. Due to these characteristics, the construction of a gauge theory of the weak interaction that respects local gauge invariance requires the introduction of new mechanisms.

Electroweak Unification

The theory of the electroweak interaction was developed by Glashow, Salam, and Weinberg in the 1960s to unify the electromagnetic and weak interactions in a single model [4–6]. The theory is constructed as a non-Abelian gauge theory based on a symmetry group referred to as $SU(2)_\text{L} \otimes U(1)_Y$, the meaning of the subscripts will be illustrated in the following.

First, a new quantum number referred to as *weak isospin* denoted by I and its component along the 3-axis, I_3, is introduced. The charged-current weak interaction violates parity symmetry maximally since it only couples to fermionic fields with left-handed chirality.[3] An appropriate weak isospin grouping of SM fermions is chosen in anticipation of an $SU(2)$ symmetry. Left-handed fermion fields are grouped into weak isospin doublets given by

$$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_\text{L}, \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_\text{L}, \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_\text{L} \qquad \begin{pmatrix}u\\ d\end{pmatrix}_\text{L}, \begin{pmatrix}c\\ s\end{pmatrix}_\text{L}, \begin{pmatrix}t\\ b\end{pmatrix}_\text{L} \qquad \text{with} \qquad I = \frac{1}{2}, I_3 = \pm\frac{1}{2},$$

[3] Chirality of a Dirac field is defined by the operator $\gamma^5 := i\gamma^0\gamma^1\gamma^2\gamma^3$. The eigenvectors of γ^5 are states of well-defined chirality with eigenvalues of +1 or −1, referred to as right- and left-handed chiral states, respectively. Any spinor can be written as a superposition of right- and left-handed chiral states using the projection operators $P_\text{R/L} = (1 \pm \gamma^5)/2$.

where the upper (lower) components correspond to $I_3 = +\frac{1}{2}$ ($I_3 = -\frac{1}{2}$), and nine singlet states for right-handed fermion fields

$$e_{\mathrm{R}}^{-}, \mu_{\mathrm{R}}^{-}, \tau_{\mathrm{R}}^{-}, u_{\mathrm{R}}, d_{\mathrm{R}}, c_{\mathrm{R}}, s_{\mathrm{R}}, t_{\mathrm{R}}, b_{\mathrm{R}} \qquad \text{with} \qquad I = 0, I_3 = 0\,,$$

where the subscripts L and R refer to the projection of the fields into their left- and right-handed chiral components, respectively. Right-handed neutrinos are omitted since they are not part of the SM. Given this grouping, $SU(2)_{\mathrm{L}}$ transformations only affect fermion fields with left-handed chirality, thus motivating the choice of subscript.

Second, a quantum number referred to as the *weak hypercharge* is defined according to

$$Y = 2(Q - I_3)\,,$$

where Q refers to the electric charge. Since the electric charge differs between the upper and lower component of the $SU(2)_{\mathrm{L}}$ doublets, an $SU(2)_{\mathrm{L}}$ transformation would violate the $U(1)_Q$ symmetry of QED. Therefore, the weak hypercharge is defined such that both components of an $SU(2)_{\mathrm{L}}$ doublet have the same value of Y. The broken $U(1)_Q$ symmetry is consequently replaced by a $U(1)_Y$ symmetry, which uses Y as the generator of the group instead. The $U(1)_Q$ symmetry of QED is later recovered in a process referred to as *electroweak symmetry breaking* (EWSB).

The principle of local gauge invariance with respect to the $SU(2)_{\mathrm{L}} \otimes U(1)_Y$ group is invoked to generate the interactions of the electroweak theory. In the following, the weak isospin doublets are denoted as χ_{L} and weak isospin singlets as ψ_{R}. The gauge transformations with space-time dependent parameters α_a ($a = 1, 2, 3$) and β transform the fields as follows

$$\begin{aligned}
\chi_{\mathrm{L}} \rightarrow \chi_{\mathrm{L}}' &= \exp\left[ig\alpha_a(\boldsymbol{x})\frac{\sigma^a}{2} + ig'\beta(\boldsymbol{x})\frac{Y}{2}\right]\chi_{\mathrm{L}} \\
\psi_{\mathrm{R}} \rightarrow \psi_{\mathrm{R}}' &= \exp\left[ig'\beta(\boldsymbol{x})\frac{Y}{2}\right]\psi_{\mathrm{R}}\,,
\end{aligned}$$

where g and g' are coupling constants and σ^a are the Pauli matrices. The gauge covariant derivative is given by

$$D_\mu = \partial_\mu + igW_\mu^a\frac{\sigma_a}{2} + ig'B_\mu\frac{Y}{2}\,,$$

where it is implied that the Pauli matrices only act on χ_{L} and not on ψ_{R}. Four gauge fields, W_μ^a ($a = 1, 2, 3$) and B_μ, associated with the $SU(2)_{\mathrm{L}}$ and $U(1)_Y$ symmetry are introduced. These fields transform, in analogy to QED and QCD, as follows

$$\begin{aligned}
W_\mu^k &\xrightarrow{SU(2)_{\mathrm{L}} \otimes SU(1)_Y} W_\mu^{k\prime} = W_\mu^k - \partial_\mu\alpha^k(\boldsymbol{x}) - g f_{ij}{}^k\alpha^i W_\mu^j \\
B_\mu &\xrightarrow{SU(2)_{\mathrm{L}} \otimes SU(1)_Y} B_\mu{}' = B_\mu - \partial_\mu\beta(\boldsymbol{x})\,,
\end{aligned}$$

where f_{ijk} are the structure constants of $SU(2)_{\mathrm{L}}$ with the generators of the group taken to be $\frac{1}{2}\sigma_a$. Substituting the gauge covariant derivative into the kinetic term of the Dirac Lagrangian yields the

interaction terms of the electroweak theory for left- and right-handed chiral fields

$$\mathcal{L}^{\mathrm{L}}_{\mathrm{int.}} = i\bar{\chi}_{\mathrm{L}}\gamma^{\mu}\left[igW^{a}_{\mu}\frac{\sigma_a}{2} + ig'B_{\mu}\frac{Y}{2}\right]\chi_{\mathrm{L}} \tag{2.3}$$

$$\mathcal{L}^{\mathrm{R}}_{\mathrm{int.}} = i\bar{\psi}_{\mathrm{R}}\gamma^{\mu}\left[ig'B_{\mu}\frac{Y}{2}\right]\psi_{\mathrm{R}}\,. \tag{2.4}$$

The four fields occurring in the interaction terms do not represent the physical fields observed in nature. Given the choice of Pauli matrices as the generators of $SU(2)_{\mathrm{L}}$, the fields W^1_μ and W^2_μ are associated with charged-current interactions, and W^3_μ and B_μ with neutral-current interactions. The physical fields of the charged-current interaction can be identified as

$$W^{\pm}_{\mu} = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)\,.$$

Furthermore, the physical fields describing the neutral-current interactions via the exchange of Z bosons and photons can be expressed as a linear combination of the W^3_μ and B_μ gauge fields. Experimental results show that the Z boson couples to both left- and right-handed chiral states, although not equally, and therefore such a mixing is required. The mixing can be described by a rotation of the fields by the weak mixing angle θ_{W} according to

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & \sin\theta_{\mathrm{W}} \\ -\sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix}\begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}\,.$$

The weak mixing angle must be chosen such that Equations (2.3) and (2.4) reproduce the coupling of QED, namely the photon must couple equally to left- and right-handed particles with a coupling constant e. This yields the condition

$$e = g\sin\theta_{\mathrm{W}} = g'\cos\theta_{\mathrm{W}}\,,$$

connecting the QED coupling constant with the coupling constants g and g' of the electroweak interaction. Finally, the kinetic term of the W^a_μ and B_μ fields is given, in analogy to QED and QCD, by

$$\mathcal{L}_{W,Z,\gamma} = -\frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\,,$$

where the field strength tensors are defined as

$$\begin{aligned} W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g f^a{}_{bc} W^b_\mu W^c_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu\,. \end{aligned}$$

Since $SU(2)_{\mathrm{L}}$ is a non-Abelian group, triple and quartic gauge boson couplings are introduced through the kinetic term of the W^a_μ fields.

The Brout–Englert–Higgs Mechanism

In the context of the electroweak theory, the non-zero masses of the gauge bosons (and fermions) have been ignored thus far. Inserting gauge field mass terms into the Lagrangian, for example for the physical Z_μ field with mass m_Z according to

$$\mathcal{L}^Z_{\text{mass}} = \frac{1}{2} m_Z^2 Z_\mu Z^\mu \,,$$

would violate $SU(2)_\text{L} \otimes U(1)_Y$ symmetry due to the transformation properties of the fields. A mechanism to dynamically generate the required mass terms was introduced into the electroweak theory by Weinberg [6] to address this issue. This mechanism traces back to Brout, Englert, and Higgs [9, 10] and is thus referred to as the Brout–Englert–Higgs (BEH) mechanism.

The BEH mechanism introduces two complex scalar fields, one electrically charged and one neutral field, arranged in a weak isospin doublet with $Y = 1$ according to

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

which can analogously be expressed as four real scalar fields ϕ_i. The Lagrangian of the complex scalar fields is given by the Klein–Gordon equation

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

with the potential term denoted by $V(\phi)$. The aim of the BEH mechanism is to embed the doublet of complex scalar fields into the electroweak theory with an $SU(2)_\text{L} \otimes U(1)_Y$ symmetry. To fulfil the gauge invariance of the electroweak theory, the potential term can only depend on $\phi^\dagger \phi$. One such choice is

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{2.5}$$

with μ^2 and λ being parameters of the potential.[4] The potential must be bound from below to have a well-defined state of minimum potential energy (vacuum state) and therefore λ must be positive. No such restrictions exist for μ^2, leaving two options. If $\mu^2 > 0$, the vacuum state is $\phi = \mathbf{0}$ and μ is related to the mass of the scalar field. If $\mu^2 < 0$, the field configuration $\phi = \mathbf{0}$ is at an unstable point and the true vacuum is given by the condition

$$\sum_{i=1}^{4} \phi_i^2 = -\frac{\mu^2}{\lambda} =: v^2 \,,$$

which represents a continuum of degenerate states at a radius of v from the origin in the space spanned by the four real scalar fields. The quantity v is referred to as the *vacuum expectation value* (VEV). Hereafter, this particular choice of potential is referred to as the Higgs potential. The shape of the potential is illustrated in Figure 2.2 for a simplified model with a single complex scalar field.

[4] In terms of the renormalisability of the theory, the largest allowed power of $\phi^\dagger \phi$ in the Lagrangian is two. See for example Ref. [27].

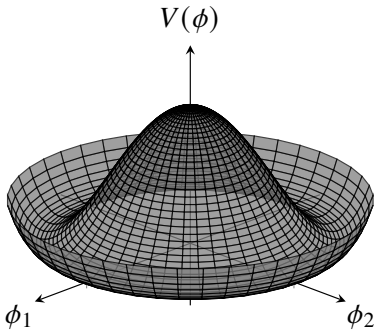


Figure 2.2: The "Mexican-hat potential" of a complex scalar field as a low dimensional example to illustrate the choice of the Higgs potential in the SM. The degenerate vacuum states lie in a circle with radius v given by the condition $\phi_1^2 + \phi_2^2 = v^2$. The figure is adapted from Ref. [28].

The fields ϕ_i will assume a configuration that minimises the potential energy, hence realising one of the infinite number of vacuum states. While the full Lagrangian still possesses an $SU(2)_\mathrm{L} \otimes U(1)_Y$ symmetry, the spontaneous choice of a vacuum state with non-vanishing VEV appears to break the symmetry, a process referred to as *spontaneous symmetry breaking*. Perturbative methods are used to find solutions to the field equations of motion, hence the fields have to be expressed as small perturbations relative to the vacuum state. Let the vacuum state be

$$\phi_\mathrm{v} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix},$$

chosen such that only the neutral component is non-vanishing to ensure that the $U(1)_Q$ symmetry of QED is recovered after spontaneous symmetry breaking.[5] Four degrees of freedom for perturbations relative to the chosen vacuum state exist. Three degrees of freedom are chosen such that they leave $\phi^\dagger\phi$, and thus $V(\phi)$, invariant. The remaining degree of freedom alters the value of $\phi^\dagger\phi$ and therefore $V(\phi)$.[6] In the Lagrangian describing the fields ϕ, these perturbations appear as three massless scalar fields and one massive scalar field. The quanta of the massless fields are referred to as Nambu–Goldstone bosons [29, 30]. An appropriate gauge transformation, referred to as the *unitary gauge*, removes the massless scalar fields from the theory. In unitary gauge, the doublet of complex scalar fields can be expressed as

$$\phi(\boldsymbol{x}) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(\boldsymbol{x}) \end{pmatrix} \tag{2.6}$$

with $H(\boldsymbol{x}) \in \mathbb{R}$.

Expressing the Higgs potential of Equation (2.5) in terms of perturbations of the vacuum state, i.e.

[5] An infinitesimal $U(1)_Q$ transformation yields $(1 + i\epsilon Q)\phi_\mathrm{v} = \phi_\mathrm{v}$, thus leaving the vacuum state unchanged. This is not the case when replacing Q with the generators of $SU(2)_\mathrm{L}$ or $U(1)_Y$.

[6] In the toy model depicted in Figure 2.2, the degrees of freedom correspond to perturbations of the vacuum state in angular and radial direction, respectively.

using Equation (2.6), and dropping terms not depending on H yields

$$V(H) = \lambda v^2 H^2 + \lambda v H^3 + \frac{\lambda}{4} H^4 \,. \tag{2.7}$$

The first term of $V(H)$ represents the mass term of the scalar field H with a mass of $m_H = \sqrt{2\lambda} v$. The quantum of the scalar field is referred to as the Higgs boson (H).[7] The terms cubic and quartic in the scalar field represent self-interactions between Higgs bosons with coupling strengths defined as[8]

$$\lambda_{HHH} \coloneqq \frac{3m_H^2}{v} \qquad \text{and} \qquad \lambda_{HHHH} \coloneqq \frac{3m_H^2}{v^2} \,.$$

The Feynman vertices of Higgs boson self-interactions are depicted in Figures 2.3(a) and 2.3(d).

The term of the Klein–Gordon equation involving the space-time derivatives yield after substituting the $SU(2)_\mathrm{L} \otimes U(1)_Y$ gauge covariant derivative and inserting the physical fields describing the W and Z bosons

$$(D_\mu \phi)^\dagger (D^\mu \phi) = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \left[\frac{g^2 v^2}{4} W_\mu^- W^{+\mu} + \frac{(g^2 + g'^2) v^2}{8} Z_\mu Z^\mu \right] \left(1 + \frac{1}{v} H \right)^2 \,. \tag{2.8}$$

The first term of Equation (2.8) represents the kinetic term of the scalar field H. Moreover, the BEH mechanism dynamically generates mass terms for the $W^\pm$ and Z bosons while leaving the photon massless. Using the four parameters of the electroweak theory (g, g', m_H, v) the masses of the bosons can be obtained from the Lagrangian such that

$$m_W = \frac{1}{2} g v \qquad\qquad m_Z = \frac{1}{2} \sqrt{g^2 + g'^2}\, v \qquad\qquad m_\mathrm{photon} = 0 \,.$$

Equation (2.8) additionally describes interaction vertices of the form WWH, ZZH, $WWHH$, and $ZZHH$, which are depicted in Figure 2.3.

Fermion Masses

The inclusion of fermion mass terms of the form

$$\mathcal{L}_\mathrm{mass}^\mathrm{fermion} = -m\bar{\psi}\psi = -m[\bar{\psi}_\mathrm{R}\psi_\mathrm{L} + \bar{\psi}_\mathrm{L}\psi_\mathrm{R}]$$

into the Lagrangian violates the gauge invariance with respect to the $SU(2)_\mathrm{L} \otimes U(1)_Y$ symmetry since fermions with left- and right-handed chirality transform differently. Instead, fermion mass terms are generated dynamically through EWSB by introducing interactions between the fermion fields and the scalar Higgs field, which are referred to as Yukawa interactions [31].

[7] Depending on the context, H refers to the Higgs boson or the scalar field that describes the Higgs field in unitary gauge.

[8] There exists no consensus regarding the definition of λ_{HHH} and λ_{HHHH}. In this thesis, a convention is adopted such that the Higgs potential can be expressed as

$$V(H) = \frac{m_H^2}{2} H^2 + \frac{\lambda_{HHH}}{3!} H^3 + \frac{\lambda_{HHHH}}{4!} H^4 \,.$$

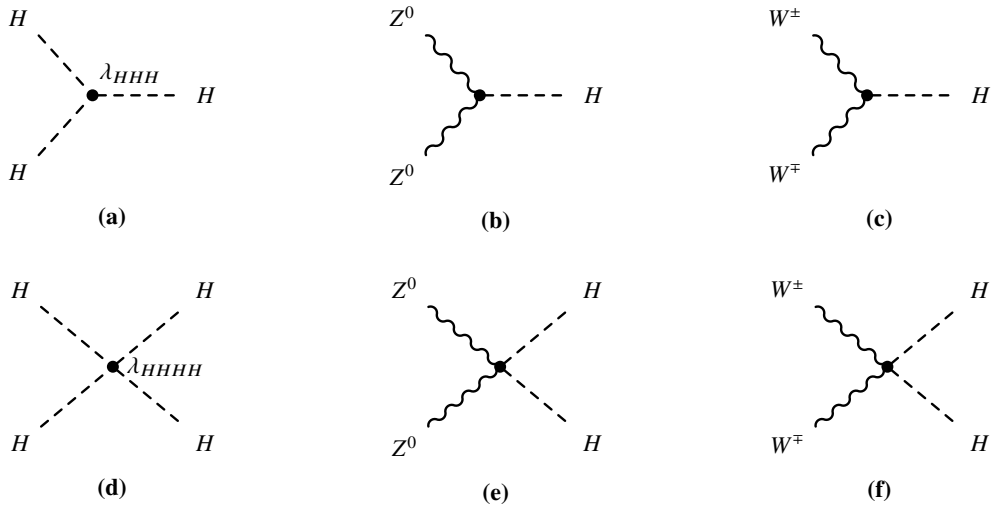


Figure 2.3: Interaction vertices between Higgs, W, and Z bosons predicted by the SM.

First, the conjugate of the Higgs field ϕ is defined as $\phi^{\mathrm{c}} = i\sigma_2\phi^*$ such that after EWSB in unitary gauge

$$\phi^{\mathrm{c}} = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(\boldsymbol{x}) \\ 0 \end{pmatrix} .$$

The Lagrangian of Yukawa interactions with $SU(2)_{\mathrm{L}} \otimes U(1)_Y$ gauge invariance is defined, here exemplary for the first generation of fermions, as

$$\mathcal{L}_{\mathrm{Yukawa}} = -y_e \bar{L}\phi e_{\mathrm{R}} - y_u \bar{Q}\phi^{\mathrm{c}} u_{\mathrm{R}} - y_d \bar{Q}\phi d_{\mathrm{R}} + \mathrm{h.c.}\,, \tag{2.9}$$

where $\bar{L}$ and $\bar{Q}$ are the $SU(2)_{\mathrm{L}}$ doublets of the first generation of leptons and quarks, respectively. The $SU(2)_{\mathrm{L}}$ singlets for electrons, up quarks, and down quarks are denoted as e_{R}, u_{R}, and d_{R}, respectively. Moreover, the coupling constants of the Yukawa interactions, which are free parameters of the theory, are given by y_e, y_u, and y_d. Lastly, h.c. represents the hermitian conjugate of the preceding terms. After EWSB and in unitary gauge Equation (2.9) reads

$$\mathcal{L}_{\mathrm{Yukawa}} \xrightarrow{\mathrm{EWSB}} \sum_{f \in \{e,u,d\}} -\frac{y_f v}{\sqrt{2}} [\bar{f}_{\mathrm{L}} f_{\mathrm{R}} + \bar{f}_{\mathrm{R}} f_{\mathrm{L}}] \left(1 + \frac{1}{v} H\right)$$

yielding fermion mass terms with masses

$$m_f = \frac{y_f v}{\sqrt{2}} ,$$

which are proportional to the Yukawa coupling constants, as well as interactions between Higgs bosons

and fermions with a coupling strength proportional to the mass of the fermion given by

$$g_{Hff} = \frac{m_f}{v}\,.$$

The Lagrangian of Equation (2.9) can be further generalised to include the remaining fermion generations and mixing between the weak and mass eigenstates of quarks as described by the Cabibbo–Kobayashi–Maskawa matrix [32, 33].

2.3 The Higgs Boson

The Higgs boson is the only scalar particle in the SM. It is massive and has neither electric nor colour charge. The Higgs boson couples to all massive particles, including itself, with coupling strengths increasing with the particle's mass. In particular, the coupling to gauge bosons, V, and fermions is proportional to m_V^2 and m_f, respectively. The detection of the Higgs boson and confirmation of its properties (e.g. spin, intrinsic parity, mass-dependent coupling strengths) would validate our current understanding of the BEH mechanism and the Glashow–Salam–Weinberg model of the electroweak interaction.

In 2012, almost half a century after the proposal of the Glashow–Salam–Weinberg model, the Higgs boson was discovered by the ATLAS [7] and CMS [8] collaborations at the LHC with a mass of about 125 GeV. Since its discovery, extensive measurements of its properties have been performed, showing remarkable agreement with the SM predictions. In 2022, the Higgs boson mass has been measured with a relative error approaching 0.1 %. At the time of writing, the most precise measurement of the Higgs boson mass by the ATLAS collaboration yields

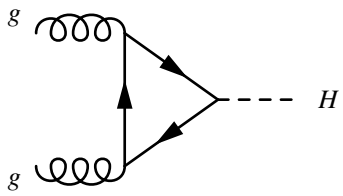
$$m_H = 124.94 \pm 0.17\,(\text{stat.}) \pm 0.03\,(\text{syst.})\,\text{GeV}$$

in the $H \to ZZ^* \to 4\ell$ channel [34] using pp collision events recorded in the period from 2011–2012 and 2015–2018. The observed Higgs boson is compatible with the scalar particle hypothesis [35, 36] and its coupling strengths are in good agreement with the SM predictions [37, 38]. The EWSB in the SM is now well-established and its free parameters, v and m_H, determined with high precision.[9]

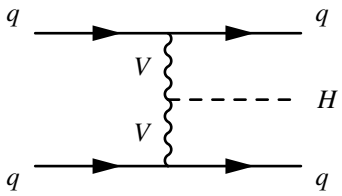
2.3.1 Production and Decay Modes

The production of Higgs bosons at the LHC occurs through different production modes. Feynman diagrams of the four dominant ones are shown in Figure 2.4, all of which have been experimentally confirmed to exist. The Higgs boson production cross section in pp collisions at a centre-of-mass energy of 13 TeV is shown in Figure 2.5(a) as a function of m_H. For $m_H = 125.0\,\text{GeV}$, the gluon–gluon fusion (ggF) production mode has the largest cross section with $\sigma(pp \to H) \approx 50\,\text{pb}$, followed by the vector boson fusion (VBF) mode with $\sigma(pp \to qqH) \approx 4\,\text{pb}$, the VH mode with $\sigma(pp \to VH) \approx 2\,\text{pb}$ for $V = W^\pm$ or Z, and finally the $t\bar{t}H$ mode with $\sigma(pp \to t\bar{t}H) \approx 0.5\,\text{pb}$ [40].

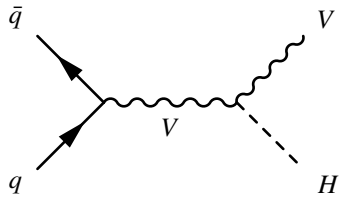
[9] The VEV was known prior to the discovery of the Higgs boson through measurements of the muon lifetime, which determines the effective coupling constant G_F of the charged-current weak interaction. With known G_F, the VEV can be calculated according to $v = \left(\sqrt{2}G_\text{F}\right)^{-1/2} \approx 246\,\text{GeV}$ [39].



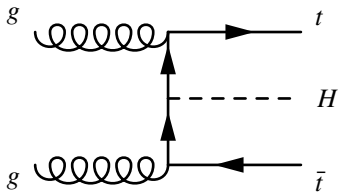
(a) Gluon–gluon fusion (ggF)



(b) Vector boson fusion (VBF)

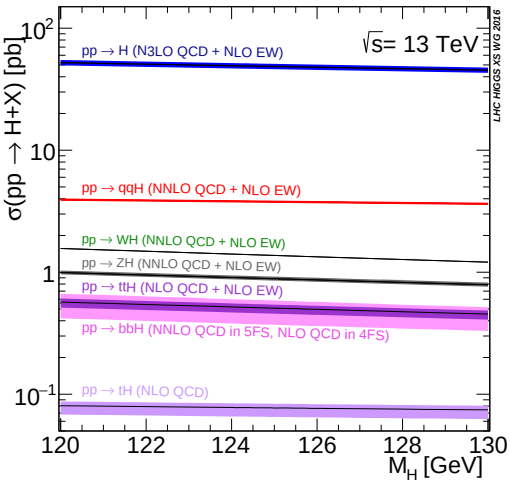


(c) Associated production with a massive vector boson (VH)



(d) Associated production with $t\bar{t}$ ($t\bar{t}H$)

Figure 2.4: The dominant Higgs boson production modes in pp collisions at centre-of-mass energies of 13 TeV.



(a) Cross sections of Higgs boson production modes as a function of m_H.

Decay mode	BR / %	Observed
$b\bar{b}$	58	✓
W^+W^-	21	✓
gg	8.2	
$\tau^+\tau^-$	6.3	✓
$c\bar{c}$	2.9	
ZZ	2.6	✓
$\gamma\gamma$	0.23	✓
$Z\gamma$	0.15	
$\mu^+\mu^-$	0.022	$*$

(b) Branching ratios (BR) of the Higgs boson. The gg, $\gamma\gamma$, and $Z\gamma$ decay modes occur via higher-order processes. $*$: First evidence of $H \to \mu^+\mu^-$ decays exists [41].

Figure 2.5: Higgs boson production cross section in pp collisions at a centre-of-mass energy of 13 TeV (a) and Higgs boson branching ratios for $m_H = 125.0$ GeV (b). The figure and branching ratios are taken from Ref. [40].

The Higgs boson is predicted to have a total decay width of about 4 MeV [40], yielding a proper lifetime of 10^{-22} s. Therefore, the Higgs boson decays almost immediately via one of the decay modes summarised in Figure 2.5(b), allowing detection only through its decay products. The table reflects the preferential coupling of the Higgs boson to heavy particles such as the massive gauge bosons[10] and third generation fermions. The vast majority of Higgs bosons decay into $b\bar{b}$ with a branching ratio of 58 %. The b quark is the heaviest fermion that can be produced in the Higgs boson decay, decays to top-quark pairs being forbidden due to $2m_t \gg m_H$. The second most abundant fermionic Higgs boson decay mode is $H \to \tau^+\tau^-$ with a branching ratio of 6.3 %. The $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ decay modes are among the most important probes of Yukawa interactions in the SM.

2.3.2 Higgs Boson Pair Production

After the discovery of the Higgs boson and the measurement of its mass, the free parameters of the EWSB are determined and thus the shape of the Higgs potential in the SM. An important test of the SM is the measurement of processes involving the Higgs boson self-coupling. These processes can be used to measure the cubic and quartic terms of the Higgs potential directly, which is instrumental to validate the consistency of the theory. Measurements of the trilinear and quartic Higgs boson self-couplings can be performed in final states with two or three Higgs bosons, respectively. The determination of the quartic self-coupling is currently infeasible due to the small cross section of triple Higgs boson production of about 80 ab in pp collisions at $\sqrt{s} = 13$ TeV [42]. Instead, the focus of this thesis lies on searching for Higgs boson pair production and, in doing so, probing the trillinear Higgs boson self-coupling.[11]

The production of Higgs boson pairs at the LHC is a rare process with cross sections about 1000 times smaller than the production of single Higgs bosons. Under the SM hypothesis, Higgs boson pair production is not yet accessible using data collected during Run 2 of the LHC; however, first evidence is likely to be obtained by the end of the LHC programme [45]. Nevertheless, probing Higgs boson pair production is valuable already. First, experimental methods can be developed and improved that might culminate in a discovery. Second, possible deviations from the SM can enhance the production of Higgs boson pairs to which searches might be sensitive already.

Production Modes

Higgs boson pair production in the SM (SM HH production) proceeds non-resonantly via different production modes, yielding final states with two on-shell Higgs bosons. The following description of the SM HH production modes assumes pp collisions at $\sqrt{s} = 13$ TeV and $m_H = 125.0$ GeV.

The ggF process is the dominant SM HH production mode. The leading order diagrams of this process are depicted in Figure 2.6. Both diagrams involve a loop of heavy quarks, the loop being dominated by top quarks due to their large coupling to the Higgs boson. The box diagram depicted in Figure 2.6(a) does not involve any Higgs boson self-interaction vertices. However, the triangle diagram in Figure 2.6(b) does involve the Higgs boson self-coupling thus making SM HH production sensitive to λ_{HHH}. The triangle and the box diagram interfere destructively leading to a small SM

[10] The decays of Higgs bosons to W^+W^- and ZZ are suppressed since $m_H < 2m_W < 2m_Z$; therefore, one of the gauge bosons has to be produced *off-shell*.

[11] The Higgs boson self-coupling can also be probed indirectly through higher-order corrections to the production of single Higgs bosons. See for example Refs. [43, 44].

HH production cross section via ggF of

$$\sigma(pp \to HH) = 31.05\,\text{fb} \quad \text{with} \quad \Delta\sigma/\sigma = {}^{+\,7\,\%}_{-23\,\%}$$

at NNLO $\text{FT}_{\text{approx}}$ [46–48].

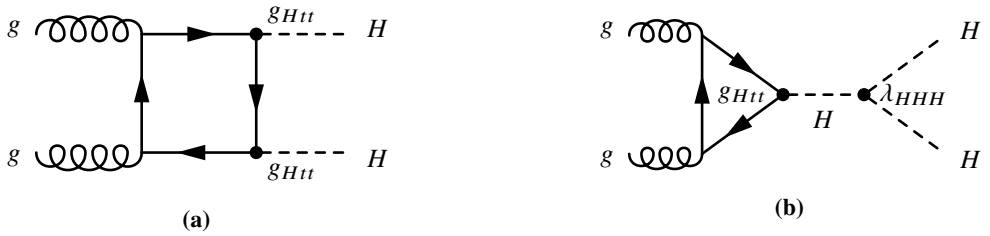


Figure 2.6: Feynman diagrams of non-resonant Higgs boson pair production via ggF at leading order. The diagrams (a) and (b) are commonly referred to as the box and triangle diagrams, respectively. The Higgs boson self-coupling is denoted by λ_{HHH} and the top-quark Yukawa coupling by g_{Htt}.

The second largest SM HH production mode is VBF, the leading order diagrams being depicted in Figure 2.7. The diagrams involve the coupling of the Higgs boson to vector bosons, g_{HVV}, the quartic $HHVV$ coupling, g_{HHVV}, and the Higgs boson self-coupling. A characteristic feature of the VBF production mode is the presence of two additional jets originating from the fragmentation and hadronisation of the final state quarks. The predicted SM HH production cross section via VBF is

$$\sigma(pp \to qqHH) = 1.726\,\text{fb} \quad \text{with} \quad \Delta\sigma/\sigma = \pm 2.1\,\%$$

at N^3LO [48, 49]. Due to the small cross section, the VBF production mode is currently of lesser importance to probe the nature of the Higgs boson self-coupling.

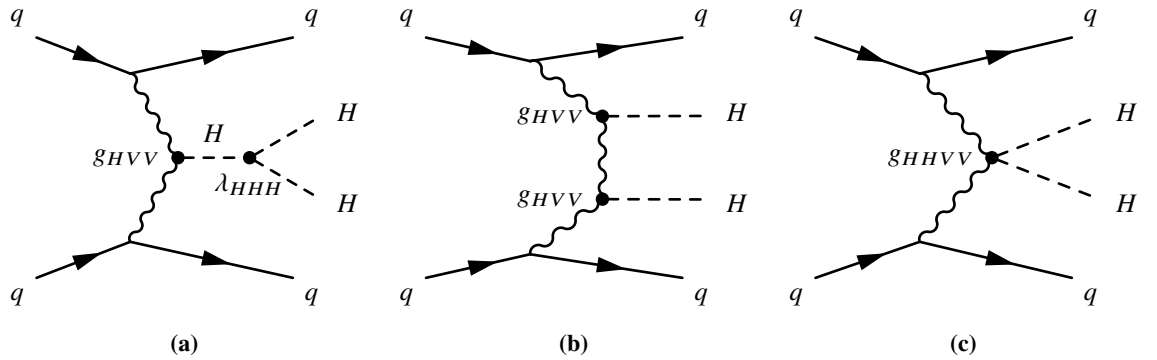


Figure 2.7: Feynman diagrams of non-resonant Higgs boson pair production via VBF at leading order.

Additional production modes of SM HH with even smaller cross sections exist. Examples of these are the associated production of HH with a massive vector boson with $\sigma(pp \to VHH) \approx 0.9\,\text{fb}$ [40] ($V = W^{\pm}$ or Z) and associated production with $t\bar{t}$ with $\sigma(pp \to t\bar{t}HH) \approx 0.8\,\text{fb}$ [40]. These production modes are not considered in this thesis.

Decay Channels of Pairs of SM Higgs Bosons

With the observed mass of the Higgs boson of about 125 GeV, many Higgs boson decay modes are of interest to probe the nature of the EWSB in the SM. When considering decays of Higgs boson pairs, this results in a plethora of final states with distinct signatures. An overview of the branching ratios of a system of two Higgs bosons is given in Figure 2.8.

	$b\bar{b}$	W^+W^-	$\tau^+\tau^-$	ZZ^*	$\gamma\gamma$
$b\bar{b}$	34 %				
W^+W^-	25 %	4.6 %			
$\tau^+\tau^-$	7.3 %	2.7 %	0.39 %		
ZZ^*	3.1 %	1.1 %	0.33 %	0.069 %	
$\gamma\gamma$	0.26 %	0.097 %	0.028 %	0.012 %	0.00052 %

Figure 2.8: Branching ratios of the decay of a system of two Higgs bosons. The Higgs boson branching ratios are taken from Ref. [40] and assume $m_H = 125.0\,\text{GeV}$. The figure is based on Ref. [50].

Experimental searches for SM HH production have to make a compromise between the branching ratio of the decay channel, the ability to select the relevant SM HH events, and the background contributions of other processes. Currently, the search channels most sensitive to SM HH production are:

The $b\bar{b}b\bar{b}$ channel has the largest branching ratio (34 %) of any final state. However, this channel is dominated by backgrounds from multi-jet production due to its fully hadronic final state. This makes searches in the $b\bar{b}b\bar{b}$ channel challenging. First, the signal acceptance is limited due to the use of b-jet triggers with strict p_T thresholds and the requirement of four b-tagged jets. Second, the reconstruction of $H \to b\bar{b}$ candidates in $b\bar{b}b\bar{b}$ final states is ambiguous, which introduces combinatorial backgrounds and a non-negligible fraction of misreconstructed signal events. Lastly, the modelling of the multi-jet background is difficult and introduces uncertainties that degrade the sensitivity of this channel.

The $b\bar{b}\tau^+\tau^-$ channel has a reduced branching ratio of 7.3 %, yet the presence of two τ leptons provides a distinct signature to select signal events and to suppress the multi-jet background. The relevant backgrounds in this channel are the production of $t\bar{t}$, Z + jets, multi-jet, and single-Higgs-boson events. Searches for SM HH in the $b\bar{b}\tau^+\tau^-$ channel often focus on final states with one leptonic and one hadronic τ-lepton decay ($\tau_\text{lep}\tau_\text{had}$), and final states with two hadronic τ-lepton decays ($\tau_\text{had}\tau_\text{had}$). These final states cover almost 90 % of $HH \to b\bar{b}\tau^+\tau^-$ events.

The $b\bar{b}\gamma\gamma$ channel has a very small branching ratio of 0.26 % but the photon pair provides an outstanding signature to select signal events. The di-photon invariant mass, $m_{\gamma\gamma}$, can be

reconstructed with a mass resolution of 1 to 2 GeV in the ATLAS and CMS detectors [51, 52].[12] Therefore, $m_{\gamma\gamma}$ provides an excellent discriminant to reject most background processes, leaving only continuum $\gamma\gamma$ production and single-Higgs-boson production as relevant backgrounds.

2.4 Physics Beyond the Standard Model

The SM is among the most precisely tested theories of physics. It had numerous successes in predicting phenomena before they were experimentally observed. For example, the SM predicted the existence of gluons, the $W^{\pm}$ and Z bosons, and the Higgs boson prior to their discovery. Despite its many successes, the SM is known to be an incomplete theory, leaving a number of phenomena unexplained:

Matter–antimatter asymmetry In current cosmological models, equal amounts of matter and antimatter are produced in the initial phase of the evolution of the universe. However, the universe observed today mostly consists of matter particles. This fact is referred to as the matter–antimatter asymmetry of the universe. According to the conditions proposed by Sakharov [53], violation of CP symmetry is required for the generation of such an asymmetry in the early universe. While CP violation has been observed in the quark sector [54] and first indications of CP violation in the lepton sector exist [55], its size might not be sufficient to explain the observed asymmetry.

Gravitation The theory of general relativity provides an accurate description of gravitation in the context of a classical field theory. However, no generally accepted approach exists that reconciles general relativity with the quantum theory underlying our current formulation of the SM. Moreover, understanding the weakness of gravitational interactions at the scale of elementary particles remains one of the open questions in particle physics.

Dark matter and dark energy Astrophysical observations [56–60] indicate that the majority of the matter content of the universe consists of *dark matter*, a form of matter that does not participate in the electromagnetic interaction. Its existence can, however, be inferred from the gravitational interaction between dark and ordinary matter. Provided dark matter is microscopic in nature, the SM does not provide a suitable dark matter candidate that is consistent with current cosmological models. Moreover, the observation of an accelerated expansion of the universe [61, 62] is hypothesised to originate from a form of energy referred to as *dark energy*. This energy is associated with the vacuum of space and its origin remains unexplained in the framework of the SM.

The electroweak hierarchy problem The mass of the Higgs boson is affected by virtual corrections involving loops of massive particles. These corrections diverge quadratically with the momentum scale Λ [63], which is the upper cut-off of the loop momentum integration. Based on arguments of *naturalness* [63–66], the radiative corrections to the Higgs boson mass should not exceed the electroweak scale of $O(100\,\text{GeV})$. Under this assumption, the SM cannot remain valid beyond a scale of $O(1\,\text{TeV})$, at which point the divergences would have to be regulated by BSM physics contributions. At the LHC, the SM is being probed at this scale and beyond; however, no direct

[12] Compared to the typical resolutions of $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ reconstruction in the $b\bar{b}b\bar{b}$ and $b\bar{b}\tau^+\tau^-$ channels, this represents an improvement of about an order of magnitude.

signs of new physics have been observed thus far. Consequently, it needs to be considered that the SM remains valid up to a larger energy scale, for example the scale of a *Grand Unified Theory* (GUT) of about 10^{16} GeV [18]. In this case, for the Higgs boson mass to remain at the electroweak scale would require excessive *fine-tuning* of the theory parameters; otherwise, radiative corrections would drag the Higgs boson mass towards, for example, the GUT scale.

Neutrino masses The observation of neutrino oscillations [19, 20] constitutes experimental evidence of neutrinos being massive particles. In the SM, it is assumed that neutrinos are massless particles, however. Extending the SM to incorporate non-vanishing neutrino masses poses the question whether neutrinos are Dirac or Majorana [67] particles. In addition, upper limits on the neutrino masses are $O(1\,\text{eV})$ for electron-based measurements [68], which are unnaturally small compared to the mass scales of other fermions (1 MeV to 100 GeV).

Even though the SM has its shortcomings, it describes many natural phenomena with excellent precision. Thus, it is often believed that the SM represents the low-energy manifestation of an extended theory that only becomes relevant at larger energy scales, for example a *Grand Unified Theory* that unifies the electroweak and strong interaction.

2.4.1 Non-Resonant Higgs Boson Pair Production

BSM phenomena might appear at energy scales beyond what can be experimentally probed using direct searches[13] at the LHC. Nevertheless, it is possible to test such models indirectly through their contributions to SM processes via virtual corrections. These corrections can, for example, alter the total or differential cross section of a given process.

A way of exploring BSM contributions to non-resonant HH production is to examine the Higgs boson self-coupling constant for possible deviations from the SM value of $\lambda^{\text{SM}}_{HHH} = 3m_H^2/v$. Such deviations can arise due to virtual corrections involving massive BSM particles as indicated in Figure 2.9. If the mass scale of the particles participating in the corrections is sufficiently large, then the dynamics of the BSM theory can be reduced to an effective interaction vertex between three Higgs bosons with a coupling constant

$$\lambda_{HHH} = \kappa_\lambda \times \lambda^{\text{SM}}_{HHH}\,,$$

where κ_λ is an arbitrary coupling modifier. A change in κ_λ would alter both the total cross section of non-resonant HH production and the kinematics of the final state particles. These effects are discussed in Chapter 7.

Current bounds on possible values of κ_λ from requirements of perturbative unitarity in $HH \to HH$ scattering allow for variations within $|\kappa_\lambda| \lessapprox 6$ [69].[14] These bounds still allow for ample variation of the Higgs boson self-coupling strength, further motivating searches for non-resonant HH production. These searches constitute a major part of Chapters 6 and 7 in which upper limits are set on the non-resonant HH production cross section of a signal with SM-like ($\kappa_\lambda = 1$) kinematics and signals with anomalous κ_λ ($\kappa_\lambda \neq 1$), respectively.

[13] i.e. searches for on-shell production of BSM particles.

[14] Similar arguments were made in the past to obtain upper limits on the Higgs boson mass from unitarity bounds in the scattering of longitudinally polarised vector bosons [70].

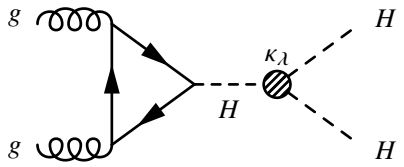


Figure 2.9: Non-resonant production of Higgs boson pairs for anomalous values of the Higgs boson self-coupling constant. Contributions of new physics, for example through loops of heavy BSM particles, are indicated as a hatched circle. The effective coupling constant in units of the self-coupling constant predicted by the SM is given by κ_λ.

2.4.2 Resonant Higgs Boson Pair Production

If BSM physics occurs at experimentally accessible energy scales, new particles could be produced directly (on-shell) in collider experiments. Further assuming that these particles are short-lived and decay into detectable SM particles, one can reconstruct the mass of such particles using the four-momenta of their decay products. The presence of BSM physics can then appear as an enhancement of the differential cross section $\mathrm{d}\sigma/\mathrm{d}m$, m referring to the invariant mass of the final state particles, in a region close to the mass of the new particle. This phenomenon is referred to as a *resonance* and the production of particles via an intermediate resonance as *resonant production*.

The Higgs sector is often used as an entry point for BSM physics. Aside from aesthetic reasons, there are currently no arguments that require nature to realise a *minimal Higgs model* with a single Higgs-doublet [71]. In fact, the Higgs sector can be readily extended with additional scalar fields with singlet and doublet representations under the SM gauge group [71]. Such extended Higgs sectors are part of many BSM theories, resulting in a phenomenology with new scalar particles. Under certain circumstances, these models allow for a sizeable production of SM-like Higgs boson pairs via intermediate scalar resonances. A possible Feynman diagram of resonant Higgs boson pair production is depicted in Figure 2.10. Two examples of models with extended Higgs sectors are given hereafter.

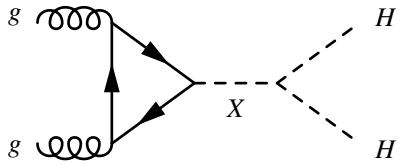


Figure 2.10: Resonant production of SM Higgs boson pairs via an intermediate scalar resonance X produced in ggF.

Additional Higgs-singlet models The simplest extension of the SM Higgs sector is the addition of a real scalar field ϕ_S that transforms as a singlet under the SM gauge group. This scalar field, being a gauge singlet, does not interact with any of the SM fermions or vector bosons. It could therefore be part of a "hidden sector" that might provide a suitable candidate for dark matter. In so-called *Higgs portal models* [72], the hidden sector can only be accessed through coupling/mixing of ϕ_S with the CP-even component of the SM Higgs field. Such models can predict resonant production of Higgs boson pairs through new scalar resonances [73–80].

A general choice for the potential of a Higgs sector extended by an additional scalar field

reads [77, 78, 80, 81]

$$V(\phi, \phi_S) = V(\phi) + \frac{a_1}{2}(\phi^\dagger\phi)\phi_S + \frac{a_2}{2}(\phi^\dagger\phi)\phi_S^2 + b_1\phi_S + \frac{b_2}{2}\phi_S^2 + \frac{b_3}{3}\phi_S^3 + \frac{b_4}{4}\phi_S^4\,,$$

where ϕ refers to the complex Higgs-doublet and $V(\phi)$ is the Higgs potential. In unitary gauge, the fields can be expanded about the vacuum state as $\phi = (0, v+H)^{\mathrm{T}}/\sqrt{2}$ and $\phi_S = v_S + S$, where v_S is the VEV of ϕ_S. After the expansion, terms bilinear in H and S appear in the potential, which indicate that the physical fields are mixtures of H and S. The physical fields H_1 and H_2 with masses m_1 and m_2, respectively, can be expressed as

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} H \\ S \end{pmatrix},$$

with a mixing angle θ. In the following, it is assumed that H_1 can be identified with the observed Higgs boson and H_2 is a new scalar with $m_2 > 2m_1$. The scalar H_2 can be produced via ggF through the admixture of H in H_2, however, suppressed by a factor of $\sin^2\theta$. Moreover, the interaction terms of the scalar potential allow for decays of H_2 into a pair of H_1. As a consequence, resonant production processes according to $pp \to H_2 \to H_1H_1$ are possible given the prior assumptions.

Two-Higgs-doublet models (2HDM) Generic 2HDM extend the Higgs sector of the SM by introducing an additional $SU(2)_{\mathrm{L}}$ doublet of complex scalar fields [71, 82]. Such extensions are motivated by theories such as supersymmetry [83], which require at least two Higgs-doublets to generate masses of up- ($I_3 = +1/2$) and down-type ($I_3 = -1/2$) fermions, or models of electroweak baryogenesis [84] in which 2HDM can provide CP-violating processes necessary to generate a matter–antimatter asymmetry in the universe.

Further discussion is restricted to flavour- and CP-conserving 2HDM, which, for example, include models describing the Higgs sector of minimal supersymmetric extensions of the SM (MSSM).[15] The particle spectrum of these models consist of five scalar particles after EWSB: two CP-even Higgs bosons H_1 and H_2, a CP-odd Higgs boson A, and two charged Higgs bosons $H^\pm$. Similar to the model with an additional Higgs-singlet, the physical fields H_1 and H_2 are mixed states of the CP-even components of both Higgs-doublets and interaction vertices of the form $H_1H_1H_2$ exist [71, 82]. This can allow for resonant production processes according to $pp \to H_2 \to H_1H_1$, which are promising search channel for heavy, CP-even Higgs boson in certain BSM scenarios [76, 86, 87].

The selected examples are not intended to be comprehensive but rather serve to illustrate how resonant HH production can arise in models with extended Higgs sectors. In this thesis, the benchmark signal process is the decay of a scalar resonance X produced via ggF into a pair of SM Higgs bosons (cf. Figure 2.10).

[15] For CP-violating 2HDM as possible explanations of electroweak baryogenesis, see for example the *complex 2HDM* (C2HDM) in Ref. [85].

2.5 Previous Searches for Higgs Boson Pair Production

The following summarises the experimental status of searches for Higgs boson pair production prior to the work performed as part of this thesis. The focus lies on results of the ATLAS and CMS collaborations obtained using pp collision datasets at $\sqrt{s} = 13\,\text{TeV}$ collected during the 2015 and 2016 data-taking periods at the LHC.

Searches for SM Higgs Boson Pair Production

The ATLAS and CMS collaborations set upper limits on the cross section of SM HH production via $gg\text{F}$, $\sigma_{gg\text{F}}$, at 95 % CL. The results of both collaborations are summarised in Figure 2.11. Searches for SM HH production were performed in various channels, the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels being the most sensitive ones. A statistical combination of SM HH searches was performed by both collaborations. The observed (expected) upper limits on $\sigma_{gg\text{F}}/\sigma_{gg\text{F}}^{\text{SM}}$ of the combination are 6.9 (10) and 22.2 (12.8) ATLAS and CMS collaboration, respectively [88, 89]. These were the most stringent limits on SM HH production at the time.

Constraints on the Strength of the Higgs Boson Self-Coupling

The ATLAS and CMS collaborations reinterpreted the searches for SM HH production in the context of anomalous values of the trilinear Higgs boson self-coupling constant. Upper limits at 95 % CL were set on the cross section of non-resonant HH production as a function of the Higgs boson self-coupling modifier, κ_λ. All other couplings were fixed to their SM values. The results of both collaborations are summarised in Figure 2.12. The κ_λ interval in which the upper limit on the cross section does not exclude the cross section predicted by theory is considered as the *allowed κ_λ interval*. The results depicted in Figure 2.12 yield allowed κ_λ intervals of

$$-5.0 < \kappa_\lambda < 12.0\,\text{(observed)} \qquad -5.8 < \kappa_\lambda < 12.0\,\text{(expected)}$$

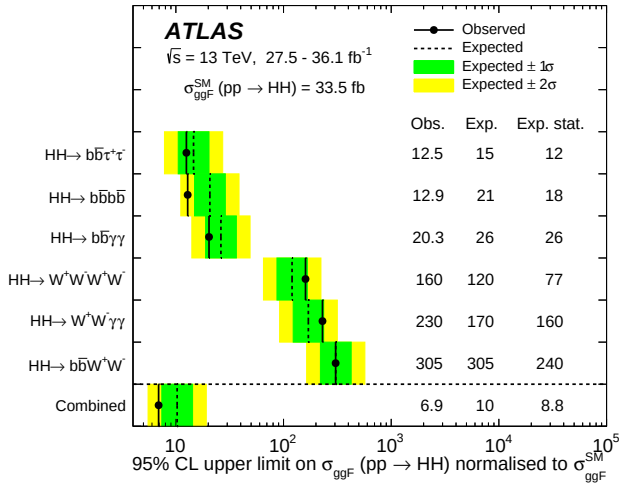
for the result of the ATLAS collaboration [88] and

$$-11.8 < \kappa_\lambda < 18.8\,\text{(observed)} \qquad -7.1 < \kappa_\lambda < 13.6\,\text{(expected)}$$

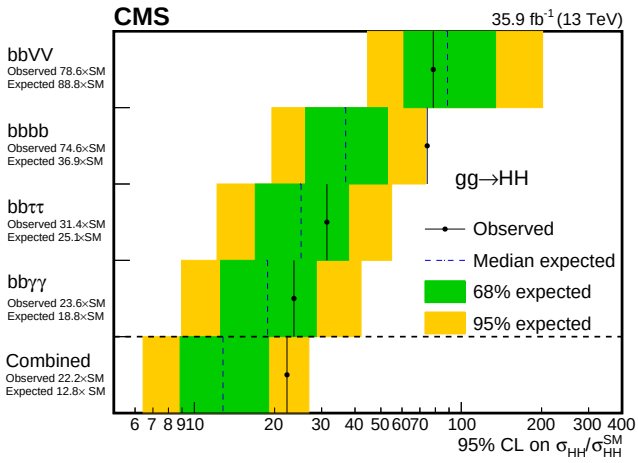
for the result of the CMS collaboration [89].

Searches for Resonant Production of Higgs Boson Pairs

The ATLAS and CMS collaborations performed searches for CP-even, scalar resonances with narrow width decaying into a pair of SM Higgs bosons. Resonance masses ranging from the HH production threshold up to 3 000 GeV are considered by both collaborations. The upper limits at 95 % CL on the production cross section of the scalar resonance as a function of its mass are shown in Figure 2.13. Neither the ATLAS nor the CMS result shows a statistically significant excess in the search for resonant HH production.

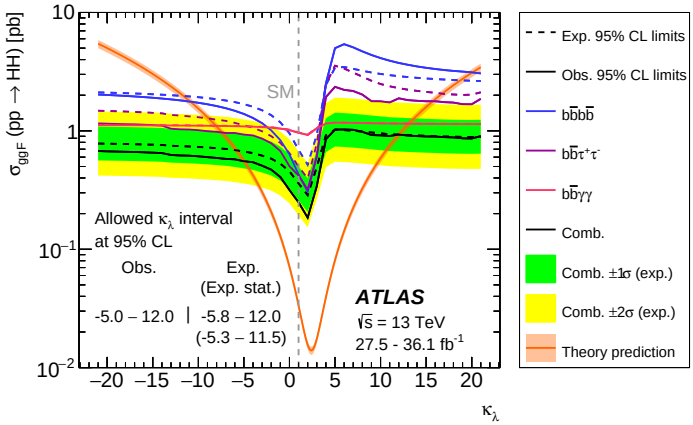


(a) Results of SM HH searches by the ATLAS collaboration. Upper limits excluding systematic uncertainties are given in the "Exp. stat." column. The figure is taken from Ref. [88].

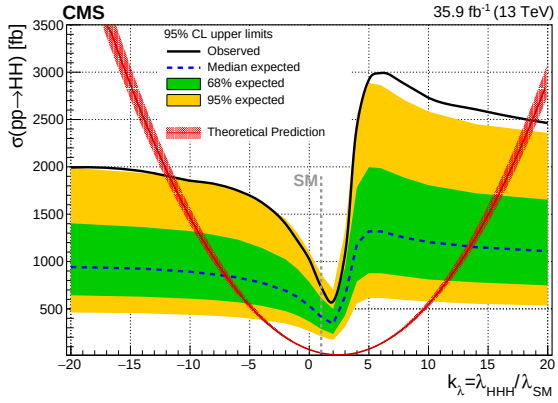


(b) Results of SM HH searches by the CMS collaboration. The $b\bar{b}VV$ ($V = Z$ or $W^{\pm}$) channel targets final states with two charged leptons. The figure is taken from Ref. [89].

Figure 2.11: Upper limits at 95 % CL on the cross section of SM HH production via ggF by the ATLAS (a) and CMS (b) collaborations. The upper limits are normalised to a SM cross section prediction of $\sigma_{ggF} = 33.5\,\text{fb}$ and given separately for the individual channels, and the statistical combination of all listed channels. In both cases, the expected limits are derived under the background-only hypothesis (i.e. no SM HH production). The results are based on pp collision data taken in 2015 and 2016.

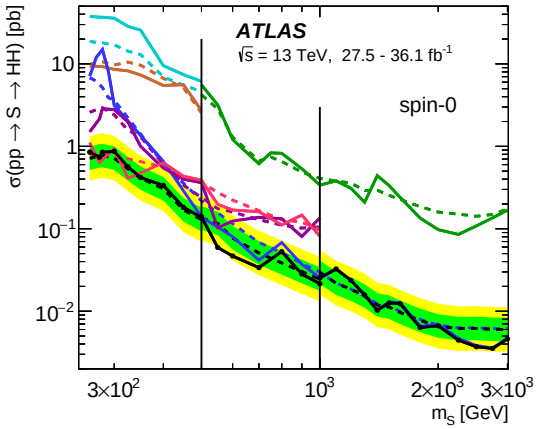


(a) Results of the ATLAS collaboration for the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels and their combination. The figure is taken from Ref. [88].

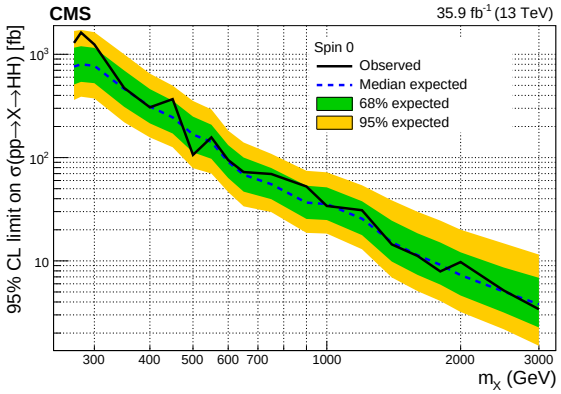


(b) Results of the CMS collaboration for the combination of the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, and $b\bar{b}VV$ channels. The figure is taken from Ref. [89].

Figure 2.12: Upper limits at 95 % CL on the cross section of non-resonant HH production as a function of κ_λ by the ATLAS (a) and CMS (b) collaborations. The expected upper limits are obtained under the background-only assumption (i.e. no non-resonant HH production). Values of κ_λ where the theoretical prediction exceeds the upper limit are excluded by the measurements. The results are based on pp collision data taken in 2015 and 2016.



(a) Results of the ATLAS collaboration for the channels: $b\bar{b}b\bar{b}$ (■), $b\bar{b}\tau^+\tau^-$ (■), $b\bar{b}\gamma\gamma$ (■), $b\bar{b}W^+W^-$ (■), $W^+W^-\gamma\gamma$ (■), and $W^+W^-W^+W^-$ (■). The statistical combination of all channels is shown in black. The observed (expected) upper limits are depicted as solid (dashed) lines. The figure is taken from Ref. [88].



(b) Results of the CMS collaboration for the statistical combination of the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, and $b\bar{b}VV$ ($V = Z$ or $W^\pm$) channels. The figure is taken from Ref. [89].

Figure 2.13: Upper limits at 95 % CL on the production cross section of CP-even, scalar resonances (S / X) decaying into a pair of SM Higgs bosons by the ATLAS (a) and CMS (b) collaboration. The expected upper limits are derived assuming the background-only hypothesis. The results are based on pp collision data taken in 2015 and 2016.

CHAPTER 3

The ATLAS Experiment at the Large Hadron Collider

This chapter describes the experimental environment surrounding the ATLAS experiment at the LHC. It provides the context for the development of algorithms for the identification of hadronic decays of τ leptons (Chapter 5) and searches for Higgs boson pair production (Chapters 6 and 7) at the ATLAS experiment. The chapter is structured as follows: The LHC is introduced in Section 3.1 including important parameters of the machine. In Section 3.2, the ATLAS detector, one of two general-purpose particle detector experiments at the LHC, is described. The chapter concludes in Section 3.3 by summarising the techniques used to reconstruct particle collision events at the LHC with the ATLAS detector.

3.1 The Large Hadron Collider at CERN

The LHC [90] is a particle accelerator located at CERN[1] at the French–Swiss border in Geneva, Switzerland. At present, the LHC is the world's highest-energy, laboratory-based particle collider, accelerating protons to energies of up to 6.8 TeV and reaching pp centre-of-mass energies ($\sqrt{s}$) of up to 13.6 TeV. The LHC is also used to accelerate heavy ions; however, the focus of this section lies in the operation of the LHC in pp collision mode.

The LHC was constructed in the former tunnel of the CERN Large Electron–Positron Collider (LEP) with a circumference of 26.7 km and became first operational in 2008. It is a synchrotron consisting of two counter-rotating beams of protons that are accelerated using alternating electric fields in superconducting radio frequency resonators. The beams are bent into a cyclic trajectory around the LHC ring using superconducting dipole magnets with field strengths of about 8 T. Along the ring, numerous quadrupole magnets are used for controlled focusing and defocusing of the proton beams. The proton beams consist of localised packages of ca. 10^{11} protons, hereafter referred to as bunches, circulating in the LHC with a minimum spacing in time of 25 ns [90].

The proton energies necessary for the injection into the LHC are achieved using a sequence of particle accelerators at CERN, schematically depicted in Figure 3.1. Protons are first accelerated in LINAC 2 after which they pass through the Proton Synchrotron Booster (BOOSTER), the Proton

[1] From the French *Conseil Européen pour la Recherche Nucléaire* referring to both the research organisation and the location of the laboratory sites.

Synchrotron (PS), and the Super Proton Synchrotron (SPS). The SPS accelerates protons to an energy of 450 GeV [90], subsequently injecting the protons into the LHC. In the LHC, the protons are further accelerated until the target energy is reached. Afterwards, the beams are brought into collision at four points, the interaction points (IPs), along the ring. Four large experiments are situated at the IPs to observe and record particle collision events: ATLAS [51], CMS [52], ALICE [91], and LHCb [92]. The ATLAS and CMS experiments are particle detector experiments targeting largely overlapping, extensive physics programmes, while the ALICE and LHCb adopt more specialised programmes. The ALICE experiment studies the production of the quark–gluon plasma, a state of matter with asymptotically free quarks and gluons occurring at temperatures similar to those right after the Big Bang, in heavy-ion collisions. The LHCb experiment investigates decays of heavy-flavour quarks to, for example, shed light on the nature of the matter–antimatter asymmetry in the universe by studying CP violation in the quark sector. Several smaller experiments are installed at the LHC to study physics processes at small angles with respect to the LHC beamline or to search for exotic particles. These experiments are LHCf [93], TOTEM [94], MoEDAL [95], FASER [96], and SND@LHC [97].

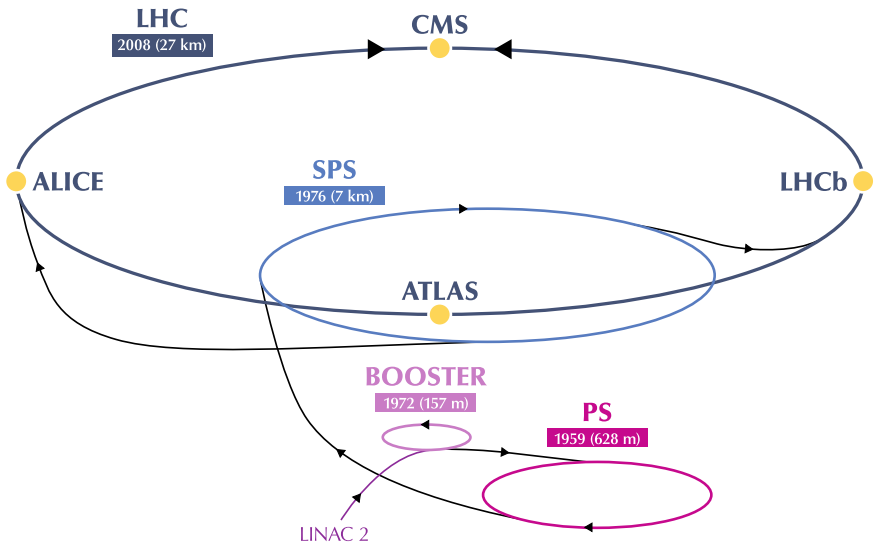


Figure 3.1: Illustration of the LHC and its pre-accelerators when operating in proton–proton collision mode during the data-taking period from 2015–2018 (Run 2). The year of first operation and circumference of the accelerators is given in solid coloured boxes. The figure is adapted from Ref. [98].

The main operating period of the LHC started in 2010 with pp collisions at $\sqrt{s} = 7$ TeV. In 2012, the centre-of-mass energy was increased to 8 TeV. The data-taking period from 2010 to 2012 is referred to as *Run 1* of the LHC. After extensive upgrades of the LHC and the detectors, the LHC restarted operation with pp collisions at $\sqrt{s} = 13$ TeV in *Run 2*, which took place from 2015 to 2018. After another shutdown of LHC and experiments, data-taking recommenced in 2022 with *Run 3*, reaching unprecedented energy scales of $\sqrt{s} = 13.6$ TeV. At present, Run 3 is foreseen to last until the end of 2025 [99].

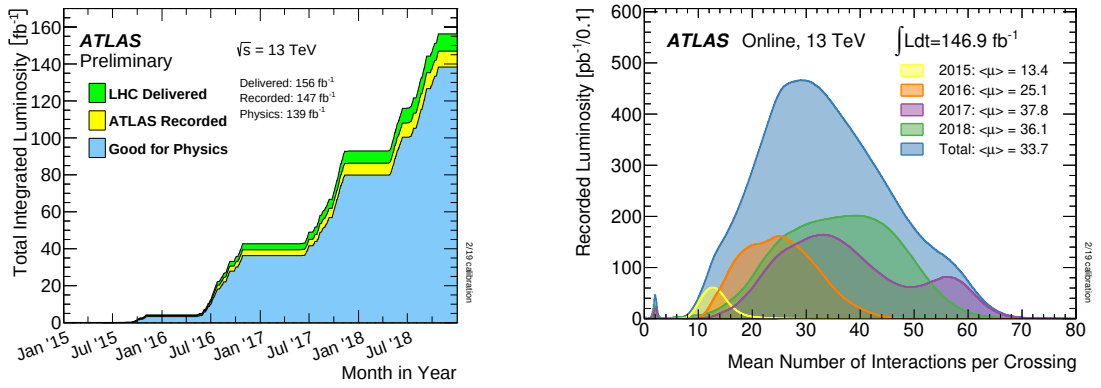
An important performance characteristic of a particle collider is the instantaneous luminosity, L, at a given interaction point. For a process p, the instantaneous luminosity relates the expected number of

events from the process per unit time, $\mathrm{d}N_p/\mathrm{d}t$, to the cross section of the process, σ_p, according to

$$\frac{\mathrm{d}N_p}{\mathrm{d}t} = L\sigma_p \,.$$

The expected number of events over a time interval, assuming constant cross section, is given by $N_p = L_{\mathrm{int}}\,\sigma_p$, where $L_{\mathrm{int}} = \int \mathrm{d}t\, L(t)$ is referred to as the integrated luminosity. When searching for rare physics processes occurring at high-energy scales, it is typically desirable to perform collisions at the largest, experimentally feasible L over extensive time periods to maximise the expected number of events from the rare process.

The integrated luminosity delivered to the ATLAS experiment by the LHC during Run 2 is shown in Figure 3.2(a). In this data-taking period, the LHC delivered pp collisions with an integrated luminosity of about 156 fb^{-1} of which 139 fb^{-1} pass the data-quality requirements of the ATLAS experiment [100]. The peak instantaneous luminosity at the IP of the ATLAS experiment ranged from $0.5 \times 10^{-34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at the beginning, to $1.9 \times 10^{-34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at the end of the Run 2 [100]. A quantity related to the instantaneous luminosity is the expected number of inelastic pp interactions per bunch crossing, μ. Due to the large cross section of inelastic pp collisions at $\sqrt{s} = 13$ TeV of about 80 mb [101], multiple interactions occur in a single crossing of the proton bunches. These interactions contaminate collision events of interest and are referred to as pile-up. The distribution of μ is depicted in Figure 3.2(b) at the IP of the ATLAS experiment during Run 2, showing that on average 33.7 inelastic collision events are expected to occur in a single bunch crossing in the combined Run 2 pp collision dataset.



(a) **(b)**

Figure 3.2: The integrated luminosity as a function of time (a) and the distribution of the mean number of interactions per bunch crossing (b) at the ATLAS experiment for pp collisions during Run 2 of the LHC. In Figure (a), the integrated luminosity delivered by the LHC (green), recorded by the ATLAS detector (yellow), and passing data-quality criteria [102] (blue) is shown. In Figure (b), the mean number of interactions per bunch crossing is calculated from the instantaneous luminosity assuming an inelastic pp collision cross section at $\sqrt{s} = 13$ TeV of 80 mb. The figures are taken from Ref. [103].

3.2 The ATLAS Detector

The ATLAS detector [51], shown in Figure 3.3, is a cylindrical particle detector surrounding the LHC beamline at one of the IPs. The detector covers most of the solid angle surrounding the IP to ensure that a large fraction of particles produced in hard scattering reactions pass through the active detector volume. The central part of the ATLAS detector is referred to as the *barrel*, while the two sections covering the solid angle close to the LHC beamline are referred to as the *end-caps*. Different layers of detector technologies are concentrically arranged around the IP that, in conjunction, allow for the detection and identification of different types of particles, enabling an almost full interpretation of collision events.

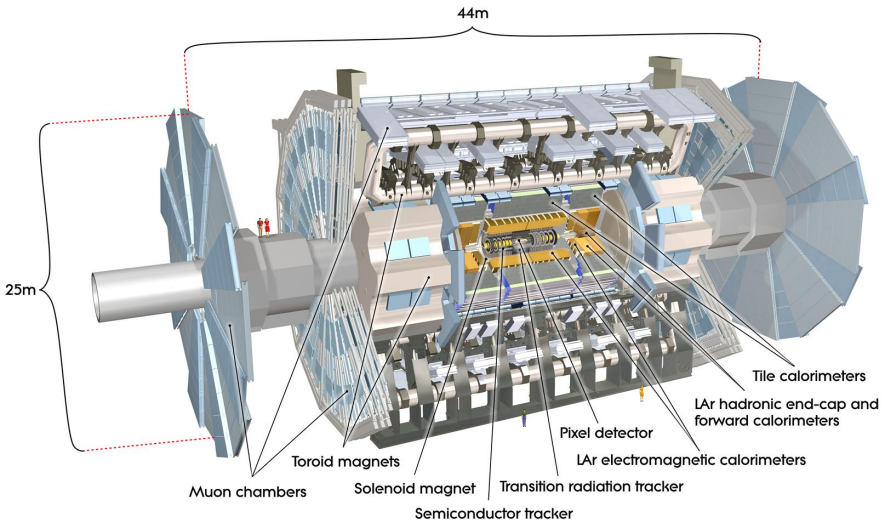


Figure 3.3: Overview of the ATLAS detector. The image is taken from Ref. [51].

The ATLAS experiment uses a right-handed Cartesian coordinate system with the origin being located in the centre of the detector at the nominal IP. The axes of the coordinate system are given as follows: the x-axis points to the centre of the LHC, the y-axis points upwards, and the z-axis points along the LHC beamline. The plane spanned by the x- and y-axes is referred to as the transverse plane. A spherical coordinate system is used to specify directions in three-dimensional space. The azimuthal angle, ϕ, of this coordinate system is defined as the angle in the transverse plane measured with respect to the x-axis, and the polar angle, θ, being the angle with respect to the z-axis. With these coordinate systems, transverse momenta and energies are defined as $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} = p \sin\theta$ and $E_{\mathrm{T}} = E \sin\theta$, respectively. At hadron colliders, the polar angle is frequently given in terms of the pseudorapidity η, which is defined as

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

Similarly, the angular separation between two particles is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}\,,$$

where $\Delta\eta$ is the difference in pseudorapidity and $\Delta\phi$ the smallest azimuthal separation between the particle momenta.

The main components of the ATLAS detector, going from the IP outwards, are the *inner detector* (ID) used for measuring the trajectories of charged particles, the *calorimeters* used to destructively measure the energy of most charged and neutral particles, and the *muon spectrometer* (MS) used to measure the trajectories of muons that can pass the calorimeters. Particles in the ID are bent in the transverse plane due to a magnetic field of 2 T field strength pointing along the z-axis that is produced by a superconducting solenoid surrounding the ID. The MS is equipped with superconducting toroid magnets that bend the trajectories of muons in the direction described by the η-coordinate. The curvature of charged-particle trajectories in the ID and MS is used to determine the sign of the electric charge and momentum of particles. The following sections summarise the sub-systems of the ATLAS detector.

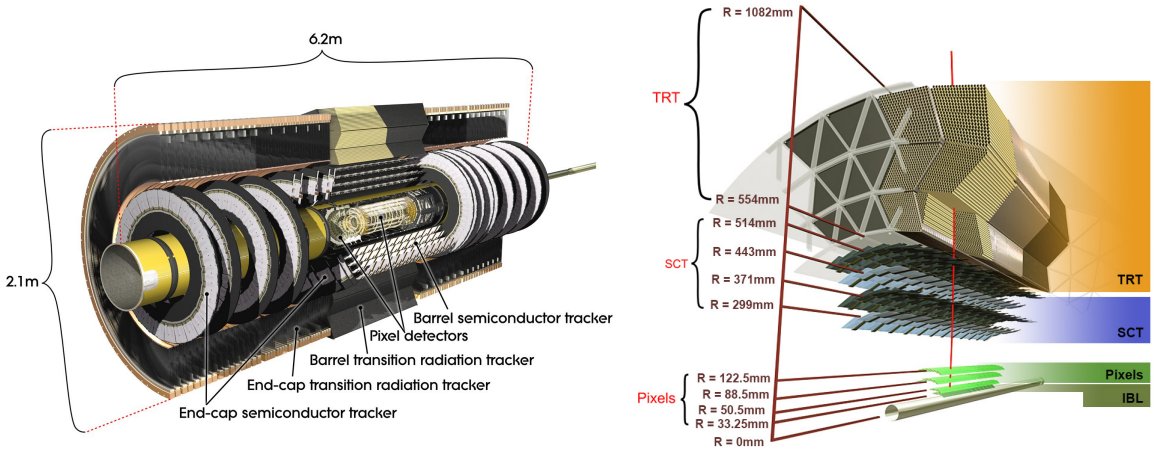
3.2.1 The Inner Detector

The ID, schematically depicted in Figure 3.4, is the innermost part of the ATLAS detector. It performs non-destructive measurements of the trajectories of charged particles within $|\eta| < 2.5$ by measuring the points where charged particles cross active detector layers. The measurement of several points on the trajectory is used to reconstruct the trajectory of the particle. The reconstructed trajectory is referred to as the *charged-particle track*, which is often abbreviated as *track*. The precise measurement of charged-particle tracks is important to reconstruct the primary vertex (PV) of the hard interaction with good spatial resolution. This facilitates the removal of tracks from pile-up, which are typically displaced from the PV along the z-axis. Moreover, the accurate reconstruction of secondary vertices, i.e. displaced vertices originating from decays of unstable particles produced in the hard interaction, is crucial to identify jets originating from b quarks.

The requirements on the tracking system vary with the distance from the IP. Closest to the IP, tracking detectors with high-granularity are required for the reconstruction of primary and secondary vertices that can operate in a high-radiation environment. At larger distances, the particle flux is significantly reduced and the requirements on the spatial resolution relaxed. Therefore, different detector technologies are used to cover the needs of the tracking system in a cost-effective manner.

The ID sub-system closest to the beampipe are the pixel detectors located at distances of 33.25 to 122.5 mm from the beamline in the barrel region of the ATLAS detector (cf. Figure 3.4). The pixel detectors are based on semiconductor technology with an active detector area that is segmented into a grid of rectangular elements, referred to as pixels. These pixels have size of 50 µm in the transverse direction and 250 to 400 µm along the beamline [51, 105]. A charged particle traversing the pixel detector ionises the active detector material leading to the deposition of electric charge in nearby pixels. The charge deposited in individual pixels can be read out to determine the point where the particle crossed the detector layer. The pixel detectors are arranged in four layers concentric with the beamline in the barrel region and three disks per end-cap region. Particles produced at the IP typically traverse four layers of pixel detectors within the acceptance of $|\eta| < 2.5$ of the tracking system.

The pixel detector is surrounded by the *semiconductor tracker* (SCT) covering radii of 299 to



(a) **(b)**

Figure 3.4: Schematic view of the ID including the two end-caps (a). A zoomed in view of the barrel region of the ID (b). The Insertable B-Layer (IBL), the innermost pixel detector layer, was installed after Run 1 of the LHC [104, 105]. The images are taken from Refs. [51, 106].

514 mm from the beamline in the barrel region (cf. Figure 3.4). Similar to the pixel detector, the SCT is based on semiconductor detector technology; however, the active detector area is segmented in long but thin strips (microstrips) with a pitch between strips of about 80 µm [51]. A single microstrip detector layer only provides a position measurement in the plane perpendicular to the strips. Therefore, SCT modules consist of two layers of microstrips that are tilted by a small angle to resolve the ambiguity in the direction parallel to the strips. The SCT is arranged in four layers in the barrel and nine disks in the end-cap region, yielding at least four measurements for tracks within the acceptance of the tracking system [51].

The *transition radiation tracker* (TRT) is the final layer of the ID covering radii of 554 to 1 082 mm (cf. Figure 3.4), however, with reduced acceptance of $|\eta| < 2.0$ compared to the semiconductor-based trackers. The TRT consists of about 300 000 straw-tube chambers, which are gaseous ionisation detectors, with a diameter of 4 mm arranged in up to 73 layers in the barrel region and 160 planes in the end-caps [51]. The straw-tube chambers determine the radius at which an ionising particle passed through the tube; however, no position information parallel to the straw-tubes is obtained. The straw-tubes of the TRT are interleaved with polypropylene fibres (foils) in the barrel (end-cap) region [51]. These fibres/foils provide interfaces between different dielectric media at which ultrarelativistic particles can emit transition radiation in the form of X-rays [107]. At the energies of particles produced at the ATLAS experiment, the emission of transition radiation is only relevant for electrons and positrons. The X-rays emitted by electrons/positrons at the interfaces are absorbed in the xenon-based gas-mixture[2] inside the straw-tubes, leading to measurably higher ionisation compared to a charged particle passing through the drift tube. This sensitivity of the TRT to transition radiation is used to identify electrons and positrons.

[2] Selected drift tubes affected by gas leaks are using a cheaper, argon-based gas mixture [108].

3.2.2 The Calorimeter System

The ATLAS calorimeter system, schematically depicted in Figure 3.5, surrounds the ID and is used for destructive measurements of the energy of most charged and neutral particles. Of the particles predicted by the SM, only muons and neutrinos can traverse the ATLAS calorimeter system without being absorbed. The calorimeter system covers a pseudorapidity range of $|\eta| < 4.9$, thus instrumenting almost the full solid angle surrounding the IP. This is important for the detection of all measurable particles originating from collision events. In addition, a hermetic detector allows for the reconstruction of the sum of transverse momenta carried away by undetectable particles such as neutrinos. This quantity is referred to as the missing transverse momentum, which is discussed in Section 3.3.7.

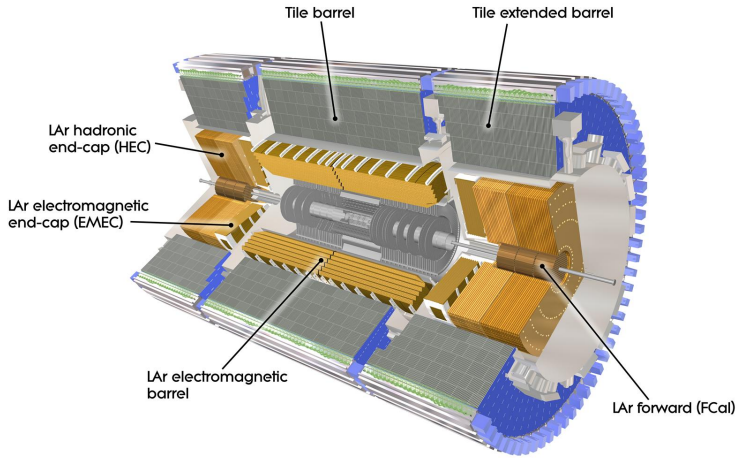


Figure 3.5: Overview of the ATLAS calorimeter system. The image is taken from Ref. [51].

The ATLAS calorimeter consists of two major parts. The *electromagnetic calorimeter*, the innermost part of the calorimeter, is used to measure the energies of electrons/positrons and photons. It is followed by the *hadronic calorimeter*, which measures the energy of charged and neutral hadrons. This design is governed by the interactions of particles with matter at the energy scales relevant for the ATLAS experiment. Highly energetic electrons/positrons and photons interact with the calorimeter material by emitting Bremsstrahlung and converting into electron–positron pairs, respectively. This process occurs repeatedly, producing a cascade of electrons, positrons, and photons, until the energy of the constituent particles is sufficiently small to be absorbed in the material. These cascades, referred to as *electromagnetic showers*, are homogeneous and compact with high energy density. The dimensions of the electromagnetic calorimeters are chosen such that they fully contain electromagnetic showers originating from promptly produced electrons/positrons and photons. The interactions of hadrons with the calorimeter material are driven by nuclear interactions, which also result in cascades of secondary particles. A variety of processes are relevant for the evolution of these so-called *hadronic showers*—for example, inelastic scattering of hadrons at nuclei, spallation, fission, or neutron capture [109]. Moreover, electromagnetic showers are an important sub-component of hadronic showers which, for example, originate from photons produced in π^0 decays. The length scales of hadronic showers are larger, both laterally and longitudinally, than the ones of electromagnetic showers. This is reflected in the design of the ATLAS calorimeter system, with the hadronic calorimeter having

almost four times the radial thickness of the electromagnetic calorimeter [110, 111].

The electromagnetic and hadronic calorimeters are sampling calorimeters, i.e. they consist of alternating layers of absorber material for shower development and active detector layers that sample the shower as it develops. Depending on the part of the calorimeter, the active detector layers are either ionisation chambers filled with liquid argon (LAr) or plastic scintillator tiles connected to photomultiplier tubes.

Electromagnetic Calorimeters

The electromagnetic calorimeter of the ATLAS detector (cf. Figure 3.5) is divided into a barrel ($|\eta| < 1.475$), end-cap ($1.375 < |\eta| < 3.2$), and forward ($3.1 < |\eta| < 4.9$) region [51]. The electromagnetic calorimeters in the barrel and end-cap region consist of lead absorbers with LAr-filled gaps that are instrumented with electrodes to form ionisation chambers. The calorimeter is segmented both laterally and longitudinally to provide information on the location and shape of showers in the calorimeter. The forward electromagnetic calorimeter uses copper absorbers and a single longitudinal layer instead.

Hadronic Calorimeters

The hadronic calorimeter (cf. Figure 3.5) is divided into a barrel ($|\eta| < 1.0$), extended barrel ($0.8 < |\eta| < 1.7$), end-cap ($1.5 < |\eta| < 3.2$), and forward ($3.1 < |\eta| < 4.9$) region [51]. The barrel and extended barrel calorimeter consists of alternating steel plates and plastic scintillator tiles. The scintillators are connected via optical fibres to photomultiplier tubes for readout. The hadronic end-cap calorimeter (HEC) and forward calorimeter consist of LAr-filled ionisation chambers with copper and tungsten absorbers, respectively. The hadronic calorimeter is also segmented laterally and longitudinally, although with reduced granularity compared to the electromagnetic calorimeter.

3.2.3 The Muon Spectrometer

The MS, shown in Figure 3.6, is the outermost and largest part of the ATLAS detector. It is a tracking detector that provides independent momentum measurements of muons that pass the calorimeters. The active detector elements of the MS are based on gaseous ionisation detectors for cost-effective instrumentation of large areas. The MS provides a coverage of $|\eta| < 2.7$ for precision tracking and $|\eta| < 2.4$ for triggering on muons.

The ATLAS detector uses multiple layers *monitored drift tube chambers*[3] for precision tracking in $|\eta| < 2.7$. To cope with the large rate of incident particles in the forward region, the innermost layer in $2.0 < |\eta| < 2.7$ is replaced by *cathode strip chambers*. The MS is complemented with *resistive plate chambers* and *thin gap chambers* in the barrel ($|\eta| < 1.05$) and end-cap ($1.05 < |\eta| < 2.7$) region, respectively, providing the necessary timing information to identify the bunch crossing a muon originates from. The MS is arranged, as depicted in Figure 3.6, in three cylindrical layers surrounding the LHC beamline in the barrel region. A gap in the instrumentation of the MS is left at $\eta = 0$ for access to the inner parts of the detector for servicing. The end-cap region is instrumented by muon chambers in the form of three wheels perpendicular to the beampipe. A *small wheel* is located in front

[3] Deformations of the muon chambers are actively monitored using an optical alignment system built into the frame of the chambers.

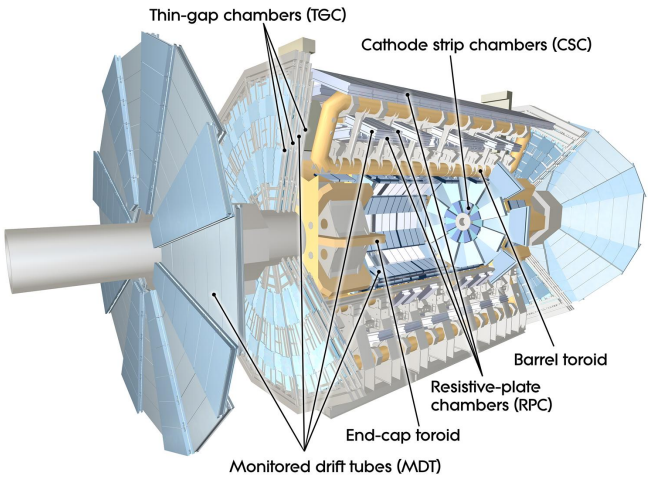


Figure 3.6: Overview of the ATLAS muon spectrometer. The image is taken from Ref. [51].

of the end-cap toroid, and two *big wheels* are located behind the end-cap toroid. Every layer/wheel consists of multiple tracking planes.

3.2.4 The ATLAS Trigger System

With a peak bunch crossing rate of 40 MHz and an expected number of inelastic pp collisions in a single bunch crossing of around 30 (cf. Figure 3.2), it is not possible to record every collision event due to limitations in the detector read-out and data storage. However, most collision events are not of particular interest to the physics programme of the ATLAS experiment. For example, events containing a weak vector boson decaying into leptons are expected to occur about once every 10^6 inelastic pp collisions at $\sqrt{s} = 13$ TeV [112]. Other processes, such as the production of Higgs bosons, are multiple orders of magnitude less frequent than the production of Z and W bosons.

The task of the ATLAS trigger system [113] is to reduce the rate at which events are recorded by performing an event pre-selection already during data-taking (*online*). This is achieved by monitoring the ATLAS detector for signatures of rare physics processes, for example indicated by the presence of charged leptons with large transverse momenta. A multitude of different event selections, usually referred to as *triggers*, are implemented in the ATLAS trigger system to target the signatures of different physics processes. The implementation of these triggers are heavily constrained due to the limited detector read-out and strict latency and bandwidth constraints. Due to the increasing instantaneous luminosity during Run 2 of the LHC, the triggers employed by the ATLAS experiment had to evolve to meet the bandwidth requirements. Therefore, the exact selections of triggers targeting particular signatures changed as Run 2 progressed. During Run 2, the ATLAS trigger system reduced the rate of selected events to about 1.2 kHz [113], which could be recorded for (*offline*) analysis.

The ATLAS trigger system consists of two stages. The first stage, called the Level-1 (L1) trigger, implements the event selection in hardware using custom electronic circuits. At the L1 trigger, only calorimeter and MS information with reduced granularity can be used for selection. The rate of accepted events is reduced to 100 kHz [113] by the L1 trigger. If an event is accepted, all sub-detectors of the ATLAS experiment are read-out and the data are sent to the second stage of the trigger system,

the so-called High-Level Trigger (HLT). The HLT is software-based and runs reconstruction algorithms similar to those applied during offline event reconstruction. Event selections at the HLT are staged into *chains* of algorithms with increasing computational complexity. A given HLT chain is only applied to events selected by particular L1 triggers, usually referred to as the L1 seed of the chain. The selections applied by the HLT reduces the mean rate of accepted events to the target of 1.2 kHz [113].

3.3 Reconstruction of Collision Events in the ATLAS Detector

Particles produced in pp collision events are reconstructed from their signals in the sub-systems of the ATLAS detector. At first, low-level objects such as charged-particle tracks, primary vertices, and clusters of calorimeter cell signals are reconstructed. These objects are used to reconstruct and identify high-level objects that correspond to the signatures of particles such as electrons, muons, τ leptons, or quarks/gluons in the form of jets. These objects, often referred to as *physics objects*, provide estimates of the four-momentum and other properties of the underlying particle. The same algorithms are used to reconstruct and identify objects in simulated pp collision events and in data recorded with the detector. Differences in the performance of the object reconstruction and selection in simulation and data are accounted for by calibration measurements of selection efficiencies, energy/momentum scales and resolutions. The following provides a summary of the object reconstruction relevant for this thesis.

3.3.1 Charged-Particle Tracks and Primary Vertices

The reconstruction of charged-particle trajectories is referred to as track reconstruction or tracking. The inputs to the track reconstruction in the ID of the ATLAS detector are *space-points* from the pixel and SCT detector, and *drift-circles* from the TRT. Space-points are measurements of location in three-dimensional space obtained by clustering the charge signals of adjacent segments in the pixel and SCT detectors [114]. Drift-circles are measurements of distance from the anode wires of individual straw-tubes in the TRT determined by the electron drift times in the straw-tubes [108].

Track reconstruction employs pattern recognition techniques to select space-points and drift-circles that are compatible with the hypothesis of a charged particle traversing through the axial magnetic field in the ID. Least-squares fits are performed, using selected space-points and drift-circles, to determine the parameters characterising the trajectory. Initially, the track parameters are given at the point of closest approach in the transverse plane (*perigee*) to the beam-spot position [114]. Five parameters, $(d_0, z_0, \phi, \theta, q/p)$, describe the track at the perigee [115]. The transverse (longitudinal) distance of the perigee from the beam-spot position is given by d_0 (z_0), also referred to as the transverse (longitudinal) impact parameter of the track. The azimuthal and polar angle of the track at the perigee is given by ϕ and θ, respectively. The ratio of electric charge and momentum of the particle is given by q/p, which is related to the curvature of the track. After locating the vertex of the hard interaction of interest, tracks are often re-parameterised using this point as a reference instead of the beam-spot position.

In the ATLAS experiment, two primary tracking algorithms are used that are referred to as the *inside-out* and *outside-in* algorithms [114, 116, 117]. The inside-out algorithm starts by reconstructing tracks in the pixel and SCT detector, then extending the track using measurements in the TRT. In contrast, the outside-in algorithm starts with the reconstruction of track segments in the TRT, which are then combined with space-points from the pixel and SCT detectors. The outside-in algorithm is used to improve the reconstruction efficiency for tracks from secondary particles, such as electrons

produced in conversions of photons in the detector material, for which the inside-out algorithm can fail to reconstruct a track. The track reconstruction algorithms provide charged-particle tracks within the ID acceptance of $|\eta| < 2.5$ and with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$.

Reconstructed tracks are used to determine the locations of inelastic scattering interactions between protons in the ATLAS detector. These locations are marked by multiple charged-particle tracks originating from the same point along the beamline and are referred to as primary vertices (PVs). The ATLAS PV reconstruction [118] finds and reconstructs PVs using an iterative fitting procedure. It starts by reconstructing a vertex at the location of the highest track density along the z-direction with respect to the beam-spot position. An adaptive vertex fitter [119] is used to iteratively determine the vertex position while weighting reconstructed tracks according to their compatibility with the vertex position. After the fit, tracks that are incompatible with the fitted vertex are considered as unassociated and the process is repeated on all unassociated tracks. After the PV reconstruction, only PVs with two or more associated tracks are kept. The PV with the largest sum of p_{T}^2 of associated tracks is selected as the PV of the hard interaction. Other vertices are considered as originating from pile-up [118].

3.3.2 Topological Clustering of Calorimeter Cell Signals

The segmentation of the calorimeters in lateral and longitudinal direction allows for the reconstruction of the three-dimensional shape of electromagnetic and hadronic showers. These showers typically extend over multiple cells in the calorimeter, thus requiring the combination of several cells to reconstruct a shower. The ATLAS experiment uses a *topological calorimeter cell clustering algorithm* [120] to combine the signals of locally connected cells passing signal and noise thresholds, thereby suppressing noise from the calorimeter electronics and pile-up. These clusters of calorimeter cells, referred to as topo-clusters, are used to reconstruct electromagnetic and hadronic showers in the calorimeters. Due to fluctuations in the shower development and calorimeter noise, multiple topo-clusters are often required to fully reconstruct the calorimeter response to a single particle.

The ATLAS calorimeter has a different response to electromagnetic and hadronic showers, i.e. the calorimeter is *non-compensating*. By default, the energies of topo-clusters are calibrated assuming the response originates from an electromagnetic shower (EM scale). However, the development of electromagnetic and hadronic showers leads to differences in their shower shapes, as previously discussed in Section 3.2.2. This is exploited by the *local hadronic calibration* [120], which uses shower shape information of topo-clusters to determine their likely origin and apply an appropriately weighted combination of calibrations for electromagnetic and hadronic showers. The energy scale of topo-clusters after the local hadronic calibration is referred to as the LC scale.

Topo-clusters are used as inputs for the reconstruction of higher-level physics objects in the ATLAS detector. For example, topo-clusters at EM and LC scale are used for the reconstruction of electrons and hadronic τ-lepton decays, respectively.

3.3.3 Electrons

The reconstruction of electrons (and positrons) in the ATLAS detector exploits their characteristic signature of a charged-particle track in the ID pointing towards a narrow cluster of energy in the electromagnetic calorimeter. The reconstruction of electrons in the region of $|\eta| < 2.47$ and excluding the transition regions between the barrel and end-caps, $1.37 < |\eta| < 1.52$, is described in the following based on Refs. [121, 122].

Electron reconstruction is seeded by topo-clusters calibrated at EM scale that have more than 50 % of their energy located in cells of the electromagnetic calorimeter, hereafter referred to as EM topo-clusters. In addition, EM topo-clusters are only considered if the transverse energy in the electromagnetic part of the calorimeter, $E_{\mathrm{T}}^{\mathrm{EM}}$, exceeds 400 MeV. A first attempt of geometrically matching the EM topo-cluster to an ID track is made. If the EM topo-cluster cannot be matched to a well-reconstructed track but has a longitudinal and lateral shower shape similar to the signature of an electron, then a second pass of tracking is performed in a region surrounding the cluster. The second pass allows for up to 30 % of energy loss at each intersection with detector material due to the emission of bremsstrahlung. After an ambiguity resolution scheme in case multiple tracks match the cluster, the cluster is required to be geometrically matched to a single track. EM topo-clusters with matched tracks are considered as seeds for a *supercluster* reconstruction algorithm if the cluster fulfils $E_{\mathrm{T}}^{\mathrm{EM}} > 1$ GeV and the matched track passes reconstruction quality criteria. *Satellite clusters*, which are EM topo-clusters in the vicinity of the supercluster seed, are included in the supercluster to account for electromagnetic showers being reconstructed as multiple topo-clusters or the formation of additional clusters from electrons emitting bremsstrahlung. After applying initial calibrations and corrections to a supercluster, the track matching is repeated using the supercluster barycentre instead of the barycentre of the EM topo-cluster seed. Finally, multivariate calibrations of the electron energy and corrections from comparisons of reconstructed electrons in data and simulation are applied [122, 123].

Electrons that are promptly produced in the hard interaction are of interest for most physics analyses. Other sources of reconstructed electron candidates are quark- or gluon-initiated jets that mimic the signature of an electron, or electrons from secondary sources such as hadron decays or photon conversions. *Identification* and *isolation* requirements can be applied to select promptly produced electrons and reject electron candidates from other sources. Electron identification is performed using a likelihood-based classifier exploiting variables sensitive to the shower shape, reconstruction quality of the matched track, information about transition radiation emission in the TRT, and spatial and momentum matching between the track and the supercluster [122]. Furthermore, in most event topologies little detector activity is expected in the vicinity of promptly produced electrons. Therefore, isolation variables are defined that quantify the activity in an area surrounding the electron candidate using reconstructed tracks and topo-clusters in the calorimeters [122]. Selections are applied on these isolation variables to further reject backgrounds originating from jets and non-prompt electrons.

3.3.4 Muons

The reconstruction of muons in the ATLAS experiment targets the signature of a charged particle that is able to traverse the calorimeters with only minimal energy deposition due to ionisation. The instrumentation of the MS allows for the reconstruction of muons up to $|\eta| < 2.7$. The following description of muon reconstruction is based on Ref. [124].

A stand-alone reconstruction of tracks in the MS is attempted by first reconstructing straight-line segments in individual layers of the MS. Track candidates are constructed from multiple track segments compatible with the trajectory of a muon produced at the IP. These candidates seed a fit to obtain the muon trajectory and the associated track parameters. Different reconstruction methods are employed yielding five types of reconstructed muons:

Combined muons are reconstructed by matching a track in the MS to a track in the ID. A combined fit of the ID and MS track is performed, accounting for the ionisation energy loss of muons in

the calorimeters, to reconstruct the muon trajectory. Within $2.5 < |\eta| < 2.7$ combined muons may be reconstructed using short segments of ID tracks instead of fully reconstructed tracks.

Inside-out muons are reconstructed by extending a track in the ID with hits in the MS. The additional hits are used for a combined fit of the muon trajectory in the ID and MS.

MS extrapolated muons are reconstructed by extrapolating a stand-alone MS track to the beamline in cases where no matching ID track is found. The MS track defines the properties of the reconstructed muon.

Segment-tagged muons are reconstructed by matching an ID track to one or more short track segments in the MS. The muon properties are determined by the parameters of the ID track.

Calorimeter-tagged muons are reconstructed from ID tracks of charged particles with a signature in the calorimeters characteristic of a minimum ionising particle. The ID track is used to define the properties of calorimeter-tagged muons.

Several muon identification working points are defined using these muon types and additional requirements on the quality of the ID and MS tracks and the compatibility of ID and MS tracks in terms of charge and momentum. In this thesis, the *loose* and *medium* working points are used, which are summarised hereafter. Of these working points, the medium working point has the most stringent requirements on the quality of reconstructed muons. It requires muons to be reconstructed as combined or inside-out muons or, alternatively, MS extrapolated muons in $2.5 < |\eta| < 2.7$ to improve the reconstruction efficiency outside the acceptance of ID track reconstruction. The *loose* working point augments the medium working point by additionally allowing segment- and calorimeter-tagged muons in $|\eta| < 0.1$, a region with a gap in the MS instrumentation, and relaxing the MS track quality requirements applied to inside-out muons. Muons passing the *loose* working point are a superset of muons passing the *medium* working point.

Isolation requirements are applied to reconstructed muon candidates to distinguish prompt from non-prompt muons. A similar approach to the one adopted for electrons, previously discussed in Section 3.3.3, is used.

3.3.5 Jets and b-tagging

The production of quarks or gluons in hard scattering interactions leads to the development of collimated sprays of particles referred to as jets. Jets are produced as a result of the colour confinement in QCD, leading to a fragmentation of the initial quark/gluon until the energy of the resulting fragments is sufficiently small to bind into colourless hadrons. The primary constituents of jets are charged/neutral hadrons and photons from hadron decays. To reconstruct the kinematic properties of the quark/gluon that initiated the jet, the four-momenta of all particles from the fragmentation and hadronisation process have to be collected. This task is addressed by jet algorithms, which combine particles, or other entities, into clusters to reconstruct jets. In the ATLAS experiment, the anti-k_t jet clustering algorithm [125] is most frequently used.

The Anti-k_t Jet Clustering Algorithm

The anti-k_t jet clustering algorithm operates on collections of entities with defined four-momenta (e.g. topo-clusters). Any set of one or more entities can be viewed as a pseudo-jet with a four-momentum

given by the sum of the constituent four-momenta. Pseudo-jets are sequentially combined if they are close according to a distance metric. The distance between pseudo-jets i and j is defined by the anti-k_t algorithm as [125]

$$d_{ij} = \min\left(p_{\mathrm{T}i}^{-2}, p_{\mathrm{T}j}^{-2}\right) \frac{\Delta R_y(i,j)^2}{R^2},$$

where $p_{\mathrm{T}i}$ refers to the transverse momentum of pseudo-jet i, $\Delta R_y = \sqrt{\Delta y^2 + \Delta\phi^2}$ is the distance of two pseudo-jets in the $y\phi$-plane, with y being the rapidity,[4] and R is the jet-radius parameter of the algorithm. A stopping criterion for the clustering algorithm is provided by the distance $d_{i\mathrm{B}} = p_{\mathrm{T}i}^{-2}$ by proceeding as follows [125]:

1. Find the minimum distance d_{ij} for all pairs of distinct pseudo-jets ($i \neq j$) and the minimum distance $d_{i\mathrm{B}}$ for all pseudo-jets.
2. If $\min_{i,j}(d_{ij}) < \min_i(d_{i\mathrm{B}})$: The pair of pseudo-jets corresponding to the minimum in d_{ij} is combined to form a new pseudo-jet.
3. If $\min_{i,j}(d_{ij}) \geq \min_i(d_{i\mathrm{B}})$: The pseudo-jet yielding the minimum $d_{i\mathrm{B}}$ is declared as a jet and removed from the procedure.

These steps are repeated until no pseudo-jets are left.

The anti-k_t algorithm produces conical jets with a typical radius of $\Delta R_y = R$ provided no high-p_{T} emissions are located in the vicinity of the jet. The algorithm has desirable theoretical properties, such as reconstructed jets being insensitive to soft or collinear emissions [125], thus making it the choice of the default jet clustering algorithm in the ATLAS experiment. In this thesis, the jets are clustered using the anti-k_t algorithm with a jet-radius parameter of $R = 0.4$. The implementation of the FastJet library [126] is used.

Jet Reconstruction using Particle Flow

Earlier analyses performed by the ATLAS collaboration used jets reconstructed by applying the anti-k_t algorithm to topo-clusters in the calorimeters, disregarding any information about charged particles from the tracking system. Techniques combining information from tracking and calorimetry are now adopted by the ATLAS experiment for the reconstruction of jets. These are based on a reconstruction algorithm referred to as *particle flow* that attempts to reconstruct individual charged and neutral particles from their detector signatures.

Jet reconstruction based on particle flow, described in Refs. [127, 128], exploits the superior energy and angular resolution of reconstructing low-p_{T} charged hadrons from their track in the ID instead of using calorimetric measurements.[5] Neutral particles have to be reconstructed from topo-clusters in the calorimeter, making it necessary to subtract the energy deposited by charged hadrons to prevent double-counting. This subtraction is performed by the particle flow algorithm to reconstruct neutral

[4] The rapidity is defined as $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$, where E is then energy of a particle and p_z the momentum component along the beamline. In the ultrarelativistic limit, the rapidity is equivalent to the pseudorapidity.

[5] The charged pion mass is assumed for the relationship between the momentum and energy of charged particles in the tracker.

particles in the form of *neutral particle flow objects*. Similarly, charged hadrons reconstructed from their track in the ID are referred to as *charged particle flow objects*. Charged and neutral particle flow objects are used as inputs to the anti-k_t jet clustering algorithm to reconstruct jets used for physics analyses.

Jet reconstruction using particle flow has several advantages compared to a calorimeter-based approach as detailed in Ref. [127]. The energy resolution for jets at low transverse momenta ($p_{\mathrm{T,\,jet}} \lessapprox 50\,\mathrm{GeV}$, see for example Ref. [128]) is improved due to the inclusion of tracking information. In part, this is due to the ability to account for charged hadrons with transverse momenta down to 500 MeV in the reconstruction of jets. In comparison, calorimeter-based jet reconstruction has shortcomings in accounting for low-p_{T} charged hadrons due to charged hadrons being bent out of the jet clustering cone or their calorimeter signals being suppressed by the noise thresholds of the topo-cluster algorithm. Moreover, jets reconstructed using particle flow are less susceptible to pile-up since charged particle flow objects can be associated to the PV of the hard interaction.

b-tagging

The production of b quarks is an important signature of many physics processes at the LHC. A b quark produced in a hard scattering interaction leads to the formation of a jet, referred to as a b-jet. A distinct feature of b-jets is the presence of a b hadron with a mass exceeding 5 GeV [18]. These b hadrons are short-lived and decay after traversing a short but often measurable distance in the ATLAS detector. For example, the lightest B mesons, the B^0 and $B^{\pm}$, have proper lifetimes of about 1.5×10^{-12} s [18] and thus, assuming a B meson momentum of 10 GeV, a mean flight path of about 1 mm before decaying. The dominant decay modes of b hadrons produce final states containing c hadrons, which also decay after traversing a short distance in the detector.

The task of identifying b-jets is referred to as b-tagging. Features resulting from displaced decays of b and c hadrons can be exploited for this purpose. Two categories of features are used by the b-tagging algorithms employed in this thesis:

Track impact parameter based features Due to the large mass of b hadrons, the daughter particles of displaced b-hadron decays are produced at an angle with respect to the initial flight direction of the b hadron. Consequently, the tracks of the daughter particles tend to have larger longitudinal and transverse track impact parameters with respect to the PV of the hard interaction compared to promptly produced charged particles. The deviation of an impact parameter from zero is quantified in terms of the *impact parameter significance* by dividing the impact parameter by its uncertainty.

Secondary vertex based features Secondary vertices resulting from the daughter particles of b- and c-hadron decays can be reconstructed using charged-particle tracks measured in the ID. Among others, the displacement of the reconstructed secondary vertex from the PV of the hard interaction or the invariant mass of the charged particles associated to the vertex can be used as features in b-tagging.

The b-tagging algorithms used at the ATLAS experiment combine the discriminants of multiple low-level algorithms. In this thesis, the DL1r b-tagging algorithm [129] is used, which uses neural networks to combine the outputs of the following algorithms:

IP2D & IP3D The IP2D and IP3D algorithms [130] provide b-tagging discriminants based on likelihood-ratio classifiers constructed from the distribution of impact parameter significances in b-, c-, and *light*-quark jets. While IP2D only considers the distribution of the transverse impact parameter significance of tracks, IP3D considers the joint distribution of transverse and longitudinal impact parameter significances. In both cases, the likelihoods of all tracks in the jet are combined by neglecting any dependencies of the impact parameters between tracks.

RNNIP The RNNIP algorithm [131] extends the idea of IP2D and IP3D by also considering dependencies of impact parameter significances between tracks associated to a jet. This is accomplished with a recurrent neural network taking the tracks of a jet as inputs. Track properties, such as the impact parameter significances, are passed to the network with every track. Based on these inputs, RNNIP estimates the probability of a jet being a b-, c-, or *light*-quark jet.

SV1 & JetFitter The SV1 [132] and JetFitter [133] algorithms reconstruct the vertices resulting from the displaced decays of b and c hadrons. While SV1 reconstructs a single secondary vertex, JetFitter attempts to reconstruct the cascade of b- and c-hadron decays with a secondary and tertiary vertex. The vertices determined by both algorithms are used to define discriminating variables that are provided to the high-level b-tagging algorithms. Examples of these variables are the invariant mass of the tracks associated to the vertices or the significance of the distance between the secondary vertex and the PV of the hard interaction.

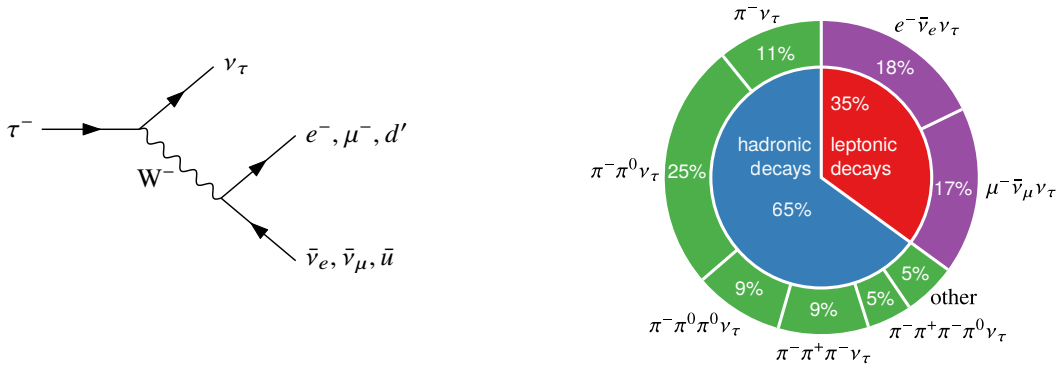
The DL1r b-tagging algorithm combines these discriminants with basic kinematic properties of the jet to yield a high-level b-tagging discriminant. Selections on this discriminant define several b-tagging working points with different probabilities of correctly identifying a b-jet, referred to as the b-tagging efficiency. In this thesis, a working point with 77 % b-tagging efficiency in simulated $t\bar{t}$ events is used. This working point has c- and *light*-quark jet rejection factors of approximately 6 and 200 in simulated $t\bar{t}$ events, respectively [129].

3.3.6 Tau Leptons

The τ lepton is the heaviest lepton in the SM with a mass of 1.777 GeV [18]. It is sufficiently massive to decay into final states with a charged lepton or hadrons. These decay modes are referred to as leptonic (τ_{lep}) and hadronic (τ_{had}) decay modes, respectively. The τ-lepton decay and its branching ratios are depicted in Figure 3.7. The majority of τ-lepton decays produce final states with one or three charged hadrons, predominately $\pi^{\pm}$, which are referred to as 1- and 3-prong τ_{had}, respectively.

At the ATLAS experiment, τ leptons are reconstructed from the signature of their visible decay products, neglecting any neutrinos produced in the decay. Leptonic τ-lepton decays are reconstructed as electrons or muons using the reconstruction techniques previously introduced in Sections 3.3.3 and 3.3.4. The remainder of this section focuses on the reconstruction of the visible decay products of hadronic τ-lepton decays, referred to as $\tau_{\text{had-vis}}$, for which dedicated reconstruction techniques are employed. The reconstruction of $\tau_{\text{had-vis}}$ at the beginning of Run 2 is described in Ref. [135]. Several improvements have been made since, for example the introduction of a multivariate track selection method [136], which are included in the following description.

The reconstruction of $\tau_{\text{had-vis}}$ is seeded by jets reconstructed using the anti-k_t algorithm with $R = 0.4$ applied to topo-clusters at LC scale. Only jets with $p_{\text{T}} > 10\,\text{GeV}$ and within $|\eta| < 2.5$ are considered



(a) **(b)**

Figure 3.7: Feynman diagram (a) and branching ratios (b) of the τ^- decay. The eigenstate coupling to the u quark via the W boson is denoted as d'. The branching ratios of the τ-lepton decay are taken from Ref. [18] rounded to the nearest integer percentage. The figures are adapted from Ref. [134].

as seeds for $\tau_{\text{had-vis}}$ reconstruction.

A PV is associated to the $\tau_{\text{had-vis}}$ candidate using tracking information. The PV association considers ID tracks that are matched to the seed jet via ghost-association [137], are within $\Delta R < 0.2$ of the jet axis, and fulfil $p_{\text{T}} > 1\,\text{GeV}$ and basic track quality criteria. For every PV, the scalar sum of p_{T} of tracks that are assigned to the vertex and pass the selection criteria is determined. The PV maximising this sum is associated to the $\tau_{\text{had-vis}}$ candidate and is referred to as the *tau vertex*.

An initial estimate of the $\tau_{\text{had-vis}}$ candidate four-momentum is derived by summing the four-momenta of topo-clusters (at LC scale) in a cone of $\Delta R < 0.2$ about the jet axis. Afterwards, the four-momentum is transformed into the coordinate system with the tau vertex at its origin. The axis defined by the tau vertex and the three-momentum of the $\tau_{\text{had-vis}}$ candidate is referred to as the $\tau_{\text{had-vis}}$ axis.

Tracks in the ID are associated to a $\tau_{\text{had-vis}}$ candidate if they are within $\Delta R < 0.4$ of the $\tau_{\text{had-vis}}$ axis. Tracks originating from charged hadrons produced in the τ-lepton decay are identified by Boosted Decision Trees [136]. These tracks are referred to as *core* tracks, since they are usually found in the *core region* of $\Delta R < 0.2$ about the $\tau_{\text{had-vis}}$ axis. The number of *core* tracks associated to a $\tau_{\text{had-vis}}$ candidate is referred to as N_{tracks}. Only candidates with $N_{\text{tracks}} = 1$ and $N_{\text{tracks}} = 3$ are considered and referred to as 1- and 3-prong $\tau_{\text{had-vis}}$ candidates, respectively. Tracks not classified as *core* are classified as *isolation* tracks if they fulfil

$$p_{\text{T}} > 1\,\text{GeV} \qquad |d_0| < 1.0\,\text{mm} \qquad |z_0 \sin\theta| < 1.5\,\text{mm}$$
$$N_{\text{pixel}} \geq 2 \qquad N_{\text{pixel}} + N_{\text{SCT}} \geq 7\,,$$

where the track parameters are given at the perigee with respect to the tau vertex. The number of hits on the reconstructed track in the pixel and SCT detector layers, counting defective sensors located on the trajectory as hits, is given by N_{pixel} and N_{SCT}, respectively. *Isolation* tracks are used to define variables sensitive to the charged-particle activity in the vicinity of the $\tau_{\text{had-vis}}$ candidate.

Calibrations of the $\tau_{\text{had-vis}}$ candidate four-momentum to the $\tau_{\text{had-vis}}$ energy scale (TES) are derived using simulated events containing hadronic τ-lepton decays. Two different calibration methods are used:

Calorimeter-based TES The calorimeter-based TES is derived from the initial estimate of the $\tau_{\text{had-vis}}$ four-momentum from topo-clusters in the core region of the jet seeding the $\tau_{\text{had-vis}}$. First, the expected energy contribution of pile-up to the $\tau_{\text{had-vis}}$ is subtracted. This subtraction is parameterised as a linear function of the number of reconstructed PVs in the event and in bins of $\tau_{\text{had-vis}}$ $|\eta|$ and N_{tracks}. Afterwards, the detector response calibration is performed as a function of the $\tau_{\text{had-vis}}$ energy after pile-up subtraction and in bins of $\tau_{\text{had-vis}}$ $|\eta|$ and N_{tracks}. The $\tau_{\text{had-vis}}$ transverse momentum after this calibration step is denoted as p_{T}^{LC}.

BRT-based TES The BRT-based TES is derived from $\tau_{\text{had-vis}}$ energy calibrations performed using Boosted Regression Trees (BRT). This calibration uses information from *Tau Particle Flow* reconstruction [138], which attempts to reconstruct individual charged and neutral hadrons produced in the τ-lepton decay, as inputs. Tau Particle Flow provides an alternative method to reconstruct the $\tau_{\text{had-vis}}$ four-momentum with improved angular resolution, and improved energy resolution for $\tau_{\text{had-vis}}$ with transverse energies below approximately 100 GeV [138]. The BRT also includes variables sensitive to the pile-up conditions, the shapes of showers in the calorimeter, and the reconstructed τ-lepton decay mode. In addition, p_{T}^{LC} is included as an input since the calorimeter-based estimate of $\tau_{\text{had-vis}}$ p_{T} is superior for high-p_{T} $\tau_{\text{had-vis}}$.

Unless otherwise noted, the BRT-based calibration is used for the remainder of this thesis. After calibration, the relative $\tau_{\text{had-vis}}$ p_{T} resolution ranges from 5 to 7 % [135].

The $\tau_{\text{had-vis}}$ reconstruction techniques presented thus far have limited ability to reject $\tau_{\text{had-vis}}$ candidates originating from sources other than hadronic τ-lepton decays. The primary source being quark- or gluon-initiated jets and electrons that are mis-reconstructed as $\tau_{\text{had-vis}}$ candidates. A separate step, referred to as tau identification, is performed to reject $\tau_{\text{had-vis}}$ candidates from these sources. A description of tau identification is given in Chapter 5.

3.3.7 Missing Transverse Momentum

The missing transverse momentum, $\boldsymbol{p}_{\text{T}}^{\text{miss}}$, is used to reconstruct the total transverse momentum of undetected particles produced in the hard interaction. In the laboratory frame, the colliding partons have negligible transverse momentum and thus it holds that

$$\boldsymbol{p}_{\text{T}}^{\text{detected}} + \boldsymbol{p}_{\text{T}}^{\text{undetected}} = \mathbf{0}\,,$$

where $\boldsymbol{p}_{\text{T}}^{\text{detected}}$ and $\boldsymbol{p}_{\text{T}}^{\text{undetected}}$ are the total transverse momenta of the detected and undetected particles produced in the hard interaction, respectively. Therefore, the missing transverse momentum is defined as

$$\boldsymbol{p}_{\text{T}}^{\text{miss}} \coloneqq \boldsymbol{p}_{\text{T}}^{\text{undetected}} = -\boldsymbol{p}_{\text{T}}^{\text{detected}}\,.$$

The purpose of $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ is to reconstruct the total transverse momentum of non- or only weakly interacting particles (*invisible particles*) such as neutrinos. An object-based definition of $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ from Ref. [139] is used that is defined as

$$\boldsymbol{p}_{\text{T}}^{\text{miss}} = -\sum \boldsymbol{p}_{\text{T}}^{e} - \sum \boldsymbol{p}_{\text{T}}^{\gamma} - \sum \boldsymbol{p}_{\text{T}}^{\mu} - \sum \boldsymbol{p}_{\text{T}}^{\tau_{\text{had-vis}}} - \sum \boldsymbol{p}_{\text{T}}^{\text{jet}} - \boldsymbol{p}_{\text{T}}^{\text{soft}}\,, \tag{3.1}$$

where the summations are over selected and calibrated electrons, photons, muons, $\tau_{\text{had-vis}}$, and jets, respectively. An additional term, $\boldsymbol{p}_{\text{T}}^{\text{soft}}$, is introduced to account for soft radiation that is not reconstructed as part of another object. This term is estimated by summing the transverse momenta of tracks in the ID that are associated with the PV of the hard interaction, pass track quality criteria, and cannot be associated with any selected electron, photon, muon, $\tau_{\text{had-vis}}$, or jet. The term $\boldsymbol{p}_{\text{T}}^{\text{soft}}$ is also referred to as the *track soft term*.

Equation (3.1) illustrates the difficulties in reconstructing the total transverse momentum of invisible particles using $\boldsymbol{p}_{\text{T}}^{\text{miss}}$. Any detectable object that fails to be detected, either due to detector acceptance effects or inefficiencies in the object reconstruction, introduces deviations from the total transverse momentum of invisible particles. Moreover, uncertainties on transverse momenta of all reconstructed objects accumulate in the calculation of $\boldsymbol{p}_{\text{T}}^{\text{miss}}$, leading to large uncertainties on $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ particularly in events with high activity.

CHAPTER 4

Statistical Methods

This chapter provides an overview of the statistical methods employed in this thesis. Section 4.1 introduces methods of statistical inference that are used for the interpretation of the results of Chapters 6 and 7. The machine learning algorithms employed in Chapters 5 to 7 are described in Section 4.2.

4.1 Statistical Inference

The following introduces the statistical inference techniques used for the interpretation of results of searches and measurements at the ATLAS experiment. Among these are methods for parameter estimation and hypothesis testing. They are used for fitting statistical models to observed data (or pseudodata) and to make frequentist statements about the presence or absence of a signal.

4.1.1 The HistFactory Model

The use of binned data is widespread for data visualisation and statistical modelling in HEP. Analyses of collider data typically consider multiple mutually disjoint regions, also referred to as *channels*, defined by event selections. Each region consists of a histogram of a discriminating variable with one or more bins. Events from different signal and background processes, hereafter referred to as *physics processes*, contribute to these regions. In this context, HistFactory [140] is the tool used in this thesis for constructing statistical models for likelihood-based inference. The HistFactory model is introduced in the following, adopting the notation of Ref. [140] with a few modifications.

Let C denote the set of channels and $\mathcal{B}_c$ the set of bins of the histogram in channel c. The probability of observing n_{cb} events in bin b of channel c is modelled by a Poisson distribution with probability mass function $\text{Pois}(n_{cb}; \nu_{cb})$, where ν_{cb} denotes the expected number of events in a given bin. The expectation ν_{cb}, which has to be inferred from the observed data, is parameterised as [140]

$$\nu_{cb}(\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\gamma}) = \sum_{s \in \mathcal{S}_c} \gamma_{csb}\, \Phi_{cs}(\boldsymbol{\phi})\, \eta_{cs}(\boldsymbol{\alpha})\, \sigma_{csb}(\boldsymbol{\alpha})\,,$$

where $\mathcal{S}_c$ is the set of physics processes contributing to channel c and the parameters of the model are denoted by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)$, and $\boldsymbol{\gamma}$.[1] The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are nuisance

[1] The vector $\boldsymbol{\gamma}$ groups the γ_{csb} parameters for all bins.

parameters (NPs) with external constraints, which will be introduced shortly. The relevance of the four factors γ_{csb}, Φ_{cs}, η_{cs}, and σ_{csb} is described in the following [140]:

- $\sigma_{csb}(\boldsymbol{\alpha})$ is referred to as the *parameterised histogram* of process s in channel c, the index b denoting the bins of the histogram. It estimates the expected number of events from process s in channel c and is usually derived from simulation or control region data. The parameterised histogram has additional degrees of freedom, parameterised by $\boldsymbol{\alpha}$, to account for uncertainties on the shape of the histogram. These degrees of freedom leave the overall normalisation $\sum_{b\in\mathcal{B}_c}\sigma_{csb}(\boldsymbol{\alpha})$ for a given channel c and process s unchanged.

- $\eta_{cs}(\boldsymbol{\alpha})$ represents a normalisation factor applied uniformly to all bins of the parameterised histogram for a given process s in channel c. The normalisation factor η_{cs} cannot vary freely, since it is a function of the constrained parameters $\boldsymbol{\alpha}$. This factor is included to account for uncertainties on the normalisation of process s in channel c.

- $\Phi_{cs}(\boldsymbol{\phi})$ also represents a normalisation factor applied uniformly to all bins of the parameterised histogram for a given process s in channel c. However, Φ_{cs} is the product of *free* (unconstrained) normalisation factors given by

$$\Phi_{cs}(\boldsymbol{\phi}) = \prod_{p\in\mathcal{N}_{cs}} \phi_p \,,$$

  where $\mathcal{N}_{cs}$ defines the normalisation factors that are to be applied to a given process s in channel c. In most cases, at least one normalisation factor is present that is applied to the physics process of interest, also referred to as the *signal*. This normalisation factor is called the *signal strength* μ and is usually considered to be the parameter of interest (POI). Normalisation factors are considered to be NPs if they are not POIs.

- γ_{csb} are parameters that introduce additional degrees of freedom for every bin. They are used for incorporating uncertainties from sources that are independent between bins. An example of such an uncertainty is the statistical uncertainty arising from the use of finite samples of events to estimate σ_{csb}.

  In this thesis, the method by Barlow and Beeston [141] is used to account for statistical uncertainties on the estimation of σ_{csb}. To reduce the number of parameters of the statistical model, the method is simplified as proposed in Ref. [142] by only considering the statistical uncertainty on $\sum_{s\in\mathcal{S}_c}\sigma_{csb}$ for a given bin and channel. Consequently, the parameters γ_{csb} can be replaced by γ_{cb}, omitting the dependence on the physics process.

A description of the exact functional form of $\sigma_{csb}(\boldsymbol{\alpha})$ and $\eta_{cs}(\boldsymbol{\alpha})$ is omitted here but can be found in Ref. [140]. The likelihood function of the statistical model is given by

$$L(\boldsymbol{\alpha},\boldsymbol{\phi},\boldsymbol{\gamma}) = \left[\prod_{c\in C}\prod_{b\in\mathcal{B}_c} \mathrm{Pois}\big(n_{cb};\nu_{cb}(\boldsymbol{\alpha},\boldsymbol{\phi},\boldsymbol{\gamma})\big)\right] \times L_{\mathrm{ext}}(\boldsymbol{\alpha},\boldsymbol{\gamma})\,, \tag{4.1}$$

where $L_{\mathrm{ext}}(\boldsymbol{\alpha},\boldsymbol{\gamma})$ is the likelihood function that defines the external constraints on the NPs $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

The likelihood function $L_{\text{ext}}(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ reads

$$L_{\text{ext}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \left[\prod_{p=1}^{n} f(a_p; \alpha_p) \right] \left[\prod_{c \in C} \prod_{b \in \mathcal{B}_c} \text{Pois}(m_{cb}; \gamma_{cb} \tau_{cb}) \right], \tag{4.2}$$

with the terms in brackets being described in the following:

- The term $f(a_p; \alpha_p)$ can be interpreted as the likelihood function of an auxiliary measurement of the parameter α_p given that a_p is observed. Conventionally, $f(a_p; \alpha_p)$ is taken to be the probability density of a Normal distribution with unit variance and mean α_p. Furthermore, it is assumed that the measurement observed $a_p = 0$ such that, considering only the auxiliary measurement, the maximum likelihood estimate (MLE) of α_p is 0 with an uncertainty of $\Delta\alpha_p = \pm 1$. This particular choice of constraint term is closely connected to the functional form of $\sigma_{csb}(\boldsymbol{\alpha})$ and $\eta_{cs}(\boldsymbol{\alpha})$. For example, in a model with one externally constrained NP the nominal histogram and normalisation is given by $\sigma_{csb}(0)$ and $\eta_{cs}(0)$, while $\sigma_{csb}(\pm 1)$ and $\eta_{cs}(\pm 1)$ correspond to $\pm 1\sigma$ variations of the histogram shape and normalisation, respectively.
- The $\text{Pois}(m_{cb}; \gamma_{cb}\tau_{cb})$ terms provide constraints for the γ_{cb} parameters introduced by the simplified Barlow–Beeston method. For every bin a constant $\tau_{cb} = (\sum_i w_i)^2 / \sum_i w_i^2$ is defined, where the sums go over events making up the background prediction in the given bin and w_i are the corresponding event weights. The constant τ_{cb} can be interpreted as an effective number of events, which measures the statistical precision of the background estimate in the bin.[2] The model incorporates the statistical uncertainty by setting up auxiliary measurements of the effective number of events, $\gamma_{cb}\tau_{cb}$, for every bin. These measurements observe $m_{cb} = \tau_{cb}$, and therefore the nominal value of the γ_{cb} parameters is 1. Supplementary information on the Barlow–Beeston method is given in Appendix A.4.1.

Given observed data, the parameters of the model can be estimated using maximum likelihood estimation. Hereafter, it is assumed that the model has one POI, namely, the signal strength μ. All other parameters of the model are collectively referred to as $\boldsymbol{\theta}$. Let Ω be the parameter space of the model with elements $(\mu, \boldsymbol{\theta})$. The MLE of the parameters is determined by the *unconditional fit*

$$(\hat{\mu}, \hat{\boldsymbol{\theta}}) = \underset{(\mu, \boldsymbol{\theta}) \in \Omega}{\arg\max}\, L(\mu, \boldsymbol{\theta})\,. \tag{4.3}$$

Often, a restricted model is constructed by fixing the POI to an arbitrary value μ^*. In this case, the MLE of the model parameters is given by the *conditional fit for* $\mu = \mu^*$

$$\hat{\hat{\boldsymbol{\theta}}}(\mu^*) = \underset{\boldsymbol{\theta} \in \{\boldsymbol{\theta}' \,|\, (\mu^*, \boldsymbol{\theta}') \in \Omega\}}{\arg\max} L(\mu^*, \boldsymbol{\theta})\,. \tag{4.4}$$

The model with the restriction $\mu = 0$ is called the *background-only model*, while the unrestricted model is called the *signal-plus-background model*.

[2] Let $Y = \sum_{i=1}^{N} W_i$ and $Y' = \sum_{i=1}^{N'} c$, where N and N' are independent Poisson random variables, W_i are i.i.d. weights that are independent of N and N', and c is a constant. The effective number of events refers to the value of $\mathrm{E}(N')$ such that $\mathrm{E}(Y) = \mathrm{E}(Y')$ and $\mathrm{Var}(Y) = \mathrm{Var}(Y')$. This condition yields $\mathrm{E}(N') = \mathrm{E}(N)\,\mathrm{E}(W)^2/\mathrm{E}(W^2)$, and after approximating expected values with sample averages $\mathrm{E}(N') \approx (\sum_i w_i)^2 / \sum_i w_i^2$.

Lastly, special datasets referred to as *Asimov datasets* [143] are introduced. For a given set of model parameters, the Asimov dataset is the dataset in which the observables n_{cb} and the global observables a_p and m_{cb} are equal to their expected values. Asimov datasets can be used, in place of the observed data, to evaluate the expected experimental sensitivity of searches and measurements in HEP. Some uses of Asimov datasets are highlighted in subsequent sections.

4.1.2 Hypothesis Testing in Particle Physics

In HEP, one is often concerned with comparing the goodness of fit of competing statistical models to observed data. These comparisons assist in making statements about values of the signal strength that are still in agreement with the observations. The framework of statistical hypothesis testing provides a principled approach to perform these comparisons.

With the statistical model introduced previously, a null hypothesis, H_0, is defined as

$$H_0 : (\mu, \boldsymbol{\theta}) \in \Omega_0 \quad \text{with} \quad \Omega_0 \subset \Omega \,,$$

where Ω_0 is the set of model parameters that specify the null hypothesis, and an alternative hypothesis, H_1, is defined as

$$H_1 : (\mu, \boldsymbol{\theta}) \in \Omega_1 \quad \text{with} \quad \Omega_1 = \Omega \setminus \Omega_0 \,.$$

A *likelihood ratio test* (LRT) can be used to compare these hypotheses. The test statistic of the LRT is given by [144]

$$\Lambda = -2 \ln \left[\frac{\sup_{(\mu,\boldsymbol{\theta}) \in \Omega_0} L(\mu, \boldsymbol{\theta})}{\sup_{(\mu,\boldsymbol{\theta}) \in \Omega} L(\mu, \boldsymbol{\theta})} \right] ,$$

where the numerator (denominator) of the term in brackets is the supremum of the likelihood for the restricted (unrestricted) model. A critical value of the test statistic, Λ_{crit}, is chosen and compared to the observed value of the test statistic. If $\Lambda > \Lambda_{\text{crit}}$ then H_0 is rejected in favour of H_1. Otherwise, H_0 cannot be rejected. The chosen value of Λ_{crit} defines the rejection region of the test and thus its significance level and power.

Discovery of a Signal

In the context of statistical hypothesis testing, the discovery of a signal implies rejecting the background-only hypothesis in favour of the signal-plus-background hypothesis. Generally, only signals with positive strength, that is $\mu > 0$, are considered as potential discoveries. The relevant hypotheses for testing the discovery of a signal are

$$H_0 : (\mu, \boldsymbol{\theta}) \in \{(\mu', \boldsymbol{\theta}') \in \Omega^+ \mid \mu' = 0\} \qquad H_1 : (\mu, \boldsymbol{\theta}) \in \{(\mu', \boldsymbol{\theta}') \in \Omega^+ \mid \mu' > 0\} \,,$$

where Ω^+ denotes the parameter space of the model with the restriction that $\mu \geq 0$. An empirical test statistic based on the LRT is defined as [143]

$$q_0 = \begin{cases} -2\ln\left[\frac{L\left(0,\hat{\hat{\boldsymbol{\theta}}}(0)\right)}{L\left(\hat{\mu},\hat{\boldsymbol{\theta}}\right)}\right], & \hat{\mu} > 0 \\ 0, & \hat{\mu} \leq 0 \end{cases} , \tag{4.5}$$

where $\hat{\mu}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\hat{\boldsymbol{\theta}}}$ are defined as in Equations (4.3) and (4.4).[3] This test statistic is referred to as the *discovery test statistic*. The asymptotic sampling distribution of q_0 under H_0 is given by the probability density function [143]

$$f(q_0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}f_{\chi^2}(q_0; 1)\,,$$

which is an equal mixture of a Dirac δ distribution and a χ^2 distribution with one degree of freedom. The discovery test statistic is often expressed in terms of an asymptotic p-value according to [143]

$$p_0 = \int_{q_0}^{\infty} \mathrm{d}q_0'\, f(q_0') = 1 - \Phi\left(\sqrt{q_0}\right)\,,$$

where Φ is the cumulative distribution function of the Standard Normal distribution. Another way of expressing the test statistic is in terms of the *discovery significance* Z_0, which is defined as [143]

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0}\,.$$

In particle physics, the conventional significance threshold that has to be exceeded to claim discovery of new physics is $Z_0 = 5$ ($p_0 = 2.87 \times 10^{-7}$), which is also called the "5σ threshold". Lastly, the *median discovery significance* of a signal with an assumed strength of μ can be estimated by determining Z_0 for an Asimov dataset constructed with the same signal strength [143]. The median discovery significance is also referred to as the *expected signal significance*.

Upper Limits on the Signal Strength

Often, one is interested in determining the largest signal strength that would still be compatible with the observed data. Formally, this constitutes estimating a one-sided confidence interval for μ that is bounded from above, hence referred to as an upper limit. The upper limit can be obtained by inverting a test of the hypotheses

$$H_0 : (\mu, \boldsymbol{\theta}) \in \{(\mu', \boldsymbol{\theta}') \in \Omega^+ \mid \mu' \geq \mu^*\} \qquad H_1 : (\mu, \boldsymbol{\theta}) \in \{(\mu', \boldsymbol{\theta}') \in \Omega^+ \mid \mu' < \mu^*\}\,,$$

where μ^* is used to parameterise the hypotheses. Let $1 - \alpha$ be the desired confidence level (CL) of the interval to be estimated. Given a test of H_0 and H_1 with significance level α, the confidence interval is the set of values of μ^* for which H_0 cannot be rejected by the test. The upper bound of this set is the upper limit on μ at $1 - \alpha$ CL.

[3] The unconditional fit is performed without the $\mu \geq 0$ constraint. Instead, the constraint is imposed in the definition of the test statistic by setting the maximum likelihood of the unrestricted model to $L\left(0, \hat{\hat{\boldsymbol{\theta}}}(0)\right)$ in the $\hat{\mu} < 0$ case.

The estimation of upper limits in HEP uses an empirical test statistic derived from the LRT that reads [143]

$$\tilde{q}_\mu = \begin{cases} -2\ln\left[\frac{L\left(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu)\right)}{L\left(0,\hat{\hat{\boldsymbol{\theta}}}(0)\right)}\right], & \hat{\mu} \in (-\infty, 0] \\ -2\ln\left[\frac{L\left(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu)\right)}{L\left(\hat{\mu},\hat{\boldsymbol{\theta}}\right)}\right], & \hat{\mu} \in (0, \mu] \\ 0, & \hat{\mu} \in (\mu, \infty) \end{cases}, \tag{4.6}$$

where for notational simplicity μ^* is replaced by μ. Moreover, the unconditional fit determining $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ in Equation (4.6) is performed without the positivity constraint on the signal strength. The asymptotic sampling distribution of $\tilde{q}_\mu$ was derived in Ref. [143] for any assumed value of the true signal strength. Let $f(\tilde{q}_\mu \mid \mu')$ be the asymptotic sampling distribution of $\tilde{q}_\mu$ under the assumption that the signal strength is μ'. Following the approach outlined before, an upper limit at $1-\alpha$ CL can be determined by finding the largest value of μ such that

$$p_\mu = \int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} \mathrm{d}\tilde{q}_\mu \, f(\tilde{q}_\mu \mid \mu) > \alpha \,,$$

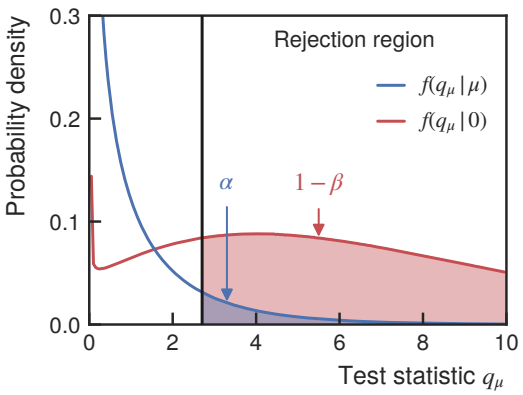
where $\tilde{q}_{\mu,\mathrm{obs}}$ denotes the observed value of the test statistic. Confidence intervals derived using this method have an asymptotic coverage probability of $1-\alpha$.

Searches for new physics use the $\mathrm{CL_s}$ technique [145–147] to set upper limits on μ instead. To motivate the use of this technique, a problem with tests based on p_μ is outlined hereafter. A test of $H_0 : \mu \geq \mu^*$ and $H_1 : \mu < \mu^*$ with rejection region $p_{\mu^*} \leq \alpha$ has an asymptotic significance level of α by construction. Since α is fixed a priori, the probability/power of correctly rejecting H_0 under the background-only hypothesis, denoted by $1-\beta$, is solely determined by the experimental sensitivity to a signal with strength μ^*. This is illustrated in Figures 4.1(a) and 4.1(b) for an experiment with high and low sensitivity, respectively. A problem arises when considering an experiment that has little sensitivity in distinguishing between the $\mu = \mu^*$ and $\mu = 0$ hypotheses. In this case, the power of the test $(1-\beta)$ approaches α, which is illustrated in Figure 4.1(b). This means that a signal to which the experiment has no sensitivity will still be excluded with a probability of at least α under the background-only hypothesis. In HEP, such *spurious exclusions* are not desired, motivating the use of an alternative method for setting upper limits that takes into account the signal sensitivity of the experiment.

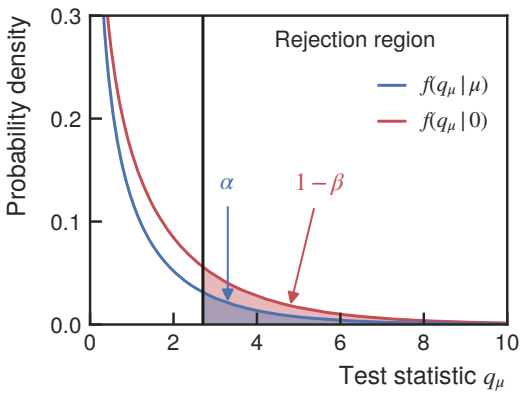
The $\mathrm{CL_s}$ technique addresses the problem of spurious exclusions by introducing the test statistic [143]

$$\mathrm{CL_s} = \frac{\int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} \mathrm{d}\tilde{q}_\mu \, f(\tilde{q}_\mu \mid \mu)}{\int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} \mathrm{d}\tilde{q}_\mu \, f(\tilde{q}_\mu \mid 0)} \,.$$

This statistic modifies p_μ (the numerator) by dividing it by the probability of $\tilde{q}_\mu$ being larger than the observed value under the background-only hypothesis. Upper limits on the signal strength are then estimated by finding the largest value of μ for which $\mathrm{CL_s} > \alpha$. In the limiting case of an experiment with no signal sensitivity the value of $\mathrm{CL_s}$ approaches unity, thus preventing spurious exclusions.



(a) Test with high statistical power ($1 - \beta = 0.8$).



(b) Test with low statistical power ($1 - \beta = 0.1$).

Figure 4.1: Sampling distribution of q_μ under the signal-plus-background (blue) and background-only (red) hypothesis. The area shaded in blue corresponds to the significance level α of the test, which is $\alpha = 0.05$ in both cases. The area shaded in red corresponds to the statistical power $1 - \beta$ of the test. The test statistic q_μ [143], which differs from $\tilde{q}_\mu$ by allowing negative values of the signal strength, is chosen for illustration purposes only.

Limits derived using this approach are also referred to as upper limits at $1 - \alpha$ CL; however, their coverage probability exceeds $1 - \alpha$ by construction due to $\mathrm{CL_s} > p_\mu$. Therefore, confidence intervals estimated using the $\mathrm{CL_s}$ technique are considered to be conservative. By convention, upper limits on μ are set at 95 % CL ($\alpha = 0.05$) in HEP. Lastly, the *median upper limit* on μ can be derived by performing the procedure using background-only Asimov data instead of the observed data [143]. The median upper limit is also referred to as the *expected upper limit*. An example of the limit setting procedure using the $\mathrm{CL_s}$ method is given in Figure 4.2. All upper limits in the remainder of this thesis are determined using the $\mathrm{CL_s}$ technique.

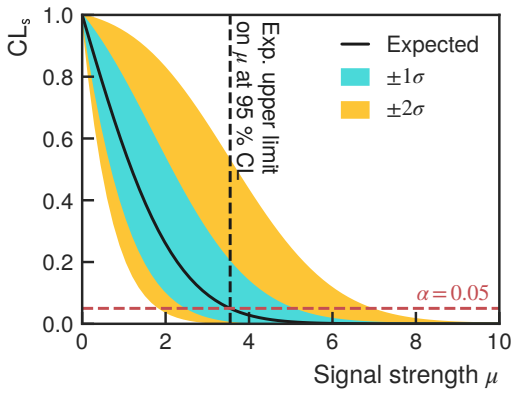


Figure 4.2: The $\mathrm{CL_s}$ test statistic as a function of μ for an event counting experiment with an expectation of 5 signal events, 50 background events, and a background systematic uncertainty of 10 %. The expected/median $\mathrm{CL_s}$ is obtained from the background-only Asimov dataset. The $\mathrm{CL_s}$ intervals containing approximately 68 % ($\pm1\sigma$) and 95 % ($\pm2\sigma$) of probability mass are indicated as coloured bands.

4.2 Machine Learning

Methods of multivariate analysis (MVA) have been used for decades in HEP to maximise the insights gained from experimental data. Previously, these analyses were often performed by hand, a difficult and time-consuming process, particularly when working with high-dimensional data. Over the past decade, automated approaches of MVA have become increasingly popular, freeing up researchers and frequently outperforming handcrafted solutions. This automation is achieved by algorithms that construct predictive models using samples of training data, a process referred to as *machine learning*.

Arguably the most common type of machine learning used in HEP falls under the category of *supervised learning*. The goal of supervised learning is to build models that predict an outcome $\boldsymbol{Y}$ using a set of predictors $\boldsymbol{X}$ by learning from a sample of labelled training data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$. Depending on the type of outcome a distinction between classification (categorical outcomes) and regression (continuous outcomes, count outcomes, etc.) is made. Supervised learning is particularly applicable in HEP since large amounts of labelled training data can be generated using Monte Carlo (MC) simulation. Nowadays, the state-of-the-art solutions to many tasks in HEP, for example particle identification, are based on machine learning.

This section gives an overview of the machine learning algorithms used in this thesis, specifically focusing on algorithms for solving binary classification tasks. Section 4.2.1 introduces *boosted decision trees*, which are used in Chapters 6 and 7 for signal-background discrimination in the search for non-resonant HH production. The basic principles of *neural networks* are introduced in Section 4.2.2, including a discussion of an architecture referred to as *recurrent neural networks*. Neural networks are used for the identification of hadronic τ-lepton decays in Chapter 5 as well as for the search for resonant HH production in Chapter 6.

4.2.1 Boosted Decision Trees

Boosted decision trees (BDT) are models used for classification and regression consisting of ensembles of *decision trees*. These ensembles are created using an algorithm called *boosting*, which iteratively fits shallow decision trees to altered versions of the training data. The training data is modified at every iteration to emphasise prediction errors made by previous boosting iterations. Finally, the predictions of the ensemble of trees are combined with the goal of providing superior classification/regression performance compared to a single decision tree. The following sections discuss the application of decision trees to regression tasks (regression trees) as well as a *gradient boosting* algorithm that builds ensembles of regression trees to perform binary classification.

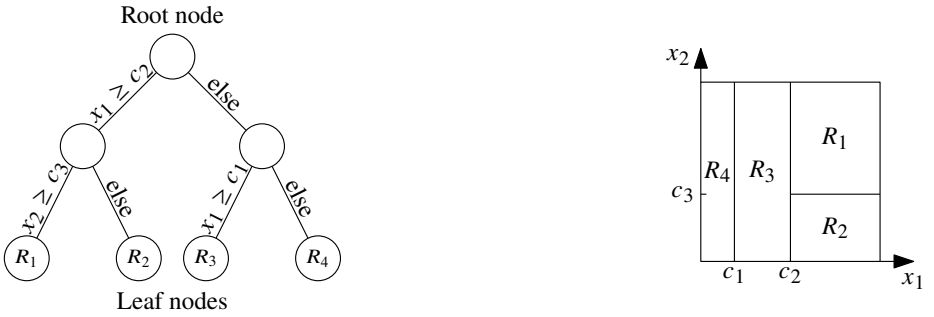
Decision Trees for Regression

A decision tree partitions an n-dimensional space with coordinates $\boldsymbol{x} = (x_1, \ldots, x_n)$ by recursively performing binary splits along the coordinate axes until a stopping criterion is met [148, 149]. The resulting binary tree structure and partitioning is illustrated in Figure 4.3 for a two-dimensional example. A decision tree with J leaf nodes splits the input space into J mutually disjoint subregions denoted by R_j for $j = 1, \ldots, J$. A constant value c_j is assigned to every region R_j such that the

prediction of a decision tree for a point $\boldsymbol{x}$ can be expressed as

$$h\big(\boldsymbol{x}; \{c_j, R_j\}_{j=1}^{J}\big) = \sum_{j=1}^{J} c_j \, \mathbf{1}(\boldsymbol{x} \in R_j) \qquad \text{with} \qquad \mathbf{1}(\boldsymbol{x} \in R_j) = \begin{cases} 1, & \boldsymbol{x} \in R_j \\ 0, & \text{else} \end{cases},$$

where $\{c_j, R_j\}_{j=1}^{J}$ fully characterises the decision tree [149].



(a) Binary tree structure of a decision tree.

(b) The partitioning resulting from the binary tree in (a).

Figure 4.3: A decision tree partitioning a two-dimensional space with coordinates (x_1, x_2). The tree has a depth of two and four leaf nodes that define the regions $R_1, \ldots, R_4$. The figures are adapted from Refs. [134, 149].

The construction of regression trees is summarised hereafter. First, constraints are set on the tree structure as a means to limit tree complexity, for example by setting a maximum tree depth. In the ideal case, regression trees are constructed by finding a partitioning $\{R_j\}_{j=1}^{J}$ and leaf node constants $\{c_j\}_{j=1}^{J}$ that minimise the mean squared error (MSE) given by

$$\text{MSE} = \frac{\sum_{i=1}^{N} w_i \left(y_i - h\big(\boldsymbol{x}_i; \{c_j, R_j\}_{j=1}^{J}\big)\right)^2}{\sum_{i=1}^{N} w_i},$$

where $\{(\boldsymbol{x}_i, y_i, w_i)\}_{i=1}^{N}$ is a sample of training data with scalar targets y_i and sample weights w_i. Generally, this optimisation problem is not easily solved; therefore, a greedy tree growth strategy is used instead [149, 150]. At any stage of the tree growing algorithm, the split (and the corresponding leaf node constants) leading to the largest reduction in MSE is chosen. The tree growing algorithm terminates once no further splits can be performed.

Boosting of Decision Trees

Boosting is an approach of solving prediction tasks by constructing models of the form

$$F_M\left(\boldsymbol{x}; \{\beta_m, \gamma_m\}_{m=1}^{M}\right) = \sum_{m=1}^{M} \beta_m b(\boldsymbol{x}; \gamma_m),$$

where β_m are coefficients and $b(\boldsymbol{x}; \gamma_m)$ are basis functions parameterised by γ_m [151, 152]. While there is some flexibility in choosing the family of basis functions, boosting is typically applied to decision trees. The following is concerned with a boosting algorithm called gradient boosting [152]. Gradient

boosting is particularly versatile, as it constructs models by minimising an arbitrary differentiable loss function. This allows gradient boosting to be applied to various kinds of prediction tasks. An example of binary classification is given hereafter.

Consider a binary classification problem with predictors $\boldsymbol{X}$ and class labels Y, which are $Y = +1$ for the positive class and $Y = -1$ for the negative class. Further, let

$$L(Y, F(\boldsymbol{X})) = \ln\left(1 + e^{-YF(\boldsymbol{X})}\right) \tag{4.7}$$

be a loss function and F be a scalar function of the predictors. The function that minimises the expected loss is given by

$$F^*(\boldsymbol{x}) = \arg\min_F \mathrm{E}[L(Y, F) \mid \boldsymbol{X} = \boldsymbol{x}] = \ln\left(\frac{\mathbb{P}(Y = +1 \mid \boldsymbol{X} = \boldsymbol{x})}{\mathbb{P}(Y = -1 \mid \boldsymbol{X} = \boldsymbol{x})}\right),$$

which are the conditional log-odds of an observation with predictors $\boldsymbol{x}$ belonging to the positive class [151]. Knowledge of F^* would solve the classification task since

$$\mathbb{P}(Y = +1 \mid \boldsymbol{X} = \boldsymbol{x}) = \frac{1}{1 + e^{-F^*(\boldsymbol{x})}} \qquad \text{and} \qquad \mathbb{P}(Y = -1 \mid \boldsymbol{X} = \boldsymbol{x}) = \frac{1}{1 + e^{+F^*(\boldsymbol{x})}}, \tag{4.8}$$

thus motivating the choice of loss function in Equation (4.7) for binary classification.

An algorithm referred to as TreeBoost proposed by Friedman [152] is outlined. TreeBoost uses gradient boosted decision trees and the loss function in Equation (4.7) to fit a binary classifier to a sample of training data. A version of this algorithm is employed in TMVA [153], which provides the BDT implementation used in this thesis. The description is adapted from Ref. [152] with some modifications.

TreeBoost algorithm for binary classification

Inputs:

- Training data $\{(\boldsymbol{x}_i, y_i, w_i)\}_{i=1}^N$ with predictors $\boldsymbol{x}_i$, class labels y_i, and weights w_i.
- Number of boosting iterations M.
- Shrinkage parameter η ($0 < \eta \leq 1$).
- Hyperparameters of the regression tree algorithm.

Algorithm:

1. The model is initialised to

$$F_0(\boldsymbol{x}) = \ln\left(\frac{1 + \bar{y}}{1 - \bar{y}}\right) \qquad \text{with} \qquad \bar{y} = \frac{\sum_{i=1}^N w_i y_i}{\sum_{i=1}^N w_i},$$

   which is the training sample estimate of the (unconditional) log-odds of $Y = +1$.

2. For $m = 1$ to M:

a) Calculate the pseudo-residual

$$r_i = -\left.\frac{\partial L(y_i, F(\boldsymbol{x}_i))}{\partial F(\boldsymbol{x}_i)}\right|_{F(\boldsymbol{x}_i)=F_{m-1}(\boldsymbol{x}_i)} \overset{\text{Eq. (4.7)}}{=} \frac{y_i}{1 + e^{y_i F_{m-1}(\boldsymbol{x}_i)}}$$

for all training examples.

b) Fit a regression tree to estimate the relationship between pseudo-residuals and predictors using $\{(\boldsymbol{x}_i, r_i, w_i)\}_{i=1}^{N}$ as the training dataset. The prediction of the fitted regression tree is denoted by $h\big(\boldsymbol{x}; \{c_{jm}, R_{jm}\}_{j=1}^{J_m}\big)$, where J_m is the number of leaf nodes, c_{jm} is the constant predicted by the j-th leaf node, and R_{jm} is the region defined by the j-th leaf node.

c) The leaf node constants of the regression tree, $\{c_{jm}\}_{j=1}^{J_m}$, are updated to minimise

$$\sum_{i=1}^{N} w_i \cdot L\Big(y_i, F_{m-1}(\boldsymbol{x}_i) + h\big(\boldsymbol{x}_i; \{c_{jm}, R_{jm}\}_{j=1}^{J_m}\big)\Big) .$$

This optimisation problem has no analytical minimiser for the loss function in Equation (4.7). Instead, the $\{c_{jm}\}_{j=1}^{J_m}$ are chosen to approximately minimise the above criterion by performing a single step of Newton's method. This yields the updated leaf node constants

$$c'_{jm} = \frac{\sum_i w_i r_i}{\sum_i w_i |r_i| (1 - |r_i|)} ,$$

where the sums go over the training data populating the j-th leaf node of the tree. This step is specific to the TreeBoost algorithm by Friedman.

d) Determine the m-th stage of the model by setting

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \eta \cdot h\big(\boldsymbol{x}; \{c'_{jm}, R_{jm}\}_{j=1}^{J_m}\big) ,$$

where η is a parameter of the boosting algorithm referred to as the *shrinkage* or *learning rate*. Generally, the shrinkage is set to values below unity such that every stage of boosting performs a suboptimal update. This serves as a form of regularisation to prevent overfitting.

3. The final prediction of the boosting procedure, $F_M(\boldsymbol{x})$, can be used to estimate the probability $\mathbb{P}(Y = +1 \mid \boldsymbol{X} = \boldsymbol{x})$ according to Equation (4.8) as

$$p(\boldsymbol{x}) = \frac{1}{1 + e^{-F_M(\boldsymbol{x})}} .$$

In HEP, this quantity is often referred to as the *BDT score*.

This algorithm is reminiscent of the classical gradient descent algorithm for minimisation. However, in the case of gradient boosting the optimisation is performed in the space of functions and not in parameter space [152]. Steps 2a) and 2b) of the algorithm estimate the negative gradient of the loss function with respect to $F(\boldsymbol{x})$, which is evaluated at the function estimate of the previous boosting iteration. Classical gradient descent algorithms often determine the step size of the parameter update by performing a line search along the direction of the steepest descent. In the TreeBoost algorithm,

this step size is determined by step 2c) on a per-leaf basis. Finally, step 2d) performs the (regularised) gradient descent update.

4.2.2 Neural Networks

Neural networks (NNs) are parametric functions defined by the composition of multiple, generally non-linear, functions. The functions composing NNs are also referred to as *layers*. An illustrative example of an NN is the *multilayer perceptron* (MLP). The basic element of MLPs, and many other types of NNs, are *densely connected layers* that apply transformations of the form

$$\boldsymbol{f}_i(\boldsymbol{x}; \boldsymbol{W}_i, \boldsymbol{b}_i) = \boldsymbol{\phi}_i(\boldsymbol{W}_i \boldsymbol{x} + \boldsymbol{b}_i)\,,$$

where $\boldsymbol{x}$ is the layer input, $\boldsymbol{W}_i$ and $\boldsymbol{b}_i$ is a weight matrix and bias vector, respectively, and $\boldsymbol{\phi}_i$ is an activation function that is applied element-wise.[4] An MLP with N hidden layers is defined by the composition of $N + 1$ densely connected layers, yielding the expression

$$\boldsymbol{f}\big(\boldsymbol{x}; \{\boldsymbol{W}_i, \boldsymbol{b}_i\}_{i=1}^{N+1}\big) = \big(\boldsymbol{f}_{N+1} \circ \cdots \circ \boldsymbol{f}_1\big)\big(\boldsymbol{x}; \{\boldsymbol{W}_i, \boldsymbol{b}_i\}_{i=1}^{N+1}\big)$$

for the MLP output. In general, the structure of NNs can be more complex than illustrated in this example. Therefore, NN architectures are frequently expressed as directed graphs with nodes representing layers and edges defining how layers are composed.

NNs are able to approximate large classes of functions [154, 155], making them attractive candidates for machine learning applications. Consider a binary classification problem with predictors $\boldsymbol{X}$ and class labels Y, which take values of $Y = 1$ and $Y = 0$ for the positive and negative class, respectively. Moreover, let $f(\boldsymbol{x}; \boldsymbol{\theta}) \in (0, 1)$ denote the prediction of an NN with inputs $\boldsymbol{x}$ and free parameters $\boldsymbol{\theta}$.[5] In this case, the canonical loss function for binary classification is

$$L(Y, f(\boldsymbol{X}; \boldsymbol{\theta})) = \begin{cases} -\ln(f(\boldsymbol{X}; \boldsymbol{\theta})), & Y = 1 \\ -\ln(1 - f(\boldsymbol{X}; \boldsymbol{\theta})), & Y = 0 \end{cases}, \tag{4.9}$$

which is called the *binary cross-entropy loss*. The function that minimises the expected binary cross-entropy loss is

$$f^*(\boldsymbol{x}) = \arg\min_f \mathrm{E}[L(Y, f) \mid \boldsymbol{X} = \boldsymbol{x}] = \mathbb{P}(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x})\,,$$

motivating the use of Equation (4.9) as a loss function for binary classification.

The binary classification task can be solved by approximating $f^*(\boldsymbol{x})$ using an NN with parameters set such that the mean loss over a sample of training data $\{(\boldsymbol{x}_i, y_i, w_i)\}_{i=1}^N$ is minimised. Formally, the parameters are given by the optimisation problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\frac{\sum_{i=1}^N w_i L(y_i, f(\boldsymbol{x}_i; \boldsymbol{\theta}))}{\sum_{i=1}^N w_i} \right).$$

[4] The element-wise application of a function $f(x)$ on a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ is defined as $\boldsymbol{f}(\boldsymbol{x}) = \big(f(x_1), \ldots, f(x_n)\big)$.

[5] Using the logistic (sigmoid) function as the activation function of the final layer constrains NN outputs to be within $(0, 1)$.

This optimisation is non-convex and analytically intractable (for non-trivial NNs). In practice, the minimisation is often performed using *stochastic gradient descent* (SGD) or derivatives thereof. SGD is a gradient-based optimisation method in which the gradient of the loss function is estimated using random subsamples of training data. These subsamples are referred to as *mini-batches* and are denoted by $\{\boldsymbol{x}_{(i)}, y_{(i)}, w_{(i)}\}_{i=1}^{M}$ for a batch of size M [156]. The mini-batch estimate of the gradient is given by

$$\boldsymbol{g}(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{M} w_{(i)} \nabla_{\boldsymbol{\theta}} L(y_{(i)}, f(\boldsymbol{x}_{(i)}; \boldsymbol{\theta}))}{\sum_{i=1}^{M} w_{(i)}} ,$$

where $\nabla_{\boldsymbol{\theta}}$ refers to the vector of partial derivatives in parameter space [156]. The training of NNs using SGD proceeds by first randomly initialising the parameters to $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Afterwards, a number of SGD iterations are performed until convergence or another stopping criterion is met. The iterations consist of drawing a mini-batch from the training data, computing the mini-batch estimate of the gradient, and then performing the parameter update

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{g}(\boldsymbol{\theta}_t) ,$$

where η_t is referred to as the learning rate [156]. The learning rate can follow a predefined schedule with respect to the iteration counter t, common ones being constant or exponentially decaying learning rates. A modification of this algorithm is *SGD with momentum*, which aims to improve the convergence properties of SGD [157–159]. The modified algorithm alters the parameter update according to

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \Delta\boldsymbol{\theta}_t - \eta_t \boldsymbol{g}(\boldsymbol{\theta}_t) \qquad \text{with} \qquad \Delta\boldsymbol{\theta}_t = \begin{cases} \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}, & t > 0 \\ 0, & t = 0 \end{cases}$$

which introduces the momentum parameter $0 \leq \alpha < 1$ [156, 158]. The update of SGD with momentum can be loosely interpreted as the movement of a massive object through parameter space with two forces acting on the object: a force proportional to $-\boldsymbol{g}(\boldsymbol{\theta})$ and a friction-like force proportional and opposed to the velocity $\Delta\boldsymbol{\theta}_t$ [156].

Recurrent Neural Networks

An appealing feature of NNs is their ability to process various forms of unstructured data (e.g. graphical images, natural language, etc.) using specialised network architectures. Among these architectures are *recurrent neural networks* (RNNs), which operate on ordered, variable-length sequences. Such sequences occur in many scenarios, an example being the representation of sentences in natural language processing. In the context of HEP, one might encode the properties of jets reconstructed in a particle collision event as a variable-length sequence. Hereafter, sequences are denoted by $(\boldsymbol{x}_t)_{t=1}^{N}$, where t is conventionally referred to as *time* or a *time step*, N is the length of the sequence, and $\boldsymbol{x}_t$ represents the feature vector associated with the t-th element of the sequence. When including

sequence data as inputs or outputs in NNs, layers that describe mappings of the form

$$\begin{aligned}
&\text{Many-to-many: } (\boldsymbol{x}_t)_{t=1}^{N} \rightarrow (\boldsymbol{y}_t)_{t=1}^{M} \\
&\text{Many-to-one: } \quad (\boldsymbol{x}_t)_{t=1}^{N} \rightarrow (\boldsymbol{y}_1) \\
&\text{One-to-many: } \quad (\boldsymbol{x}_1) \quad \rightarrow (\boldsymbol{y}_t)_{t=1}^{M} ,
\end{aligned}$$

are of particular interest. Additionally, for certain machine learning tasks it can be beneficial to exploit the context in which an element of the input/output sequence occurs. These capabilities are provided by recurrent NN layers, which are described in the remainder of this section. The focus of the following description lies on the *many-to-one* and *many-to-many* ($N = M$) cases.

The ability of RNNs to perform computations involving variable-length sequences while exploiting contextual information is enabled by two concepts: *parameter sharing* and the inclusion of an *internal state* in the network. These concepts are illustrated using the example of the *long short-term memory* (LSTM) layer [160, 161]. Consider an input and output sequence denoted by $(\boldsymbol{x}_t)_{t=1}^{N}$ and $(\boldsymbol{y}_t)_{t=1}^{N}$, respectively, both sequences having the same length. In addition, let $\boldsymbol{y}_0 = \boldsymbol{0}$. An LSTM layer computes the output sequence by iterating over time steps starting at $t = 1$. Three quantities referred to as *gate activations* are defined as [156]

$$\begin{aligned}
\boldsymbol{f}_t &= \boldsymbol{\sigma}\big(\boldsymbol{U}^{\mathrm{f}}\boldsymbol{x}_t + \boldsymbol{W}^{\mathrm{f}}\boldsymbol{y}_{t-1} + \boldsymbol{b}^{\mathrm{f}}\big) \\
\boldsymbol{i}_t &= \boldsymbol{\sigma}\big(\boldsymbol{U}^{\mathrm{i}}\boldsymbol{x}_t + \boldsymbol{W}^{\mathrm{i}}\boldsymbol{y}_{t-1} + \boldsymbol{b}^{\mathrm{i}}\big) \\
\boldsymbol{o}_t &= \boldsymbol{\sigma}\big(\boldsymbol{U}^{\mathrm{o}}\boldsymbol{x}_t + \boldsymbol{W}^{\mathrm{o}}\boldsymbol{y}_{t-1} + \boldsymbol{b}^{\mathrm{o}}\big) ,
\end{aligned}$$

which are the gate activations of the *forget gate*, *input gate*, and the *output gate* at time t, respectively. The gate activations are parameterised by weight matrices $\boldsymbol{U}^{\mathrm{f/i/o}}$ and $\boldsymbol{W}^{\mathrm{f/i/o}}$, as well as bias vectors $\boldsymbol{b}^{\mathrm{f/i/o}}$. For a given gate, the weights and biases are the same for all time steps, which is an implementation of parameter sharing. The function $\boldsymbol{\sigma}$ refers to the element-wise application of the logistic (sigmoid) function; thus, the components of the gate activation vectors take values in $(0, 1)$. Moreover, an LSTM layer includes an internal state that propagates and evolves forward through time. At time t, the state of an LSTM is denoted by $\boldsymbol{s}_t$ and the initial state is set to $\boldsymbol{s}_0 = \boldsymbol{0}$. The internal state of an LSTM at time t is given by the following recurrence relation [156, 162]

$$\boldsymbol{s}_t = \boldsymbol{f}_t \circ \boldsymbol{s}_{t-1} + \boldsymbol{i}_t \circ \boldsymbol{\phi}(\boldsymbol{U}\boldsymbol{x}_t + \boldsymbol{W}\boldsymbol{y}_{t-1} + \boldsymbol{b}) , \tag{4.10}$$

where $\circ$ is the element-wise multiplication of two vectors, $\boldsymbol{\phi}$ is an activation function defined by the element-wise application of the hyperbolic tangent,[6] and $\boldsymbol{U}$, $\boldsymbol{W}$, and $\boldsymbol{b}$ are weights and biases shared across time steps. Finally, the output of the LSTM layer at time t is given by [156]

$$\boldsymbol{y}_t = \boldsymbol{o}_t \circ \boldsymbol{\phi}(\boldsymbol{s}_t) , \tag{4.11}$$

using the definitions introduced before.

The computational graph of an LSTM layer for a single time step is depicted in Figure 4.4, illustrating the role of the gates in controlling the flow of information. The forget gate is used to selectively forget/remember aspects about the internal state at time $t - 1$ by multiplying the state vector $\boldsymbol{s}_{t-1}$ by

[6] The choice of activation function in Equation (4.10) is specific to the implementation of LSTM layers in Keras [162].

the activation of the forget gate. Similarly, the input gate controls the inclusion of information derived from the current element of the input sequence, $\boldsymbol{x}_t$, and the output of the preceding time step, $\boldsymbol{y}_{t-1}$, into the internal state at time t. Lastly, the output gate controls which parts of the internal state (after application of $\boldsymbol{\phi}$) are presented as the output $\boldsymbol{y}_t$. An input sequence of length N can be processed by repeating the computations depicted in Figure 4.4 until the input sequence is exhausted, thus yielding an output sequence of the same length. Finally, a many-to-one mapping can be achieved by discarding the first $N-1$ outputs of an LSTM and using only $\boldsymbol{y}_N$ for further computation. A concrete example of LSTM layers being used in HEP is given in Chapter 5.

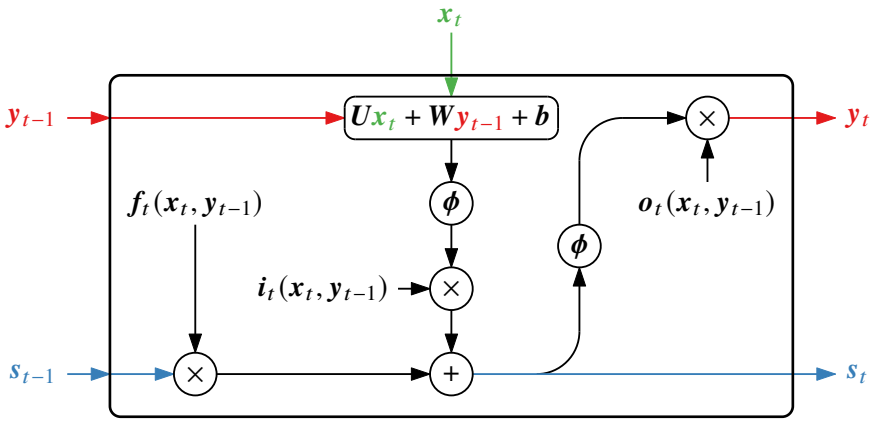


Figure 4.4: Computational graph of an LSTM layer for time step t. Operations in circles represent element-wise multiplication ($\times$), element-wise addition (+), and element-wise application of the tanh activation function ($\boldsymbol{\phi}$). The graph illustrates the computations performed as part of Equations (4.10) and (4.11). The computations involved in the calculation of the gate activation vectors are omitted; however, the dependency of the gate activation vectors on $\boldsymbol{x}_t$ and $\boldsymbol{y}_{t-1}$ is indicated in parenthesis.

CHAPTER 5

Tau Identification with Neural Networks

This chapter describes a novel algorithm used at the ATLAS experiment to identify the visible decay products of hadronic τ-lepton decays. The algorithm is applied to $\tau_{\text{had-vis}}$ candidates passing $\tau_{\text{had-vis}}$ reconstruction and aims to differentiate between candidates originating from τ_{had} (true $\tau_{\text{had-vis}}$) and those originating from non-τ_{had} sources (fake $\tau_{\text{had-vis}}$). This step is necessary since $\tau_{\text{had-vis}}$ reconstruction is not optimised to reject fake $\tau_{\text{had-vis}}$ candidates but rather to correctly reconstruct true $\tau_{\text{had-vis}}$ candidates and maintain high $\tau_{\text{had-vis}}$ reconstruction efficiencies.

The dominant source of fake $\tau_{\text{had-vis}}$ candidates at the ATLAS experiment are quark- or gluon-initiated jets due to their similarity to the hadronic and jet-like signature of $\tau_{\text{had-vis}}$. Electrons can be another, less abundant source of fake $\tau_{\text{had-vis}}$ candidates that have to be distinguished from true $\tau_{\text{had-vis}}$ candidates using a dedicated algorithm (*electron veto*). This chapter is concerned with the former source of fake $\tau_{\text{had-vis}}$ candidates and, in particular, with classifying the source of $\tau_{\text{had-vis}}$ candidates as either originating from τ_{had} or from quark- or gluon-initiated jets. This process is referred to as tau identification hereafter.

A number of features can be exploited to differentiate between $\tau_{\text{had-vis}}$ candidates originating from τ_{had} and quark- or gluon-initiated jets:

τ-lepton mass The τ lepton has a mass of 1.777 GeV [18] and is therefore sufficiently massive to decay hadronically while still having a small mass compared to the energy scales typically studied at the ATLAS experiment.

The τ-lepton mass can be used as a feature directly by considering the invariant mass of the visible daughter particles of hadronic τ-lepton decays. Ignoring reconstruction effects, this invariant mass is bounded by the mass of the τ lepton. This is not the case for $\tau_{\text{had-vis}}$ candidates originating from quark- or gluon-initiated jets, which do not have a strict upper bound.

The features described hereafter are consequences of, or closely related to the mass of the τ lepton.

Particle multiplicity Hadronic decays of τ leptons produce few (visible) daughter particles. Most decays produce one or three charged hadrons (most frequently $\pi^{\pm}$) and zero to two neutral pions. In contrast, the average multiplicity of charged and neutral particles in jets originating from the fragmentation of partons produced in hard scattering interactions is large and increases with the momentum of the jet [163, 164]. Therefore, particle multiplicity requirements are effective at

rejecting $\tau_{\text{had-vis}}$ candidates originating from quark- or gluon-initiated jets.[1]

Collimated daughter particles The $\tau_{\text{had-vis}}$ candidates typically considered by analyses at the ATLAS experiment have transverse momenta exceeding 20 GeV. At these momentum scales, the decay products of τ leptons are collimated due to the Lorentz boost of the τ lepton. This leads to the characteristic detector signature of a narrow jet with few visible particles. Requirements on the isolation[2] of $\tau_{\text{had-vis}}$ candidates can be used to reject candidates originating from quark- or gluon-initiated jets, which have a wider angular distribution of hadrons.

τ-lepton lifetime The τ lepton has a proper lifetime of 2.9×10^{-13} s ($c\tau = 87\,\mu\text{m}$) [18]; therefore, τ leptons with momenta in the range of 20 to 100 GeV typically travel for a few millimetres inside the beampipe before decaying. The distance traversed by the τ lepton before its decay results in a decay vertex that is displaced from the PV. For τ-lepton decay modes with three charged hadrons, this secondary vertex can be reconstructed and its displacement from the PV determined. For decay modes with only one charged hadron, the secondary vertex cannot be reconstructed directly. However, the longitudinal and transverse impact parameters of the reconstructed charged-hadron track can be used to gauge the incompatibility of the track with the PV.

Features sensitive to the τ-lepton lifetime can be used to distinguish τ_{had} from light-quark- or gluon-initiated jets in which hadrons are produced promptly at the PV. An exception are jets originating from b and c quarks. These also contain displaced decays of b- or c-flavoured hadrons. Nevertheless, the other features remain effective in discerning τ_{had} from b- and c-jets.

Prior to the introduction of the method described in this chapter, the ATLAS collaboration used BDTs as binary classifiers using high-level discriminating variables, i.e. variables purposefully constructed for the classification task, as inputs.

A method of performing tau identification using NNs that combines the information of high-level discriminating variables with information from reconstructed charged-particle tracks and topo-clusters in the calorimeters is presented. Tracks and topo-clusters in the vicinity of $\tau_{\text{had-vis}}$ candidates and their associated features are included as inputs. Since the number of tracks and topo-clusters associated to $\tau_{\text{had-vis}}$ candidates varies, an RNN architecture is used that can operate on sequences of varying length. The method is referred to as the RNN tau identification hereafter.

The RNN tau identification algorithm was initially proposed in Ref. [134] motivated by a similar approach developed for track impact parameter based b-tagging [131]. The algorithm was implemented in the reconstruction software of the ATLAS collaboration [165] and some of the results presented in this chapter were published in Ref. [11].

This chapter is structured as follows: The simulated events used for the development and performance evaluation of tau identification are introduced in Section 5.1. The identification method based on RNN is described in Section 5.2. Its performance is estimated based on simulation and compared to the BDT-based approach in Section 5.3. Section 5.4 concludes and gives an outlook on possible future developments.

[1] Gluon-initiated jets have, on average, a larger particle multiplicity and a broader angular distribution of particles compared to quark-initiated jets due to the larger effective colour charge of gluons [163]. Consequently, quark-initiated jets are more likely to be reconstructed and misidentified as $\tau_{\text{had-vis}}$ candidates.

[2] Isolation of a reconstructed object refers to a lack of activity in the vicinity of the object. The activity is often quantified using reconstructed charged-particle tracks or topo-clusters in a cone or annulus surrounding the object.

5.1 Simulated Event Samples

The tau reconstruction and identification algorithms employed at the ATLAS experiment for Run 2 of the LHC were developed using simulated events that provide samples of $\tau_{\text{had-vis}}$ candidates. For tau identification, simulated $\gamma^* \to \tau^+\tau^-$ and dijet events are used, yielding samples of true and fake $\tau_{\text{had-vis}}$ candidates, respectively.

An artificial $\gamma^* \to \tau^+\tau^-$ event sample was generated using Pythia 8.212 [166] for the matrix element calculation at leading order (LO), parton showering, hadronisation, and τ-lepton decays. The contribution of the Z boson propagator to the hard scattering process was removed to provide an unpolarised sample of τ leptons. In addition, the cross section of the process was modified at generator-level to enhance the number of events with high invariant $\tau^+\tau^-$ masses to increase the number of $\tau_{\text{had-vis}}$ candidates with large transverse momenta. Both τ leptons are enforced to decay hadronically to minimise statistical uncertainties from the size of the true-$\tau_{\text{had-vis}}$ sample.

Dijet events are generated using Pythia 8.186 [166] for the matrix element calculation at LO, parton showering, and hadronisation. The generation is performed in slices of p_{T} of the leading jet (anti-k_t with $R = 0.6$) constructed from generator-level particles. Slices with large jet transverse momenta are oversampled to increase the number of events with jets (fake $\tau_{\text{had-vis}}$) of large transverse momentum.

The $\gamma^* \to \tau^+\tau^-$ and dijet samples use the A14 set of tuned parameters for Pythia 8 [167] and the NNPDF2.3lo [168] set of parton distribution functions (PDFs). Decays of hadrons containing b or c quarks are simulated using EvtGen v1.2.0 [169]. The contamination of the hard scattering interaction with soft, inelastic proton–proton collision events is accounted for by overlaying the event with additional minimum-bias events. The response of the ATLAS detector is simulated for all generated events [170]. Subsequently, events are reconstructed from the simulated detector response using the Athena software suite [165].

5.1.1 $\tau_{\text{had-vis}}$ Candidate Selection

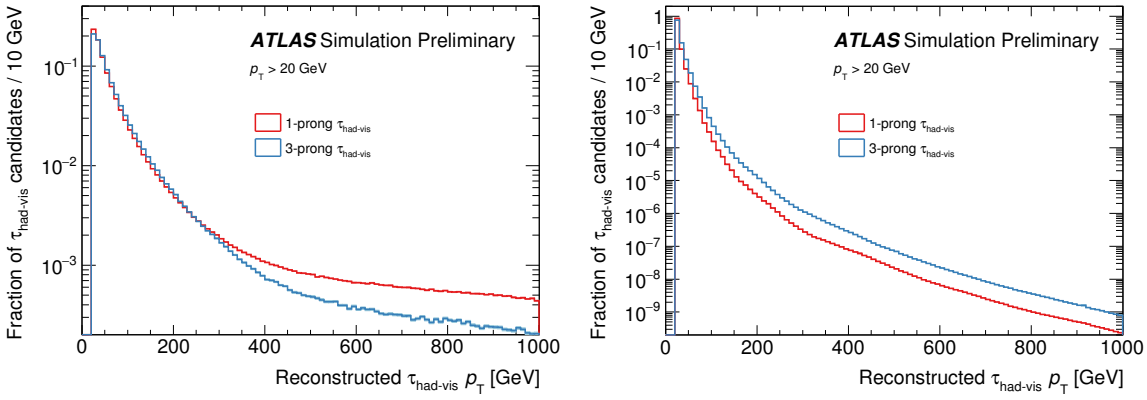
The simulated events are used to construct samples of $\tau_{\text{had-vis}}$ candidates for the development and performance evaluation of tau identification algorithms. Only candidates passing the $\tau_{\text{had-vis}}$ reconstruction are considered and the following selections are applied in addition:

- The number of *core* tracks (N_{tracks}) of the $\tau_{\text{had-vis}}$ candidate is either one or three. These are referred to as 1- or 3-prong $\tau_{\text{had-vis}}$ candidates, respectively.
- The (visible) transverse momentum of the candidate needs to fulfil $p_{\text{T}} > 20\,\text{GeV}$.
- The $\tau_{\text{had-vis}}$ candidate needs to be within $|\eta| < 2.5$ but outside the transition region between barrel and end-cap electromagnetic calorimeters given by $1.37 < |\eta| < 1.52$.

In addition, reconstructed $\tau_{\text{had-vis}}$ candidates from $\gamma^* \to \tau^+\tau^-$ events are required to be geometrically matched to a τ_{had} at generator-level ($\Delta R < 0.2$). All efficiencies and rejection factors given in the remainder of this chapter do not include the effects of $\tau_{\text{had-vis}}$ reconstruction or the selections outlined above.

The $\gamma^* \to \tau^+\tau^-$ and dijet events provide samples of true and fake $\tau_{\text{had-vis}}$ candidates with a size of 20 million and 46 million candidates, respectively. The distributions of $\tau_{\text{had-vis}}$ candidate p_{T} is shown for both samples and separately for 1- and 3-prong candidates in Figure 5.1. The difference in p_{T} spectra between 1- and 3-prong true $\tau_{\text{had-vis}}$ in Figure 5.1(a) results from a reduction in track

association efficiency for 3-prong $\tau_{\text{had-vis}}$ candidates with increasing candidate p_{T} due to the decrease in angular separation of charged hadrons from the τ-lepton decay. In contrast, the p_{T} spectrum of $\tau_{\text{had-vis}}$ candidates from dijet events, depicted in Figure 5.1(b), shows a heavier tail towards large p_{T} for 3-prong candidates due to the increase of the hadron multiplicity in jets with jet p_{T}. For the development and performance evaluation of tau identification algorithms, the sample of fake $\tau_{\text{had-vis}}$ candidates is re-weighted, separately for 1- and 3-prong candidates, to match the p_{T} spectrum of true $\tau_{\text{had-vis}}$ candidates from $\gamma^* \rightarrow \tau^+\tau^-$.



(a) True-$\tau_{\text{had-vis}}$ from $\gamma^* \rightarrow \tau^+\tau^-$ events

(b) Fake-$\tau_{\text{had-vis}}$ from dijet events

Figure 5.1: Transverse momentum distributions of 1- and 3-prong $\tau_{\text{had-vis}}$ candidates in $\gamma^* \rightarrow \tau^+\tau^-$ (a) and dijet events (b). Statistical uncertainties are shown as coloured bands surrounding the central value. The figures are taken from Ref. [11].

5.2 Tau Identification with Recurrent Neural Networks

The RNN tau identification exploits the discrimination power of both high- and low-level inputs to distinguish τ_{had} from quark- or gluon-initiated jets. This approach aims to avoid a possible loss of information relevant to tau identification when manually constructing high-level variables from low-level inputs. Specifically, charged-particle tracks and topo-clusters in the calorimeters, hereafter abbreviated as tracks and clusters, are included as low-level inputs to the algorithm. Their inclusion targets the differences in charged- and neutral-hadron multiplicities and track- and calorimeter-based isolation between true and fake $\tau_{\text{had-vis}}$. Tau identification is performed separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates due to their distinct signatures.

5.2.1 Input Variables

The input variables included in the RNN tau identification algorithm are summarised in Table 5.1. Three categories of inputs are considered: high-level, track, and cluster inputs. High-level inputs are observables directly associated to $\tau_{\text{had-vis}}$ candidates. Track and cluster inputs refer to observables of tracks and clusters that are associated to a $\tau_{\text{had-vis}}$ candidate. In the following, a description of the input variable selection and track/cluster association is given.

Table 5.1: Summary of input variables used for the RNN tau identification. The local hadronic calibration [120] is used to calibrate jets, clusters, and $\tau_{\text{had-vis}}$ candidates unless otherwise noted. Definitions of geometrical topo-cluster moments measuring the location and shape of clusters (λ, $\langle\lambda^2\rangle$, $\langle r^2\rangle$) are given in Ref. [120]. Variables using cell-level calorimeter information only consider cells that are part of topo-clusters for noise suppression. †: Energy depositions in the pre-sampler and first two layers of the electromagnetic calorimeters that are part of topo-clusters are abbreviated as "EM clusters". The table is adapted from Ref. [11].

	Variable	Description
High-level inputs	p_{T}^{LC}	Calorimeter-based estimate of $\tau_{\text{had-vis}}$ candidate p_{T}.
	f_{cent}	Ratio of E_{T} deposited in calorimeter cells (at EM scale) in cones of $\Delta R < 0.1$ and $\Delta R < 0.2$ about the $\tau_{\text{had-vis}}$ axis.
	$f_{\text{leadtrack}}^{-1}$	Ratio of E_{T} deposited in calorimeter cells (at EM scale) in a cone of $\Delta R < 0.2$ about the $\tau_{\text{had-vis}}$ axis and the p_{T} of the p_{T}-leading *core* track.
	ΔR_{max}	Maximum ΔR between *core* tracks and the $\tau_{\text{had-vis}}$ axis.
	$\lvert S_{\text{leadtrack}}\rvert$	Transverse impact parameter significance of the p_{T}-leading track. Only considered for 1-prong $\tau_{\text{had-vis}}$ candidates.
	$S_{\text{T}}^{\text{flight}}$	Transverse flight path significance. Only considered for 3-prong $\tau_{\text{had-vis}}$ candidates.
	$f_{\text{iso}}^{\text{track}}$	Ratio of the scalar sum of p_{T} of *isolation* tracks and the scalar sum of p_{T} of *core* and *isolation* tracks.
	$f_{\text{track}}^{\text{EM}}$	Ratio of the energy in EM clusters† and the scalar sum of momenta of *core* tracks.
	$p_{\text{T}}^{\text{EM+track}}/p_{\text{T}}$	p_{T} of the $\tau_{\text{had-vis}}$ estimated from the momenta of *core* tracks and the two most energetic EM clusters† divided by the p_{T} of the calorimetric measurement.
	$m^{\text{EM+track}}$	Invariant mass of the system of *core* tracks and the two most energetic EM clusters†.
	m^{track}	Invariant mass of the system of *core* tracks. Only considered for 3-prong $\tau_{\text{had-vis}}$ candidates.
Track inputs	$p_{\text{T}}^{\text{jet seed}}$	p_{T} of the jet seeding the $\tau_{\text{had-vis}}$ candidate.
	$p_{\text{T}}^{\text{track}}$	p_{T} of the track.
	$\Delta\eta^{\text{track}}$	Difference in η between track and $\tau_{\text{had-vis}}$ axis.
	$\Delta\phi^{\text{track}}$	Angle between track and $\tau_{\text{had-vis}}$ axis in the transverse plane.
	$\lvert d_0^{\text{track}}\rvert$	Absolute value of the transverse track impact parameter.
	$\lvert z_0^{\text{track}}\sin\theta\rvert$	Absolute value of the product of longitudinal track impact parameter and the sine of the polar angle of the track.
	$N_{\text{IBL hits}}$	Number of hits on the track in the IBL.
	$N_{\text{Pixel hits}}$	Number of hits on the track in pixel detector layers (excl. IBL).
	$N_{\text{SCT hits}}$	Number of hits on the track in SCT layers.
Cluster inputs	$p_{\text{T}}^{\text{jet seed}}$	p_{T} of the jet seeding the $\tau_{\text{had-vis}}$ candidate.
	$E_{\text{T}}^{\text{cluster}}$	E_{T} of the cluster.
	$\Delta\eta^{\text{cluster}}$	Difference in η between cluster and $\tau_{\text{had-vis}}$ axis.
	$\Delta\phi^{\text{cluster}}$	Angle between cluster and $\tau_{\text{had-vis}}$ axis in the transverse plane.
	λ_{cluster}	Longitudinal distance of the cluster barycentre from the calorimeter front face.
	$\langle\lambda^2\rangle_{\text{cluster}}$	Second longitudinal cluster moment.
	$\langle r^2\rangle_{\text{cluster}}$	Second radial cluster moment.

High-Level Input Variables

The selection of high-level input variables is based on the variables used in the BDT-based tau identification algorithm developed by the ATLAS collaboration [171], which was updated in Ref. [134] for the new $\tau_{\text{had-vis}}$ reconstruction techniques deployed during Run 2 of the LHC. The BDT tau identification uses only the high-level variables summarised in Table 5.1 as inputs and serves as a baseline for comparison with the RNN-based algorithm.

Variables sensitive to the lifetime and mass of the τ lepton ($|S_{\text{T}}^{\text{flight}}|$, $|S_{\text{leadtrack}}|$, m^{track}), the isolation of $\tau_{\text{had-vis}}$ in the tracking system ($f_{\text{iso}}^{\text{track}}$) and the calorimeters ($f_{\text{cent}}$), and combinations of track- and calorimeter-based isolation ($f_{\text{track}}^{\text{EM}}$, $p_{\text{T}}^{\text{EM+track}}/p_{\text{T}}$) are among the most important high-level variables included in the tau identification algorithms. Three exemplary distributions are shown in Figure 5.2.

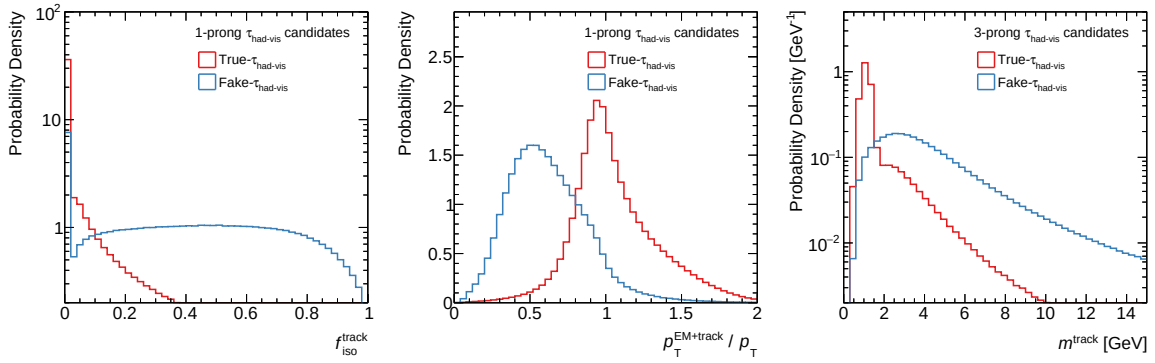
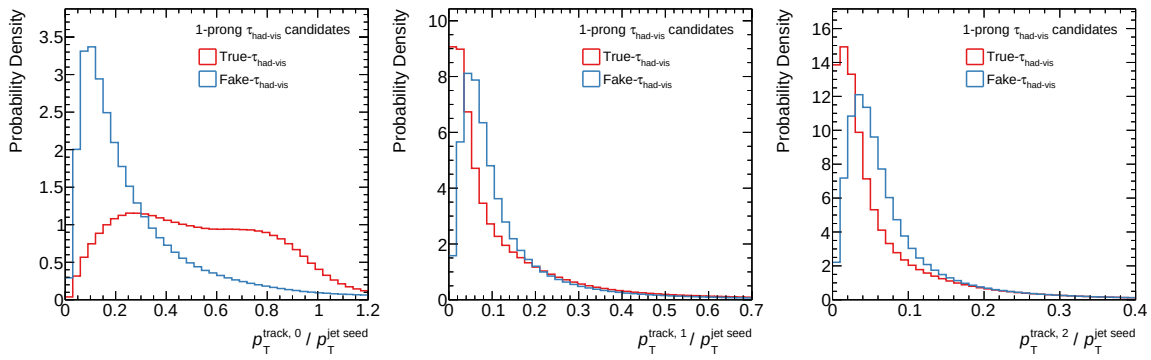


Figure 5.2: Distributions of exemplary high-level input variables used for tau identification. (a): The ratio of the scalar sum of p_{T} of tracks classified as *isolation* with respect to tracks classified as *core* or *isolation*. (b): The ratio of $\tau_{\text{had-vis}}$ candidate p_{T} estimated using a simplified particle flow approach and the purely calorimeter-based measurement (cf. Table 5.1). (c): The mass of the system of *core* tracks for 3-prong $\tau_{\text{had-vis}}$ candidates.

Track Input Variables

Reconstructed tracks with $p_{\text{T}} > 500\,\text{MeV}$ and within a cone of $\Delta R < 0.4$ about the $\tau_{\text{had-vis}}$ candidate axis are considered as inputs to the RNN tau identification. No selections are applied on the quality and impact parameters of reconstructed tracks, thus the inputs include tracks from the τ-lepton decay as well as fake tracks and tracks from pile-up. Instead, track quality criteria ($N_{\text{IBL hits}}$, $N_{\text{Pixel hits}}$, $N_{\text{SCT hits}}$) and track impact parameters ($|d_0^{\text{track}}|$, $|z_0^{\text{track}} \sin\theta|$) are included as observables of tracks.

In addition, several other track-level observables, summarised in Table 5.1, are included. Among the most important variables are the transverse momenta of reconstructed tracks ($p_{\text{T}}^{\text{track}}$) and their angular separation from the $\tau_{\text{had-vis}}$ candidate axis ($\Delta\eta^{\text{track}}$, $\Delta\phi^{\text{track}}$). These variables are included to probe the isolation properties of $\tau_{\text{had-vis}}$ candidates. A special case is the $p_{\text{T}}^{\text{jet seed}}$ variable, the p_{T} of the jet seeding the $\tau_{\text{had-vis}}$ candidate, which is not a track property but is still included as an observable for every track. This is done to provide an approximate p_{T}-scale of the jet already at the level of individual input tracks. Exemplary distributions of the transverse momenta of the three highest-p_{T} tracks normalised to $p_{\text{T}}^{\text{jet seed}}$ are shown in Figure 5.3 for 1-prong $\tau_{\text{had-vis}}$ candidates.



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

Figure 5.3: Distributions of the transverse momenta of the three highest-p_{T} tracks associated to 1-prong $\tau_{\text{had-vis}}$ candidates. For illustration purposes, the track p_{T} are normalised to the p_{T} of the jet seeding the $\tau_{\text{had-vis}}$ candidate.

The discrimination power of the RNN tau identification saturates after including the ten highest-p_{T} tracks; therefore, the sequence of tracks is truncated at this point to reduce the computational resources required for training and evaluation of the networks.

Cluster Input Variables

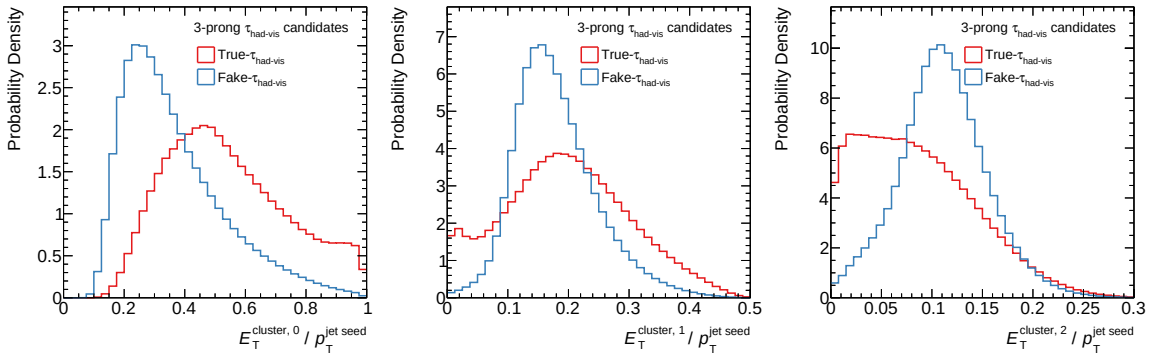
Topo-clusters in the calorimeters are considered as inputs to the RNN tau identification if they are constituents of the jet seeding the $\tau_{\text{had-vis}}$ reconstruction. All clusters are calibrated using the local hadronic calibration to account for the non-compensating nature of the calorimeters, energy deposition in calorimeter cells not part of the cluster, and energy loss in inactive material [120].

The inclusion of the $E_{\mathrm{T}}^{\text{cluster}}$, $\Delta\eta^{\text{cluster}}$, $\Delta\phi^{\text{cluster}}$, and $p_{\mathrm{T}}^{\text{jet seed}}$ observables (cf. Table 5.1) follow from considerations similar to those for charged-particle tracks. In addition, information on the position and shape of showers in the calorimeters is included in the form of cluster moments [120], targeting the differences between electromagnetic and hadronic showers. These cluster moments include the longitudinal location of the cluster barycentre, λ_{cluster}, and the lateral and longitudinal extension of the cluster, $\langle r^2 \rangle_{\text{cluster}}$ and $\langle \lambda^2 \rangle_{\text{cluster}}$, respectively. For illustration, the E_{T} of the three highest-E_{T} clusters is shown in Figure 5.4 for 3-prong $\tau_{\text{had-vis}}$ candidates.

The classification performance of the RNN tau identification saturates after the inclusion of the six highest-E_{T} clusters, thus the sequence of input clusters is truncated at this point.

5.2.2 Network Architecture

The network architecture adopted for the RNN tau identification is shown schematically in Figure 5.5. The network consists of three branches, each one dedicated to one type of input. The high-level variables, track inputs, and cluster inputs are first processed independently in their respective branches. These branches are then merged and reduced to a single output node that is used to define the tau identification working points. The network is implemented using the Keras library [162] with



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

Figure 5.4: Distributions of the transverse energies of the three highest-E_{T} clusters associated to 3-prong $\tau_{\text{had-vis}}$ candidates. For illustration purposes, the cluster E_{T} are normalised to the p_{T} of the jet seeding the $\tau_{\text{had-vis}}$ candidate.

the TensorFlow backend [172].[3] Layers use the default configurations of Keras unless indicated otherwise. A description of the network architecture is given in the following.

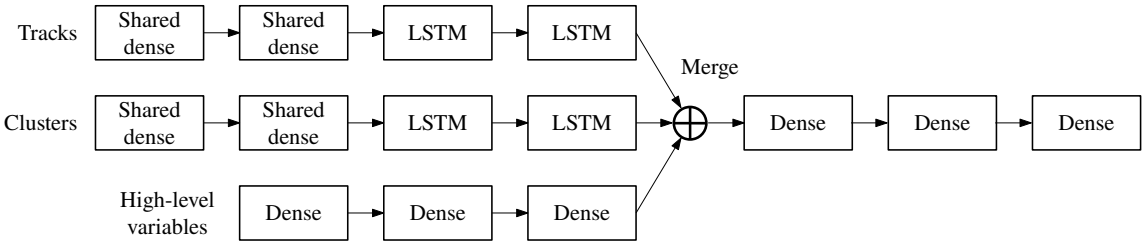


Figure 5.5: Network architecture used for the RNN tau identification algorithm. Individual layers of the network are depicted as rectangles. The figure is adapted from Ref. [11].

The branch of the network operating on high-level input variables consists of three fully-connected (*dense*) layers with 128, 128, and 16 units each. The ReLU [173] activation function is used in all layers.

The branches operating on track and cluster inputs are structured identically. Tracks (clusters) are provided to the network as sequences of vectors, each vector consisting of the values of the input variables for a given track (cluster). The track and cluster sequences are given in descending $p_{\text{T}}^{\text{track}}$ and E_{T}^{cluster} order, respectively. First, the sequences are passed through two fully-connected layers with shared weights (*shared dense*). These layers have 32 units each and use the ReLU activation function. *Shared dense* layers map input sequences, $(\boldsymbol{x}_t)_{t=1}^{N}$, to output sequences of the same length, $(\boldsymbol{y}_t)_{t=1}^{N}$, using transformations of the form

$$\boldsymbol{y}_t = \boldsymbol{\phi}(\boldsymbol{W}\boldsymbol{x}_t + \boldsymbol{b}) \,,$$

where $\boldsymbol{W}$ and $\boldsymbol{b}$ are trainable weight matrices and bias vectors, respectively, and $\boldsymbol{\phi}$ is the activation

[3] Keras v2.2.0 and TensorFlow v1.8.0 are used.

function. Notably, the weights and biases do not depend on the index t, meaning the weights and biases are the same (i.e. shared) for all elements of the sequence. The *shared dense* layers produce intermediate representations of the track and cluster sequences for further computation.

The transformed sequences of tracks and clusters pass through two recurrent layers based on an LSTM architecture (cf. Section 4.2.2). The first LSTM layer maps the input sequence to an output sequence of the same length. In contrast to *shared dense* layers, LSTM layers have an internal state that is updated as elements of the sequence are processed. Therefore, information about elements occurring in the sequence can be exploited when processing subsequent elements. The second LSTM layer repeats the process, however, all except the last element of the output sequence are discarded, thereby encoding the input sequence into an output of fixed size. The size of the internal state and outputs of the LSTM layers are chosen to be the same and correspond to 32 units for the first and 24 units for the second LSTM layer.

Finally, the outputs of the three branches are concatenated (*Merge*) and passed through three fully-connected layers with sizes of 64, 32, and 1 units. The ReLU (sigmoid) activation function is used for the hidden layers (output layer). The entire model consists of approximately 56 000 trainable parameters.

5.2.3 Network Training and Evaluation

Two separate networks are trained for the classification of 1- and 3-prong $\tau_{\text{had-vis}}$ candidates. The samples of $\tau_{\text{had-vis}}$ candidates are partitioned into a training (40 %), validation (10 %), and testing datasets (50 %). The trainable parameters of the networks are determined by minimising the binary cross-entropy loss on the training dataset. The loss of the network is monitored on the validation datasets and is used to steer the training procedure.

SGD with momentum is used for loss minimisation. Prior to the training, transformations are applied to input variables for better conditioning of the minimisation problem. The learning rate of the optimiser is reduced when the validation loss did not improve for four successive training epochs, where an epoch refers to a single pass over the training data. Similarly, the training is stopped after ten epochs without improvements of the validation loss in which case the model parameters resulting in the best validation loss are restored.

The trained NNs are implemented for evaluation during offline event reconstruction in the Athena software suite [165] using the lwtnn library [174].

The classification score computed by the RNN tau identification is transformed to be approximately uniformly distributed for true $\tau_{\text{had-vis}}$ candidates. The transformed score, referred to as the RNN score and shown in Figure 5.6, allows the definition of working points for a specific true-$\tau_{\text{had-vis}}$ identification efficiency target by applying a threshold to the score. Moreover, the transformation is derived in bins of $\tau_{\text{had-vis}}$ p_{T} (calorimeter-based p_{T} estimate at LC scale) and the average number of interactions per bunch crossing, μ, to ensure that the true-$\tau_{\text{had-vis}}$ efficiencies of working points remains approximately constant with changing $\tau_{\text{had-vis}}$ p_{T} or pile-up conditions. The same procedure is applied for the BDT-based tau identification algorithm.

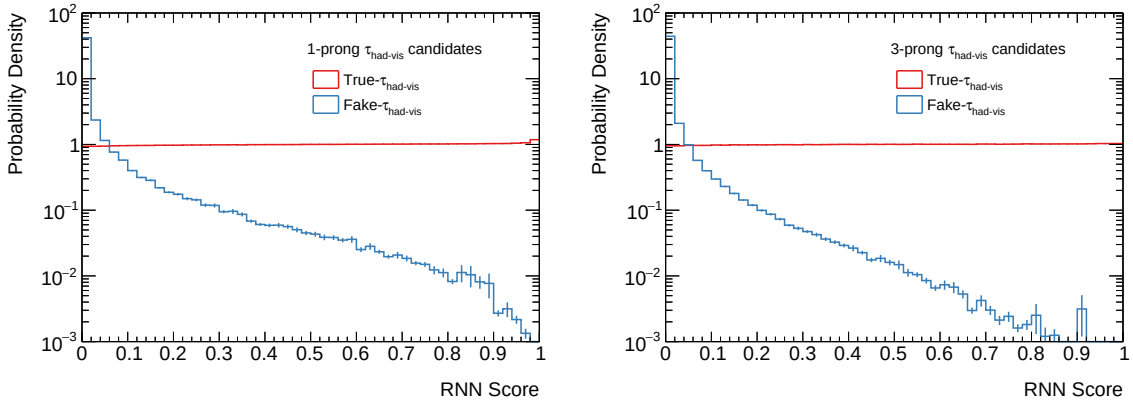


Figure 5.6: Distributions of the RNN score for 1-prong (a) and 3-prong (b) $\tau_{\text{had-vis}}$ candidates. The RNN scores are shown after transformations that ensure that the scores of true $\tau_{\text{had-vis}}$ are approximately uniformly distributed. The figures are adapted from Ref. [11].

5.3 Expected Performance of Tau Identification with RNN

Thresholds on the RNN scores of 1- and 3-prong $\tau_{\text{had-vis}}$ candidates define working points of varying $\tau_{\text{had-vis}}$ identification efficiencies and fake-$\tau_{\text{had-vis}}$ rejection factors.[4] Four named working points are defined to cover the needs of most physics analyses at the ATLAS experiment. These working points are, in order of increasing fake-$\tau_{\text{had-vis}}$ rejection: *very loose*, *loose*, *medium*, and *tight*.

The targeted $\tau_{\text{had-vis}}$ identification efficiencies of the working points are summarised in Table 5.2. The rejection of fake $\tau_{\text{had-vis}}$ candidates from simulated dijet events is compared between the BDT- and RNN-based tau identification algorithms for working points targeting the same true-$\tau_{\text{had-vis}}$ efficiency. The RNN tau identification improves the fake-$\tau_{\text{had-vis}}$ rejection by about 80 % for 1-prong and between 40 to 80 % for 3-prong $\tau_{\text{had-vis}}$ candidates over the BDT-based method. The majority of these improvements stem from the inclusion of charge-particle track information in addition to the high-level variables also used in the BDT. The *receiver operating characteristic curves* of both algorithms are shown in Figure 5.7.

The $\tau_{\text{had-vis}}$ identification efficiencies are shown in Figure 5.8 for all four working points in bins of $\tau_{\text{had-vis}}$ p_{T}, $\tau_{\text{had-vis}}$ $|\eta|$, and the average number of interactions per bunch crossing, μ. The efficiency remains approximately constant as $\tau_{\text{had-vis}}$ p_{T} and μ is varied due to the definition of the working points using the transformed output of the neural network. An exception is the first $\tau_{\text{had-vis}}$ p_{T} bin which does not meet the target efficiency. This is an effect of different $\tau_{\text{had-vis}}$ momentum calibrations being used for the transformation of the RNN output and for the performance evaluation shown in Figures 5.8(a) and 5.8(b). The transformation was derived using the calorimeter-based $\tau_{\text{had-vis}}$ p_{T} estimate at LC scale, which is now superseded by a multivariate method combining information from the calorimeters and tracking systems (shown in the figures).

The rejection of fake $\tau_{\text{had-vis}}$ from simulated dijet events by the *medium* tau identification working

[4] The $\tau_{\text{had-vis}}$ identification efficiency is the fraction of true $\tau_{\text{had-vis}}$ from simulated $\gamma^* \to \tau^+\tau^-$ events that pass a given working point. The fake-$\tau_{\text{had-vis}}$ rejection is the reciprocal of the fraction of fake $\tau_{\text{had-vis}}$ from simulated dijet events that pass a given working point.

Table 5.2: Comparison of working points defined for the BDT- and RNN-based tau identification. Only the targeted $\tau_{\text{had-vis}}$ efficiency (target $\tau_{\text{had-vis}}$ eff.) of the working points is given, which can deviate by ca. 1 % from the efficiency observed in simulated $\gamma^* \to \tau^+\tau^-$ events. The fake-$\tau_{\text{had-vis}}$ rejection is evaluated using the $\tau_{\text{had-vis}}$ candidate sample from dijet events. The table is adapted from Ref. [11].

Working point	Target $\tau_{\text{had-vis}}$ eff.		Fake-$\tau_{\text{had-vis}}$ rejection			
			BDT		RNN	
	1-prong	3-prong	1-prong	3-prong	1-prong	3-prong
Tight	60 %	45 %	40	420	72	770
Medium	75 %	60 %	20	160	36	260
Loose	85 %	75 %	12	66	21	99
Very loose	95 %	95 %	5.4	12	10	17

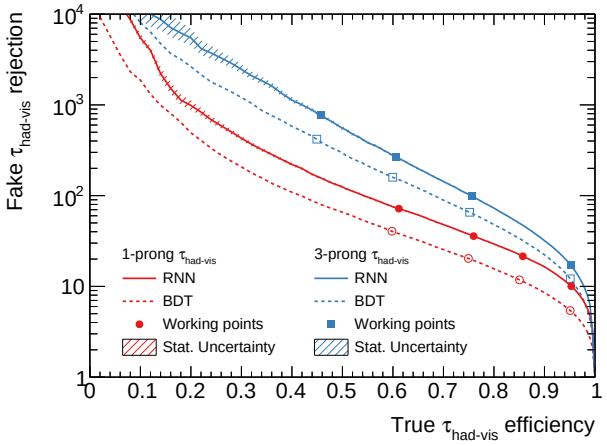


Figure 5.7: Receiver operating characteristic curves of the tau identification algorithms based on RNN (solid) and BDT (dashed) shown separately for 1-prong (red) and 3-prong (blue) $\tau_{\text{had-vis}}$ candidates. The true- and fake-$\tau_{\text{had-vis}}$ efficiencies are evaluated using samples of $\tau_{\text{had-vis}}$ candidates from $\gamma^* \to \tau^+\tau^-$ and dijet events, respectively. The figure is adapted from Ref. [11].

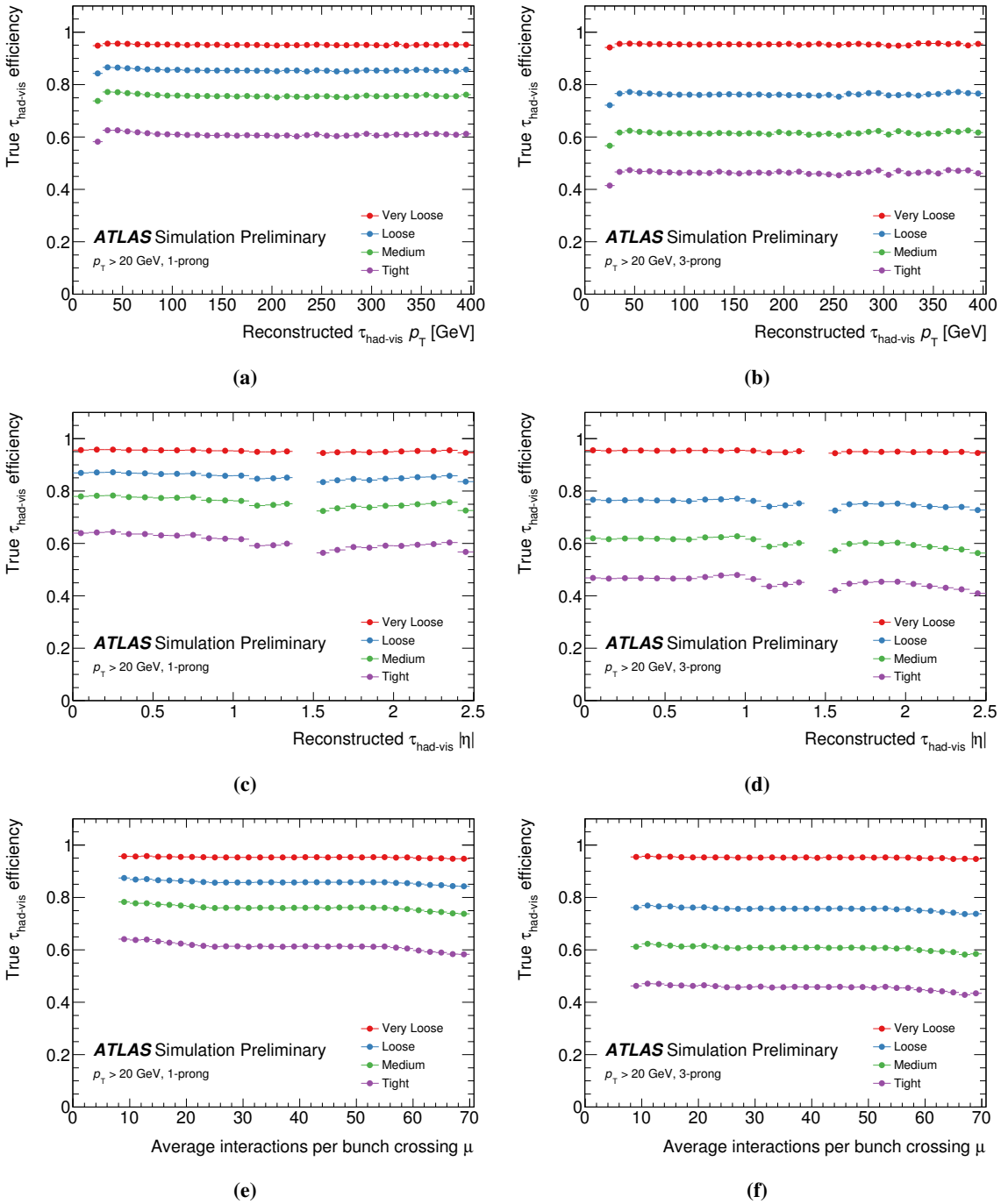


Figure 5.8: True-$\tau_{\text{had-vis}}$ efficiencies of the RNN tau identification working points. The efficiencies are estimated using $\tau_{\text{had-vis}}$ candidates from simulated $\gamma^* \rightarrow \tau^+\tau^-$ events separately for 1-prong (left) and 3-prong (right) $\tau_{\text{had-vis}}$. Efficiencies are shown in bins of $\tau_{\text{had-vis}}$ p_{T} (top), $\tau_{\text{had-vis}}$ $|\eta|$ (middle), and the average number of interactions per bunch crossing (bottom). All figures are taken from Ref. [11].

points of the BDT- and RNN-based methods is compared in Figure 5.9. The comparison is performed separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates and in bins of $\tau_{\text{had-vis}}$ p_{T}, $\tau_{\text{had-vis}}$ $|\eta|$, and μ. In general, the fake-$\tau_{\text{had-vis}}$ rejection of tau identification is analysis and process dependent[5] and thus the comparison mainly serves as a benchmark of the identification algorithms.

The fake-$\tau_{\text{had-vis}}$ rejection of tau identification increases with $\tau_{\text{had-vis}}$ candidate p_{T}, as is shown in Figures 5.9(a) and 5.9(b). This trend is partially explained by the increasing discrimination power of variables sensitive to the isolation of $\tau_{\text{had-vis}}$ candidates as larger momentum scales are considered. This is a consequence of the increasing collimation of true $\tau_{\text{had-vis}}$ candidates due to the Lorentz boost of the τ lepton, while fake $\tau_{\text{had-vis}}$ candidates become less isolated due to the increase in the average charged and neutral hadron multiplicity in jets.

The RNN tau identification improves the fake-$\tau_{\text{had-vis}}$ rejection by a factor of two for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates with $p_{\text{T}} > 70\,\text{GeV}$ over the BDT-based approach at comparable working points. Similar behaviour of the fake-$\tau_{\text{had-vis}}$ rejection is observed for both methods with respect to the pseudorapidity of $\tau_{\text{had-vis}}$ candidates. However, the RNN tau identification is more susceptible to pile-up compared to the BDT-based method. This is depicted in Figures 5.9(e) and 5.9(f) showing a degradation in fake-$\tau_{\text{had-vis}}$ rejection as the average number of interactions per bunch crossing increases.

Data-driven measurements of the $\tau_{\text{had-vis}}$ identification efficiencies of selected RNN tau identification working points (*loose*, *medium*, and *tight*) were performed by the ATLAS collaboration. The measurements used the tag-and-probe method in $Z \to \tau_{\mu}\tau_{\text{had}}$ events and have provided calibrations of the $\tau_{\text{had-vis}}$ selection efficiencies in simulation. A relative agreement within 5 % is observed between the efficiency predicted by simulation and the central value of the data-driven measurement. Both agree within the statistical and systematic uncertainties of the measurement, showing acceptable modelling of the tau identification decisions thresholds in simulation even after including low-level information from charged-particle tracks and topo-clusters.

5.4 Conclusion and Outlook

This chapter introduced a tau identification algorithm that exploits the discrimination power of high-level input variables in conjunction with low-level detector signatures of charged and neutral particles in the vicinity of $\tau_{\text{had-vis}}$ candidates. The main feature of the algorithm is the use of an RNN architecture, which allows it to operate on sequences of charged-particle tracks and clusters of energy in the calorimeters.

The new algorithm improves the rejection of fake $\tau_{\text{had-vis}}$ reconstructed as 1- and 3-prong $\tau_{\text{had-vis}}$ candidates by about 80 % and 40 to 80 %, respectively, compared to the BDT-based method previously employed at the ATLAS experiment. For $\tau_{\text{had-vis}}$ candidates with large transverse momenta ($p_{\text{T}} > 70\,\text{GeV}$) the improvement in fake-$\tau_{\text{had-vis}}$ rejection exceeds 100 %. This improvement is mostly driven by the inclusion of charged-particle tracks in the network to exploit the track-based isolation of $\tau_{\text{had-vis}}$.

The method is considered by the ATLAS collaboration as the recommended tau identification algorithm for analyses of the $139\,\text{fb}^{-1}$ pp collision dataset recorded during Run 2 of the LHC. In addition, the approach was adopted for tau identification at the high-level trigger in 2018 [175]. The

[5] The ratio of quark- to gluon-initiated jets affects the fake-$\tau_{\text{had-vis}}$ rejection since gluon-initiated jets are more easily rejected by tau identification. Additionally, the rejection depends on the p_{T} and N_{tracks} of fake $\tau_{\text{had-vis}}$ and thus on the selections applied by analyses.

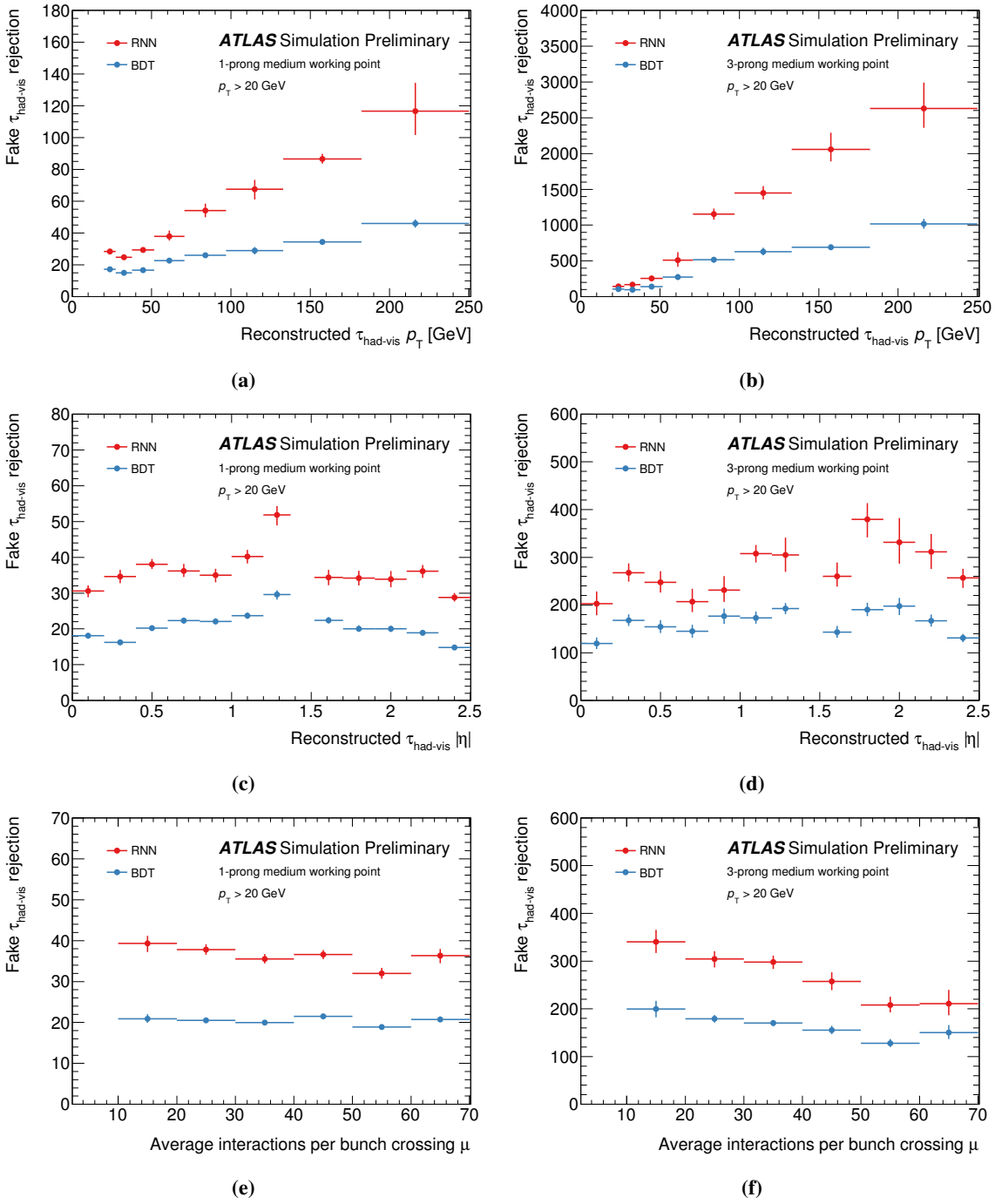


Figure 5.9: Fake-$\tau_{\text{had-vis}}$ rejection of the *medium* RNN (red) and BDT (blue) tau identification working points. The rejection is shown separately for 1-prong (left) and 3-prong (right) $\tau_{\text{had-vis}}$ candidates in bins of p_{T} (top), $|\eta|$ (middle) of $\tau_{\text{had-vis}}$ candidates, and the average number of interactions per bunch crossing, μ, for the event (bottom). The rejection is estimated using a sample of $\tau_{\text{had-vis}}$ candidates from simulated dijet events. All figures are taken from Ref. [11].

superior background rejection of the RNN tau identification allowed for an increase in the efficiency of $\tau_{\text{had-vis}}$-triggers while remaining within the allocated trigger bandwidth [175].

The improvements in tau identification, both at the high-level trigger and during offline event reconstruction, are utilised in Chapter 6 for the search for Higgs boson pair production. Due to the improved background rejection of the RNN, it is possible to loosen the identification requirements for $\tau_{\text{had-vis}}$ in the search without incurring large increases in backgrounds from the misidentification of quark- or gluon-initiated jets as $\tau_{\text{had-vis}}$.

The tau identification algorithms employed by the CMS collaboration underwent a similar evolution as the one presented in this chapter. A BDT-based discriminant based on high-level features [176] was replaced by a deep neural network that combines high-level information with information from reconstructed particles in the vicinity of the $\tau_{\text{had-vis}}$ [177]. Similarly, the CMS collaboration observed a large reduction in the fake-$\tau_{\text{had-vis}}$ misidentification efficiency by factors exceeding 1.8 compared to the previous method [177], which only relied on high-level features. While the concept is similar to the approach presented here, the technical implementation adopted by the CMS collaboration differs and is documented in Ref. [177].

The RNN tau identification continues to be used by the ATLAS experiment at the beginning of Run 3 of the LHC for identification at the high-level trigger and for offline event reconstruction [12]. In the long term, changes to the tau identification strategy are expected. A new RNN-based $\tau_{\text{had-vis}}$ track classification algorithm [178] was deployed for $\tau_{\text{had-vis}}$ reconstruction in Run 3 [12]. This algorithm could be integrated into or combined with the tau identification procedure to provide isolation information directly. Such an approach is motivated by the large overlap in track observables used by both algorithms and the inherent connection between $\tau_{\text{had-vis}}$ track selection and tau identification. Second, other types of networks operating on variable-size collections of inputs are explored by the collaboration. For tau identification, DeepSets [179] are an alternative to the RNN-based approach. DeepSets operate on unordered sets of inputs and therefore, unlike RNN, do not require a specific ordering of the inputs. Preliminary results show that tau identification based on DeepSets yields results that are competitive with the implementation using RNN while requiring significantly less time for training and prediction.[6]

[6] Similar findings were obtained for DeepSets-based b-tagging algorithms in Ref. [180].

CHAPTER 6

Search for Higgs Boson Pair Production in the $b\bar{b}\tau^+\tau^-$ Final State

This chapter presents a search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state using 139 fb^{-1} of pp collision data recorded at the ATLAS experiment in Run 2 of the LHC. Non-resonant production of Higgs boson pairs predicted by the SM (SM HH production) and resonant production via scalar resonances of narrow width are considered in the search. Final states with two hadronic τ-lepton decays ($b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$) and with one hadronic and one leptonic τ-lepton decay ($b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$) are targeted. These cover the majority of decays of the $\tau^+\tau^-$ system with fractions of $\mathcal{B}(\tau^+\tau^- \to \tau_{\text{had}}\tau_{\text{had}}) = 42.0\,\%$ and $\mathcal{B}(\tau^+\tau^- \to \tau_{\text{lep}}\tau_{\text{had}}) = 45.6\,\%$ [181]. Final states with two leptonic τ-lepton decays, which have a branching fraction of $\mathcal{B}(\tau^+\tau^- \to \tau_{\text{lep}}\tau_{\text{lep}}) = 12.4\,\%$ [181], are not considered in this search. The primary focus of this chapter lies on the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

The $b\bar{b}\tau^+\tau^-$ channel is among the most promising channels to probe the nature of SM HH production. This is due to a substantial fraction of decays of Higgs boson pairs to $b\bar{b}\tau^+\tau^-$ final states of about 7.3 % (cf. Figure 2.8) and the distinct signature of τ leptons. Previous searches in the $b\bar{b}\tau^+\tau^-$ channel were performed by the ATLAS and CMS collaborations using data recorded in Run 1 [182, 183] and at the beginning of Run 2 of the LHC [184, 185]. Other searches for Higgs boson pair production using the full Run 2 pp collision datasets were performed by the ATLAS and CMS collaborations in the $b\bar{b}\tau^+\tau^-$ [186], $b\bar{b}b\bar{b}$ [187, 188], and $b\bar{b}\gamma\gamma$ [189, 190] channels. The work presented in this chapter contributes to Refs. [191, 192].

The search for SM HH production aims to set upper limits on the signal strength and cross section of SM HH production via ggF and VBF provided no significant deviation from the background-only hypothesis is observed. Under the SM expectation, the size of the pp collision dataset recorded during Run 2 of the LHC is not sufficient to obtain evidence for SM HH production. Nevertheless, BSM physics could enhance the non-resonant production of Higgs boson pairs, which can be probed in this search. Interpretations of the SM HH search in terms of anomalous values of the Higgs boson self-coupling strength are not subject of this chapter but rather of Chapter 7. The search for resonant HH production targets scalar resonances with masses ranging from 251 to 1 600 GeV produced via ggF and aims to, in the absence of significant deviations from the SM expectation, set upper limits on the production cross section of the scalar resonance.

This chapter is structured as follows: Section 6.1 describes the data and simulated event samples used for this search. The reconstruction of physics objects and selection of events is discussed

in Sections 6.2 and 6.3, respectively. The estimation of background processes contributing to the search is described in Section 6.4 with a focus on the estimation of backgrounds with quark- or gluon-initiated jets that are misidentified as $\tau_{\text{had-vis}}$. The multivariate analysis used to distinguish signal from background events is discussed in Section 6.5. Systematic uncertainties and the statistical interpretation are described in Sections 6.6 and 6.7. This chapter concludes with a discussion of the results in Section 6.8 including a comparison with other searches for Higgs boson pair production performed by the ATLAS and CMS collaborations as well as an outlook on future prospects of SM HH searches.

6.1 Data and Simulated Event Samples

This search uses pp collision data at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ collected by the ATLAS experiment during Run 2 of the LHC. All recorded events have to pass data quality criteria [102], requiring a fully operational ATLAS detector and stable beams at the LHC. The integrated luminosity of events passing the quality criteria corresponds to $139\,\text{fb}^{-1}$ with an uncertainty of 1.7 % [100]. Events recorded by the ATLAS detector are reconstructed using the Athena software suite [165].

MC event generators are used to estimate the contributions of signal and most background processes in this search. The response of the ATLAS detector to generated events is obtained either from a *full simulation* of the detector based on Geant4 [170, 193] or from a hybrid approach referred to as *fast simulation* that replaces the simulation of the calorimeter response with a parametric description thereof [170]. Events are reconstructed from simulated detector responses using the same algorithms used to reconstruct collision events recorded by the ATLAS detector. The effect of pile-up is accounted for by overlaying all generated events with additional inelastic pp collisions obtained from simulation. Simulated events are then re-weighted to ensure that the pile-up conditions match those of the recorded data.

In the following, a description of the MC event generators used for the simulation of signal processes is given. The generator configurations used for the simulation of background processes are summarised in Table 6.1.

SM HH production via ggF is simulated using Powheg Box v2 [194–196] at NLO accounting for the finite top-quark mass in real and virtual corrections [224–228].[1] The generator uses the PDF4LHC15nlo set of PDFs [197] and is interfaced to Pythia 8 [166] with the A14 set of tuned parameters [167] for parton showering, hadronisation, and simulation of the underlying event. The sample of simulated events is normalised using a $pp \to HH$ cross section of 31.05 fb at $\text{NNLO}_{\text{FTapprox}}$ [46], which is a combination of the full-theory prediction at NLO with additional NNLO corrections derived in the large top-quark mass limit. The theoretical uncertainties on the cross section prediction are $^{+\,6\,\%}_{-23\,\%}$ from scale variations and the treatment of the finite top-quark mass[2] [47] and ±3 % from uncertainties on PDFs and α_s [48].

SM HH production via VBF is simulated using MadGraph5_aMC@NLO [198] at LO using the NNPDF3.0nlo set of PDFs [199]. The matrix element generator is interfaced to Pythia 8 with the A14 tune for parton showering, hadronisation, and simulation of the underlying event. A cross section of 1.726 fb at N^{3}LO [48, 49] is used to normalise the sample of simulated events. The

[1] As opposed to earlier calculations in the $m_t \to \infty$ limit using effective field theory approximations to simplify top-quark loops to effective couplings. See for example Ref. [229] (NLO) and Ref. [230] (NNLO).

[2] This uncertainty is defined by a comparison of the pole mass and the $\overline{\text{MS}}$ scheme for the mass of virtual top-quarks.

Table 6.1: Summary of generators used to simulate signal and background processes for the search for Higgs boson pair production. The order of the perturbative expansion in α_s is given unless qualified by "EW", which indicates higher order electroweak corrections. $*$: V + jets (diboson) event generation with Sherpa 2.2.1 merges matrix elements with NLO accuracy for up to two (one) final state partons and LO accuracy for up to four (three) final state partons. †: $q\bar{q}/qg$ induced production of ZH is normalised using the total $pp \to ZH$ cross section (NNLO+NLO EW) and subtracting the $gg \to ZH$ cross section (NLO+NLL) using predictions from Ref. [40]. The table is adapted from Ref. [192].

Process	ME generator	ME PDF set	ME order	Parton shower	Tune	Cross section order
Signals (m_H = 125 GeV)						
SM HH (ggF)	Powheg Box v2 [194–196]	PDF4LHC15nlo [197]	NLO	Pythia 8.244 [166]	A14 [167]	$\text{NNLO}_{\text{FTapprox}}$ [46]
SM HH (VBF)	MadGraph5_aMC@NLO 2.7.3 [198]	NNPDF3.0nlo [199]	LO	Pythia 8.244	A14	N^{3}LO [49]
$gg \to X \to HH$	MadGraph5_aMC@NLO 2.6.1	NNPDF2.3lo [168]	LO	Herwig 7.1.3 [200, 201]	H7.1-Default	–
Top-quark						
$t\bar{t}$	Powheg Box v2 [202]	NNPDF3.0nlo	NLO	Pythia 8.230	A14	NNLO+NNLL [203–209]
Single top	Powheg Box v2 [210, 211]	NNPDF3.0nlo	NLO	Pythia 8.230	A14	NLO [212]
tW	Powheg Box v2 [213]	NNPDF3.0nlo	NLO	Pythia 8.230	A14	NNLO approx. [212, 214, 215]
$t\bar{t}Z$	Sherpa 2.2.1 [216]	NNPDF3.0nnlo	NLO	Sherpa 2.2.1 [217]	Sherpa	NLO (NLO EW) [40]
$t\bar{t}W$	Sherpa 2.2.8	NNPDF3.0nnlo	NLO	Sherpa 2.2.8	Sherpa	NLO
Vector boson + jets						
$W \to \ell\nu$ $Z/\gamma^* \to \ell\ell$	Sherpa2.2.1	NNPDF3.0nnlo	NLO*	Sherpa2.2.1	Sherpa	NNLO [218, 219]
Diboson						
WW, WZ, ZZ	Sherpa2.2.1	NNPDF3.0nnlo	NLO*	Sherpa2.2.1	Sherpa	NLO
Single Higgs boson (m_H = 125 GeV)						
$H \to \tau^+\tau^-$ (ggF)	Powheg Box v2 [220, 221]	NNPDF3.0nnlo	NNLO	Pythia 8.212	AZNLO [222]	N^{3}LO (NLO EW) [40]
$H \to \tau^+\tau^-$ (VBF)	Powheg Box v2 [223]	NNPDF3.0nlo	NLO	Pythia 8.230	AZNLO	NNLO approx. (NLO EW) [40]
$H \to \tau^+\tau^-/b\bar{b}$ (WH)	Powheg Box v2	NNPDF3.0nlo	NLO	Pythia 8.212	AZNLO	NNLO (NLO EW) [40]
$H \to \tau^+\tau^-/b\bar{b}$ ($q\bar{q}/qg \to ZH$)	Powheg Box v2	NNPDF3.0nlo	NLO	Pythia 8.212	AZNLO	NNLO (NLO EW)† [40]
$H \to \tau^+\tau^-/b\bar{b}$ ($gg \to ZH$)	Powheg Box v2	NNPDF3.0nlo	LO	Pythia 8.244/8.212	AZNLO	NLO+NLL [40]
$H \to \tau^+\tau^-/b\bar{b}$ ($t\bar{t}H$)	Powheg Box v2	NNPDF3.0nlo	NLO	Pythia 8.230	A14	NLO (NLO EW) [40]

theoretical uncertainties on the cross section prediction are $^{+0.03\,\%}_{-0.04\,\%}$ from scale variations and $\pm 2.1\,\%$ from uncertainties on PDFs and α_s [48].

Higgs boson pair production via scalar resonances produced in ggF is simulated using MadGraph5_aMC@NLO at LO using the NNPDF2.3LO set of PDFs [168]. The matrix element generator is interfaced to Herwig 7.1 [200, 201] with the default tune for parton shower, hadronisation, and simulation of the underlying event. Twenty benchmark signals are generated for resonances with masses m_X ranging from 251 to 1 600 GeV and a decay width of 10 MeV. The interference between resonant HH production and SM HH production is neglected.

Lastly, decays of hadrons containing b and c quarks are simulated using EvtGen [169] for all signal, top-quark, and single-Higgs-boson processes (cf. Table 6.1). Moreover, full simulation of the ATLAS detector is used for all processes except for resonant HH production with $m_X \leq 1\,000$ GeV and alternative samples used for the derivation of uncertainties. In these cases, fast simulation is used instead.[3]

6.2 Object Reconstruction and Selection

The reconstruction and selection of physics objects is described in this section. The most relevant objects for this search are electrons, muons, $\tau_{\text{had-vis}}$, jets, b-tagged jets, and the missing transverse momentum. These objects are reconstructed and identified using algorithms that target their distinct signatures in the ATLAS detector (cf. Section 3.3). Differences between the performance of the object reconstruction and selection in simulation and data are accounted for by dedicated calibration measurements of efficiencies, energy/momentum scales and resolutions.

The baseline selection of objects used in this search is described in the following. More restrictive selections are applied as part of the event selection, which is discussed in Section 6.3.

Electrons are required to have $p_T > 7$ GeV and to be reconstructed within the acceptance of the tracking detectors, $|\eta| < 2.47$. Electrons in the transition region between the barrel and end-cap electromagnetic calorimeters, $1.37 < |\eta| < 1.52$, are rejected. All reconstructed electrons have to pass the loose working point of a likelihood-based electron identification algorithm to reject non-prompt electrons and electron candidates originating from jets. This working point has a target identification efficiency of 93 % for electrons with transverse momenta of about 40 GeV [122, 231].

Electron candidates with high activity in the vicinity of the electron are rejected by making a loose requirement on both calorimeter- and track-based isolation variables [122]. The electron selection efficiency of the isolation requirement exceeds 95 % for electrons with $p_T > 20$ GeV, quickly approaching efficiencies close to 100 % at higher transverse momenta [122]. This requirement is inverted as part of the fake-$\tau_{\text{had-vis}}$ background estimation in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel to provide a control region (CR) enhanced in multi-jet events.

Muons are required to have $p_T > 7$ GeV and pass the loose identification working point (cf. Section 3.3.4). The muon identification efficiency of the loose working point exceeds 97 % for muons with $p_T > 10$ GeV in simulated $t\bar{t}$ events [124]. In addition, muons are required to pass the loose isolation working point based on isolation requirements combining information

[3] The fast detector simulation was only validated for scalar resonances with $m_X \leq 1\,000$ GeV.

on charged and neutral particle signatures using a particle flow approach [124]. The muon selection efficiency of the isolation working point exceeds 97 % for muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ in simulated $t\bar{t}$ events [124]. The isolation requirement is inverted to provide a CR for the fake-$\tau_{\text{had-vis}}$ background estimation in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

$\tau_{\text{had-vis}}$ are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$. In addition, $\tau_{\text{had-vis}}$ candidates in the transition region between the barrel and end-cap calorimeters, $1.37 < |\eta| < 1.52$, are rejected. Only $\tau_{\text{had-vis}}$ candidates with one or three associated charged-particle tracks are considered. The total electric charge of these tracks is required to be ± 1.

A BDT-based electron veto algorithm is applied to 1-prong $\tau_{\text{had-vis}}$ candidates to reject cases where electrons are reconstructed as 1-prong $\tau_{\text{had-vis}}$. This algorithm has an efficiency of 95 % in selecting $\tau_{\text{had-vis}}$ candidates originating from hadronic τ-lepton decays.

The loose working point of the RNN-based tau identification algorithm introduced in Chapter 5 is used to reject $\tau_{\text{had-vis}}$ candidates originating from quark- and gluon-initiated jets. This working point targets a true $\tau_{\text{had-vis}}$ selection efficiency of 85 % (75 %) in simulated $\gamma^* \rightarrow \tau_{\text{had}}\tau_{\text{had}}$ events for 1-prong (3-prong) $\tau_{\text{had-vis}}$.

Anti-$\tau_{\text{had-vis}}$ are used to define CRs for the fake-$\tau_{\text{had-vis}}$ background estimation. They follow the same selection criteria as $\tau_{\text{had-vis}}$ except for a partial inversion of the tau identification requirement. Specifically, $\tau_{\text{had-vis}}$ candidates are considered to be anti-$\tau_{\text{had-vis}}$ if they fail the loose tau identification working point but pass a very loose selection on the tau identification score of RNN score > 0.01.

Anti-$\tau_{\text{had-vis}}$ are only considered in events with fewer than two (one) $\tau_{\text{had-vis}}$ passing the selection criteria of the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{had}}$) channel. In this case, anti-$\tau_{\text{had-vis}}$ are randomly selected until the required number of (anti-)$\tau_{\text{had-vis}}$ candidates, two in the $\tau_{\text{had}}\tau_{\text{had}}$ and one in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, is achieved. Anti-$\tau_{\text{had-vis}}$ that are not selected by the random anti-$\tau_{\text{had-vis}}$ selection are removed as part of an overlap removal procedure that is discussed in Section 6.2.1.

Jets are reconstructed using the anti-k_t algorithm with a radius parameter of $R = 0.4$. The inputs to the jet algorithm are provided by the particle flow reconstruction algorithm previously described in Section 3.3.5.

Jets in the central (forward) region of the detector defined by $|\eta| < 2.5$ ($2.5 < |\eta| < 4.5$) are required to fulfil $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ ($p_{\mathrm{T}} > 30\,\mathrm{GeV}$). Jets originating from pile-up are suppressed by ensuring that all jets pass Jet Vertex Tagger (JVT) requirements. In the central region, jets have to pass the tight JVT working point [232]. Jets in the forward regions are required to fulfil the loose working point of a dedicated JVT for forward jets [233, 234].

A multivariate b-tagging algorithm is applied to all jets in the central region to identify jets originating from b quarks. The DL1r algorithm [129] is used, which combines the scores of several low-level taggers that exploit the signature of displaced decays of b hadrons using information on the track impact parameters (IP2D, IP3D & RNNIP) and secondary vertices (SV1 & JetFitter). A b-tagging working point with an efficiency of 77 % of correctly tagging b-jets in simulated $t\bar{t}$ events is used.

The missing transverse momentum ($\boldsymbol{p}_{\mathrm{T}}^{\text{miss}}$) is reconstructed using the object-based definition of $\boldsymbol{p}_{\mathrm{T}}^{\text{miss}}$ previously introduced in Section 3.3.7. The $\boldsymbol{p}_{\mathrm{T}}^{\text{miss}}$ reconstruction uses the definitions of

electrons, muons, $\tau_{\text{had-vis}}$, and jets introduced in this section. Selected anti-$\tau_{\text{had-vis}}$ are treated as $\tau_{\text{had-vis}}$ in the reconstruction of the missing tranverse momentum.

6.2.1 Overlap Removal

The object reconstruction and selection described previously does not ensure that detector signatures can be unambiguously assigned to exactly one reconstructed object. Any ambiguities are resolved by performing an overlap removal procedure. This procedure rejects objects that can be associated, either geometrically or by shared charged-particle tracks, with an object of higher priority. In the case of geometric matching, the association is performed on the basis of the angular distance

$$\Delta R_y = \sqrt{\Delta y^2 + \Delta\phi^2}\,,$$

where Δy and $\Delta\phi$ refers to the differences in rapidity and azimuthal angle of two objects, respectively.

A sequential procedure is used to resolve the overlaps of reconstructed physics objects. The procedure is summarised in Table 6.2 following recommendations developed by the ATLAS collaboration. The last three steps serve to resolve overlaps between jets and anti-$\tau_{\text{had-vis}}$. This procedure establishes the following priority between jets, $\tau_{\text{had-vis}}$, and selected anti-$\tau_{\text{had-vis}}$:

$$\tau_{\text{had-vis}} > b\text{-tagged jet} > \text{anti-}\tau_{\text{had-vis}} > \text{untagged jet}.$$

An alternative priority given by

$$b\text{-tagged jet} > \tau_{\text{had-vis}} > \text{anti-}\tau_{\text{had-vis}} > \text{untagged jet}$$

was also considered but found to reduce the signal acceptance when selecting events with 2 b-tagged jets due to the limited $\tau_{\text{had-vis}}$ rejection of the DL1r b-tagging algorithm at the 77 % efficiency working point. The alternative leads to a relative reduction in SM HH signal acceptance by 13 % (8 %) in the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{had}}$) channels. Therefore, $\tau_{\text{had-vis}}$ take precedence over b-tagged jets in the overlap removal procedure as a means to maximise the signal acceptance.

6.2.2 $H \to b\bar{b}$ Candidate Reconstruction

The $H \to b\bar{b}$ candidate is reconstructed using jets in the central region of the detector. In events with two b-tagged jets, the four-momentum of the Higgs boson candidate is reconstructed by summing the four-momenta of both b-tagged jets. In regions with fewer than two b-tagged jets, which are used as control and validation regions, the Higgs candidate four-momentum is reconstructed as the sum of four-momenta of either the b-tagged jet and the p_{T}-leading untagged jet (1 b-tag regions) or the sum of the two p_{T}-leading untagged jets (0 b-tag regions). Regions with more than two b-tagged jets are not considered in this search. The reconstructed Higgs boson candidate is used to calculate the invariant mass of the $b\bar{b}$-system, m_{bb}, which is an important variable in distinguishing signal from background events.

Corrections are applied to b-tagged jets in order to improve the scale and resolution of the b-jet momentum reconstruction. These corrections are applied in addition to the standard jet calibration [128], targeting the effects of out-of-cone radiation and semi-muonic b- or c-hadron decays inside the b-jet on the four-momentum of reconstructed jets. Two separate corrections, which were previously used by the ATLAS collaboration in Refs. [236–238], are applied:

Table 6.2: Summary of the sequential overlap removal algorithm with rows representing steps of the procedure. Steps are executed from top to bottom, rejecting objects in the "Reject" column in favour of objects in the "Accept" column if the condition is fulfilled.

Reject	Accept	Condition
e_1	e_2	e_1 and e_2 share the charged particle track and $p_{\mathrm{T}}(e_1) < p_{\mathrm{T}}(e_2)$.
(anti-)$\tau_{\text{had-vis}}$	e	$\Delta R_y < 0.2$ and e passes the loose likelihood-based electron identification.
(anti-)$\tau_{\text{had-vis}}$	μ	$\Delta R_y < 0.2$ and one of the two conditions: - $\tau_{\text{had-vis}}$ $p_{\mathrm{T}} \leq 50\,\text{GeV}$ and $p_{\mathrm{T}}(\mu) > 2\,\text{GeV}$ - $\tau_{\text{had-vis}}$ $p_{\mathrm{T}} > 50\,\text{GeV}$, $p_{\mathrm{T}}(\mu) > 2\,\text{GeV}$, and μ is a combined muon.
μ	e	μ is calorimeter-tagged and shares inner detector track with e.
e	μ	Both share the inner detector track.
jet	e	$\Delta R_y < 0.2$.
e	jet	$\Delta R_y < 0.4$.
jet	μ	The ID track of the muon is ghost-associated [137, 235] to the jet and the jet has fewer than three ghost-associated ID tracks with $p_{\mathrm{T}} > 500\,\text{MeV}$.
μ	jet	$\Delta R_y < 0.4$.
jet	$\tau_{\text{had-vis}}$	$\Delta R_y < 0.2$.
anti-$\tau_{\text{had-vis}}$	jet	$\Delta R_y < 0.2$ and jet is b-tagged.
jet	anti-$\tau_{\text{had-vis}}$	$\Delta R_y < 0.2$ and the anti-$\tau_{\text{had-vis}}$ is selected by the random anti-$\tau_{\text{had-vis}}$ selection.
anti-$\tau_{\text{had-vis}}$	jet	$\Delta R_y < 0.2$.

Muon-in-jet correction Muons produced in decays of b or c hadrons inside b-jets deposit little energy in the calorimeters, thus leading to an underestimation of the jet four-momentum. The muon-in-jet correction is applied when a muon with $p_{\mathrm{T}} > 4\,\text{GeV}$ passing the *medium* selection working point can be found within a variable-size cone centred on the jet axis. No requirements are applied on the isolation of muons entering the muon-in-jet correction. In case one or more muons fulfil the criteria, the muon closest to the jet axis is selected. The four-momentum of the selected muon is added to the four-momentum of the jet, after correcting for the energy it deposited in the calorimeters.

$p_{\mathrm{T}}^{\text{reco}}$ correction The $p_{\mathrm{T}}^{\text{reco}}$ correction accounts for out-of-cone radiation as well as momentum carried away by neutrinos produced in semi-leptonic decays of hadrons in the jet. It is derived using b-tagged jets in simulated $t\bar{t}$ events as a function of the p_{T} of the reconstructed jet, $p_{\mathrm{T}}^{\text{reco}}$, separately for jets containing a muon (according to the criteria of the muon-in-jet correction) and jets that do not contain a muon. The correction is applied as a multiplicative factor on the jet momentum scale and is derived as the ratio $p_{\mathrm{T}}^{\text{truth}}/p_{\mathrm{T}}^{\text{reco}}$ in bins of $p_{\mathrm{T}}^{\text{reco}}$, where $p_{\mathrm{T}}^{\text{truth}}$ is the transverse momentum of the truth jet, a jet constructed from generator-level particles, that is geometrically matched to the jet at reconstruction-level.

Both corrections are performed at the level of individual jets and do not use information about the remainder of the event.

The effect of these corrections on the reconstructed m_{bb} in simulated SM HH events in the $\tau_{\text{had}}\tau_{\text{had}}$ channel is depicted in Figure 6.1. The improvement in the b-jet momentum scale and resolution from

the muon-in-jet correction and the $p_{\mathrm{T}}^{\mathrm{reco}}$ correction lead to a relative reduction of the width of the Higgs boson mass peak in the m_{bb} distribution of about 13 % as measured by the full width at half maximum.

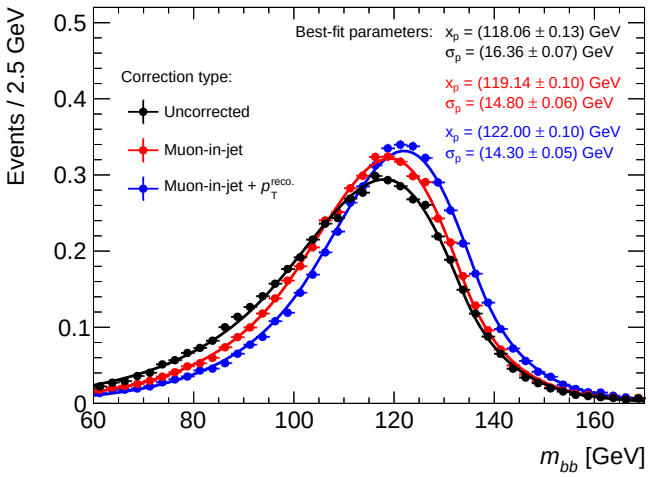


Figure 6.1: The effect of b-jet momentum corrections on the reconstructed m_{bb} in simulated SM HH events in the 2 b-tag region of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. Bukin functions [239] are fitted to the m_{bb} distributions. The location of the peak maximum is given by x_{p} and the full width at half maximum divided by $2\sqrt{2\ln 2}$ is given by σ_{p}. The uncorrected distribution is obtained using the standard jet calibration without further corrections for b-tagged jets.

6.2.3 $H \to \tau^+\tau^-$ Candidate Reconstruction

Reconstruction of the $H \to \tau^+\tau^-$ four-momentum is challenging due to the presence of at least two neutrinos originating from decays of τ leptons that cannot be measured directly. The starting point of the reconstruction are the visible decay products of τ-lepton decays. In the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, these are given by the two $\tau_{\text{had-vis}}$ candidates, while in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel the visible decay products are the $\tau_{\text{had-vis}}$ candidate and either an electron or a muon. To obtain the four-momentum of the di-τ system, additional information on the neutrinos produced from the τ-lepton decay is required. The allowed configurations of neutrinos from τ-lepton decays can be restricted using the visible decay products and kinematic constraints from the known mass of the τ lepton and the measurement of $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}$ under the assumption that its sole source are neutrinos produced from decays of τ leptons. However, the resulting system of kinematic equations is underdetermined leaving multiple degrees of freedom for the unobserved neutrinos.

The Missing Mass Calculator (MMC) technique [240] is used to estimate the four-momentum of the di-τ system and thus its invariant mass, $m_{\tau\tau}^{\mathrm{MMC}}$. The MMC uses the fact that not all configurations of neutrinos are equally likely and assigns a probability density to every configuration conditional on the τ-lepton decay mode and other properties of the visible τ-lepton decay products. In addition, the constraint on the total neutrino transverse momentum from the $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}$ measurement is relaxed to allow for errors within the experimental $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, the resolution being parameterised as a function

of the event activity[4] and jet multiplicity. The probability density functions used by the MMC are derived from simulation of resonant production of τ-lepton pairs.

For any single event, the MMC estimates the conditional distribution[5] of an observable of interest that depends on properties of the unobserved neutrinos, for example the mass of the di-τ system, using simulation. A sequence of kinematically allowed neutrino configurations is generated using Markov Chain Monte Carlo according to the conditional distribution of neutrino configurations. For every configuration, the value of the observable of interest is calculated and filled into a histogram, which serves as a binned approximation of the observable's conditional distribution. Finally, the most probable value of the observable is used as an event-level estimator.

Frequently, the most probable di-τ invariant mass is used as the estimator when the mass is of primary interest, such as in measurements of $H \to \tau^+\tau^-$ [241]. However, in searches for Higgs boson pair production it is desirable to estimate the total four-momentum of the di-τ system, which is required to calculate the invariant mass of the HH-system, m_{HH}. Instead of the most probable mass estimator, an estimator based on the most probable neutrino momenta is used, which can be used to reconstruct the four-momentum of the di-τ system.

A description of the technical implementation of the MMC used by the ATLAS collaboration is given in Ref. [242], which is used as the basis of the four-momentum reconstruction for $H \to \tau^+\tau^-$ candidates in this search.

6.3 Event Selection

Events consistent with the signature of $b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$ and $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ final states are selected. A loose event selection is applied that is largely determined by the triggers employed in this search. The discrimination of signal and background events is not the primary goal of the event selection but rather of a multivariate analysis that is introduced in Section 6.5.

All events considered in this analysis are required to have a reconstructed primary vertex. Moreover, events containing one or more jets that are classified as originating from non-collision backgrounds or calorimeter noise according to a loose jet cleaning working point [243] are rejected.

The search is divided into different channels depending on the decay mode of the τ-lepton pair and the type of trigger that selected an event. The $\tau_{\text{lep}}\tau_{\text{had}}$ channel targets semi-leptonic decay modes using single-lepton triggers (SLTs) and lepton-plus-$\tau_{\text{had-vis}}$ triggers (LTTs). Each trigger defines a corresponding sub-channel referred to as the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel, respectively. The $\tau_{\text{had}}\tau_{\text{had}}$ channel selects events with two $\tau_{\text{had-vis}}$ candidates using single-$\tau_{\text{had-vis}}$ triggers (STTs) and di-$\tau_{\text{had-vis}}$ triggers (DTTs). Events selected by STTs and DTTs are combined into a single channel.

Orthogonality between the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channel is ensured by selections on the number of electrons, muons, and $\tau_{\text{had-vis}}$. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, events are required to have exactly two $\tau_{\text{had-vis}}$ and events with electrons or muons are rejected. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, events are required to have exactly one $\tau_{\text{had-vis}}$ and exactly one electron or muon. Additionally, electrons (muons) are required to pass the tight (medium) identification working point to reduce backgrounds with non-prompt leptons or jets misidentified as electrons (muons).

[4] The event activity, also referred to as $\sum E_{\text{T}}$, is the scalar sum of transverse momenta of all hard objects produced in an event and the track soft term.

[5] The distribution is conditional on the observed properties of the event such as the reconstructed four-momenta of the visible τ-lepton decay products, the τ-lepton decay mode, $\boldsymbol{p}_{\text{T}}^{\text{miss}}$, $\sum E_{\text{T}}$, and the jet multiplicity.

Electrons, muons, and $\tau_{\text{had-vis}}$ have to be geometrically matched to the corresponding object at the HLT that fulfilled the criteria of the trigger that selected the event. This requirement is referred to as *trigger-matching*. Trigger-dependent p_{T} thresholds are applied to electrons, muons, and $\tau_{\text{had-vis}}$ to ensure that the triggers operate in regimes in which they are well-calibrated. The p_{T} thresholds of SLTs and STTs increased with increasing instantaneous luminosity of the LHC during Run 2. For LTTs and DTTs, the p_{T} thresholds on electrons, muons, and $\tau_{\text{had-vis}}$ remained constant during Run 2, the trigger-rates were instead limited by requiring additional jets at the L1 trigger.

An overview of the signal region (SR) event selection is given in Table 6.3. A more detailed description of the $\tau_{\text{had}}\tau_{\text{had}}$ channel trigger selection is given in Section 6.3.1. Further selections applied at event-level to define the SRs are discussed in Section 6.3.2.

Table 6.3: SR event selection for the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ SLT, and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel. Trigger-dependent thresholds are applied to the p_{T} of electrons, muons, and $\tau_{\text{had-vis}}$. Where applicable, the range of these thresholds is listed. Selections applied to p_{T} sub-leading objects are given in parenthesis. The trigger-dependent selections applied to $\tau_{\text{had-vis}}$ and jets in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are described in Section 6.3.1. Forward jets are not used for event selection purposes. The table is adapted from Ref. [192].

$\tau_{\text{had}}\tau_{\text{had}}$ channel		**$\tau_{\text{lep}}\tau_{\text{had}}$ channels**	
STT	**DTT**	**SLT**	**LTT**
e / μ selection			
No loose e or μ		Exactly one loose e or one loose μ	
		e is passes tight identification or μ passes medium identification and $\lvert\eta\rvert < 2.5$	
		$p_{\text{T}}(e) > 25$–$27\,\text{GeV}$ $p_{\text{T}}(\mu) > 21$–$27\,\text{GeV}$	$p_{\text{T}}(e) > 18\,\text{GeV}$ $p_{\text{T}}(\mu) > 15\,\text{GeV}$
			Lepton p_{T} below SLT threshold
$\tau_{\text{had-vis}}$ selection			
Exactly two loose $\tau_{\text{had-vis}}$		Exactly one loose $\tau_{\text{had-vis}}$	
		$\lvert\eta\rvert < 2.3$	
$p_{\text{T}} > 100$–$180\,(25)\,\text{GeV}$	$p_{\text{T}} > 40\,(30)\,\text{GeV}$		$p_{\text{T}} > 30\,\text{GeV}$
Central jet selection ($\lvert\eta\rvert < 2.5$)			
≥ 2 jets			
≥ 1 jet with $p_{\text{T}} > 45\,\text{GeV}$	Trigger-dependent	≥ 1 jet with $p_{\text{T}} > 45\,\text{GeV}$	Trigger-dependent
Event-level selection			
Event is selected by a trigger and trigger requirements (cf. Section 6.3.1) are fulfilled			
Exactly 2 b-tagged jets			
Opposite-sign electric charge between $\tau_{\text{had-vis}}$ and e / μ / $\tau_{\text{had-vis}}$			
$m_{\tau\tau}^{\text{MMC}} > 60\,\text{GeV}$			
		$m_{bb} < 150\,\text{GeV}$	

6.3.1 Trigger Selection in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

The triggers used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel depend on the data-taking period due to the changing instantaneous luminosity and introduction of new trigger algorithms during Run 2 of the LHC. A list of triggers considered in this search is given in Table 6.4. Further explanations of the items in the table are given in the remainder of this section.

As Run 2 of the LHC progressed, several improvements were made to the HLT algorithms used in $\tau_{\text{had-vis}}$-triggers. Consequently, three different HLT chains are used in this analysis, the differences between them are described in the following:

`medium1_tracktwo` This chain is the primary HLT chain for $\tau_{\text{had-vis}}$-triggers in the data-taking period from 2015 to 2017 [244–246]. A brief summary based on Ref. [247] is given hereafter.

First, a purely calorimeter-based reconstruction of the $\tau_{\text{had-vis}}$ candidate is performed in the region of interest (ROI) provided by the L1 trigger. The topo-cluster algorithm is applied to the calorimeter cells in the ROI and the resulting topo-clusters are calibrated using the local hadronic calibration. The energy of $\tau_{\text{had-vis}}$ candidates is determined from clusters in a core region ($\Delta R < 0.2$) around the barycentre of topo-cluster energy in the ROI. Subsequently, a $\tau_{\text{had-vis}}$-specific energy calibration is applied. The HLT threshold on the p_{T} of $\tau_{\text{had-vis}}$ candidates is applied after these steps.

Second, a two-stage tracking approach using *fast tracking*, instead of the more time-consuming *precision tracking* algorithm used during offline event reconstruction (cf. Section 3.3.1) and at later stages of the trigger, is employed [247–249]. The first stage performs tracking in a narrow region surrounding the $\tau_{\text{had-vis}}$ candidate close to the calorimeters but extended over a large section of the beamline. The p_{T}-leading track resulting from this stage is used to narrow down the search space along the beamline by only considering the beamline section within $|\Delta z| < 10\,\text{mm}$ of this track for the second stage of tracking, thus allowing for an expansion of the search region close to the calorimeters. After the second stage of tracking, core and isolation tracks are defined according to the conditions $\Delta R(\text{track}, \tau_{\text{had-vis}}) < 0.2$ and $0.2 < \Delta R(\text{track}, \tau_{\text{had-vis}}) < 0.4$, respectively. Subsequently, track multiplicity selections are applied by requiring $\tau_{\text{had-vis}}$ candidates to have one to three core tracks and at most one isolation track.

Lastly, a $\tau_{\text{had-vis}}$ selection similar to the offline $\tau_{\text{had-vis}}$ selection is performed. The tracks resulting from the two-stage tracking are used as seeds for precision tracking. The precision tracks are then used to calculate discriminating variables used for tau identification. A BDT-based tau identification algorithm is applied and $\tau_{\text{had-vis}}$ are required to pass the *medium* working point.[6]

`medium1_tracktwoEF` This chain was introduced for data-taking in 2018 [175] and differs from the previous item by delaying the track multiplicity selections to a later stage of the HLT chain. Instead of counting tracks from the two-stage fast track finding, the track multiplicities are defined using precision tracks. This change circumvents a reduction in efficiency for 3-prong $\tau_{\text{had-vis}}$ in high pile-up conditions due to the fast track finding being more susceptible to fake tracks [175].

[6] The medium working point of the HLT tau identification applies less stringent requirements than the loose working point of the RNN-based tau identification algorithm applied during offline event reconstruction.

Table 6.4: Summary of STTs and DTTs used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The trigger naming scheme follows the conventions by the ATLAS collaboration. An explanation of these triggers is given in the main body. The "offline event selection" column summarises selections applied to $\tau_{\text{had-vis}}$ and jets during the offline event reconstruction to ensure that the triggers operate close to their trigger-efficiency plateau. These requirements are specified in terms of the p_{T} leading or sub-leading $\tau_{\text{had-vis}}$ (τ_0 or τ_1) and the p_{T} leading or sub-leading central jet (j_0 or j_1). All events selected by DTTs have to fulfil $p_{\text{T}}(\tau_0) > 40\,\text{GeV}$ and $p_{\text{T}}(\tau_1) > 30\,\text{GeV}$ (omitted in the table). The last column specifies the data-taking period in which a trigger was used.

HLT chain		**L1 trigger**	**Offline event selection**	**Period**
Single-$\tau_{\text{had-vis}}$ triggers				
tau80	medium1_tracktwo	TAU60	$p_{\text{T}}(\tau_0) > 100\,\text{GeV}$	15 A – 16 A
tau125	medium1_tracktwo	TAU60	$p_{\text{T}}(\tau_0) > 140\,\text{GeV}$	16 B – 16 D3
tau160	medium1_tracktwo	TAU60	$p_{\text{T}}(\tau_0) > 180\,\text{GeV}$	16 D4 – 17 B4
tau160	medium1_tracktwo	TAU100	$p_{\text{T}}(\tau_0) > 180\,\text{GeV}$	17 B5 – 17 N
tau160	medium1_tracktwoEF	TAU100	$p_{\text{T}}(\tau_0) > 180\,\text{GeV}$	18 A – 18 R
tau160	mediumRNN_tracktwoMVA	TAU100	$p_{\text{T}}(\tau_0) > 180\,\text{GeV}$	18 K – 18 R
Di-$\tau_{\text{had-vis}}$ triggers				
tau35 + tau25	medium1_tracktwo	TAU20IM_2TAU12IM	$p_{\text{T}}(j_0) > 80\,\text{GeV}$	15 A – 15 J
tau35 + tau25	medium1_tracktwo	TAU20IM_2TAU12IM_J25_2J20_3J12	$p_{\text{T}}(j_0) > 80\,\text{GeV}$	16 A – 17 B4
tau35 + tau25	medium1_tracktwo	TAU20IM_2TAU12IM_4J12	$p_{\text{T}}(j_1) > 45\,\text{GeV}$	17 A – 17 N
tau35 + tau25	medium1_tracktwo	DR-TAU20ITAU12I-J25	$p_{\text{T}}(j_0) > 80\,\text{GeV}$, $\Delta R(\tau_0, \tau_1) < 2.5$	17 B5 – 17 N
tau35 + tau25	medium1_tracktwoEF	TAU20IM_2TAU12IM_4J12.0ETA23	$p_{\text{T}}(j_1) > 45\,\text{GeV}$	18 A – 18 R
tau35 + tau25	medium1_tracktwoEF	DR-TAU20ITAU12I-J25	$p_{\text{T}}(j_0) > 80\,\text{GeV}$, $\Delta R(\tau_0, \tau_1) < 2.5$	18 A – 18 R
tau35 + tau25	mediumRNN_tracktwoMVA	TAU20IM_2TAU12IM_4J12.0ETA23	$p_{\text{T}}(j_1) > 45\,\text{GeV}$	18 K – 18 R
tau35 + tau25	mediumRNN_tracktwoMVA	DR-TAU20ITAU12I-J25	$p_{\text{T}}(j_0) > 80\,\text{GeV}$, $\Delta R(\tau_0, \tau_1) < 2.5$	18 K – 18 R

mediumRNN_tracktwoMVA This chain started operation in period K of 2018 data-taking [175]. The integrated luminosity from period K to the end of Run 2 corresponds to about 37 fb^{-1}. Several changes were implemented on top of the `medium1_tracktwoEF` chain.

First, the $\tau_{\text{had-vis}}$ energy calibrations as part of the calorimeter-based $\tau_{\text{had-vis}}$ reconstruction are replaced by multivariate methods. Second, the HLT tau identification is replaced by a method based on the RNN tau identification algorithm introduced in Chapter 5. The improved background rejection of this algorithm allowed relaxing the track multiplicity requirements from 1–3 precision tracks to 0–3 precision tracks in the core region. Selecting $\tau_{\text{had-vis}}$ candidates without precision tracks in the core region recovers cases in which the fast track finding does not yield good quality seeds for the precision tracking. After the offline event reconstruction, a fraction of these events can be correctly reconstructed, thus improving the selection efficiency of the trigger.

The `mediumRNN_tracktwoMVA` HLT chain is intended to be used in conjunction with the `medium1_tracktwoEF` chain by combining both using a logical *or*. Calibrations for this combination are provided by the ATLAS collaboration.

STT Selection

The STTs used in this analysis have varying p_{T} thresholds applied to $\tau_{\text{had-vis}}$ candidates depending on the data-taking period. At the HLT, these thresholds range from 80 GeV to 160 GeV. An event is considered to pass the STT selection if the following conditions are true: the event was selected by an STT, it has at least one $\tau_{\text{had-vis}}$ candidate geometrically matched to a $\tau_{\text{had-vis}}$ at the HLT ($\Delta R < 0.2$), and the p_{T} of the trigger-matched $\tau_{\text{had-vis}}$ candidate exceeds the HLT p_{T} threshold by 15 to 20 GeV. The exact offline event selection requirements are given in Table 6.4.

DTT Selection

The $\tau_{\text{had-vis}}$ p_{T} requirements of the DTTs used in this search remain unchanged throughout Run 2 of the LHC. At the HLT, the p_{T} thresholds on the leading and sub-leading $\tau_{\text{had-vis}}$ candidate are 35 GeV and 25 GeV, respectively. Events are considered as possible DTT events if they fulfil the following requirements: First, both $\tau_{\text{had-vis}}$ candidates from the offline event reconstruction have to be trigger-matched. Second, the trigger-matched $\tau_{\text{had-vis}}$ have to exceed the corresponding HLT p_{T} threshold by at least 5 GeV.

The primary limitations of DTTs arise at the L1 trigger, requiring changes in L1 seeds as the instantaneous luminosity increased during Run 2 data-taking. This is reflected in the DTT trigger selection described in the following:

2015 data-taking period The DTT used in 2015 (cf. Table 6.4) had no additional requirements beyond two isolated $\tau_{\text{had-vis}}$ at the L1 trigger. However, a requirement of $p_{\text{T}} > 80$ GeV is applied to the p_{T}-leading central jet from the offline event reconstruction. This selection is not strictly necessary and is applied to unify the selection with the DTT used in 2016.

2016 data-taking period Requirements on the presence of additional jets at the L1 trigger were introduced in 2016 to limit the trigger-rate of DTTs. Three jets are required, two of which are overlapping with $\tau_{\text{had-vis}}$ ROIs since no disambiguation between $\tau_{\text{had-vis}}$ and jets is performed

at the L1 trigger.[7] The E_T thresholds applied to these jets are 25 GeV, 20 GeV, and 12 GeV, the two lowest thresholds matching the $\tau_{\text{had-vis}}$ requirements of the L1 trigger (cf. Table 6.4). The effective E_T threshold on the jet not overlapping with the $\tau_{\text{had-vis}}$ ROIs can be lower than 25 GeV if a $\tau_{\text{had-vis}}$ candidate exists that is also reconstructed as a jet with $E_T > 25$ GeV. Nevertheless, a requirement of $p_T > 80$ GeV is applied to the p_T-leading central jet from the offline event reconstruction to ensure that the $E_T > 25$ GeV L1 jet trigger operates close to its trigger-efficiency plateau.

2017 data-taking period Two DTTs based on different L1 seeds are used for data recorded in 2017 (cf. Table 6.4). If the sub-leading central jet of the event fulfils $p_T > 45$ GeV, then the DTT based on the `TAU20IM_2TAU12IM_4J12` L1 seed is used. This trigger requires two additional jets at the L1 trigger with $E_T > 12$ GeV. The p_T threshold of 45 GeV ensures that the L1 jet trigger operates close to its trigger-efficiency plateau. This trigger is hereafter referred to as the 4J12 DTT. If the condition for the 4J12 DTT is not fulfilled but the event has a central jet with $p_T > 80$ GeV and an angular distance between the $\tau_{\text{had-vis}}$ candidates of $\Delta R(\tau, \tau) < 2.5$, then the DTT based on the `DR-TAU20ITAU12I-J25` L1 seed is considered. This trigger is based on the L1 topological trigger system [250], which disambiguates jet and $\tau_{\text{had-vis}}$ ROIs and applies other topological requirements. At the L1 trigger, a jet with $E_T > 25$ GeV that is not overlapping with a $\tau_{\text{had-vis}}$ ROI is required. Additionally, the $\tau_{\text{had-vis}}$ ROIs at the L1 trigger have to fulfil $\Delta R(\tau, \tau) < 2.8$. This trigger is referred to as the L1Topo DTT.

The 4J12 DTT is introduced to improve the acceptance of signal events with small HH invariant masses, for example events from resonant HH production via low-mass resonances. The L1Topo DTT has limited acceptance for such events due to the high p_T thresholds on the leading jet and the $\Delta R(\tau, \tau)$ requirement. For resonances with masses larger than 325 GeV and for SM HH production, the improvement in signal acceptance from including the 4J12 DTT is negligible.

2018 data-taking period New HLT algorithms were introduced for $\tau_{\text{had-vis}}$ triggers in 2018, updating the existing L1Topo- and 4J12-based DTTs. In addition, the L1 seed of the 4J12 DTT was changed to require at least four jets with $E_T > 12$ GeV in $|\eta| < 2.3$. This change introduced a mismatch in the selection of $\tau_{\text{had-vis}}$ ($|\eta| < 2.5$) and jet ROIs ($|\eta| < 2.3$) at the L1 trigger. As a result, the jet multiplicity requirement was not specified as intended, potentially requiring more than two additional jets if a $\tau_{\text{had-vis}}$ is reconstructed in $2.3 < |\eta| < 2.5$. This mismatch lead to a relative reduction in trigger efficiency for signal processes of about 5 % compared to the 4J12 DTT without the $|\eta| < 2.3$ requirement on jets. This issue is resolved in the trigger menu of the ATLAS experiment for Run 3 of the LHC.

Summary of the Trigger Selection

A flowchart summarising the trigger selection is depicted in Figure 6.2. Two features of the trigger selection are not illustrated in the figure. First, the L1Topo DTT only started operation with period B5 in 2017. In the intermittent period where L1Topo was not available, the $\tau_{\text{had-vis}}$ trigger chain based on `TAU20IM_2TAU12IM_J25_2J20_3J12` is used instead. The $\Delta R(\tau, \tau) < 2.5$ selection necessary for the L1Topo DTT is still applied in this case. Second, for three runs in 2017 the L1 topological trigger system was disabled due to issues with the trigger firmware. For these runs the DTT chain based on

[7] At the L1 trigger, a $\tau_{\text{had-vis}}$ candidate with transverse energy of E_T^τ is also reconstructed as a jet ROI with $E_T^{\text{jet}} \geq E_T^\tau$.

TAU20IM_2TAU12IM_J25_2J20_3J12 was used as a backup. This chain was almost unprescaled during the affected runs, leading to a loss of about 60 pb^{-1} of integrated luminosity in the L1Topo DTT category due to the prescale.

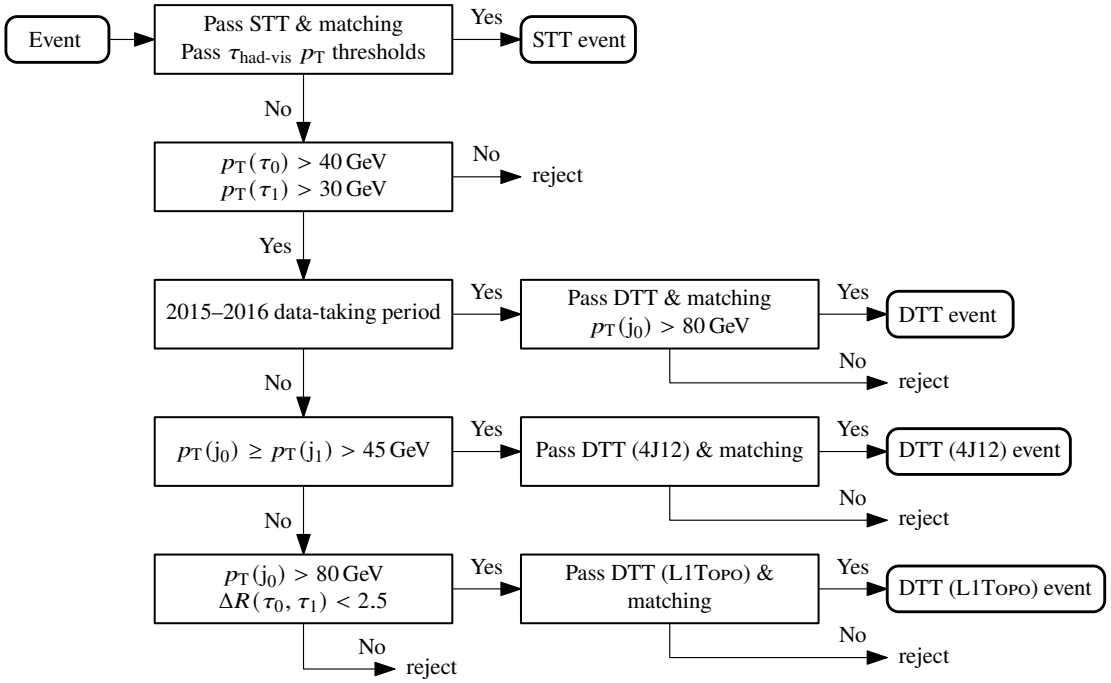


Figure 6.2: Flowchart of the $\tau_{\text{had}}\tau_{\text{had}}$ channel trigger selection. The leading and sub-leading $\tau_{\text{had-vis}}$ candidate (jet) from the offline event reconstruction are abbreviated as τ_0 and τ_1 (j_0 and j_1), respectively.

The signal selection efficiency of the trigger selection varies with the considered signal hypothesis. After an event pre-selection,[8] the probability of an event to be selected by the trigger, to pass the trigger-matching requirement, and to pass the trigger-dependent $\tau_{\text{had-vis}}$ p_{T} thresholds is about 40 % for the SM HH signal. For resonant production of Higgs boson pairs, the efficiency ranges from 30 % for $m_X = 300\,\text{GeV}$ up to 75 % for $m_X = 1\,600\,\text{GeV}$.

6.3.2 Signal Region Event Selection

Events passing the electron, muon, $\tau_{\text{had-vis}}$, and jet selections (cf. Table 6.3) as well as the trigger selection are used to define the SRs. Only regions with exactly two b-tagged jets (2 b-tag regions) are considered as SRs. Regions with fewer b-tagged jets are dominated by background processes and thus would not improve the signal sensitivity significantly. Instead, 0 and 1 b-tag regions are used as control and validation regions.

The electric charge of the electron, muon, or $\tau_{\text{had-vis}}$ candidate has to be reconstructed with opposite sign (OS) with respect to the charge of the other $\tau_{\text{had-vis}}$ candidate in the event. Events from processes producing τ-lepton pairs, such as the signal processes, Z + jets, $H \to \tau^+\tau^-$, and $t\bar{t}$, are expected to be reconstructed with OS electric charge. Events with same-sign (SS) electric charge of the visible

[8] Electron/muon veto and a $\tau_{\text{had-vis}}$ candidate pre-selection. Details are given in Table 6.7 of Section 6.3.2.

τ-lepton decay products predominately originate from events in which a quark- or gluon-initiated jet is misidentified as a $\tau_{\text{had-vis}}$. Therefore, the SS region is used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel for the estimation of multi-jet backgrounds.

All events considered in this search are required to successfully pass the di-τ mass reconstruction using the MMC. Drell–Yan processes producing τ-lepton pairs with low invariant mass are rejected by requiring $m_{\tau\tau}^{\text{MMC}} > 60\,\text{GeV}$. In addition, the SRs of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channel only consider events fulfilling $m_{bb} < 150\,\text{GeV}$. This selection allows defining an orthogonal $t\bar{t}$ CR by inverting the selection on m_{bb}. This region is used for measurements related to the fake-$\tau_{\text{had-vis}}$ background estimation.

The expected event yields after the SR selection in all three channels are summarised in Table 6.5. The bulk of events entering the SRs are from top-quark backgrounds ($t\bar{t}$ and single-t), Z + jets, and backgrounds in which a quark- or gluon-initiated jet is reconstructed as a $\tau_{\text{had-vis}}$ (jet $\rightarrow$ fake $\tau_{\text{had-vis}}$).

Table 6.5: Event yields in the SRs prior to the fit. The expected yields are shown including all statistical and systematic uncertainties. The SM HH event yields are given for the SM expectation. The fake-$\tau_{\text{had-vis}}$ background estimation technique employed in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels does not distinguish between different sources of fake $\tau_{\text{had-vis}}$. The category "other backgrounds" combines minor contributions from $Z \rightarrow \tau^+\tau^- + (bl, cl, ll)$, $Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$, W + jets, diboson and $t\bar{t}V$. The background estimation and systematic uncertainties are discussed in detail in Sections 6.4 and 6.6.

	Signal region event yield		
Process	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
SM HH (ggF)	5.4 ± 1.1	5.9 ± 1.2	1.42 ± 0.29
SM HH (VBF)	0.167 ± 0.022	0.200 ± 0.027	0.0547 ± 0.0066
SM HH (ggF + VBF)	5.6 ± 1.1	6.1 ± 1.2	1.47 ± 0.29
Top-quark	3 850 ± 330	65 300 ± 5 600	4 400 ± 460
$Z \rightarrow \tau^+\tau^- + (bb, bc, cc)$	1 200 ± 210	1 210 ± 130	406 ± 67
Single Higgs boson	74 ± 15	154 ± 20	24.4 ± 5.0
Jet $\rightarrow$ fake $\tau_{\text{had-vis}}$ (combined)	–	33 900 ± 6 500	1 750 ± 510
Jet $\rightarrow$ fake $\tau_{\text{had-vis}}$ (multi-jet)	1 350 ± 150	–	–
Jet $\rightarrow$ fake $\tau_{\text{had-vis}}$ ($t\bar{t}$)	2 490 ± 330	–	–
Other backgrounds	228 ± 42	1 090 ± 210	119 ± 21
Total background	9 200 ± 640	101 700 ± 8 600	6 700 ± 700
Observed data	8 380	98 456	6 351

The acceptance times efficiency ($\mathcal{A} \times \varepsilon$) for SM HH events in the SRs is summarised in Table 6.6 for all three channels. Compared to the previously published result in this channel [184], this analysis shows an increase in the $\mathcal{A} \times \varepsilon$ for SM HH events by a factor of about 2 (1.5) for the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{had}}$) channel. This increase is a consequence of improved $\tau_{\text{had-vis}}$ reconstruction techniques and loosened identification requirements for $\tau_{\text{had-vis}}$ and b-tagged jets. The reason for loosening the identification criteria is twofold: First, the $\tau_{\text{had-vis}}$ identification and b-tagging algorithms were significantly improved [11, 129], yielding reduced mistag rates at working points with tagging efficiencies similar to the working points used previously. Consequently, the identification criteria can be loosened while maintaining background rates similar to the ones of the previous analysis. Second, Higgs boson pair production has distinct kinematic features that can be used to distinguish

signal events and events containing jets that are misidentified as b-jets or $\tau_{\text{had-vis}}$, for example, using multivariate methods. When exploiting these features, the signal sensitivity of this analysis improves with a less stringent object selection.

Table 6.6: Acceptance times efficiency for SM HH events in the SRs. The $\mathcal{A} \times \varepsilon$ is given as the fraction of selected events with respect to all generated $pp \to HH \to b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ ($pp \to HH \to b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$) events in the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{had}}$) channel. The signal acceptance of the previous iteration of the $HH \to b\bar{b}\tau^+\tau^-$ search by the ATLAS collaboration using 36.1 fb^{-1} of pp collision data is shown in the last row. †: The $\mathcal{A} \times \varepsilon$ is given for the combination of $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channel.

	Acceptance × Efficiency / %		
Process	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
SM HH (ggF)	4.1	4.1	0.99
SM HH (VBF)	2.3	2.5	0.68
SM HH (ggF + VBF)	4.0	4.0	0.97
SM HH (ggF) in Ref. [184]	1.9	3.2†	

The majority of the increase in the $\mathcal{A} \times \varepsilon$ for SM HH events with respect to Ref. [184] can be explained by the following improvements:

$\tau_{\text{had-vis}}$ track association The introduction of a multivariate method for $\tau_{\text{had-vis}}$ track association in Ref. [136], superseding a cut-based method, lead to an overall improvement in the $\tau_{\text{had-vis}}$ selection efficiency due to the $N_{\text{tracks}} \in \{1, 3\}$ requirement on $\tau_{\text{had-vis}}$. The relative improvement in efficiency with respect to the cut-based method is 20 to 30 % for 1-prong $\tau_{\text{had-vis}}$ in the $\tau_{\text{had-vis}}$ p_{T} range relevant for the SM HH search. The efficiency for 3-prong $\tau_{\text{had-vis}}$ remains largely unchanged for $\tau_{\text{had-vis}}$ p_{T} below 100 GeV compared to the cut-based method. However, for 3-prong $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 100$ GeV, the track association efficiency is reduced by up to 10 %.

Tau identification With the introduction of the RNN-based tau identification (cf. Chapter 5), the rejection of $\tau_{\text{had-vis}}$ candidates originating from quark- or gluon-initiated jets increased by 40 to 80 % compared to the BDT-based algorithm. This allows for a change of the identification working point from the *medium BDT* to the *loose RNN* working point while maintaining similar rates of fake-$\tau_{\text{had-vis}}$ backgrounds. As a result, a relative increase in $\tau_{\text{had-vis}}$ identification efficiency of about 13 % (25 %) is achieved for 1-prong (3-prong) $\tau_{\text{had-vis}}$.

Additionally, an improved algorithm to reject 1-prong $\tau_{\text{had-vis}}$ candidates originating from electrons based on a BDT discriminant (e-veto) is used in this search. This algorithm improves the tagging efficiency by 4 to 5 % for 1-prong $\tau_{\text{had-vis}}$ compared to the previous method. No e-veto is applied to 3-prong $\tau_{\text{had-vis}}$ candidates.

***b*-tagging** Improved b-tagging algorithms allow for the use of working points with higher b-jet tagging efficiency while limiting the increase in background due to mistagged jets. Previously, the MV2c10 tagger [251] with a target efficiency of 70 % for b-jets from $t\bar{t}$ was used. This tagger is replaced by the DL1r tagger [129] operating at the the 77 % efficiency working point. Consequently, a 10 % relative improvement in b-jet tagging efficiency is expected.

These object-level efficiency improvements compound when considering the efficiencies at event-level and explain the majority of the increase in $\mathcal{A} \times \varepsilon$ for SM HH events with respect to the previous search in this channel.

The $\mathcal{A} \times \varepsilon$ of the SR selection for signals from resonant HH production are shown in Figure 6.3 as a function of the resonance mass. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the acceptance of signal events from decays of low-mass resonances is limited by the trigger selection. The signal acceptance increases quickly with increasing m_X, reaching a maximum of about 10 % at $m_X = 1\,000$ GeV. The $\mathcal{A} \times \varepsilon$ decreases for resonances with masses greater than 1 000 GeV due to the Higgs boson decay products becoming increasingly collimated, at which point the object reconstruction employed in this search becomes less effective at resolving the decay products. The $m_X > 1\,000$ GeV regime is typically covered by dedicated searches using specialised reconstruction techniques for highly boosted Higgs bosons (for a search in the $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ final state, see for example Ref. [252]).

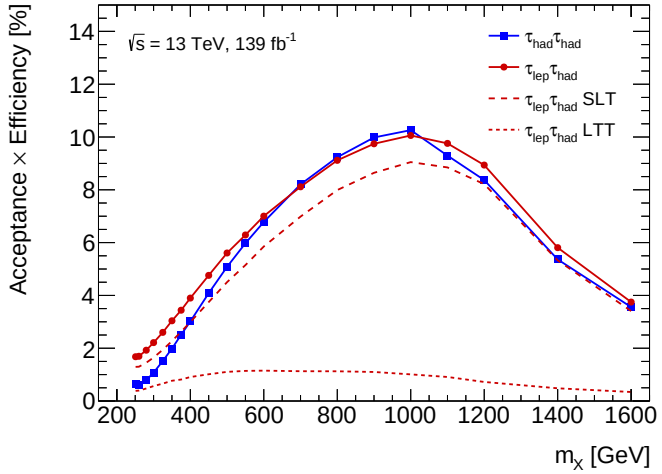


Figure 6.3: The $\mathcal{A} \times \varepsilon$ of the analysis selection for events from scalar resonances decaying into Higgs boson pairs. It is shown as a function of the resonance mass, m_X, and separately for the three channels as well as the combination of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channel. The $\mathcal{A} \times \varepsilon$ is given as the fraction of selected events with respect to all generated $pp \to X \to b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ ($pp \to X \to b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$) events in the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{had}}$) channel. The figure is adapted from Ref. [192].

Lastly, the signal event yield after different steps of the SR event selection of the $\tau_{\text{had}}\tau_{\text{had}}$ channel is shown in Table 6.7. Most events are lost due to the trigger selection, the $\tau_{\text{had-vis}}$ candidate selection, or the requirement of two b-tagged jets in the SR.

6.4 Background Estimation

The dominant background processes in the search for Higgs boson pair production in $b\bar{b}\tau^+\tau^-$ final states are Z + jets, $t\bar{t}$, and backgrounds with quark- or gluon-initiated jets that are misidentified as $\tau_{\text{had-vis}}$ (fake-$\tau_{\text{had-vis}}$ backgrounds). Minor backgrounds originate from single-top-quark, $t\bar{t}V$, diboson, and single-Higgs-boson production. The single-Higgs-boson production modes considered in this search are ggF, VBF, VH and $t\bar{t}H$ for $H \to \tau^+\tau^-$; VH and $t\bar{t}H$ for $H \to b\bar{b}$. Single-Higgs-boson production via $b\bar{b}H$ is found to be negligible.

Table 6.7: Event yields after different selection steps in the $\tau_{\text{had}}\tau_{\text{had}}$ channel for the SM HH signal and four exemplary signals from decays of scalar resonances. The expected number of events are normalised using the cross sections predicted by the SM for the SM HH production and using $\sigma(pp \to X \to HH) = 10\,\text{fb}$ for the $X \to HH$ signals.

	Expected number of events					
Selection step	SM HH (ggF)	SM HH (VBF)	$X(300\,\text{GeV})$	$X(500\,\text{GeV})$	$X(1\,000\,\text{GeV})$	$X(1\,600\,\text{GeV})$
HH	4 310	240	1 390	1 390	1 390	1 390
$HH \to b\bar{b}\tau^+\tau^-$	315	17.5	102	102	102	102
$HH \to b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$	132	7.36	42.6	42.6	42.6	42.6
Generator filters†	103	5.00	26.5	32.6	37.5	39.5
Derivation skim*	88.3	4.24	21.7	28.4	33.3	23.6
Electron and muon veto	84.4	4.03	20.9	27.5	31.8	22.0
$\tau_{\text{had-vis}}$ candidate pre-selection‡	40.4	1.89	9.25	13.6	17.5	10.2
Trigger, trigger-matching, and $\tau_{\text{had-vis}}$ p_{T} cuts	17.3	0.744	2.79	6.56	12.4	7.70
≥ 2 central jets	16.0	0.641	2.52	6.17	12.1	7.35
2 b-tagged jets	7.47	0.265	1.04	2.79	5.96	2.30
2 $\tau_{\text{had-vis}}$ candidates passing loose ID	6.26	0.221	0.833	2.37	4.85	1.75
OS $\tau_{\text{had-vis}}$ electric charge	6.18	0.217	0.822	2.34	4.76	1.71
Sub-leading $\tau_{\text{had-vis}}$ $p_{\text{T}} > 25\,\text{GeV}$	6.11	0.215	0.822	2.31	4.61	1.64
$m_{\tau\tau}^{\text{MMC}} > 60\,\text{GeV}$	5.98	0.210	0.814	2.27	4.44	1.55
Jet p_{T} and $\Delta R(\tau,\tau)$ cuts for triggers	5.40	0.167	0.459	2.17	4.37	1.52

†: Two $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 35\,\text{GeV}$ (15 GeV) in $|\eta| < 3.0$. For SM HH (VBF) the $\tau_{\text{had-vis}}$ p_{T} thresholds are 25 GeV (15 GeV).

∗: At least two $\tau_{\text{had-vis}}$ candidates with $p_{\text{T}} > 33\,\text{GeV}$ (23 GeV), $1 \leq N_{\text{tracks}} + N_{\text{tracks}}^{\text{iso.}} \leq 8$, and one or more $\tau_{\text{had-vis}}$ passing the loose identification working point.

‡: At least two (anti-)$\tau_{\text{had-vis}}$ candidates passing the $\tau_{\text{had-vis}}$ selection (cf. Section 6.2) with a relaxed tau identification requirement of RNN score > 0.01.

The Z + jets and $t\bar{t}$ backgrounds are estimated using templates obtained from simulation, with their normalisations being determined by a simultaneous likelihood fit in all analysis channels. A CR enriched in Z boson production in association with jets from quarks of heavy flavour (Z + HF) is defined in Section 6.4.1, which provides constraints on the normalisation of this background. The normalisation of the $t\bar{t}$ background is constrained by the inclusion of the SR of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel and the Z + HF CR in the simultaneous fit, both having a large contribution of events from $t\bar{t}$ production.

Major fake-$\tau_{\text{had-vis}}$ backgrounds are estimated using (semi-)data-driven methods, while minor ones are estimated using simulation. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the primary sources of fake $\tau_{\text{had-vis}}$ are $t\bar{t}$ and multi-jet production for which separate estimation techniques are used. The fake-$\tau_{\text{had-vis}}$ background from $t\bar{t}$ ($t\bar{t}$ + fake-$\tau_{\text{had-vis}}$) is estimated using simulation after applying corrections of jet $\rightarrow \tau_{\text{had-vis}}$ misidentification efficiencies measured in a CR. This method is described in Section 6.4.2. The multi-jet background is estimated using a fully data-driven *fake factor method*, which is introduced in Section 6.4.3. Both methods were developed as part of this thesis and differ from the approach adopted in the previous publication in this channel [184]. Lastly, the estimation of the fake-$\tau_{\text{had-vis}}$ backgrounds in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels uses a *combined fake factor method* that simultaneously estimates fake-$\tau_{\text{had-vis}}$ backgrounds from multi-jet and $t\bar{t}$ processes. This method is briefly summarised in Section 6.4.4.

Minor background contributions are estimated using simulation and are normalised to the integrated luminosity of the pp collision dataset using cross section predictions from theory. These backgrounds are not discussed in detail in this section; however, theoretical uncertainties on the modelling of these processes using simulation are discussed in Section 6.6.2.

6.4.1 Associated Production of Z $\rightarrow \ell\ell$ with Quarks of Heavy Flavour

The Z + jets background is estimated using events simulated with Sherpa 2.2.1 [216] interfaced to the matrix element generators OpenLoops [253–255] and Comix [256] (cf. Table 6.1). This generator configuration merges hard-scatter matrix elements at NLO for final states with up to two partons and matrix elements at LO for up to four partons. Inclusive Z + jets cross sections at NNLO [219] are used for the (pre-fit) normalisation of the background prediction.

Simulated Z + jets events are categorised according to a generator-level flavour label assigned to the pair of selected b-jet candidates. Reconstructed jets are labelled as either b, c, or light (l), depending on the presence of heavy-flavour hadrons within a cone of $\Delta R < 0.3$ about the jet axis. If a b- or c-flavoured hadron with $p_{\text{T}} > 5\,\text{GeV}$ is located within the cone, the jet is labelled b or c, respectively. Jets that are not matched to any b- or c-flavoured hadrons are labelled as light. Six categories are defined based on the flavour label of the b-jet candidate pair: $Z + bb$, $Z + bc$, $Z + cc$, $Z + bl$, $Z + cl$, and $Z + ll$. Contributions from $Z + bb$, $Z + bc$, and $Z + cc$ are combined and collectively referred to as Z + HF. The remaining Z + jets events, i.e. events with at least one jet labelled as light, are referred to as Z + LF.

The requirement of having two b-tagged jets in the SRs leads to an enhancement of events from Z boson production in association with quarks of heavy flavour. The normalisation of this background is known to be underestimated by the generator configuration chosen for this search [257]. To control for this mismodelling, the normalisation of the Z + HF background is measured in a dedicated CR. This approach is adopted with a few modifications [258] from the previous publication in this channel [184], which built on findings from searches for VH ($H \rightarrow b\bar{b}$) production [236].

A dedicated CR is defined that targets the production of $Z \rightarrow \ell^+\ell^-$ ($\ell = e, \mu$) in association with

b-jets. The definitions of selected physics objects and event quality criteria remain the same as previously described in Sections 6.2 and 6.3. Events with same-flavour lepton pairs are recorded using single- and di-lepton triggers. Thresholds are applied to the p_{T} of electrons and muons after offline reconstruction to ensure that the triggers operate close to their trigger efficiency plateau. Depending on the run conditions of the LHC, the p_{T} thresholds range from 25 to 27 GeV for single-electron and 21 to 28 GeV for single-muon triggers. Events selected by di-electron triggers need to pass symmetric p_{T} thresholds on both the leading and sub-leading electron ranging from 13 to 25 GeV. Events selected by di-muon triggers are required to pass asymmetric thresholds of 19 to 24 GeV on the leading and 10 GeV on the sub-leading muon.

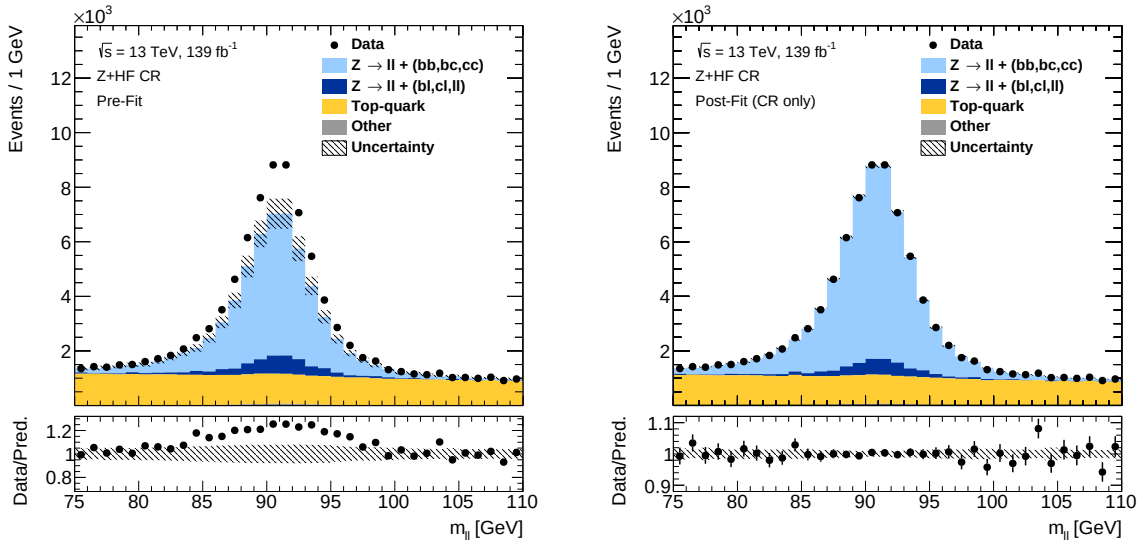
All events are required to be consistent with the decay of a Z boson into electrons or muons in association with b-jets. Leptons have to be of the same flavour with opposite electric charges and a di-lepton invariant mass, $m_{\ell\ell}$, falling into a Z boson mass window of $75\,\mathrm{GeV} < m_{\ell\ell} < 110\,\mathrm{GeV}$. Lastly, events are required to have exactly two b-tagged jets with an invariant mass fulfilling $m_{bb} \notin [40\,\mathrm{GeV}, 210\,\mathrm{GeV}]$. The m_{bb} requirement is necessary to ensure orthogonality with SRs of searches for Higgs boson pair production in $b\bar{b}\ell^+\ell^-$ ($\ell = e, \mu$) final states performed by the ATLAS collaboration. After the event selection, the electron and muon channels are combined for further analysis.

The pre-fit event yields in the Z + HF CR are given in Table 6.8. The majority of events in the CR originate from Z + HF or $t\bar{t}$ production. To distinguish between the Z + HF and $t\bar{t}$ contributions in the likelihood fit, the invariant di-lepton mass is used as a discriminant. The $m_{\ell\ell}$ distribution prior to the fit is depicted in Figure 6.4(a), showing the expected discrepancy between data and the pre-fit prediction.

Table 6.8: Event yields in the Z + HF CR before (pre-fit) and after (post-fit) a fit of the $m_{\ell\ell}$ distribution to CR data. The "Other" category summarises smaller backgrounds and largely consists of events from di-boson processes. The uncertainties on the event yields include all experimental and systematic uncertainties.

	Event yield	
Process	Pre-fit	Post-fit
$Z \to \ell^+\ell^-$ + HF	41 200 ± 3 200	55 700 ± 1 300
Top-quark	36 600 ± 1 400	35 260 ± 370
$Z \to \ell^+\ell^-$ + LF	5 300 ± 1 800	4 500 ± 1 300
Other	541 ± 94	528 ± 90
Total prediction	83 600 ± 5 200	96 030 ± 320
Observed data	96 032	

The Z + HF CR is included in the simultaneous likelihood fit of SRs and CRs to provide constraints on the normalisation of the Z + HF background. Details on systematic uncertainties and the fit model are discussed in Sections 6.6 and 6.7. Restricting the fit to the CR yields estimates of the normalisation factors of 1.39 ± 0.12 and 0.97 ± 0.04 for Z + HF and $t\bar{t}$, respectively. The quoted normalisation factors include all statistical and systematic uncertainties. Table 6.8 and Figure 6.4(b) show the event yields and $m_{\ell\ell}$ distribution after the fit.



(a) Pre-fit

(b) Post-fit (Z + HF CR only)

Figure 6.4: Distribution of the invariant di-lepton mass for the combination of electron- and muon-channel in the Z + HF CR before (a) and after (b) performing a fit of the $m_{\ell\ell}$ distribution to CR data. All statistical and systematic uncertainties are included.

6.4.2 Fake-$\tau_{\text{had-vis}}$ Background from $t\bar{t}$ Production in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

Top-quark pair production events in which at least one $\tau_{\text{had-vis}}$ candidate originates from a misidentified quark- or gluon-initiated jet are the second-largest background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. A fraction of 85 % of events from this background stem from semi-leptonic decay modes of $t\bar{t}$. In these events, the selected $\tau_{\text{had-vis}}$ candidates consist of one true and one fake $\tau_{\text{had-vis}}$. The remaining 15 % are $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events with two fake $\tau_{\text{had-vis}}$. The $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background is estimated using simulation after applying data-driven corrections in the form of fake-$\tau_{\text{had-vis}}$ scale factors (SFs) measured in a CR enriched in $t\bar{t}$ events. The SFs correct the jet $\rightarrow \tau_{\text{had-vis}}$ misidentification efficiencies predicted by simulation, which are not centrally calibrated by the ATLAS collaboration due to their process dependency.

Before proceeding with the description of the method, differences between fake-$\tau_{\text{had-vis}}$ backgrounds from $t\bar{t}$ and multi-jet are highlighted that motivate the use of different background estimation techniques:

- The majority of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background events consist of only one fake $\tau_{\text{had-vis}}$, while in multi-jet events both candidates are fake $\tau_{\text{had-vis}}$.
- The probability that a quark- or gluon-initiated jet reconstructed as a $\tau_{\text{had-vis}}$ passes tau identification, also called the *fake rate*, depends on the type of parton that initiated the jet. In $t\bar{t}$ events, fake $\tau_{\text{had-vis}}$ predominately originate from quarks produced in decays of W bosons, which have larger fake rates than gluon-initiated jets. As a result, fake-$\tau_{\text{had-vis}}$ background estimation is inherently process dependent.
- In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, no suitable $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ CR can be defined that separates fake-$\tau_{\text{had-vis}}$ backgrounds from $t\bar{t}$ and multi-jet while maintaining sufficient statistical precision for background

estimation. This necessitates defining a CR in $\ell + \tau_{\text{had-vis}}$ ($\ell = e, \mu$) final states in which the multi-jet contribution can be neglected.

The main disadvantage of separately estimating the fake-$\tau_{\text{had-vis}}$ background from $t\bar{t}$ and multi-jet is the inflation of systematic uncertainties on the estimate of the total fake-$\tau_{\text{had-vis}}$ background compared to a combined approach.[9] However, these inflated uncertainties have little effect on the signal sensitivity for two reasons: First, the targeted signal processes have distinct kinematic properties that differentiate them from fake-$\tau_{\text{had-vis}}$ backgrounds. Second, the search is limited by statistical uncertainties as opposed to systematic ones.

Control Region Definition

The CR for the SF measurement (SF-CR) targets final states with an electron or muon and a $\tau_{\text{had-vis}}$ candidate. The region definition is based on the selections applied in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, where a similar CR is used for fake-$\tau_{\text{had-vis}}$ background estimation. Minor changes are applied to its definition to ensure consistency with the SR selection of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The most important selection criteria and differences are briefly summarised:

- The $\tau_{\text{had-vis}}$ selection is adapted to follow the selection of the $\tau_{\text{had}}\tau_{\text{had}}$ channel more closely by selecting candidates with $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 2.5$ (instead of $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.3$ in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels).
- All events are required to have exactly one $\tau_{\text{had-vis}}$ candidate passing tau identification, exactly one electron or muon passing their respective isolation and identification criteria, and exactly two b-tagged jets. In addition, only events passing the SLT selection are considered.
- The electron/muon and the $\tau_{\text{had-vis}}$ candidate are required to be reconstructed with electric charges of opposite sign.
- Orthogonality between the SF-CR and the SR of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel is ensured by requiring $m_{bb} > 150\,\text{GeV}$.

The dominant process populating the SF-CR is $t\bar{t}$ production, which is selected with a purity of 94 %. About 66 % of SF-CR events are from di-leptonic decay modes of $t\bar{t}$ that yield an electron/muon and a $\tau_{\text{had-vis}}$ in the final state. The production of $t\bar{t}$ events with $\tau_{\text{had-vis}}$ candidates originating from quark- or gluon-initiated jets, which is the process of interest for this measurement, constitutes 28 % of selected events. Minor backgrounds in this region are single-top-quark (4 %) and V + jets (2 %) production. The multi-jet background is assumed to be negligible. The distribution of transverse momenta and pseudorapidity of $\tau_{\text{had-vis}}$ candidates in the SF-CR prior to the fit is shown in Figure 6.5.

Jet $\rightarrow \tau_{\text{had-vis}}$ misidentification efficiencies depend on the identification requirements applied to $\tau_{\text{had-vis}}$ candidates. In this search, such requirements are imposed on $\tau_{\text{had-vis}}$ candidates at trigger-level and during offline event reconstruction. This two-stage selection of $\tau_{\text{had-vis}}$ candidates is taken into account by measuring separate sets of SFs for every relevant combination of tau identification requirements. One set of SFs is measured for fake $\tau_{\text{had-vis}}$ after offline tau identification but without

[9] An example of a combined approach is the combined fake factor method employed in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. This method is summarised in Section 6.4.4.

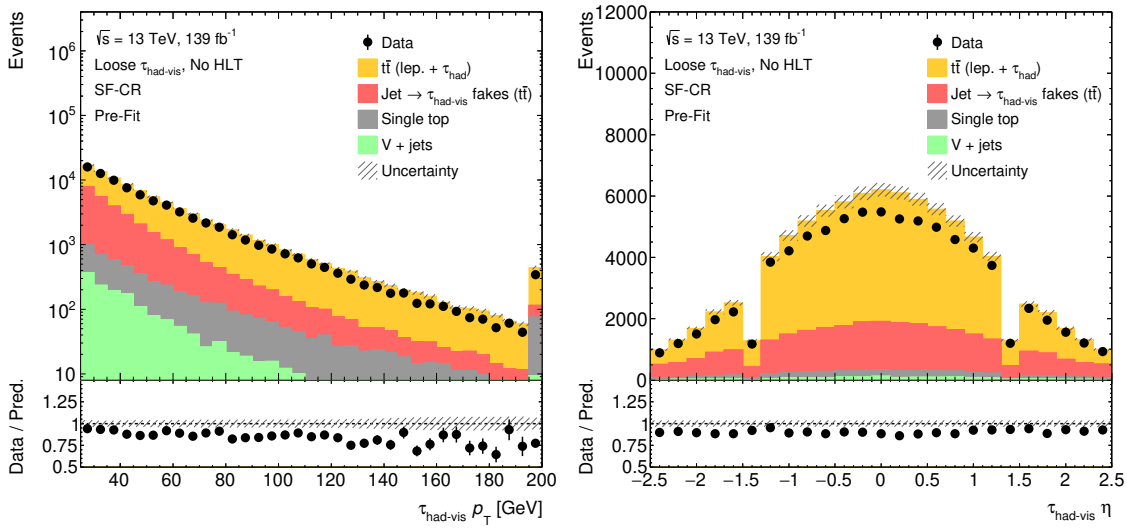


Figure 6.5: Distribution of $\tau_{\text{had-vis}}$ p_{T} and η in the SF-CR prior to the fit. Events with $\tau_{\text{had-vis}}$ candidate p_{T} larger than 200 GeV are included in the last bin. All statistical and systematic uncertainties are included.

requirements at trigger-level. Three sets of SFs are measured to account for identification requirements applied at trigger-level and during offline event reconstruction.

The measurement of SFs without requirements at trigger-level can proceed using events in the SF-CR without additional selections. For measurements of SFs that include trigger-level identification requirements, the selections applied by $\tau_{\text{had-vis}}$-triggers need to be emulated. This is achieved by requiring that events in the SF-CR are also selected by appropriately chosen single-$\tau_{\text{had-vis}}$ triggers. In addition, the reconstructed $\tau_{\text{had-vis}}$ candidate has to be geometrically matched ($\Delta R < 0.2$) to a $\tau_{\text{had-vis}}$ candidate at the HLT that fulfilled the trigger criteria. The SF measurement is performed for three different $\tau_{\text{had-vis}}$-triggers (cf. Section 6.3.1):

- `HLT_tau25_medium1_tracktwo`
- `HLT_tau25_medium1_tracktwoEF`
- `HLT_tau25_medium1_tracktwoEF or HLT_tau25_mediumRNN_tracktwoMVA`

During Run 2 of the LHC, these triggers had to be prescaled to limit their trigger rates.[10] However, events passing the SF-CR selection were already recorded using unprescaled single-lepton triggers. This allows single-$\tau_{\text{had-vis}}$ triggers to be re-run during offline event reconstruction without application of a prescale. The `HLT_tau25_medium1_tracktwo` trigger chain was active for the entirety of Run 2 of the LHC. Trigger decisions for the `HLT_tau25_medium1_tracktwoEF` and `HLT_tau25_mediumRNN_tracktwoMVA` trigger chains are only available for partial datasets with integrated luminosities of 58 fb^{-1} and 37 fb^{-1}, respectively.

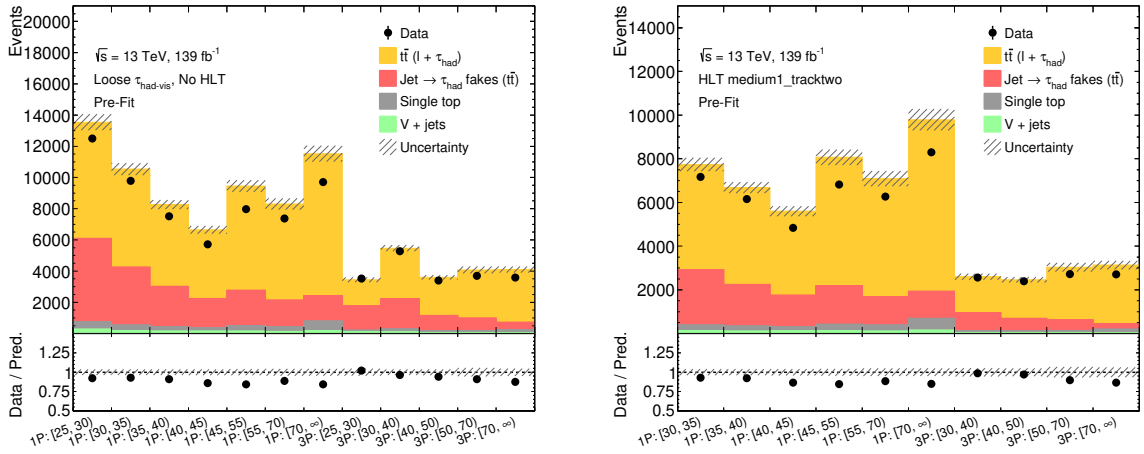
[10] A trigger with a prescale value of n accepts events satisfying the trigger conditions with a probability of $1/n$ [113].

Scale Factor Measurement

The jet $\rightarrow \tau_{\text{had-vis}}$ misidentification efficiencies strongly depend on the charged-particle multiplicity and transverse momentum of reconstructed $\tau_{\text{had-vis}}$ candidates. This might also be reflected in corrections of the jet $\rightarrow \tau_{\text{had-vis}}$ misidentification efficiencies in simulation. Therefore, the SF measurement is performed in regions of N_{tracks} and p_{T} of the $\tau_{\text{had-vis}}$ candidate given by:

- $N_{\text{tracks}} = 1$ and p_{T}/GeV: $[25, 30)$, $[30, 35)$, $[35, 40)$, $[40, 45)$, $[45, 55)$, $[55, 70)$, $[70, \infty)$.
- $N_{\text{tracks}} = 3$ and p_{T}/GeV: $[25, 30)$, $[30, 40)$, $[40, 50)$, $[50, 70)$, $[70, \infty)$.

These regions are chosen such that their size allows for a determination of the corrections with limited impact of statistical uncertainties while extracting potential p_{T}-dependencies of the correction factors. In cases where events are required to pass single-$\tau_{\text{had-vis}}$ triggers, the p_{T} intervals from 25 to 30 GeV are omitted to ensure that the $\tau_{\text{had-vis}}$-triggers operate in a regime where they are well-calibrated. This is analogous to the selections applied in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Two examples of event yields in the regions entering the SF measurement are shown in Figure 6.6.



(a) Events without trigger-level tau identification requirements.

(b) Events passing $\tau_{\text{had-vis}}$ trigger-matching and the `HLT_tau25_medium1_tracktwo` trigger.

Figure 6.6: Expected and observed event yields in regions of $\tau_{\text{had-vis}}$ candidate N_{tracks} and p_{T} in the SF-CR. The bins are labelled as "1P" and "3P" for $N_{\text{tracks}} = 1$ and $N_{\text{tracks}} = 3$, respectively, and the p_{T} intervals are given in units of GeV. The background model is shown prior to the fit.

A discriminant is required to distinguish between $t\bar{t}$ with and without fake $\tau_{\text{had-vis}}$ since only the former is sensitive to jet $\rightarrow \tau_{\text{had-vis}}$ misidentification efficiencies. The transverse mass of the electron/muon and $p_{\text{T}}^{\text{miss}}$ is used for this purpose, which is defined as

$$m_{\text{T}}^{W} = \sqrt{2|\boldsymbol{p}_{\text{T}}^{\ell}||\boldsymbol{p}_{\text{T}}^{\text{miss}}|\,(1 - \cos\Delta\phi)}\,,$$

where $\Delta\phi$ is the angle between the lepton transverse momentum, $\boldsymbol{p}_{\text{T}}^{\ell}$, and the missing transverse momentum, $\boldsymbol{p}_{\text{T}}^{\text{miss}}$. The m_{T}^{W} discriminant targets the differences between di- and semi-leptonic decay modes of $t\bar{t}$ that are the primary sources of events with true and fake $\tau_{\text{had-vis}}$, respectively. The

differences are illustrated in Figure 6.7, with di-leptonic decay modes of $t\bar{t}$ showing a heavy tail towards large m_{T}^{W} due to the presence of additional neutrinos, while semi-leptonic decay modes have a more pronounced peak close to the mass of the W boson.

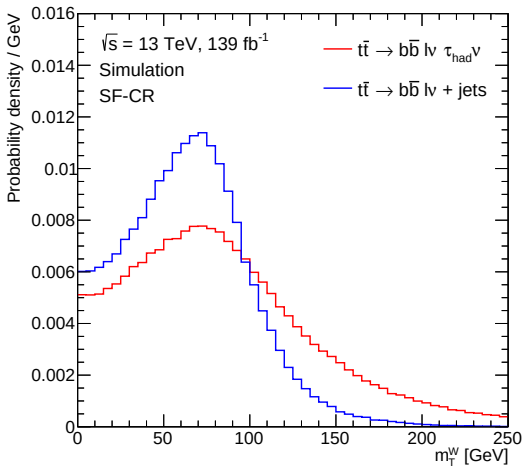


Figure 6.7: Distribution of the transverse mass of the lepton and $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}$ for simulated $t\bar{t}$ events in the SF-CR. The distributions are inclusive in p_{T} and N_{tracks} of the $\tau_{\text{had-vis}}$ candidate.

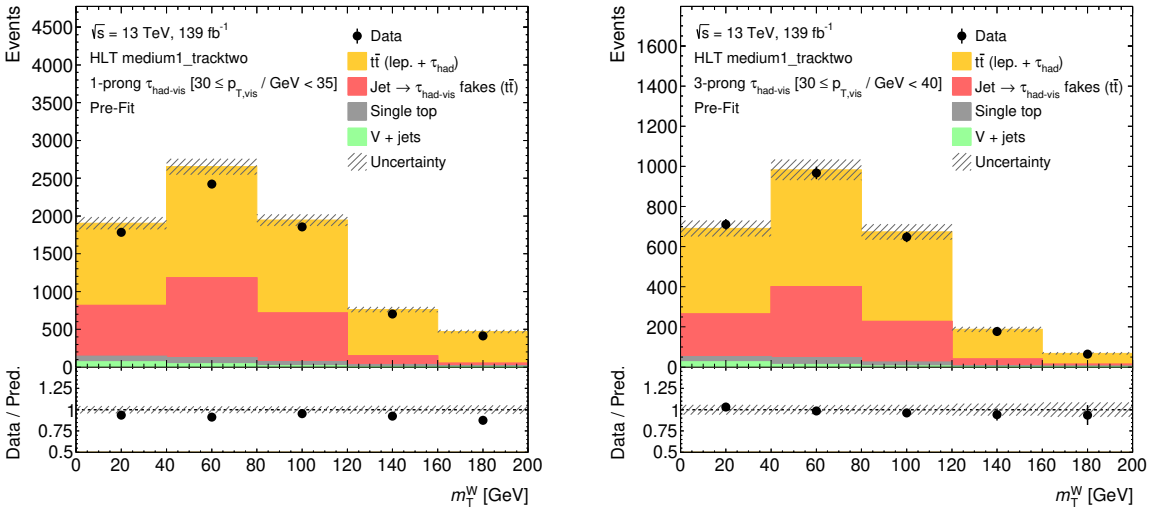
The fake-$\tau_{\text{had-vis}}$ SFs are measured using a simultaneous likelihood fit of the binned m_{T}^{W} distributions in all regions of $\tau_{\text{had-vis}}$ candidate p_{T} and N_{tracks}. The fit model is constructed using simulated $t\bar{t}$, single-top-quark, and V + jets events. The sample of $t\bar{t}$ events is split by whether the reconstructed $\tau_{\text{had-vis}}$ candidate is a fake $\tau_{\text{had-vis}}$ or not. The overall normalisation of the $t\bar{t}$ background, irrespective of whether events contain a fake $\tau_{\text{had-vis}}$, is free to vary in the model. In every region of $\tau_{\text{had-vis}}$ candidate N_{tracks} and p_{T}, an unconstrained SF is introduced that changes the normalisation of the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background in this region. These fake-$\tau_{\text{had-vis}}$ SFs are considered as the POIs of the measurement.

The pre-fit expectation of the model in two exemplary regions is shown in Figure 6.8 for SF-CR events passing the `HLT_tau25_medium1_tracktwo` trigger and trigger-matching. The binning of the m_{T}^{W} discriminants is the same in all $\tau_{\text{had-vis}}$ candidate p_{T} and N_{tracks} regions.

Uncertainties in the Scale Factor Measurement

Several experimental and theoretical uncertainties are considered in the SF measurement. In general, these uncertainties can affect the normalisation and shape of the expected m_{T}^{W} distribution for a given process in all regions entering the fit. The uncertainties included in the SF measurement and their treatment closely follows the approach taken in Sections 6.6 and 6.7 to interpret the results of the search for Higgs boson pair production. Therefore, the uncertainties included in the SF measurement are only briefly summarised.

Experimental uncertainties affecting the reconstruction and selection efficiencies of electrons, muons, $\tau_{\text{had-vis}}$, and jets are accounted for in the SF measurement, including uncertainties on the efficiencies of b-tagging. Uncertainties on the reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ are propagated to the m_{T}^{W} distributions in all



(a) $N_{\text{tracks}} = 1$ and $30\,\text{GeV} \leq p_{\text{T}} < 35\,\text{GeV}$.

(b) $N_{\text{tracks}} = 3$ and $30\,\text{GeV} \leq p_{\text{T}} < 40\,\text{GeV}$.

Figure 6.8: Pre-fit m_{T}^{W} distribution in two exemplary $\tau_{\text{had-vis}}$ candidate N_{tracks} and p_{T} regions of the SF measurement after requiring events to pass the `HLT_tau25_medium1_tracktwo` trigger and trigger-matching. Events with $m_{\text{T}}^{W} > 200\,\text{GeV}$ are included in the last bin of the histograms.

regions. Trigger efficiency uncertainties are considered for single-lepton and single-$\tau_{\text{had-vis}}$ triggers.[11] Uncertainties on the re-weighting of the pile-up conditions in simulation and the integrated luminosity used to normalise simulated event samples are included in the measurement as well. Lastly, statistical uncertainties from the finite size of the simulation samples are included according to the simplified Barlow–Beeston method [141, 142].

Theory uncertainties on the modelling of $t\bar{t}$ production using simulation are derived for this measurement according to prescriptions developed by the ATLAS collaboration, which are summarised in Appendix A.3.1. An uncertainty on the simulation of the hard interaction and matching to the parton shower is derived by comparison with an alternative matrix element generator and matching scheme. Uncertainties on the modelling of the parton shower, hadronisation, and underlying event are determined by comparison with an alternative parton shower program. The effect of missing higher orders in the truncated perturbative expansion in α_{s} is probed by performing variations of renormalisation and factorisation scales. Finally, uncertainties on the modelling of additional emissions are derived by performing variations of the simulated initial- and final-state radiation. Modelling uncertainties are derived separately for $t\bar{t}$ events with and without fake $\tau_{\text{had-vis}}$ but are regarded as correlated in the fit model if they originate from the same source. Effects of the $t\bar{t}$ modelling uncertainties on the shape of the m_{T}^{W} discriminants and the expected number of events in different $\tau_{\text{had-vis}}$ candidate N_{tracks} and p_{T} regions are considered in the fit model.

Reduced sets of theory uncertainties are considered for minor backgrounds in the SF measurement. Uncertainties on the cross sections of single-top-quark and V + jets production are included in the model. Due to the known normalisation discrepancy of V + jets production in the presence of jets

[11] Uncertainties on single-$\tau_{\text{had-vis}}$ trigger efficiencies are only considered for SF measurements taking into account trigger-level tau identification requirements.

originating from heavy-flavour quarks, an additional normalisation uncertainty of 30 % is assigned to the V + jets background.

Results of the Scale Factor Measurement

The measured fake-$\tau_{\text{had-vis}}$ SFs are shown in Figure 6.9. The size of the corrections described by the SFs can reach up to 55 % for fake $\tau_{\text{had-vis}}$ of high p_{T}, where simulation overestimates the contribution of fake $\tau_{\text{had-vis}}$ in $t\bar{t}$. SFs for fake $\tau_{\text{had-vis}}$ reconstructed as 1-prong (3-prong) candidates with $p_{\text{T}} < 70\,\text{GeV}$ are within 20 % (40 %) of unity. Only small differences are observed between SFs measured for different tau identification requirements at trigger-level, as is shown in Figures 6.9(c) and 6.9(d).

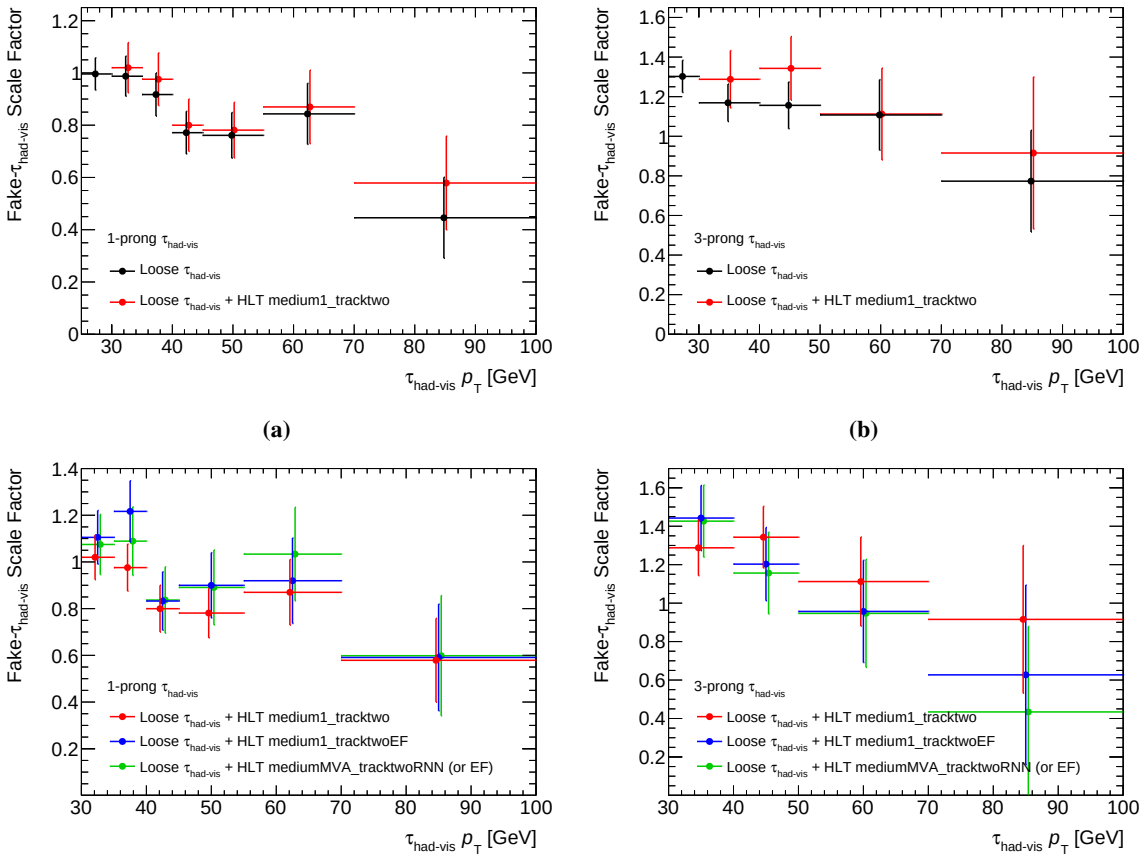


Figure 6.9: Fake-$\tau_{\text{had-vis}}$ SFs for different tau identification criteria. Figures (a) and (b) compare the SFs with and without tau identification at the HLT. The effect of different $\tau_{\text{had-vis}}$-triggers on the extracted SFs is shown in Figures (c) and (d). In all cases, the last bin summarises the SFs for $\tau_{\text{had-vis}}$ candidates with $p_{\text{T}} \geq 70\,\text{GeV}$. The markers are shifted from their geometrical bin centres for illustration purposes only.

The fit model and the corresponding fit results are checked by comparing the post-fit predictions of the model with the observed data in all $\tau_{\text{had-vis}}$ candidate N_{tracks} and p_{T} regions. In addition, pulls and constraints of NPs as well as correlations between NPs and the POIs are inspected. Exemplary post-fit predictions of the model are shown in Figure 6.10 in terms of the p_{T} of $\tau_{\text{had-vis}}$ candidates

and m_{T}^{W} in the SF-CR after requiring events to pass the `HLT_tau25_medium1_tracktwo` trigger and trigger-matching.

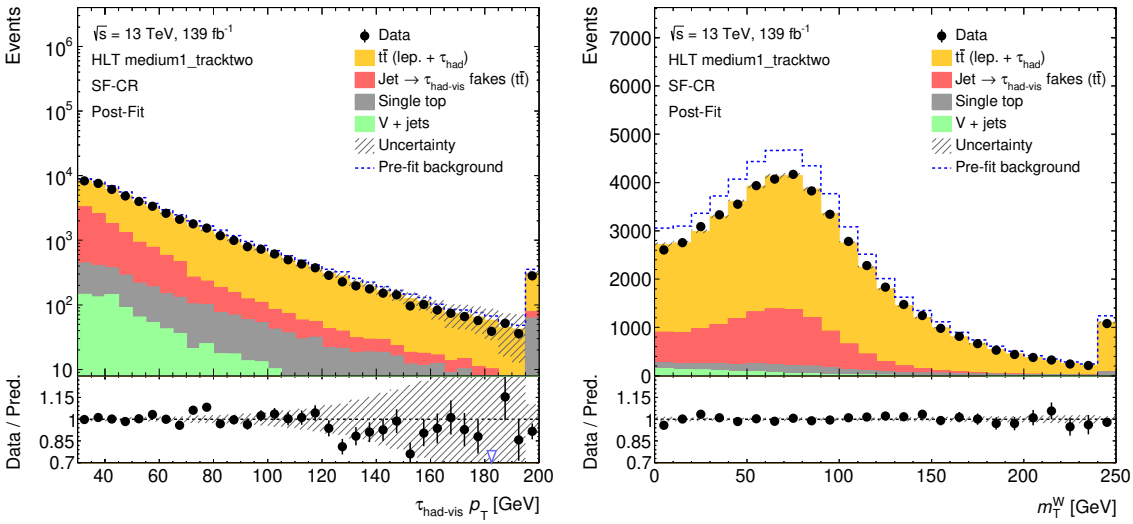


Figure 6.10: Post-fit distributions of $\tau_{\text{had-vis}}$ candidate p_{T} and m_{T}^{W} in the SF-CR for events passing the `HLT_tau25_medium1_tracktwo` trigger and trigger-matching. Both distributions are inclusive in N_{tracks} and p_{T} of $\tau_{\text{had-vis}}$ candidates. Events with $\tau_{\text{had-vis}}$ p_{T} (m_{T}^{W}) larger than 200 GeV (250 GeV) are included in the last bin. The uncertainty band contains all statistical and systematic uncertainties, including post-fit uncertainties on fake-$\tau_{\text{had-vis}}$ SFs and on the overall $t\bar{t}$ normalisation factor.

The post- and pre-fit values of all NPs agree within their uncertainties. Few instances are observed where the SF measurement puts more stringent constraints on the values of NPs than expected from the prior estimation of the associated uncertainty. These cases are the $p_{\mathrm{T}}^{\text{miss}}$ scale uncertainty, the $t\bar{t}$ modelling uncertainties resulting from comparison with alternative ME/PS generators, and the $\tau_{\text{had-vis}}$ energy scale uncertainty. The constraints on these parameters tend to be moderate with ratios of post- to pre-fit uncertainties above 70 %. Due to the sensitivity of the m_{T}^{W} discriminant to the modelling of $p_{\mathrm{T}}^{\text{miss}}$ and the abundance of $t\bar{t}$ events in the SF-CR, the fit is expected to have some power to constrain the associated NPs. Moreover, the $\tau_{\text{had-vis}}$ energy scale uncertainties are derived from $Z \to \tau_{\mu}\tau_{\text{had}}$ tag-and-probe measurements [135], which provide probe $\tau_{\text{had-vis}}$ with smaller transverse momenta than the ones produced in $t\bar{t}$ events. With the SF measurement being performed in N_{tracks} and p_{T} bins of the $\tau_{\text{had-vis}}$ candidate and targeting $t\bar{t}$ events, constraints on the $\tau_{\text{had-vis}}$ energy scale in the SF measurement are expected.

The limited discrimination power of m_{T}^{W} in distinguishing $t\bar{t}$ events with and without fake $\tau_{\text{had-vis}}$ leads to large anti-correlations between fake-$\tau_{\text{had-vis}}$ SFs and the $t\bar{t}$ normalisation factor. Due to this coupling, positive correlations are induced between the SFs themselves. This is illustrated in Figure 6.11 for an exemplary SF measurement.

Correlations between the measured SFs need to be taken into account when propagating the uncertainties on the SFs to the estimate of the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background in the $\tau_{\text{had}}\tau_{\text{had}}$ SR. This is achieved by providing a set of decorrelated variations of the measurement that explains the total uncertainty of all measured SFs. These variations can be used to propagate the uncertainties to the background estimate without having to account for correlations.

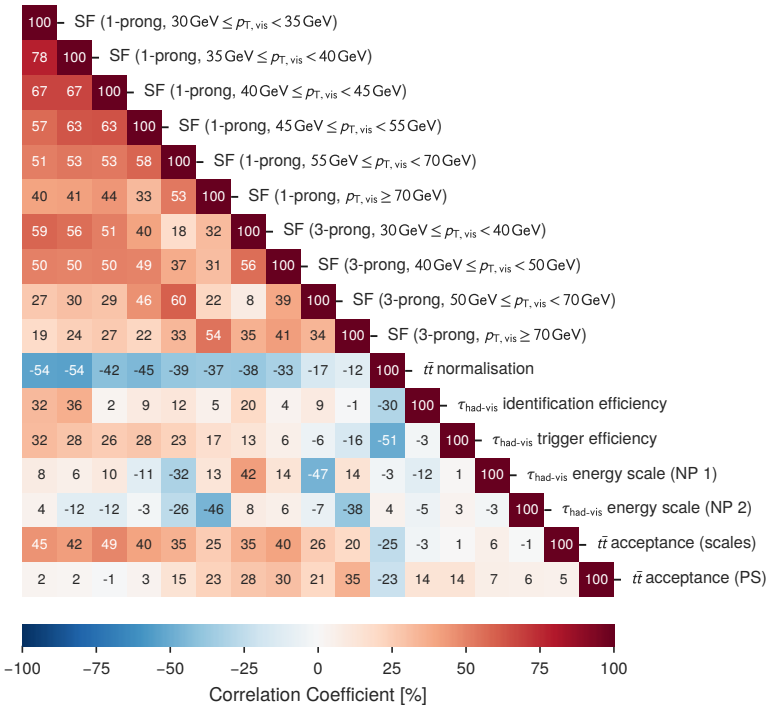
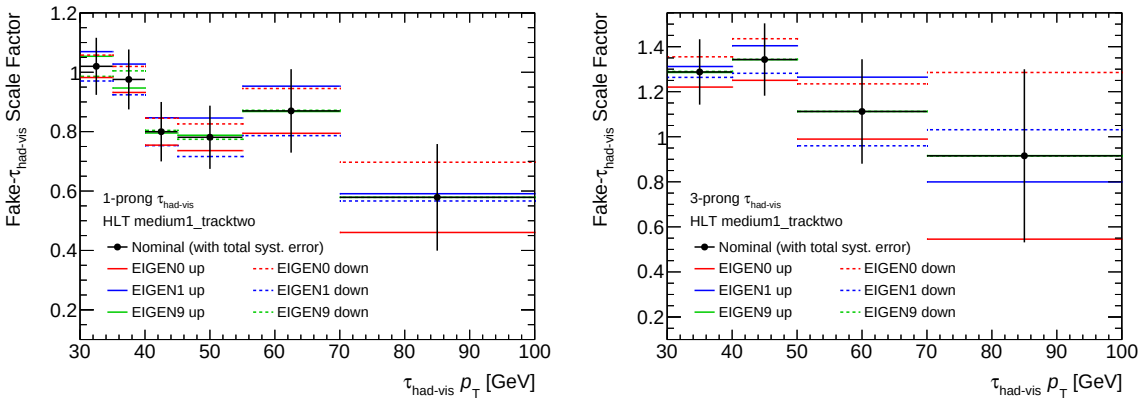


Figure 6.11: Post-fit correlation matrix between selected parameters of the fake-$\tau_{\text{had-vis}}$ SF measurement for the `HLT_tau25_medium1_tracktwo` trigger. A reduced number of NPs is shown for illustration purposes. NPs are included if the absolute value of its correlation coefficient with at least one POI exceeds 30 %.

A decorrelated set of variations is obtained by performing a linear transformation of the N measured SFs. The transformation is obtained by diagonalising the post-fit SF covariance matrix, yielding a set of eigenvectors and eigenvalues. The eigenvectors provide an alternative basis in which the measurement is described by N linear combinations of SFs with diagonal covariance matrix. The eigenvalues correspond to the variance explained by certain linear combinations of SFs. In the frame with diagonal covariance matrix, the SF measurement is varied by performing $\pm 1\sigma$ variations, then transforming the resulting variations back to the original, physically interpretable frame. This procedure yields N systematic variations of the SF measurement, each with an up- and down-variation. The variations are ordered by descending variance in the diagonal frame, yielding variations that are roughly decreasing in their impact on the total $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background estimate. An example of the results of the decorrelation procedure is shown in Figure 6.12. The effect of large correlations between SFs can be observed in the leading variation (`EIGEN0`) as a systematic shift of the variation in all bins with respect to the nominal result. In contrast, the variation explaining the least variance (`EIGEN9`) only alters the SFs for 1-prong fake $\tau_{\text{had-vis}}$ with low transverse momenta.

Application of Fake-$\tau_{\text{had-vis}}$ Scale Factors in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

The $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel is estimated by applying fake-$\tau_{\text{had-vis}}$ SFs to $t\bar{t}$ events from simulation with at least one fake $\tau_{\text{had-vis}}$. These events are required to pass the SR selection criteria of the $\tau_{\text{had}}\tau_{\text{had}}$ channel, including the trigger selection described in Section 6.3.1.



(a) 1-prong $\tau_{\text{had-vis}}$ **(b)** 3-prong $\tau_{\text{had-vis}}$

Figure 6.12: Fake-$\tau_{\text{had-vis}}$ SF variations resulting from the decorrelation procedure applied to the measurement for the `HLT_tau25_medium1_tracktwo` trigger. The two variations leading in the explained variance (`EIGEN0` and `EIGEN1`) as well as the variation explaining the least variance (`EIGEN9`) are shown.

The SFs are chosen depending on the trigger category and whether the fake $\tau_{\text{had-vis}}$ is the $\tau_{\text{had-vis}}$ candidate leading in p_{T}, sub-leading in p_{T}, or both selected candidates are fake $\tau_{\text{had-vis}}$. When both $\tau_{\text{had-vis}}$ candidates are originating from quark- or gluon-initiated jets, the SF correction is assumed to factorise and the product of SFs is assigned as an event-level weight.

In events selected by DTTs, both $\tau_{\text{had-vis}}$ candidates have to fulfil the trigger-level identification requirements. In this case, the set of SFs is chosen according to the trigger chain that selected the event, independent of which $\tau_{\text{had-vis}}$ candidate is the fake $\tau_{\text{had-vis}}$. In contrast, only one $\tau_{\text{had-vis}}$ candidate has to fulfil the trigger-level requirements in events selected by STTs. For STT events, it is assumed that the $\tau_{\text{had-vis}}$ candidate leading in p_{T} is the one satisfying the trigger conditions. This assumption is correct for more than 99 % of $t\bar{t}$ events containing fake $\tau_{\text{had-vis}}$ in the STT category. Therefore, SFs measured for fake $\tau_{\text{had-vis}}$ after trigger-matching are applied when the leading $\tau_{\text{had-vis}}$ candidate is the fake $\tau_{\text{had-vis}}$; SFs derived without trigger-matching are applied when the sub-leading $\tau_{\text{had-vis}}$ candidate is the fake $\tau_{\text{had-vis}}$. Similar to the DTT case, if the $\tau_{\text{had-vis}}$ candidate leading in p_{T} is the fake $\tau_{\text{had-vis}}$, then the set of SFs that corresponds to the trigger chain that selected the event is used. The event weight calculation is summarised in Table 6.9.

Table 6.9: Event weights for the application of SFs to $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events in simulation. Events are categorised by whether the leading $\tau_{\text{had-vis}}$ candidate ($\tau_{\text{lead.}}$), sub-leading $\tau_{\text{had-vis}}$ candidate ($\tau_{\text{subl.}}$), or both $\tau_{\text{had-vis}}$ candidates are fake $\tau_{\text{had-vis}}$. SFs for fake $\tau_{\text{had-vis}}$ without identification at trigger-level are denoted by SF_{loose}. SFs for fake $\tau_{\text{had-vis}}$ with both offline and trigger-level identification requirements are denoted by $\text{SF}_{\text{loose+trig.}}$.

$\tau_{\text{lead.}}$	$\tau_{\text{subl.}}$	Event weight (STT)	Event weight (DTT)
true	fake	$1 \times \text{SF}_{\text{loose}}(\tau_{\text{subl.}})$	$1 \times \text{SF}_{\text{loose+trig.}}(\tau_{\text{subl.}})$
fake	true	$\text{SF}_{\text{loose+trig.}}(\tau_{\text{lead.}}) \times 1$	$\text{SF}_{\text{loose+trig.}}(\tau_{\text{lead.}}) \times 1$
fake	fake	$\text{SF}_{\text{loose+trig.}}(\tau_{\text{lead.}}) \times \text{SF}_{\text{loose}}(\tau_{\text{subl.}})$	$\text{SF}_{\text{loose+trig.}}(\tau_{\text{lead.}}) \times \text{SF}_{\text{loose+trig.}}(\tau_{\text{subl.}})$

Uncertainties on the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ Background in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

In addition to the uncertainties originating from the SF measurement, two other sources of uncertainties are considered. First, an uncertainty accounting for a possible bias in the estimated SFs due to trigger efficiency turn-on effects arising from $\tau_{\text{had-vis}}$ p_{T} (E_{T}) thresholds at the HLT (L1 trigger) is determined. Second, an uncertainty on the extrapolation of the measured SFs from $\ell + \tau_{\text{had-vis}}$ final states to final states with two $\tau_{\text{had-vis}}$ is derived.

A systematic uncertainty accounting for the effect of $\tau_{\text{had-vis}}$ p_{T} (E_{T}) thresholds at the HLT (L1 trigger) on the fake-$\tau_{\text{had-vis}}$ SFs is estimated. The nominal SF measurement is performed using $\tau_{\text{had-vis}}$-triggers with thresholds of $p_{\text{T}} > 25\,\text{GeV}$ at the HLT and $E_{\text{T}} > 12\,\text{GeV}$ at the L1 trigger (`tau25`). However, triggers with higher thresholds are also employed in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. For example, thresholds of $p_{\text{T}} > 35\,\text{GeV}$ at the HLT and $E_{\text{T}} > 20\,\text{GeV}$ at the L1 trigger (`tau35`) are applied to the leading $\tau_{\text{had-vis}}$ candidate selected by DTTs. The application of SFs measured for `tau25` triggers to fake $\tau_{\text{had-vis}}$ that are required to pass a `tau35` trigger can introduce a bias due to differences in selection efficiency between both triggers. This effect is only relevant for fake $\tau_{\text{had-vis}}$ with transverse momenta close to the p_{T} threshold of the `tau35` trigger, i.e. 40–50 GeV for 1-prong candidates and 40–60 GeV for 3-prong candidates. For fake $\tau_{\text{had-vis}}$ with larger transverse momenta, the differences between triggers become negligible.

An uncertainty is assigned to SFs applied to fake $\tau_{\text{had-vis}}$ that are required to pass a `tau35` trigger if the transverse momentum of the fake $\tau_{\text{had-vis}}$ is close to the p_{T} threshold of the trigger. In particular, the uncertainty is only applied for 1-prong (3-prong) fake $\tau_{\text{had-vis}}$ with transverse momenta below 50 GeV (60 GeV). The size of the uncertainty is estimated by repeating the SF measurement for the `tau35` triggers and comparing with the nominal set of SFs. This comparison is performed for all trigger chains employed in the analysis, resulting in a relative uncertainty of approximately 6 % for all triggers considered in this search.

The measurement of the SFs is performed in the SF-CR and applied to events in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. An uncertainty is assigned to account for the extrapolation of the SFs from the SF-CR to the $\tau_{\text{had}}\tau_{\text{had}}$ SR. The uncertainties are derived by performing variations of the $t\bar{t}$ modelling in simulation and comparing the acceptance of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events in the SF-CR and the $\tau_{\text{had}}\tau_{\text{had}}$ SR. Variations of the matrix element generator, the parton shower simulation, the renormalisation and factorisation scales, and the modelling of initial and final state radiation are considered (cf. Appendix A.3.1).

The comparison of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ acceptances are performed separately for events in the $\tau_{\text{had}}\tau_{\text{had}}$ SR in which the leading $\tau_{\text{had-vis}}$ candidate is the fake $\tau_{\text{had-vis}}$ (FT events), the sub-leading $\tau_{\text{had-vis}}$ candidate is the fake $\tau_{\text{had-vis}}$ (TF events), and cases in which both candidates are fake $\tau_{\text{had-vis}}$ (FF events). This comparison yields extrapolation uncertainties of 14 % for TF events, 7 % for FT events, and 39 % for the FF events. In all cases, the uncertainty is dominated by the comparison of Pythia 8 and Herwig 7 for the simulation of the parton shower. The large uncertainty on FF events is expected since the measurement in the SF-CR can only target events with exactly one fake $\tau_{\text{had-vis}}$.

In the following, the impact of the uncertainties on the total $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background prediction in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel is summarised. The relative uncertainties on the prediction split by uncertainty source are:

- Uncertainties from the SF measurement: ±8.5 %.
- Extrapolation uncertainties (SF-CR → $\tau_{\text{had}}\tau_{\text{had}}$ SR): ±9.7 %.
- Uncertainty on SFs due to different p_{T} (E_{T}) thresholds of $\tau_{\text{had-vis}}$-triggers: ±0.2 %.

- Statistical uncertainty from finite number of simulated events: ±0.9 %.

The dominant sources of uncertainty on the prediction of the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ background are the uncertainties on the SFs from the SF measurement and the extrapolation uncertainties. Due to the small expected number of FF events in the $\tau_{\text{had}}\tau_{\text{had}}$ SR (cf. Table 6.10), the large extrapolation uncertainty for FF events has limited impact on the uncertainty of the total $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ prediction. Table 6.10 summarises the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ prediction in the $\tau_{\text{had}}\tau_{\text{had}}$ SR and compares it to the prediction from simulation without application of fake-$\tau_{\text{had-vis}}$ SFs.

Table 6.10: Expected number of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events in the $\tau_{\text{had}}\tau_{\text{had}}$ SR with (right) and without (left) application of the fake-$\tau_{\text{had-vis}}$ SFs. Only MC statistical uncertainties are shown for the estimate without application of the fake-$\tau_{\text{had-vis}}$ SFs. The background estimate using the fake-$\tau_{\text{had-vis}}$ SFs includes statistical uncertainties and all systematic uncertainties related to the SF method. Other experimental uncertainties are omitted.

	Expected number of events (pre-fit)	
Process	Simulation	Simulation with fake-$\tau_{\text{had-vis}}$ SF
$t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ (TF)	1 428 ± 16	1 430 ± 230
$t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ (FT)	854 ± 13	699 ± 88
$t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ (FF)	423 ± 12	360 ± 160
$t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ (total)	2 705 ± 23	2 490 ± 320

Estimation of Fake-$\tau_{\text{had-vis}}$ Scale Factors for Anti-$\tau_{\text{had-vis}}$

The estimation of the multi-jet background in the $\tau_{\text{had}}\tau_{\text{had}}$ SR, which is described in Section 6.4.3, requires a large subtraction of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events in a region defined by the presence of an anti-$\tau_{\text{had-vis}}$ (cf. Section 6.2). The SF method is extended to anti-$\tau_{\text{had-vis}}$ to provide uncertainties on this subtraction.

The SF measurement is repeated using the same SF-CR definition except for requiring an anti-$\tau_{\text{had-vis}}$ instead of a $\tau_{\text{had-vis}}$ candidate passing loose identification. This inversion of the tau identification requirement rejects most $t\bar{t}$ events with true-$\tau_{\text{had-vis}}$, yielding a region with $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ purity of about 80 %. Trigger-matching of the anti-$\tau_{\text{had-vis}}$ to a $\tau_{\text{had-vis}}$ at the HLT reduces the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ yield by about 80 % due to the trigger-level identification requirements; however, only a mild decrease in $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ purity by about 5 percentage points is observed.

A reduced set of experimental uncertainties is used for this measurement due to practical limitations in the dataset preparation for the ℓ + anti-$\tau_{\text{had-vis}}$ region. Uncertainties varying the four-momentum of reconstructed objects are omitted except uncertainties on the $\tau_{\text{had-vis}}$ energy scale. Other uncertainties that can be expressed as alternative event weights, for example uncertainties on tagging efficiencies, are considered. Due to this constraint, the total deviation of the central value of the SFs from unity is assigned as an additional uncertainty when applying the measured SFs to anti-$\tau_{\text{had-vis}}$.

The measured SFs for anti-$\tau_{\text{had-vis}}$ are shown in Figure 6.13. The SFs are generally within 20 % of unity. The same decorrelation technique is used to propagate the measurement uncertainties

when applying the SFs to anti-$\tau_{\text{had-vis}}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The context and impact of the SFs for anti-$\tau_{\text{had-vis}}$ on the multi-jet estimate is discussed in Section 6.4.3.

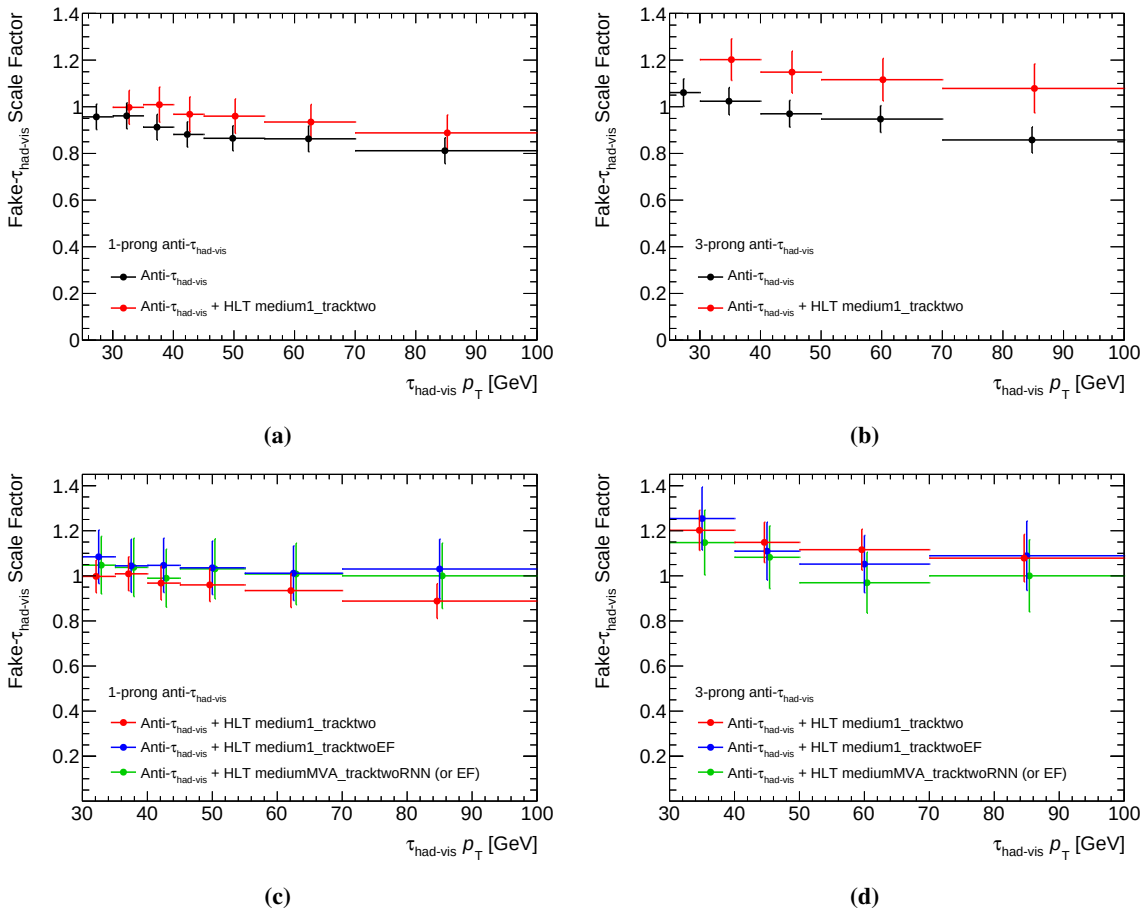


Figure 6.13: Fake-$\tau_{\text{had-vis}}$ SFs for anti-$\tau_{\text{had-vis}}$ and different tau identification criteria applied at trigger-level. Figures (a) and (b) compare the SFs with and without tau identification at the HLT. The effect of different $\tau_{\text{had-vis}}$-triggers on the extracted SFs is shown in (c) and (d). The last bin summarises events with $\tau_{\text{had-vis}}$ $p_{\text{T}} \geq 70\,\text{GeV}$. The markers are shifted from the geometrical bin centre for illustration purposes only. The depicted SFs include all statistical and systematic uncertainties.

6.4.3 Fake-$\tau_{\text{had-vis}}$ Background from Multi-jet Production in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

Multi-jet production is a source of background in the $\tau_{\text{had}}\tau_{\text{had}}$ SR in which both $\tau_{\text{had-vis}}$ candidates originate from quark- or gluon-initiated jets. It represents the second-largest background with fake $\tau_{\text{had-vis}}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ SR after the dominant $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ contribution.

The Fake Factor Method

The multi-jet background is estimated using the data-driven fake factor method. This method is applicable in cases where two observables exist that are statistically independent for the background

process to be estimated while also being strong discriminators between the background and other processes (i.e. signal and non-multi-jet processes). Four disjoint regions can be defined, three background-enriched CRs and a signal-like region, by categorising events based on both observables. The assumption of statistical independence allows relating the expected number of events for the background process between CRs and the signal-like region. This can be used to estimate the background in the signal-like region using data observed in the CRs.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, two observables that allow for the definition of CRs enriched in multi-jet events are the identification requirements fulfilled by the $\tau_{\text{had-vis}}$ candidates and the sign of the electric charges of both candidates. Two disjoint regions are defined based on the tau identification requirement: the ID region and the Anti-ID region. The ID region consists of events in which both $\tau_{\text{had-vis}}$ candidates pass the loose tau identification working point. The identification criterion is partially inverted to define a multi-jet enriched CR by requiring exactly one loose $\tau_{\text{had-vis}}$ and exactly one anti-$\tau_{\text{had-vis}}$ (cf. Section 6.2). This region is referred to as the Anti-ID region. Similarly, two regions are defined based on the charge signs of the two $\tau_{\text{had-vis}}$ candidates: an OS region and an SS region. Lastly, the tau identification and charge sign criterion are combined to define four distinct regions: the OS ID region (the signal-like region), the OS Anti-ID region, the SS ID region, and the SS Anti-ID region.

With these region definitions and the independence assumption, the expected multi-jet contribution in the OS ID region can be estimated according to

$$N_{\text{multi-jet}}^{\text{OS, ID}} = N_{\text{multi-jet}}^{\text{OS, Anti-ID}} \cdot \underbrace{\frac{N_{\text{multi-jet}}^{\text{SS, ID}}}{N_{\text{multi-jet}}^{\text{SS, Anti-ID}}}}_{=:\text{FF}_{\text{SS}}} ,$$

where $N_{\text{multi-jet}}^{r}$ is the expected number of multi-jet events in region r. The ratio of multi-jet events in the ID and Anti-ID region is referred to as the fake factor (FF). Generally, the CRs do not provide pure samples of multi-jet events; therefore, the number of multi-jet events is estimated according to

$$N_{\text{multi-jet}}^{r} = N_{\text{data}}^{r} - N_{\text{non-multi-jet}}^{r} ,$$

where N_{data}^{r} is the observed number of events in region r and $N_{\text{non-multi-jet}}^{r}$ the expected number of non-multi-jet events in region r estimated using simulation. Lastly, the probability of misidentifying a quark- or gluon-initiated jet as a $\tau_{\text{had-vis}}$ depends on properties of reconstructed $\tau_{\text{had-vis}}$ candidates, especially on N_{tracks} and p_{T}. To control for this dependence, FFs are measured in bins of $\tau_{\text{had-vis}}$ candidate properties. The choice of binning is discussed at a later point.

In Table 6.11, the expected multi-jet and non-multi-jet event yields in the regions relevant for the fake-$\tau_{\text{had-vis}}$ estimation are summarised. The 2 b-tag region, while most similar to the SR, is not well suited to estimate FFs:

- The 2 b-tag SS regions have large contributions from non-multi-jet sources, primarily $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$, that have to be subtracted. The large size of the subtraction leads to a degradation of the statistical precision of the FFs and an increase in systematic uncertainties from modelling uncertainties on the subtracted components.

- The strict b-tagging requirement suppresses the multi-jet contribution in the CRs preventing a

measurement of FFs in bins of $\tau_{\text{had-vis}}$ candidate properties.

- The multi-jet estimate cannot be validated in the 2 b-tag region due to the absence of a region with high multi-jet purity that is similar to the SR.

These issues are partially addressed by performing the FF measurement in the 1 b-tag region, which has a higher abundance and purity of multi-jet events, and extrapolating the measurement to the 2 b-tag region to obtain a multi-jet background estimate in the SR. Distributions of the p_{T} of the leading and sub-leading $\tau_{\text{had-vis}}$ candidates in the regions relevant to the FF measurement are shown in Figure 6.14.

Table 6.11: Expected number of multi-jet and non-multi-jet events in regions relevant to the FF method in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The expected number of multi-jet events is estimated by subtracting the expected number of non-multi-jet events from the observed number of events in a given region. The breakdown is shown after the 1 b-tag requirement in (a); after the 2 b-tag requirement in (b). The SR (2 b-tag OS ID) is omitted. Only statistical uncertainties on the expected event yields are shown.

(a) 1 b-tag regions

Region		$N_{\text{multi-jet}}$	$N_{\text{non-multi-jet}}$	Multi-jet purity
SS	ID	14 040 ± 130	1 970 ± 30	88 %
	Anti-ID	78 400 ± 300	5 710 ± 70	93 %
OS	ID	16 070 ± 210	16 440 ± 100	49 %
	Anti-ID	91 580 ± 340	13 680 ± 80	87 %

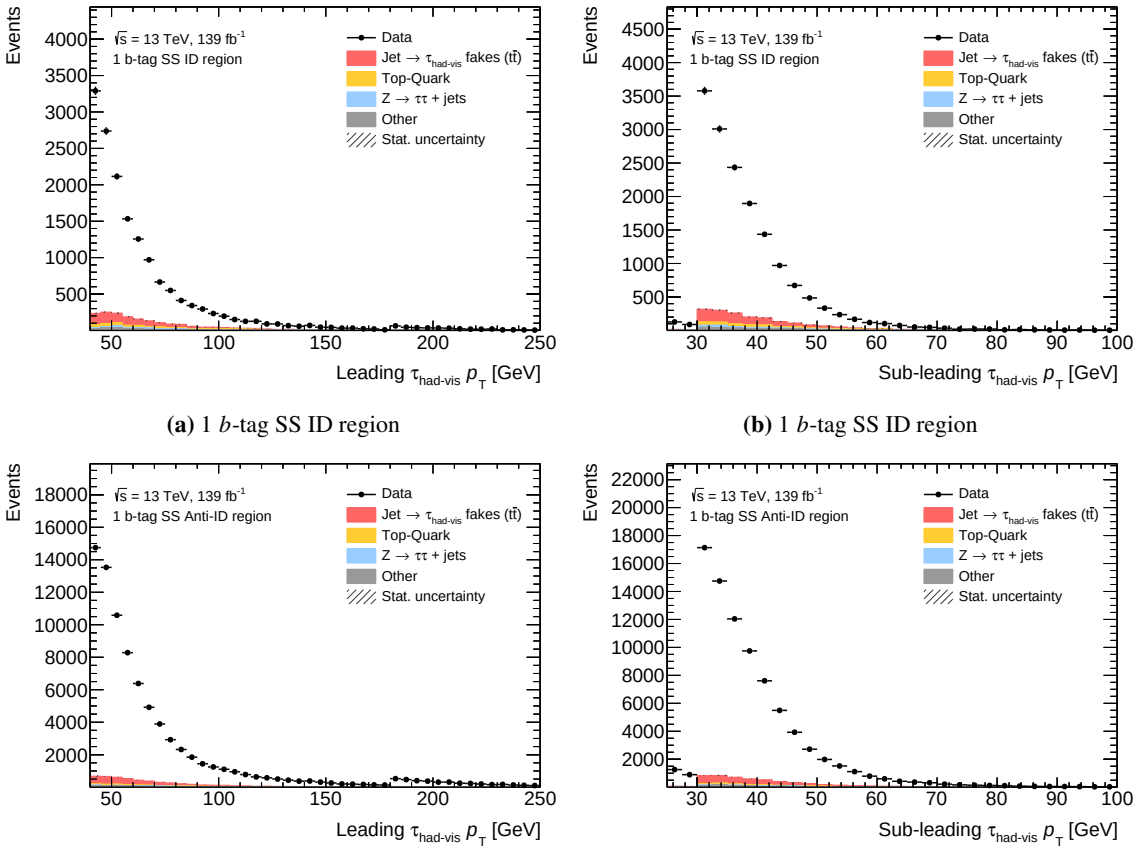
(b) 2 b-tag regions

Region		$N_{\text{multi-jet}}$	$N_{\text{non-multi-jet}}$	Multi-jet purity
SS	ID	1 300 ± 60	1 000 ± 20	56 %
	Anti-ID	7 650 ± 110	3 340 ± 30	70 %
OS	ID	———	Signal Region	———
	Anti-ID	8 430 ± 140	8 860 ± 50	49 %

A schematic illustration of the approach is given in Figure 6.15. FFs measured in the 1 b-tag SS regions ($\text{FF}_{\text{SS}}^{\text{1-tag}}$) are applied to events in the 2 b-tag OS Anti-ID region after subtraction of non-multi-jet contributions to obtain an estimate of the multi-jet background in the SR. Multiplicative transfer factors ($\text{TF}_{1\to 2\,b\text{-tag}}$) are applied to $\text{FF}_{\text{SS}}^{\text{1-tag}}$ when used in 2 b-tag regions, accounting for possible differences between FFs measured in 1 and 2 b-tag regions and the uncertainties associated with this extrapolation. The 1 b-tag OS ID region serves as a validation region (VR) to check the agreement of the background prediction with the observed data. Any non-closure observed in this region indicates either a violation of the assumptions of the FF method or dependencies that are not captured by the parameterisation of the FFs.

Binning of the Fake Factor Measurement

The FF measurement is performed separately for events selected by STTs and DTTs to account for the differences in selection. In addition, the changes in $\tau_{\text{had-vis}}$-triggers during Run 2 of the



(a) 1 *b*-tag SS ID region

(b) 1 *b*-tag SS ID region

(c) 1 *b*-tag SS Anti-ID region

(d) 1 *b*-tag SS Anti-ID region

Figure 6.14: Distribution of the leading and sub-leading $\tau_{\text{had-vis}}$ p_{T} for observed data and non-multi-jet backgrounds in regions used for the FF measurement. The 1 *b*-tag SS ID region is shown in (a,b) and the 1 *b*-tag SS Anti-ID region in (c,d). Coloured histograms depict the contributions of non-multi-jet processes that are subtracted when estimating the FFs. The difference between the observed data and the non-multi-jet background estimate is attributed to the missing multi-jet background estimate.

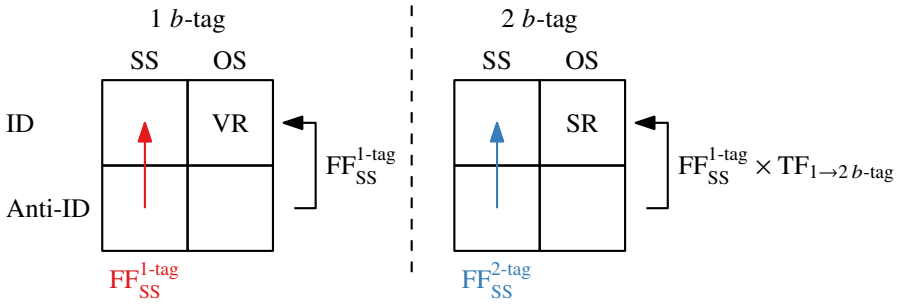


Figure 6.15: Schematic description of the FF method used to estimate the multi-jet background in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The squares represent the multi-jet events ($N_{\text{multi-jet}} = N_{\text{data}} - N_{\text{non-multi-jet}}$) in a particular region. The red / blue arrows correspond to FFs calculated as the ratio of multi-jet events in ID and Anti-ID regions. Black arrows correspond to the application of FFs in the OS Anti-ID region to obtain the multi-jet template in the OS ID region.

LHC are accounted for by performing the FF measurement separately for three major data-taking periods: 2015–2016, 2017, and 2018.

Dependencies of FFs on properties of reconstructed $\tau_{\text{had-vis}}$ candidates are accounted for by further categorisation based on properties of the (anti-)$\tau_{\text{had-vis}}$ that distinguishes the ID from the Anti-ID regions. The FF measurement is performed separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates for both trigger categories. For events selected by DTTs, the FFs are additionally measured in bins of the $\tau_{\text{had-vis}}$ candidate p_{T} and separately for candidates in the barrel ($|\eta| < 1.37$) and end-cap ($|\eta| \geq 1.52$) regions of the detector. Few multi-jet events are selected by STTs due to the large p_{T} thresholds on $\tau_{\text{had-vis}}$ candidates, preventing a fine binning of the FF measurement. In this case, FFs are measured separately for cases where the anti-$\tau_{\text{had-vis}}$ is leading in p_{T} and sub-leading in p_{T}. This accounts for the tau identification applied at the HLT and the large difference in transverse momentum between the leading and sub-leading $\tau_{\text{had-vis}}$ candidates.

Measurement of Fake Factors for Events Selected by DTTs

The Anti-ID region can be split into two subregions: one where the anti-$\tau_{\text{had-vis}}$ is the leading and one where it is the sub-leading $\tau_{\text{had-vis}}$ candidate. Provided the conditions for the FF method are fulfilled, both regions can be used to obtain separate estimates of the multi-jet background in the OS ID region. The following notation is used to describe the FF measurement:

$\boldsymbol{\tau_0}$ **(**$\boldsymbol{\tau_1}$**)** The $\tau_{\text{had-vis}}$ candidate leading (sub-leading) in p_{T}.

$\boldsymbol{x_\tau}$ A vector of categorical observables of the reconstructed $\tau_{\text{had-vis}}$ candidate that specifies the bin of the FF measurement.

$\boldsymbol{N^{\text{SS(OS), loose }\tau_i}_{\text{multi-jet}}(x_\tau)}$ The estimated number of multi-jet events in the SS (OS) ID region where τ_i falls into the bin specified by $\boldsymbol{x_\tau}$.

$\boldsymbol{N^{\text{SS(OS), anti-}\tau_i}_{\text{multi-jet}}(x_\tau)}$ The estimated number of multi-jet events in the SS (OS) Anti-ID region where τ_i is the anti-$\tau_{\text{had-vis}}$ and falls into the bin specified by $\boldsymbol{x_\tau}$.

With these definitions, two sets of FFs can be defined as

$$\mathrm{FF}_i(\boldsymbol{x}_\tau) = \frac{N_{\text{multi-jet}}^{\text{SS, loose } \tau_i}(\boldsymbol{x}_\tau)}{N_{\text{multi-jet}}^{\text{SS, anti-}\tau_i}(\boldsymbol{x}_\tau)} \quad \text{for} \quad i = 0, 1\,,$$

where FF_i is the FF relating the ID region with the part of the Anti-ID region in which τ_i is the anti-$\tau_{\text{had-vis}}$. These can be used to obtain two multi-jet estimates in the OS region given by

$$N_{\text{multi-jet}}^{\text{OS, loose } \tau_i}(\boldsymbol{x}_\tau) = \mathrm{FF}_i(\boldsymbol{x}_\tau) \cdot N_{\text{multi-jet}}^{\text{OS, anti-}\tau_i}(\boldsymbol{x}_\tau) \quad \text{for} \quad i = 0, 1\,.$$

The average of both estimates also yields an estimate of the multi-jet background. The process of averaging both multi-jet estimates can be expressed by an alternative set of FFs that do not distinguish in whether the anti-$\tau_{\text{had-vis}}$ is the leading or sub-leading $\tau_{\text{had-vis}}$ candidate, thus acting on the entirety of the Anti-ID region instead of a subregion thereof. This inclusive FF is given by

$$\mathrm{FF}_{\text{incl.}}(\boldsymbol{x}_\tau) = \frac{1}{2}\left[f_0(\boldsymbol{x}_\tau) \cdot \mathrm{FF}_0(\boldsymbol{x}_\tau) + f_1(\boldsymbol{x}_\tau) \cdot \mathrm{FF}_1(\boldsymbol{x}_\tau)\right]\,,$$

with $f_i(\boldsymbol{x}_\tau)$ being the fraction of anti-$\tau_{\text{had-vis}}$ in the bin specified by $\boldsymbol{x}_\tau$ that are leading ($i = 0$) or sub-leading ($i = 1$) in p_{T}, formally given by

$$f_i(\boldsymbol{x}_\tau) = \frac{N_{\text{multi-jet}}^{\text{SS, anti-}\tau_i}(\boldsymbol{x}_\tau)}{N_{\text{multi-jet}}^{\text{SS, anti-}\tau_0}(\boldsymbol{x}_\tau) + N_{\text{multi-jet}}^{\text{SS, anti-}\tau_1}(\boldsymbol{x}_\tau)}\,.$$

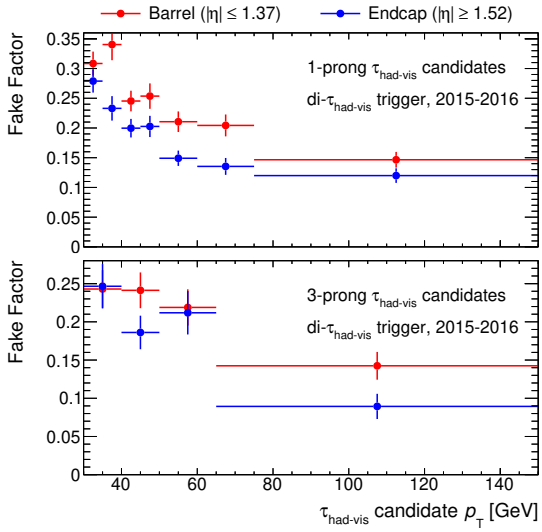
The inclusive FF can be measured directly using the following relationship

$$\mathrm{FF}_{\text{incl.}}(\boldsymbol{x}_\tau) = \frac{1}{2}\frac{N_{\text{multi-jet}}^{\text{SS, loose } \tau_0}(\boldsymbol{x}_\tau) + N_{\text{multi-jet}}^{\text{SS, loose } \tau_1}(\boldsymbol{x}_\tau)}{N_{\text{multi-jet}}^{\text{SS, anti-}\tau_0}(\boldsymbol{x}_\tau) + N_{\text{multi-jet}}^{\text{SS, anti-}\tau_1}(\boldsymbol{x}_\tau)}\,, \tag{6.1}$$

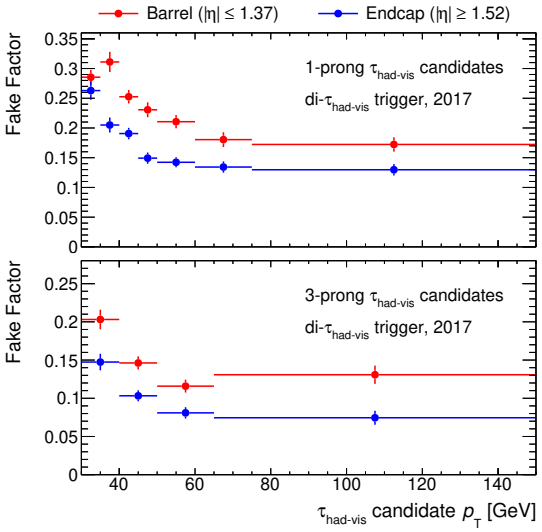
which can be applied to events in the inclusive Anti-ID region to obtain the multi-jet estimate in the ID region.

The motivation of using inclusive FFs is twofold: First, it allows for all events in the Anti-ID region to be used, independent of whether the anti-$\tau_{\text{had-vis}}$ is leading or sub-leading in p_{T}, thus improving the statistical precision of the background estimate. Second, the FFs can be parameterised in properties of the anti-$\tau_{\text{had-vis}}$, thus targeting the key differences between the ID and Anti-ID regions. This represents a change with respect to the previous publication in Ref. [184] in which FFs were parameterised in the properties of both $\tau_{\text{had-vis}}$ candidates simultaneously. Previously, the high dimensionality of the FF parameterisation lead to large statistical uncertainties on the measured FFs.

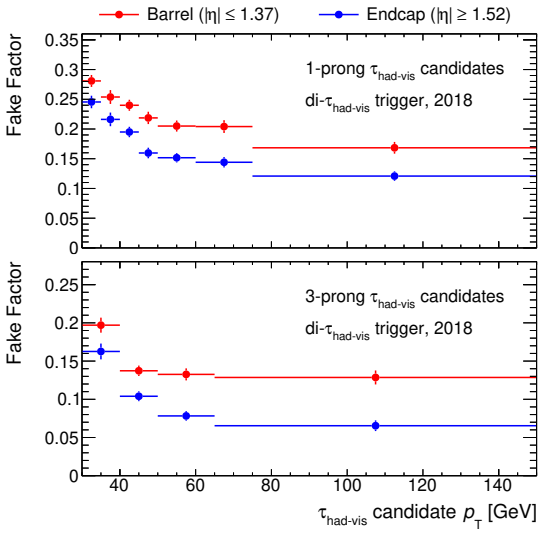
The result of the FF measurement for DTTs is summarised in Figure 6.16. Qualitatively, the behaviour of the FFs with respect to $\tau_{\text{had-vis}}$ candidate properties is the same between data-taking periods. Minor differences can be observed when comparing FFs between years. No attempt was made to combine the measurements for different years as the statistical precision of the FFs is not a limiting factor in the analysis.



(a) 2015–2016 data-taking period



(b) 2017 data-taking period



(c) 2018 data-taking period

Figure 6.16: FFs measured for events selected by DTTs in the 1 b-tag SS region. The measurement is performed separately for the three major data-taking periods (a-c), for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates (upper/lower panels), and for $\tau_{\text{had-vis}}$ in the barrel (red) and end-cap (blue) regions of the ATLAS detector. Events with (anti-)$\tau_{\text{had-vis}}$ $p_{\text{T}} > 150\,\text{GeV}$ are included in the last FF bin. Only statistical uncertainties are shown. Systematic uncertainties originating from the non-multi-jet subtraction are assumed to be negligible due to the small size of the subtraction in the 1 b-tag SS region.

Measurement of Fake Factors for Events Selected by STTs

The measurement of FFs for events selected by STTs can proceed using Equation (6.1). The measured STT FFs are shown in Figure 6.17 for the three major data-taking periods. Each period is divided into four categories depending on N_{tracks} and whether the anti-$\tau_{\text{had-vis}}$ is leading or sub-leading in p_{T}.

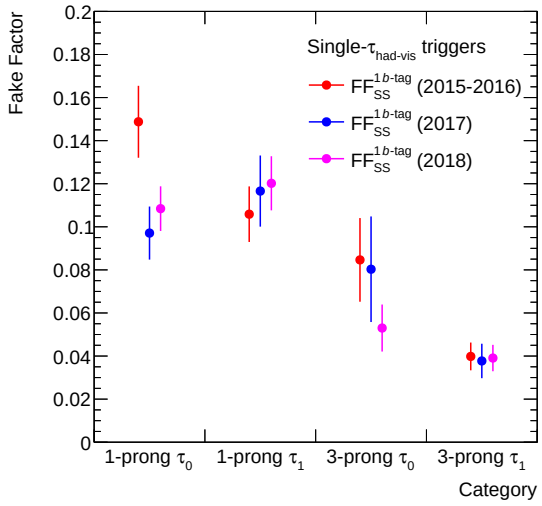


Figure 6.17: FFs measured for events selected by STTs in the 1 b-tag SS region. The measurement is performed in bins of the $\tau_{\text{had-vis}}$ candidate N_{tracks} (1- and 3-prong), separately for cases where the anti-$\tau_{\text{had-vis}}$ is leading (τ_0) and sub-leading in p_{T} (τ_1), and separately for the three major data-taking periods. Only statistical uncertainties are shown.

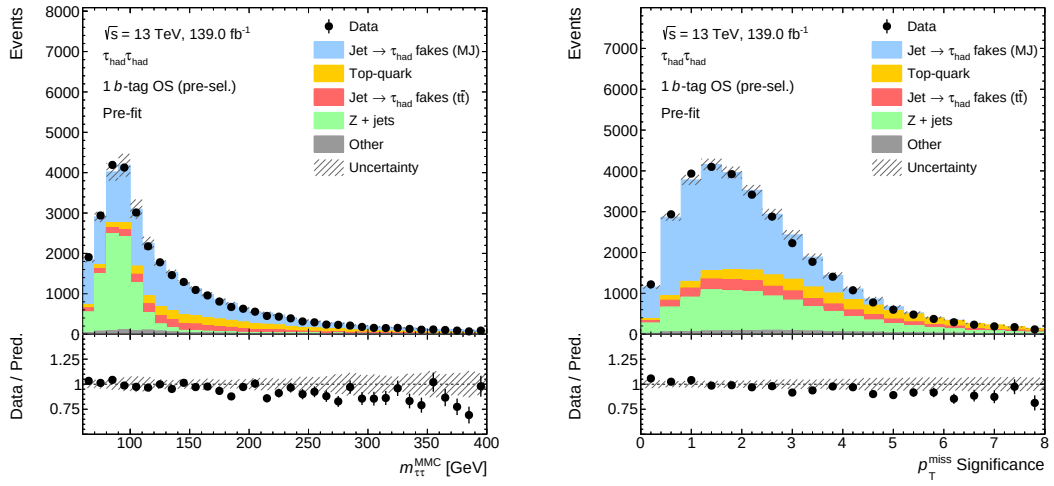
Validation of the Multi-jet Estimate in the 1 b-tag OS Region

An independent validation of the background estimate is performed in the 1 b-tag OS ID region (cf. Figure 6.15). This region has a multi-jet purity of about 50 %, with the dominant non-multi-jet contributions originating from Z + jets and $t\bar{t}$. A multi-jet VR is defined by requiring events to fulfil

$$m_{\tau\tau}^{\text{MMC}} > 110\,\text{GeV} \quad \text{and} \quad \mathcal{S} < 3\,,$$

where $\mathcal{S}$ is the object-based $p_{\text{T}}^{\text{miss}}$ significance [259]. The $p_{\text{T}}^{\text{miss}}$ significance measures the statistical significance of a test comparing the hypothesis that the reconstructed $p_{\text{T}}^{\text{miss}}$ is compatible with zero within the expected measurement errors to the alternative hypothesis of $p_{\text{T}}^{\text{miss}}$ primarily originating from undetected particles. The distributions of $m_{\tau\tau}^{\text{MMC}}$ and $\mathcal{S}$ prior to the multi-jet VR selection are shown in Figure 6.18. After the selection, the multi-jet purity in the VR increases to 75 % with a multi-jet selection efficiency of about 50 % with respect to the inclusive 1 b-tag OS ID region.

The multi-jet background prediction in the VR is obtained by applying the measured FFs to events in the OS Anti-ID region after subtracting non-multi-jet contributions. The non-multi-jet backgrounds in the OS ID region are estimated using simulation. The background prediction in the multi-jet VR is compared to data in Figure 6.19 for several observables of the leading and sub-leading $\tau_{\text{had-vis}}$



(a) **(b)**

Figure 6.18: Distributions of the di-τ invariant mass estimated with the MMC (a) and the object-based $p_{\mathrm{T}}^{\mathrm{miss}}$ significance (b) in the 1 b-tag OS ID region. The estimate of the multi-jet background (blue) is obtained using the FF method (cf. Figure 6.15). Fake-$\tau_{\text{had-vis}}$ originating from $t\bar{t}$ (red) are estimated using simulation. The background prediction is shown pre-fit, including statistical and detector-related systematic uncertainties.

candidate. Decent agreement between the background prediction and data is observed in the VR for $\tau_{\text{had-vis}}$-related observables. A small mismodelling of the background is observed in the p_{T} distribution of the sub-leading $\tau_{\text{had-vis}}$ candidate in the range of 60 to 70 GeV where the background appears to be overestimated by about 15 %. An uncertainty will be assigned to account for the non-closure in the multi-jet VR.

The validity of the FF method can be tested by comparing FFs measured in the 1 b-tag SS region with FFs obtained from measurements in the 1 b-tag OS multi-jet VR. Under the assumptions of the FF method, the OS and SS FFs are expected to agree. Differences between both sets of FFs can arise from a violation of the assumptions or from a mismodelling of the subtracted non-multi-jet backgrounds.

Exemplary comparisons of FFs measured in the OS and SS regions are shown in Figure 6.20 for DTTs and STTs. A test of the compatibility of OS and SS FFs for DTTs is tabulated in Table 6.12, showing good agreement with one exception. A significant deviation of about 50 % between OS and SS FFs is observed for a single FF bin for DTTs in 2015–2016.[12] Except for this bin and a tension in OS and SS FFs for 3-prong $\tau_{\text{had-vis}}$ candidates selected by STTs, no large differences between OS and SS FFs are observed. To account for non-closure between the OS and SS FFs, the full difference between both sets of FFs is assigned as an additional systematic uncertainty and propagated to the multi-jet estimate.

[12] The FF bin corresponds to 3-prong $\tau_{\text{had-vis}}$ candidates with p_{T} from 50 to 65 GeV in the end-cap of the detector. The OS FF for this bin is 0.09 ± 0.02 and the SS FF 0.21 ± 0.03.

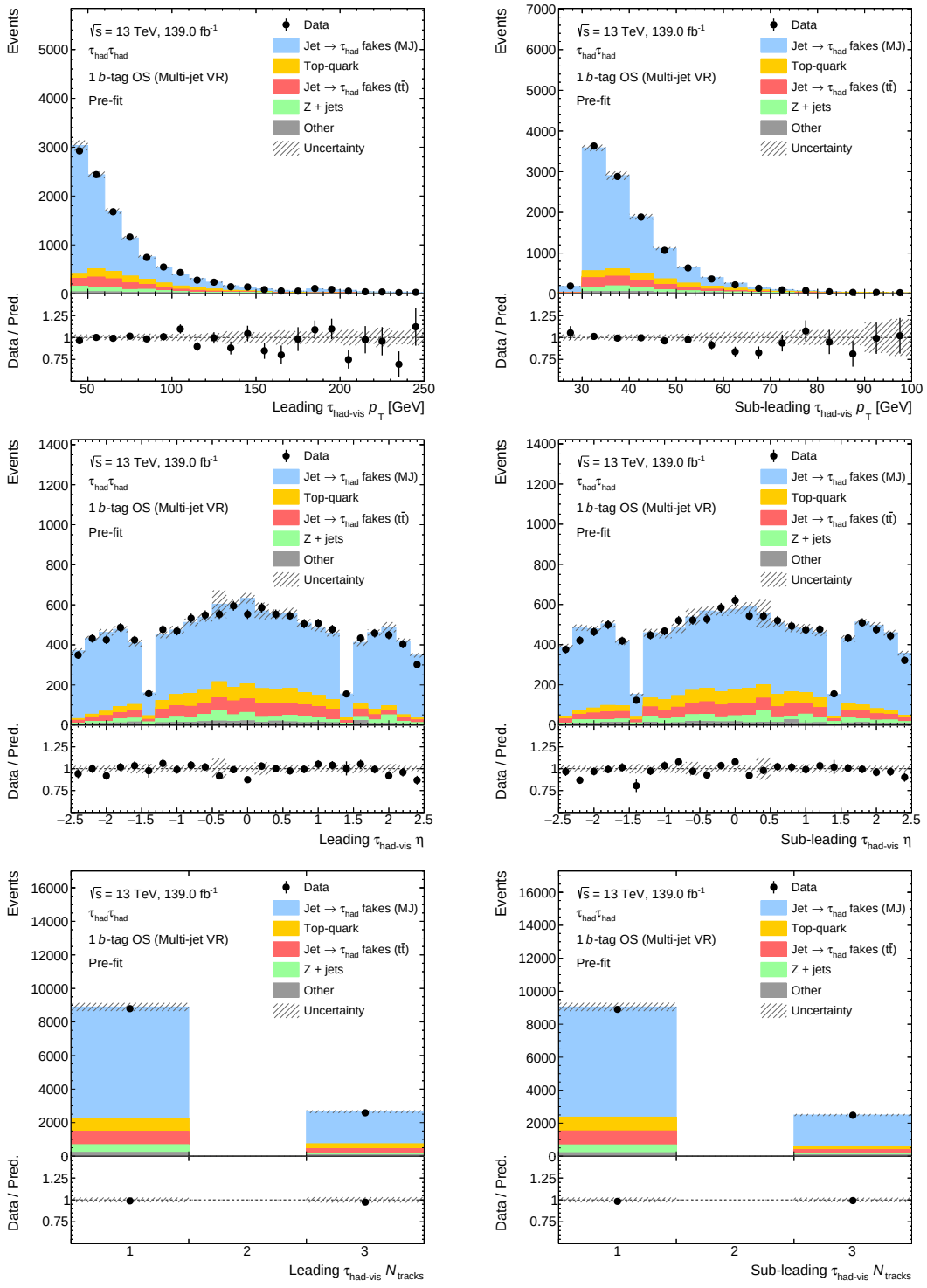
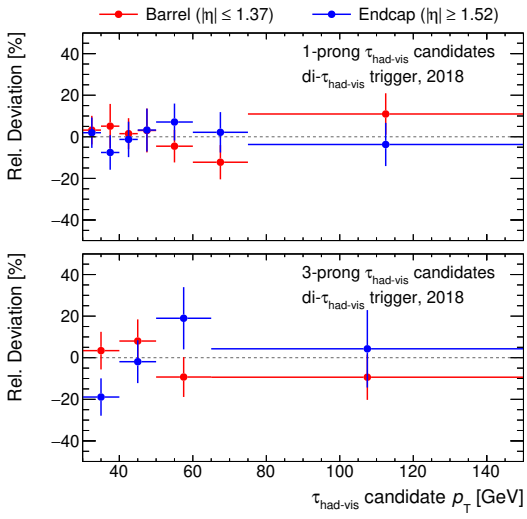
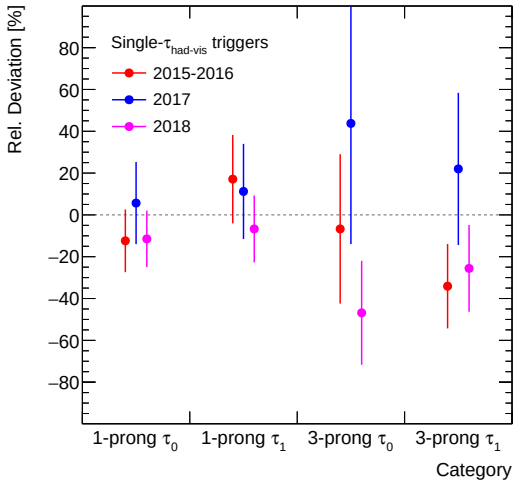


Figure 6.19: Distributions of $\tau_{\text{had-vis}}$ observables in the multi-jet VR. The multi-jet prediction (blue) is obtained using the FF method. The $\tau_{\text{had-vis}}$ observables p_{T} (top), η (centre), and N_{tracks} (bottom) are shown for the leading (left) and sub-leading $\tau_{\text{had-vis}}$ (right). The background prediction is shown pre-fit and includes statistical and detector-related systematic uncertainties. Systematic uncertainties on the multi-jet estimate are not included.



(a) Comparison of OS and SS FFs for events selected by DTTs. FFs for the 2018 data-taking period are shown.



(b) Comparison of OS and SS FFs for events selected by STTs for all major data-taking periods.

Figure 6.20: Relative deviation of FFs measured in the 1 b-tag OS multi-jet VR from the nominal set of FFs measured in the 1 b-tag SS region (cf. Figures 6.16 and 6.17). The relative deviation is measured as $\mathrm{FF_{OS}}/\mathrm{FF_{SS}} - 1$ and is used to define a non-closure uncertainty that is propagated to the multi-jet background estimate when applying SS FFs to events in OS regions. Statistical uncertainties from the finite number of observed events and the non-multi-jet subtraction are shown.

Table 6.12: Comparison of OS and SS FFs for DTTs using χ^2-tests to summarise the statistical compatibility of both sets of FFs over all $\tau_{\text{had-vis}}$ p_{T} bins. The barrel and end-cap detector regions correspond to $\tau_{\text{had-vis}}$ $|\eta| \leq 1.37$ and $|\eta| \geq 1.52$, respectively.

		$N_{\text{tracks}} = 1$		$N_{\text{tracks}} = 3$	
Period	Detector region	χ^2/NDF	p-value	χ^2/NDF	p-value
2015–2016	Barrel	4.7 / 7	69 %	3.7 / 4	45 %
	Endcap	7.5 / 7	38 %	14.8 / 4	< 1 %
2017	Barrel	6.1 / 7	53 %	4.0 / 4	41 %
	Endcap	6.2 / 7	52 %	2.3 / 4	68 %
2018	Barrel	4.2 / 7	75 %	2.3 / 4	68 %
	Endcap	1.8 / 7	97 %	5.7 / 4	22 %

Estimation of Multi-jet Backgrounds in the $\tau_{\text{had}}\tau_{\text{had}}$ SR

The multi-jet background in the $\tau_{\text{had}}\tau_{\text{had}}$ SR (2 b-tag OS ID) is estimated by applying FFs from the 1 b-tag SS region to events in the 2 b-tag OS Anti-ID region after subtraction of non-multi-jet contributions. In addition, these FFs are multiplied by a 1 to 2 b-tag transfer factor to account for possible differences between FFs for 1 and 2 b-tag regions (cf. Figure 6.15). The change in b-tagging requirement is not expected to affect the FFs; thus, the transfer factors mainly serve to provide an estimate of the uncertainty on the extrapolation.

The transfer factors are determined by comparing FFs measured in the 2 b-tag SS region to the ones from the 1 b-tag SS region. Due to the large multi-jet rejection of the 2 b-tag requirement, the comparison is performed using FFs measured inclusively in the trigger category, $\tau_{\text{had-vis}}$ p_{T}, and $\tau_{\text{had-vis}}$ η but separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates, for cases where the anti-$\tau_{\text{had-vis}}$ is leading and sub-leading in p_{T}, and for the three major data-taking periods. The 1 to 2 b-tag transfer factor is defined as the ratio

$$\text{TF}_{1\to 2\,b\text{-tag}} = \frac{\text{FF}_{\text{SS}}^{2\,b\text{-tag}}}{\text{FF}_{\text{SS}}^{1\,b\text{-tag}}} \,.$$

The measured transfer factors are depicted in Figure 6.21, showing no significant difference between FFs derived in the 1 and 2 b-tag regions. However, the power of this comparison is limited due to large uncertainties on the FF estimate in the 2 b-tag region. Therefore, extrapolation uncertainties based on the uncertainties of the transfer factor measurement are still assigned.

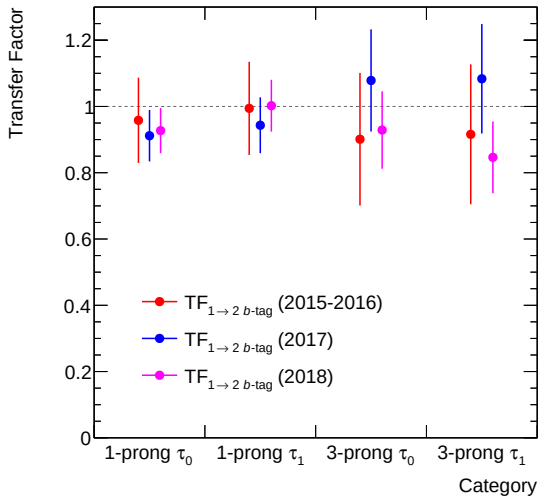
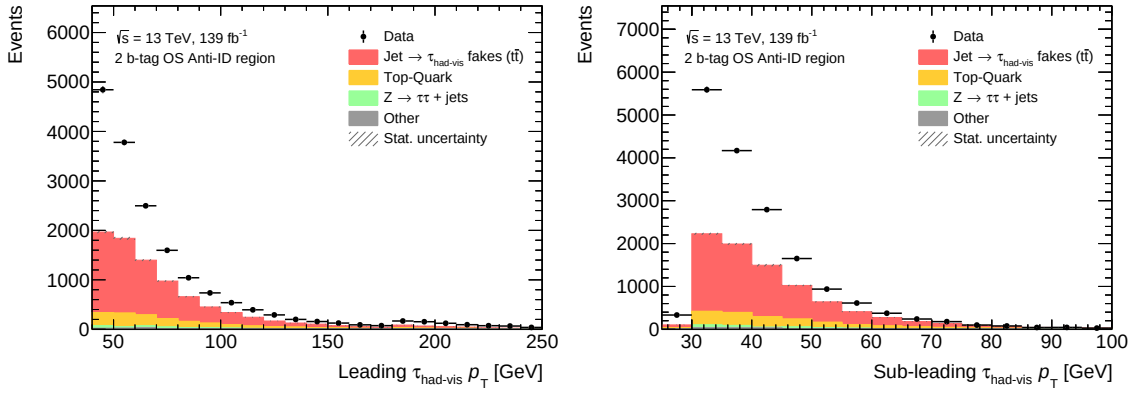


Figure 6.21: Transfer factors for the extrapolation of FFs measured in 1 b-tag regions to 2 b-tag regions. The transfer factors are shown separately for 1- and 3-prong $\tau_{\text{had-vis}}$, for cases where the anti-$\tau_{\text{had-vis}}$ is leading (τ_0) and sub-leading (τ_1) in p_{T}, and for the three major data-taking periods. The statistical uncertainties on the transfer factors are shown.

A disadvantage of applying the FF method in the 2 b-tag region is the low multi-jet purity of about 50 % in the 2 b-tag OS Anti-ID region (cf. Table 6.11). Consequently, a large subtraction of

non-multi-jet processes has to be performed when applying FFs to obtain the multi-jet prediction in the SR. The size of the subtraction in the 2 b-tag OS Anti-ID is illustrated in Figure 6.22, showing that $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ is the dominant source of non-multi-jet events in this region. Due to the large size of the subtracted non-multi-jet contributions, any uncertainties on the subtracted components have a large impact on the multi-jet estimate in the ID region. This is the largest source of systematic uncertainty on the multi-jet background estimate.



(a) **(b)**

Figure 6.22: Distribution of the leading (a) and sub-leading (b) $\tau_{\text{had-vis}}$ candidate p_{T} in the 2 b-tag OS Anti-ID region. Non-multi-jet backgrounds are display as coloured histograms and include the statistical uncertainty of the prediction. The difference between the observed data and the non-multi-jet background prediction is attributed to the missing multi-jet background estimate.

Uncertainties on the Multi-jet Prediction in the $\tau_{\text{had}}\tau_{\text{had}}$ SR

The following systematic uncertainties are considered and propagated to the multi-jet estimate in the $\tau_{\text{had}}\tau_{\text{had}}$ SR:

- Statistical uncertainties on the FFs.
- Non-closure uncertainty between FFs estimated in 1 b-tag OS and SS regions.
- Uncertainties on the extrapolation of FFs derived in the 1 b-tag region to the 2 b-tag region.
- Uncertainties on the subtraction of $t\bar{t}$ and other processes in the 2 b-tag OS Anti-ID CR.
- Statistical uncertainties due to the finite number of events in the 2 b-tag OS Anti-ID CR.

The effect of FF statistical uncertainties on the multi-jet prediction in the SR is estimated by performing variations of FFs separately for all bins of the FF measurement. The discriminants used in this analysis are insensitive to variations of FFs in a single bin; therefore, all variations are combined into a single normalisation uncertainty. The resulting uncertainty on the multi-jet estimate in the SR is ±1.4 %.

The non-closure uncertainty from the comparison of FFs measured in the 1 b-tag SS and OS regions is estimated by assigning the full difference between the central values of both measurements as an

uncertainty on the FFs. The impact of this variation on the shape of distributions in the SR is taken into account in the background model. The differences between OS and SS FFs are generally small; thus, the non-closure uncertainty has only a minor impact of ±1.0 % on the normalisation of the multi-jet estimate.

The uncertainty on the extrapolation of FFs from the 1 b-tag to 2 b-tag regions is estimated by performing variations of the transfer factors within their statistical uncertainties. The variations are performed separately for the four categories of the transfer factor measurement. In a given category, the transfer factors for the three data-taking periods are varied coherently. As a result, the model of the multi-jet background receives four degrees of freedom from extrapolation uncertainties and any shape effects of these variations are propagated to the distributions of interest. The total uncertainty of these variations on the multi-jet normalisation in the SR is ±5.5 %.

Systematic uncertainties from the subtraction of non-multi-jet events in the 2 b-tag OS Anti-ID region are estimated by performing variations of the subtracted components. The subtractions of $t\bar{t}$ + true-$\tau_{\text{had-vis}}$, $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$, and other non-multi-jet contributions are varied separately. In all cases, the effect of the variations on the shape of the multi-jet prediction in the SR is taken into account in the background model.

The uncertainty resulting from the subtraction of $t\bar{t}$ + true-$\tau_{\text{had-vis}}$, which constitutes about 13 % of the total subtraction, is obtained by varying the normalisation of the subtracted $t\bar{t}$ + true-$\tau_{\text{had-vis}}$ template by its uncertainty. This uncertainty is estimated by performing variations of the modelling of $t\bar{t}$ in simulation (cf. Appendix A.3.1). The considered variations and their impact on the normalisation of the $t\bar{t}$ + true-$\tau_{\text{had-vis}}$ template in the 2 b-tag OS Anti-ID region are:

- Hard scatter and PS+NLO matching: ±6.0 %.
- Parton shower and hadronisation model: ±11 %.
- Renormalisation and factorisation scale: $^{+9.9}_{-9.4}$ % and $^{+2.8}_{-2.2}$ %.
- Initial-state and final-state radiation: $^{+0.53}_{-0.68}$ % and $^{+\ 5.4}_{-10}$ %.

The combination of these sources yields a normalisation uncertainty on the $t\bar{t}$ + true-$\tau_{\text{had-vis}}$ template of $^{+17}_{-19}$ %. Propagating this uncertainty to the multi-jet prediction in the SR yields an uncertainty of ±2.8 %. This uncertainty is small due to the small relative size of the $t\bar{t}$ + true-$\tau_{\text{had-vis}}$ subtraction.

The subtraction of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ accounts for 78 % of the total non-multi-jet subtraction. The uncertainty on the subtraction is defined using the SF measurement for anti-$\tau_{\text{had-vis}}$ previously described in Section 6.4.2. The SF variations explaining most of the variance of the SF measurement are used to define the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ subtraction uncertainty, yielding a total of four variations. Other variations of the measured SFs have negligible impact on the multi-jet prediction and are therefore omitted to reduce the number of parameters in the background model. An additional uncertainty is assigned according to the difference between the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ prediction with and without application of SFs to account for the reduced set of experimental systematic uncertainties available for the measurement of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ SF in the Anti-ID region. These five variations have a combined effect on the normalisation of the multi-jet prediction in the SR of ±7.4 %.

Other non-multi-jet processes account for approximately 9 % of the non-multi-jet subtraction in the 2 b-tag OS Anti-ID region. The dominant contributions are events from V + jets and single-top production, both processes contributing similarly to the subtraction. As a conservative estimate, the normalisation of the subtracted components is varied by ±50 % and propagated to the multi-jet

prediction in the SR. This variation changes the normalisation of the multi-jet estimate in the SR by ±5.9 %.

In addition to these systematic uncertainties, statistical uncertainties originating from the finite number of events in the 2 b-tag OS Anti-ID region and the statistical precision of the simulation-based subtraction of non-multi-jet processes are considered. This uncertainty is implemented in the fit model using the simplified Barlow–Beeston method [141, 142]. The effect of this uncertainty on the multi-jet normalisation in the SR is ±1.9 %.

A summary of all uncertainties affecting the multi-jet background prediction in the $\tau_{\text{had}}\tau_{\text{had}}$ SR is given in Table 6.13. The dominant sources of uncertainty on the multi-jet normalisation are the subtraction of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$, the subtraction of other non-multi-jet processes (V + jets and single-top), and the extrapolation of 1 b-tag FFs to 2 b-tag regions. The total uncertainty on the normalisation of the multi-jet estimate is approximately 12 %.

Table 6.13: Uncertainties on the normalisation of the multi-jet background prediction in the $\tau_{\text{had}}\tau_{\text{had}}$ SR. For a given source of uncertainty, the number of independent NPs in the background model is given in the "Components" column. Whether a given uncertainty affects the normalisation (N) and/or shape (S) of the multi-jet prediction is given in parentheses. †: Statistical uncertainties from the finite number of simulated and CR events are combined for all background processes.

Source	Components	Uncertainty
FF statistical uncertainty	1 (N)	±1.4 %
Non-closure of OS and SS FFs	1 (NS)	±1.0 %
1 to 2 b-tag extrapolation	4 (NS)	±5.5 %
$t\bar{t}$ + true-$\tau_{\text{had-vis}}$ subtraction	1 (NS)	±2.8 %
$t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ subtraction	5 (NS)	±7.4 %
Other subtraction	1 (NS)	±5.9 %
Statistical uncertainty	–†	±1.9 %
Total		±12 %

6.4.4 Fake-$\tau_{\text{had-vis}}$ Backgrounds in the $\tau_{\text{lep}}\tau_{\text{had}}$ Channel

The estimation of fake-$\tau_{\text{had-vis}}$ backgrounds in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels is outlined in the following. A data-driven background estimation technique yielding a combined estimate of the multi-jet and $t\bar{t}$ background with fake $\tau_{\text{had-vis}}$ is adopted. The method is an extension of the FF method that accounts for multiple sources of fake $\tau_{\text{had-vis}}$, differing in their process-specific FFs.

Events in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel in which the selected $\tau_{\text{had-vis}}$ candidate is an anti-$\tau_{\text{had-vis}}$ define the Anti-ID region used for the FF method. Two CRs are defined that are enhanced in multi-jet and $t\bar{t}$ events, respectively. Both regions can be divided into an ID and an Anti-ID region. These CRs are used to determine FFs specifically for fake $\tau_{\text{had-vis}}$ from multi-jet and $t\bar{t}$ events. The CR definitions and FF measurements are described in the following:

Multi-jet fake factors are measured in a region defined by the requirement that the electron/muon fails the loose isolation working point. Moreover, the $m_{bb} < 150\,\text{GeV}$ requirement is dropped. The remainder of the CR selection is identical to the SR selection of the $\tau_{\text{lep}}\tau_{\text{had}}$ channels. This

CR has high multi-jet purity and allows calculating multi-jet FFs, $\mathrm{FF}_{\text{multi-jet}}$, as the ratio of multi-jet events in the ID and Anti-ID region. The number of multi-jet events is estimated by subtracting the expected number of non-multi-jet events, which is estimated using simulation, from the observed number of events.

$t\bar{t}$ fake factors are measured in a region defined by requiring $m_{bb} > 150\,\text{GeV}$ with the other selections remaining identical to the SR selection. This CR has high $t\bar{t}$ purity but is not pure in $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events due to contributions of $t\bar{t}$ events with a true $\tau_{\text{had-vis}}$ that have to be subtracted. FFs for $t\bar{t}$, $\mathrm{FF}_{t\bar{t}}$, are calculated as the ratio of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events in ID and Anti-ID regions. The number of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events is estimated, assuming negligible contribution of multi-jet events, by subtracting the expected number of non-$t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events, which is estimated using simulation, from the observed number of events.

The FF measurement is performed separately for the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channels, and separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates. Moreover, the FFs are measured in bins of $\tau_{\text{had-vis}}$ candidate p_{T}.

When estimating the fake-$\tau_{\text{had-vis}}$ background in one of the $\tau_{\text{lep}}\tau_{\text{had}}$ SRs, it needs to be considered that the corresponding Anti-ID region consists of a mixture of multi-jet and $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events. Let $r_{\text{multi-jet}}$ be the fraction of fake-$\tau_{\text{had-vis}}$ backgrounds that originate from multi-jet events in the Anti-ID region. *Combined fake factors* are defined as the weighted combination of $\mathrm{FF}_{\text{multi-jet}}$ and $\mathrm{FF}_{t\bar{t}}$:

$$\mathrm{FF}_{\text{comb.}} = r_{\text{multi-jet}}\,\mathrm{FF}_{\text{multi-jet}} + (1 - r_{\text{multi-jet}})\,\mathrm{FF}_{t\bar{t}}\,.$$

The combined FFs can be applied to events with fake $\tau_{\text{had-vis}}$, irrespective of whether the event originates from multi-jet or $t\bar{t}$ processes, to yield the background estimate in the ID region. The determination of $r_{\text{multi-jet}}$ proceeds according to

$$r_{\text{multi-jet}} = \frac{N_{\text{multi-jet}}}{N_{\text{multi-jet}} + N_{t\bar{t}\text{ + fake-}\tau_{\text{had-vis}}}} = \frac{N_{\text{data}} - N_{\text{non-multi-jet}}}{N_{\text{data}} - N_{\text{non-(multi-jet or }t\bar{t}\text{ + fake-}\tau_{\text{had-vis}})}}\,, \tag{6.2}$$

where N_{data} is the number of events observed in the Anti-ID region and N_p the number of events expected from process p in the Anti-ID region. The subtractions on the right-hand side of Equation (6.2) use the expected number of events predicted using simulation. This includes the subtraction of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events in the numerator.

The determination of $r_{\text{multi-jet}}$ is performed separately for the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channels, 1- and 3-prong $\tau_{\text{had-vis}}$ candidates, and events containing electrons and muons. In addition, $r_{\text{multi-jet}}$ is determined in bins of p_{T} of the $\tau_{\text{had-vis}}$ candidate. In the SLT channel, $r_{\text{multi-jet}}$ is typically small or zero showing that the majority of fake-$\tau_{\text{had-vis}}$ backgrounds originate from $t\bar{t}$. The multi-jet contribution in the LTT channel is larger with $r_{\text{multi-jet}}$ ranging from 10 to 30 % depending on $\tau_{\text{had-vis}}$ candidate p_{T} and N_{tracks}. Uncertainties on the $r_{\text{multi-jet}}$ estimate have little impact on the fake-$\tau_{\text{had-vis}}$ background prediction since $\mathrm{FF}_{\text{multi-jet}}$ and $\mathrm{FF}_{t\bar{t}}$ tend to be of similar size in most bins.

The use of a similar method in the $\tau_{\text{had}}\tau_{\text{had}}$ channel would be preferred compared to separately estimating the fake-$\tau_{\text{had-vis}}$ background from multi-jet and $t\bar{t}$. The combined FF method does not need to distinguish between events with fake $\tau_{\text{had-vis}}$ from multi-jet and $t\bar{t}$ when applying the FFs to events in the Anti-ID region. In contrast, the multi-jet estimate in the $\tau_{\text{had}}\tau_{\text{had}}$ channel requires a large subtraction of $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ events, which is a dominant source of systematic uncertainty. In the combined FF method, this uncertainty is restricted to an uncertainty on $r_{\text{multi-jet}}$. Despite possibly large

uncertainties on $r_{\text{multi-jet}}$, the uncertainty on the fake-$\tau_{\text{had-vis}}$ background estimate from the combined FF method would be small since $\text{FF}_{\text{multi-jet}}$ and $\text{FF}_{t\bar{t}}$ are of similar size. The search presented in this thesis is not limited by uncertainties related to the fake-$\tau_{\text{had-vis}}$ background estimation and therefore this approach was not pursued. In the future, systematic uncertainties will become more relevant at which point the combined FF method should also be considered for the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

6.5 Multivariate Analysis

The event selection described in Section 6.3 serves to select events compatible with $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ or $b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$ final states and ensures that reconstructed objects fulfil basic kinematic and identification requirements. The signal-to-background ratio at this level of selection, with background being three orders of magnitude more abundant than the expected SM HH signal in the $\tau_{\text{had}}\tau_{\text{had}}$ channel (cf. Table 6.5), is not sufficient to provide any signal sensitivity without further signal-background discrimination.

Events from SM HH or resonant HH production have distinct kinematic properties that can be used for signal-background discrimination. A number of reconstructed quantities are defined that are sensitive to the differences between signal and background processes. These quantities are exploited using multivariate methods for event classification. Depending on the HH production mode and analysis channel, different methods are used.

The search for SM HH production uses BDTs and NNs to distinguish signal from background events in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels, respectively. In the search for resonant HH production, multiple hypotheses are considered for the mass of the scalar resonance. The kinematic properties of events from resonant HH production depend on the mass of the resonance; therefore, the classification task varies continuously with m_X. This is in contrast to the SM HH case in which the event kinematics follow from a fixed distribution. Classification tasks that vary as a function of a parameter, for example the mass of a resonance, can be performed by *Parameterised Neural Networks* (PNNs) [260]. PNNs are therefore used as discriminants in the search for resonant HH production in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels.

The scores provided by the multivariate classification methods, hereafter called MVA scores, are used in Section 6.7 as discriminants in maximum likelihood fits to extract the signal of interest and to set upper limits on signal strengths and cross sections. During the development of the analysis, regions at high MVA scores were blinded.

In Section 6.5.1, the choice of discriminating variables used to classify signal and background events is motivated. Section 6.5.2 introduces a method used to train, optimise, and evaluate classifiers that ensures that the predicted MVA scores are unbiased and can be used in the statistical interpretation of the search results. This method is employed in Section 6.5.3 and Section 6.5.4 to obtain the BDT- and PNN-based classifiers for signal-background discrimination in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, a similar approach of multivariate analysis is adopted, which is documented in Ref. [192] and is not discussed in detail in this thesis.

6.5.1 Discriminating Variables

The set of variables provided to multivariate classification methods is critical to their performance in distinguishing between classes. The initial choice of variables considered in this search is based on

the previous publication by the ATLAS collaboration in the same analysis channel [184]. Only minor changes to the input variable selection are performed for this search.

The four-momenta of $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ candidates are reconstructed[13] and used to define discriminating variables. Among the most important variables are H and HH invariant masses:

m_{bb} The invariant mass of the $H \to b\bar{b}$ candidate. It is reconstructed using the four-momenta of both b-tagged jets in the SRs after b-jet momentum corrections. The Higgs boson mass is reconstructed with a resolution of 13 to 18 GeV for the signal processes considered in this search.

$m_{\tau\tau}^{\text{MMC}}$ The invariant mass of the $H \to \tau^+\tau^-$ candidate reconstructed by the MMC. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the mass resolution of the MMC ranges from 15 to 18 GeV depending on the signal process.

m_{HH} The invariant mass of the pair of Higgs bosons in signal events. It is determined from the sum of four-momenta of the b-jet candidates after momentum corrections and the $\tau^+\tau^-$-system four-momentum reconstructed using the MMC. For signal events in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the relative mass resolution is 8 to 10 %.

The bias and resolution of the H and HH mass reconstruction is summarised in Figure 6.23 for events from resonant HH production as a function of the resonance mass. The three invariant masses are used as inputs to the classifiers in all channels.

Higgs bosons from SM HH production or resonant HH production (with $m_X \gg 2m_H$) are typically produced with large momentum in the HH rest frame. As a result, the angular separation of the Higgs boson decay products tends to be small. The distances $\Delta R(\tau, \tau)$ and $\Delta R(b, b)$ between the visible decay products of the τ leptons (e, μ, $\tau_{\text{had-vis}}$) and the b-jet candidates, respectively, are reconstructed. These variables provide discrimination power against backgrounds from multi-jet and top-quark pair production in which $\Delta R(\tau, \tau)$ and $\Delta R(b, b)$ are typically larger. Both variables are used as inputs to the classifiers in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel.

In a given analysis channel, the same set of input variables is used for the search for SM HH and resonant HH production. The choice of variables differs between channels and is summarised in Table 6.14. The five discriminating variables used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are shown in Figure 6.24.

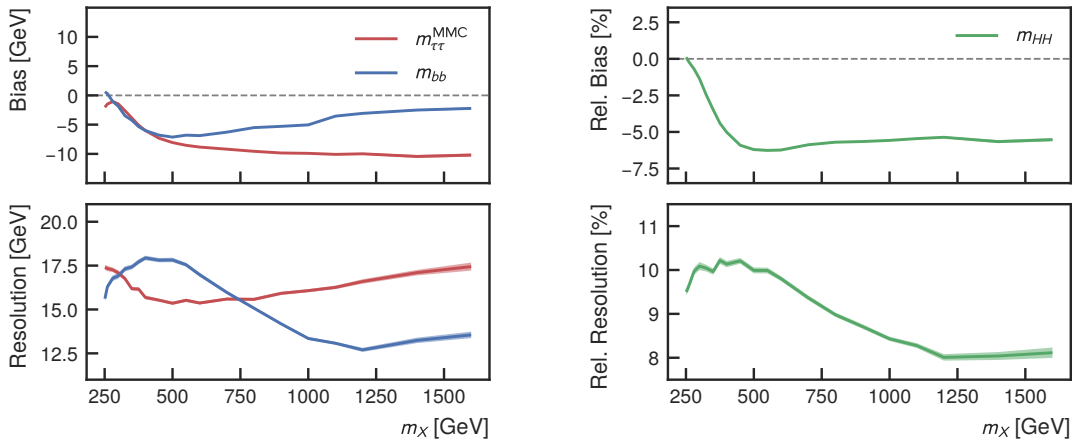
6.5.2 Cross-Validation Method

Machine learning algorithms are often susceptible to fitting statistical fluctuations in the data that are used to train a model. As a result, predictions of performance characteristics of the model based on the training data might not generalise to unseen data. In extreme cases, frequently called overfitting, the performance of the model evaluated on an independent dataset starts to degrade when further increasing the capacity of the model [149].

When using machine learning methods in searches for new physics, it has to be ensured that the methods are evaluated on datasets that were not used for training or model selection,[14] thus providing

[13] In the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the fraction of events with a misreconstructed $H \to b\bar{b}$ or $H \to \tau^+\tau^-$ candidate, i.e. b-jet candidates or $\tau_{\text{had-vis}}$ not being matched to Higgs boson decay products at generator-level, are about 2 % and 0.2 %, respectively.

[14] Model selection refers to the process of choosing a model from a set of models based on an evaluation metric.



(a) H mass reconstruction

(b) HH mass reconstruction

Figure 6.23: Performance of the H (a) and HH (b) invariant mass reconstruction in the $\tau_{\text{had}}\tau_{\text{had}}$ SR for simulated $pp \to X \to HH$ events. The top panel shows the bias of the mass reconstruction in terms of the median deviation of the reconstructed mass from its true value. The bottom panel shows the mass resolution estimated as half of the length of the central interval containing 68 % of deviations between the reconstructed and true masses. For the HH invariant mass reconstruction, the relative deviation is used to define the relative bias and relative resolution.

Table 6.14: Input variables of the classifiers used in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels. The same variables are used for the search for SM HH and resonant HH production. Definitions of the input variables used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels are given in Appendix A.2.1. The table is adapted from Ref. [192].

	Analysis channel		
Variable	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
$m_{\tau\tau}^{\text{MMC}}$	✓	✓	✓
m_{bb}	✓	✓	✓
m_{HH}	✓	✓	✓
$\Delta R(\tau,\tau)$	✓	✓	✓
$\Delta R(b,b)$	✓	✓	
$\Delta p_{\text{T}}(\ell, \tau_{\text{had-vis}})$		✓	✓
Sub-leading b-jet p_{T}		✓	
m_{T}^{W}		✓	
$p_{\text{T}}^{\text{miss}}$		✓	
$p_{\text{T}}^{\text{miss}}$ ϕ centrality		✓	
$\Delta\phi(\ell\tau_{\text{had-vis}}, bb)$		✓	
$\Delta\phi(\ell, \boldsymbol{p}_{\text{T}}^{\text{miss}})$			✓
$\Delta\phi(\boldsymbol{p}_{\tau\tau}^{\text{MMC}}, \boldsymbol{p}_{\text{T}}^{\text{miss}})$			✓
s_{T}			✓

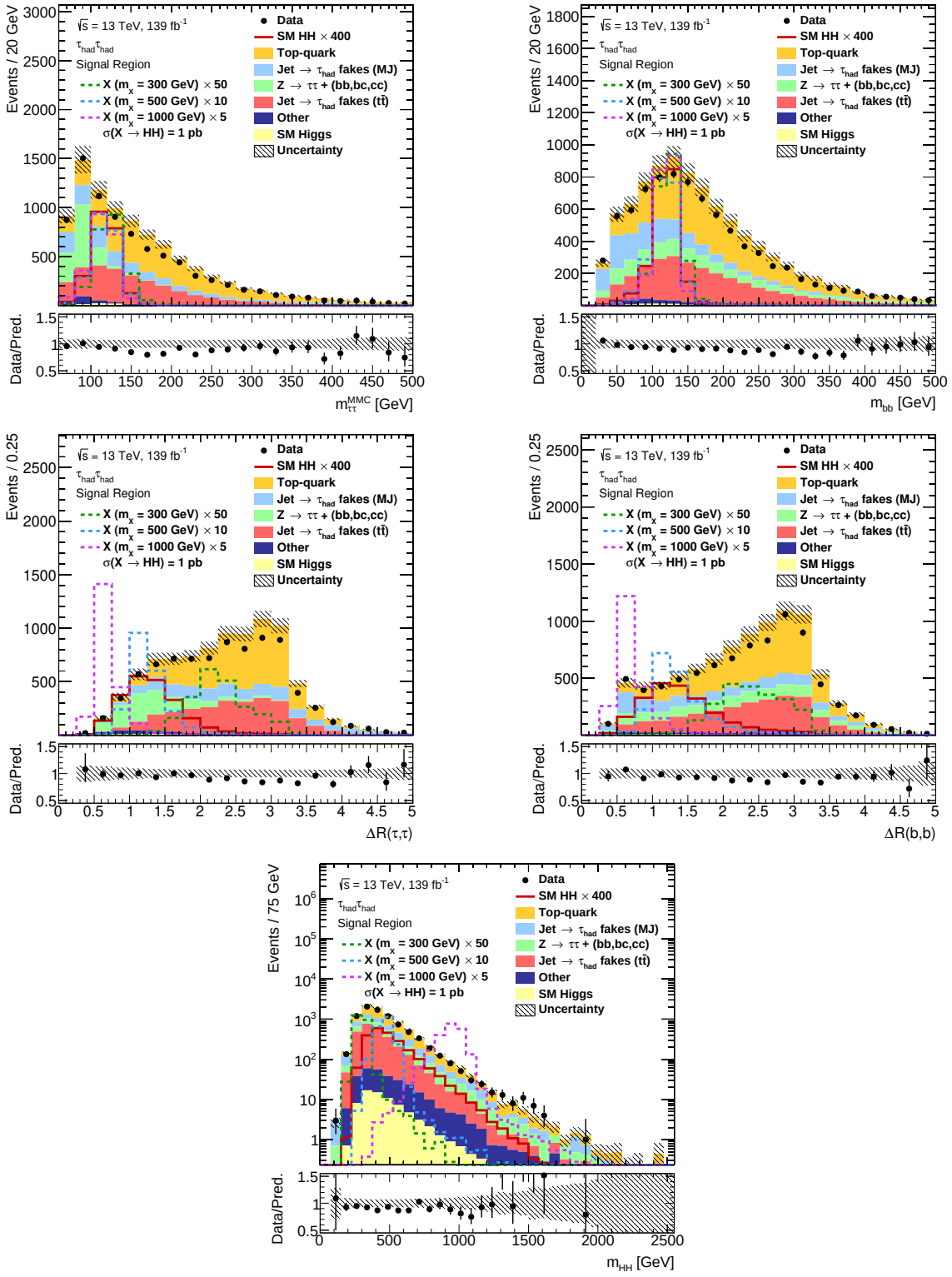


Figure 6.24: Distributions of the MVA input variables in the $\tau_{\text{had}}\tau_{\text{had}}$ SR prior to the fit. The uncertainty bands include all statistical and systematic uncertainties. The normalisation of signals from SM HH and resonant HH production are scaled for visibility.

an estimate of the generalised performance. In this analysis, events are categorised according to their event number into even- and odd-numbered events. This two-fold split of events yields an even- and odd-fold, respectively. The training and model selection can proceed by using one of the two folds, withholding the other fold for later evaluation. This procedure is applied twice by using each fold for training and model selection once.

Predictions of the MVA scores are obtained by evaluating the model trained on the even- on odd-numbered events and the model trained on the odd- on even-numbered events. This approach, which is called 2-fold *cross-validation* [149, 261], provides unbiased predictions of the MVA scores for the entirety of the available dataset. The same evaluation method is applied to SR data recorded by the ATLAS detector. After assigning MVA scores to all events, no distinction is made between even- and odd-numbered events for the remainder of the analysis.

Similar to the biased predictions obtained when evaluating machine learning methods on datasets used for training, the process of model selection needs to be performed on a dataset that is independent of the one used for final evaluation [262]. In the case of this analysis, model selection primarily applies to the determination of the hyperparameters of the classification algorithms based on a performance metric. Making this choice dependent on the performance on the withheld dataset can introduce a selection bias when using the same dataset for the statistical interpretation.

In this search, model selection is performed using 5-fold cross-validation (CV), a generalisation of the 2-fold approach to a larger number of subdivisions, separately on even- and odd-numbered events. This approach effectively nests 5-fold CV inside 2-fold CV and is therefore called *nested cross-validation* [262, 263].

Figure 6.25 shows a schematic description of one iteration of the nested cross-validation approach. The inner 5-fold CV randomly partitions events into five folds. For every choice of evaluation-fold, the model is trained on the remaining four folds and subsequently evaluated on the evaluation-fold. A decision between two competing models can be made by comparing the average and standard deviation of an evaluation metric over the five iterations of the inner CV. Only after the best-performing model is selected, usually after re-fitting the model on the combination of all five folds, it is evaluated on the hold-out dataset.

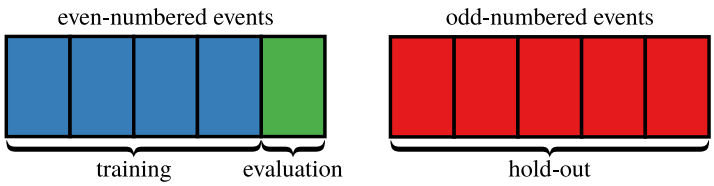


Figure 6.25: Illustration of the 5-fold cross-validation approach for model selection on even-numbered events. The separation of events into disjoint subsets (folds) is indicated by rectangles. A single step out of a total of five, the number of possible assignments of the evaluation-fold, is shown. The hold-out dataset consisting of odd-numbered events is not used when performing model selection on even-numbered events.

6.5.3 Extraction of Signals from SM HH Production in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

The search for SM HH production in the $\tau_{\text{had}}\tau_{\text{had}}$ channel uses BDTs to separate signal from background events. Events passing the SR selection of the $\tau_{\text{had}}\tau_{\text{had}}$ channel are used to train BDT. The training considers non-resonant HH production via ggF as the signal class; the combination of all backgrounds

as the background class. The individual background processes, which are estimated from simulation or CR data (multi-jet background), are weighted according to their relative cross sections. SM HH production via VBF is not included in the training. The BDT implementation of TMVA [153] is used.

Hyperparameter Optimisation

The BDT configuration is optimised using a random search over a grid of parameter values. For every hyperparameter, a set of values to test is defined. All possible combinations of hyperparameter values define a grid from which configurations are drawn randomly. The performance of the model configuration is estimated using 5-fold cross-validation separately on even- and odd-numbered events.

The hyperparameter values considered for the optimisation are listed in Table 6.15. The total weight of signal and background events in the BDT training are ensured to be equal by rescaling of the event weights prior to training. For every tree branching, 400 possible cuts on input variables are considered. Other parameters remain at their default values.

Table 6.15: Hyperparameter values considered in the optimisation of the BDT for the SM HH search. The underlined values show the configuration after optimisation.

Hyperparameter	Values considered
Number of trees	200, 400, 800, 1000, 1500, 2000
Tree depth	1, 2, 3
Minimum node size	0.01 %, 0.1 %, 1 %, 5 %
Boosting algorithm	Gradient Boosting, AdaBoost [264]
Learning rate	0.01, 0.02, 0.04, 0.08, 0.1, 0.15, 0.2, 0.3, 0.4
Ignore negatively weighted events in training	Yes, No

The metric used for optimisation is the area under the receiver operating characteristic curve (ROC-AUC). Here, the ROC curve is defined as the parametric curve given by $(x, y) = \big(\varepsilon_\mathrm{s}(t), 1 - \varepsilon_\mathrm{b}(t)\big)$, where t is a threshold applied to the score of a classifier and ε_s (ε_b) the signal (background) efficiency of this selection. A value of the ROC-AUC of 1 indicates perfect classification, while a value of 0.5 corresponds to an uninformative classifier. The ROC-AUC is chosen as the metric to be optimised as it summarises the classifier's performance over all possible working points, i.e. choices of thresholds on the classifier output [265].

About 1600 BDT configurations are tested during the model selection process. The steps described in the following are applied separately to the datasets containing even- and odd-numbered events. The 5-fold CV approach described in Section 6.5.2 is used to evaluate the performance of a given hyperparameter configuration. For every BDT configuration, the ROC-AUC average and standard deviation is calculated over the five iterations of the inner CV.

Out of all evaluated configurations, the ROC-AUC of the 200 best performing model configurations are statistically indistinguishable based on the CV results. In the absence of a clearly preferred configuration, the highest ranking configuration with the smallest maximum tree depth is selected. A smaller tree depth, effectively limiting the number variable interactions used in the classifier [149], is chosen to be less reliant on the quality of the modelling of higher-order variable interactions in the training data.

The selected configuration, shown underlined in Table 6.15, is the 5$^{\text{th}}$ and 7$^{\text{th}}$ highest ranking one with a ROC-AUC of 0.9803 ± 0.0008 and 0.9787 ± 0.0012 in CV on even- and odd-numbered events, respectively. In both cases, the same configuration ranks highest among models with a maximum tree depth of 2.[15] After choosing the configuration, the classifiers are re-trained on all five folds of the inner CV, providing the final classifiers to extract the SM HH signal.

Evaluation and Variable Importance

In Figure 6.26, the combined distribution of both BDTs, which are evaluated on events withheld from training and model selection, is shown in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The BDT score provides good separation power between the SM HH signal and most background processes. The ROC-AUC of the final evaluation is 0.9772 ± 0.0004 (0.9781 ± 0.0004) when including (excluding) SM HH production via VBF, similar to the estimates from cross-validation during model selection.

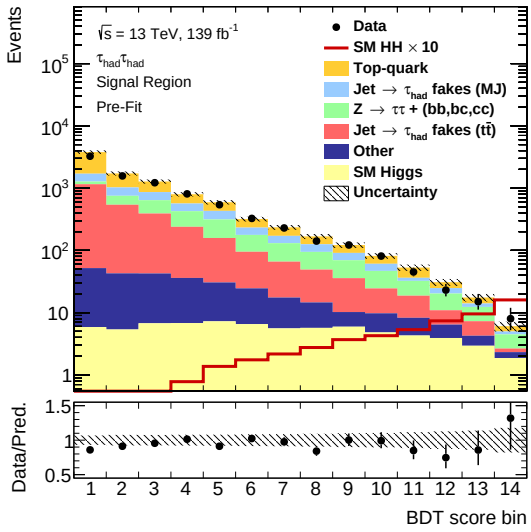


Figure 6.26: Distribution of the BDT discriminant in the $\tau_{\text{had}}\tau_{\text{had}}$ SR prior to the fit. The uncertainty bands include all statistical and systematic uncertainties. The normalisation of the SM HH signal is scaled by a factor of 10 for illustration. The choice of binning is discussed in Section 6.7.1.

The sensitivity to the SM HH signal is driven by the last bins of the BDT score histogram. The expected number of signal (background) events in the two most signal-like bins is 2.6 (24) out of 5.6 (9200) events entering the SR. The selection of the two most signal-like bins provides a background rejection of $1/\varepsilon_{\text{b}} \approx 380$ while selecting almost half of the signal events.

The single largest background in the last two BDT score bins, with an expectation of 6.9 events, is the associated production of $Z \to \tau^+\tau^-$ with jets originating from quarks of heavy flavour. The production of single Higgs bosons represents the second most abundant background in the last two bins of the final discriminant with an expectation of 4.8 events. The primary source of single-Higgs-boson backgrounds is $H \to \tau^+\tau^-$ with approximately equal contribution from the ggF, ZH, and $t\bar{t}H$ production modes. A

[15] Model optimisation using the nested cross-validation method outlined in the text typically yields two different hyperparameter configurations. One configuration for the even- and one for the odd-fold.

small fraction of 15 % is originating from $H \to b\bar{b}$ in associated production with a Z boson. Other backgrounds populating the two most signal-like bins in BDT score are originating from $t\bar{t}$ (with true $\tau_{\text{had-vis}}$) with an expectation of 4.6 events and jet $\to \tau_{\text{had-vis}}$ from multi-jet and $t\bar{t}$ production with 2.2 and 3.4 expected events, respectively. The signal and background yields, including all experimental and theoretical uncertainties, are summarised in Section 6.7.

The importance of input variables in the BDT can be estimated using the *permutation importance* technique. It is a method to inspect the importance of input variables for predictions of a black box estimator, derived from an importance measure introduced in Ref. [266] for random forests. To inspect the importance of a feature in a given model, the feature's values are permuted over all events and classes, breaking the relationship between the feature, class labels, and other correlated input variables. The importance of a given feature is estimated by measuring the degradation in the quality of the model's predictions after permuting a given feature.

This technique measures the importance of a variable in a given model, which does not necessarily correspond to the importance of the variable in solving the underlying predictive problem. In the presence of highly collinear features, this means that some importance can be assigned to multiple related variables even if a strict subset of variables contains the information relevant to the problem. This effect needs to be considered when interpreting rankings based on the permutation importance.

In Table 6.16, a ranking of the BDT input variables is shown based on the change in ROC-AUC using the permutation importance technique. The H-system masses are the most important inputs to the BDT, contributing with approximately equal importance due to similar mass reconstruction performance of $m_{\tau\tau}^{\text{MMC}}$ and m_{bb}. For the non-resonant SM HH search, the HH-system mass is of lesser importance due to the similarities of the m_{HH} spectra between signal and background (cf. Figure 6.24).

Table 6.16: Importance of the BDT input variables measured as the change in ROC-AUC when permuting the values of a single variable over all events. The mean ΔROC-AUC over 10 permutations is displayed. The statistical uncertainty is below 0.001 and therefore omitted. Variables are ordered from most to least important.

Variable	ΔROC-AUC
$m_{\tau\tau}^{\text{MMC}}$	−0.090
m_{bb}	−0.085
m_{HH}	−0.034
$\Delta R(\tau, \tau)$	−0.033
$\Delta R(b, b)$	−0.012

6.5.4 Extraction of Signals from Resonant HH Production in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel

The search for resonant HH production considers scalar resonances with masses ranging from 251 to 1 600 GeV, probing a wide range of kinematic configurations of final state particles. Consequently, the joint distribution of the discriminating variables for signal processes varies with the assumed mass of the resonance. This is particularly visible in the marginal distributions of m_{HH}, $\Delta R(\tau, \tau)$, and $\Delta R(b, b)$, previously shown in Figure 6.24 for three m_X hypotheses, where the overlap between the signal and background spectra changes as m_X is varied. For optimal signal sensitivity for all assumed m_X, this dependency should be exploited.

Using multivariate classifiers for signal extraction, one possible method of incorporating the m_X dependency of the classification problem is to train a classifier for every signal hypothesis. In this approach, multiple classification tasks are solved in isolation, ignoring the more general classification problem. This was previously explored in Ref. [184] using BDTs.

An alternative approach is provided by parameterised classifiers where the dependency of the prediction problem is incorporated during training, thus solving it in a broader context. In particular, PNNs are used due to the ability of NNs to smoothly approximate large classes of continuous functions [260]. As a result of these properties, it is shown in Ref. [260] that PNNs are able to interpolate the classification task to parameter values not seen in training. It is also suggested that parameterised classifiers may outperform approaches of using one classifier per parameter value due to their ability to solve the more general, parameter-dependent classification problem.

In this search, PNNs are implemented as feedforward NNs where, in addition to the five discriminating variables, the parameter value specifying the classification task is provided as an input to the network. During training, the value of the parameter is assigned to be the generator-level m_X for signal events and random, uninformative values for background events. Otherwise, PNNs are amenable to the methods commonly employed to train non-parameterised feedforward NNs. At evaluation time the parameter value is held fixed according to the classification problem to be solved.

Implementation and Hyperparameter Optimisation

The PNNs are trained using simulated signal events of 19 different mass hypotheses of the scalar resonance. An additional point at $m_X = 375\,\text{GeV}$ was included at a later stage of the analysis and was not part of the training and optimisation process. The background processes considered in the training are the same as the ones used for the SM HH search. Moreover, SM HH production is not included as a background. The signal event samples are combined, ensuring that every m_X hypothesis contributes with the same total event weight to the combined sample, subsequently normalising the combined signal sample and the background sample to have equal weight. The 5-fold CV approach is used for model selection.

The training proceeds by minimising the binary cross-entropy loss using SGD with momentum and exponential learning rate decay. As part of the training process, the discriminating variables are centred and scaled by subtracting the median and dividing by the interquartile range of the variable for better conditioning of the loss minimisation. Similarly, the values of the mass parameter are (linearly) transformed into $[0, 1]$. The values of the mass parameter for background events is sampled from the m_X distribution in the combined signal sample and is re-sampled after every training epoch. The PNNs consist of multiple fully-connected layers with ReLU activation [173], except for the final layer which uses sigmoid activation. The training is implemented in Keras [162] using the TensorFlow [172] backend. Trained PNNs are evaluated using lwtnn [174].

The PNN hyperparameters are optimised, following the approach previously employed for the BDT, using a random grid search with nested cross-validation. The parameter grid is defined in Table 6.17. To reduce the dimensionality of the hyperparameter space, the number of nodes in the hidden layers except for the first and last hidden layers are required to be the same.

No clear choice of optimisation metric exists to evaluate a continuum of related classification task. For simplicity, the ROC-AUC of the PNN when performing binary classification of a signal with $m_X = 325\,\text{GeV}$ against the background is chosen. This choice is motivated by the strength of the $X \to HH \to b\bar{b}\tau^+\tau^-$ search channel in an intermediate mass range of the scalar resonance from

Table 6.17: Hyperparameter values considered in the optimisation of the PNN for the search for resonant HH production. Parameters marked with $*$ and † are only applicable when the number of hidden layers is larger than 1 and 2, respectively. The underlined values show the configuration after optimisation.

Hyperparameter	Values considered
Epochs	50, 100, <u>200</u>, 400
Batch size	64, <u>128</u>, 256
Learning rate	0.01, 0.02, 0.05, 0.1, <u>0.2</u>
Learning rate decay	10^{-6}, <u>10^{-5}</u>, 10^{-4}, 10^{-3}
Number of hidden layers	1, 2, 3, <u>4</u>, 5
Layer size (first hidden layer)	16, 32, 64, <u>128</u>
Layer size (last hidden layer*)	<u>16</u>, 32, 64, 128
Layer size (other hidden layers†)	16, 32, 64, <u>128</u>

approximately 300 GeV to 800 GeV compared to the $b\bar{b}\gamma\gamma$ and $b\bar{b}b\bar{b}$ channels, which are expected dominate the low and high m_X signal sensitivity, respectively. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, background events with m_{HH} around 325 GeV are most abundant (cf. Figure 6.24) and therefore a value of m_X at the lower end of the range is selected for optimisation.

About 1600 configurations of the PNNs are tested, the 400 highest ranking ones in ROC-AUC showing compatible performance in terms of the performance estimate obtained from CV. A single configuration is obtained by choosing the best performing parameter set in CV on even-numbered events and using the same configuration for odd-numbered events. The chosen configuration, underlined in Table 6.17, has a ROC-AUC in 5-fold CV of 0.9764 ± 0.0009 on even- and 0.9754 ± 0.0009 on odd-numbered events.[16] The PNNs are then re-fit on all five folds of the inner CV separately for even- and odd-numbered events.

Evaluation and Variable Importance

Figure 6.27 shows the PNN score after evaluation on withheld data in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel for four different values of the mass parameter. In each case, the signal of interest, resonant HH production with m_X equal to the PNN mass parameter, is populating the high PNN score bins. The background processes contributing to the most signal-like bins depend on the classification task that is solved by the PNN. At low mass ($m_X \approx 300$ GeV) the dominant background is $t\bar{t}$ with both true and fake $\tau_{\text{had-vis}}$ constituting about 80 % of the total background with a small contribution from multi-jet. The background composition in the intermediate mass range ($m_X \approx 500$ GeV) is similar to the case of non-resonant HH production in the SM, which has a mean reconstructed m_{HH} of about 500 GeV after the SR selection. For high mass resonances, the most relevant backgrounds are the production of $Z \to \tau^+\tau^-$ in association with jets from b or c quarks.

The previously discussed properties of PNN, the ability to solve a continuously varying classification task parameterised by m_X and the ability to interpolate to values of m_X not seen during training,

[16] On the dataset with odd-numbered events the chosen configuration is the 54$^{\text{th}}$ highest ranking one in terms of the ROC-AUC. For comparison, the highest ranking model on odd-numbered event has a ROC-AUC of 0.9761 ± 0.0020, showing compatible performance with the chosen configuration.

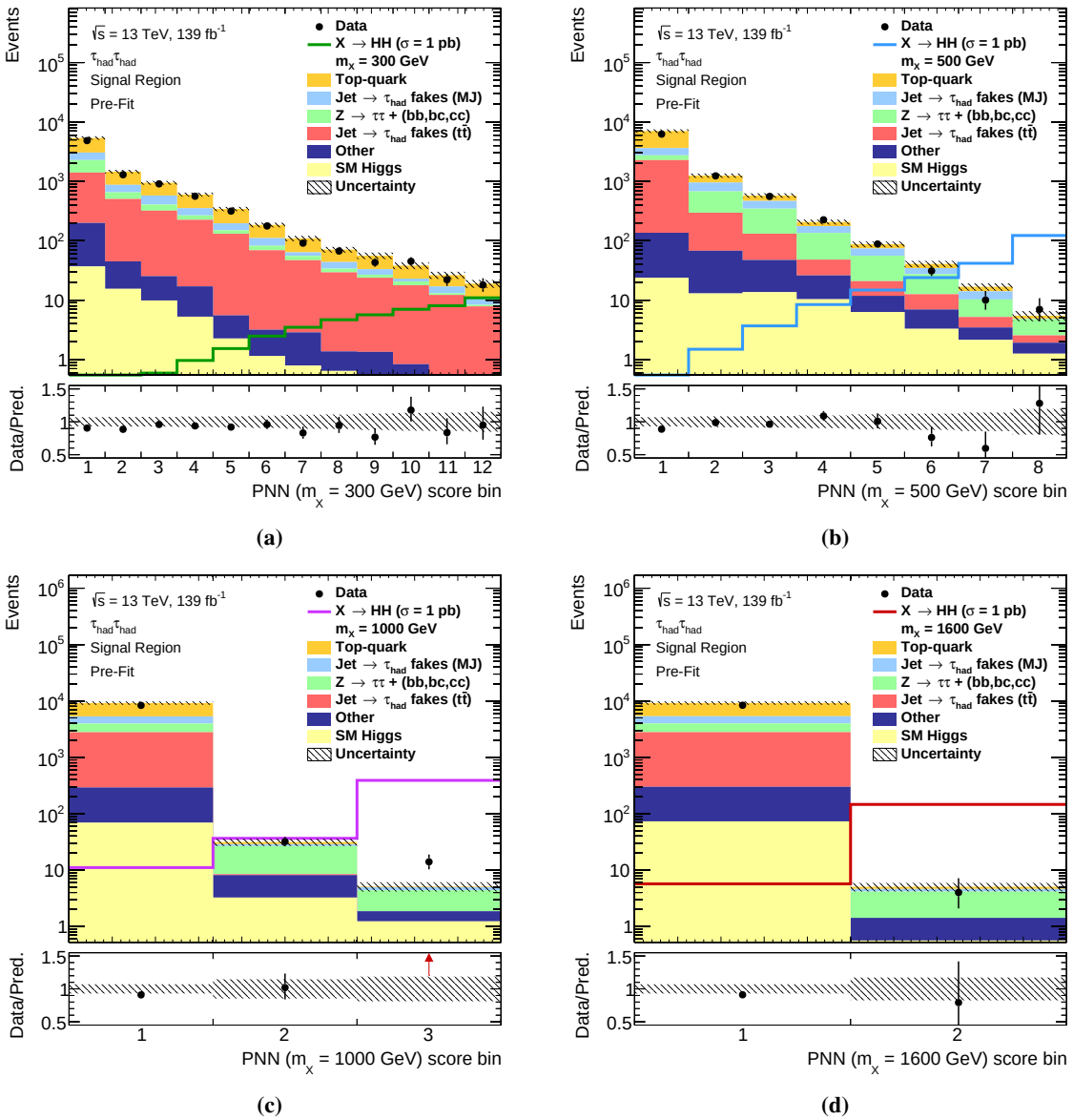
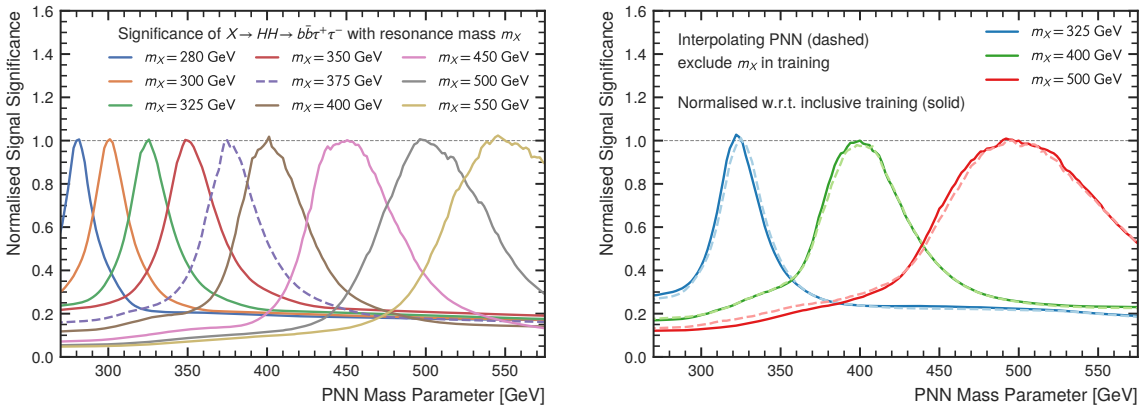


Figure 6.27: Distributions of the PNN discriminants evaluated with PNN mass parameters set to 300 GeV (a), 500 GeV (b), 1 000 GeV (c), and 1 600 GeV (d) in the $\tau_{\text{had}}\tau_{\text{had}}$ SR prior to the fit. The signal overlay is normalised to $\sigma(pp \to X \to HH) = 1$ pb. The choice of binning and excess of data in the most signal-like bin of the PNN ($m_X = 1\,000$ GeV) score is discussed in Section 6.7.

are investigated. For this, a performance metric is defined by binning the PNN score for a given value of the mass parameter using an algorithm that optimises the signal sensitivity while obeying certain constraints. The binning algorithm is formally introduced in Section 6.7.1. The expected signal significance is approximated, assuming independent Poisson counting experiments without uncertainties on the background predictions, by $\nu_s/\sqrt{\nu_b}$ for every bin, where ν_s and ν_b is the expected number of signal and background events, respectively. A combined significance is obtained by adding the bin-wise significances in quadrature.

The combined significance is used in Figure 6.28(a) to inspect the change in signal sensitivity for a given benchmark signal as the PNN mass parameter is varied. The largest expected significance is obtained when the mass parameter is set close to the resonance mass of the signal hypothesis. In Figure 6.28(b), the ability of PNNs to interpolate to classification tasks that were not part of the training is shown. The signal significance is compared between a PNN excluding and including a given signal hypothesis in training. The comparison shows similar performance in both cases. This property motivates the use of PNNs also for resonance masses that were not part of the training, which was exploited to include an additional signal with $m_X = 375\,\text{GeV}$ without retraining of the PNNs. Finally, the performance of the PNNs is compared to a strategy of using dedicated BDTs for every signal mass hypothesis. The comparison is performed for resonance masses of 300, 500, and 1 000 GeV using BDTs that are optimised following the approach used for the BDTs in the SM HH search. For these three values of m_X, the use of PNNs improves the expected signal significance by 22 %, 9 %, 4 %, respectively, over the approach of using dedicated classifiers.



(a) Response of PNNs used in the search for resonant HH production.

(b) Comparison of PNNs before hyperparameter optimisation when including (solid) and excluding (dashed) certain resonance masses in training.

Figure 6.28: Expected signal significance of a scalar resonance with mass m_X as a function of the PNN mass parameter. The significance is estimated by binning the PNN score for a given value of the parameter and adding the signal significance of all bins in quadrature. The curves are normalised such that the significance is 1 when the PNN mass parameter is equal to m_X of the hypothesis under test. Dashed lines correspond to signal hypotheses not included in the PNN training.

A ranking of the variable importance is provided in Table 6.18 based on the permutation importance technique. The H-system masses continue to be important discriminants to reject backgrounds relevant to searches of resonances in the low to intermediate mass range. The reconstructed mass of the

HH-system provides an important discriminant over the entire mass range, becoming the highest ranked PNN input for $m_X > 500\,\text{GeV}$.

Table 6.18: Importance of the PNN input variables measured as the change in ROC-AUC when permuting the values of a single variable over all events. The mean ΔROC-AUC over 10 permutations is displayed. The statistical uncertainty is below 0.002 for (a) and 0.001 for (b) and (c) and thus omitted. Variables are ordered by descending importance.

(a) $m_X = 300\,\text{GeV}$

Variable	ΔROC-AUC
$m_{\tau\tau}^{\text{MMC}}$	−0.335
m_{bb}	−0.317
m_{HH}	−0.167
$\Delta R(\tau,\tau)$	−0.123
$\Delta R(b,b)$	−0.029

(b) $m_X = 500\,\text{GeV}$

Variable	ΔROC-AUC
m_{bb}	−0.183
$m_{\tau\tau}^{\text{MMC}}$	−0.176
m_{HH}	−0.171
$\Delta R(\tau,\tau)$	−0.011
$\Delta R(b,b)$	−0.005

(c) $m_X = 1\,000\,\text{GeV}$

Variable	ΔROC-AUC
m_{HH}	−0.187
m_{bb}	−0.008
$\Delta R(\tau,\tau)$	−0.007
$m_{\tau\tau}^{\text{MMC}}$	−0.006
$\Delta R(b,b)$	−0.002

The sensitivity of the signal extraction method to the mass of the resonance decreases with increasing m_X, as is indicated by the width of the curves in Figure 6.28(a). This decrease in mass sensitivity has two primary sources. First, the m_{HH} resolution decreases approximately linearly with m_X. Second, the PNN has high separation power between events from background processes and resonant HH production at high mass.[17] Due to the large separation power, the mass sensitivity in the high mass regime is limited by the width of bins in the high PNN score region, the bin width being driven by constraints imposed on the binning algorithm. In these cases, the signal extraction method employed in this search shows worse mass sensitivity than suggested by the resolution of the m_{HH} reconstruction itself. Consequently, the approach of using the score of a multivariate classifier as a final discriminant in this search is optimised for the discovery of a signal as opposed to perform a measurement of the resonance mass.

6.6 Systematic Uncertainties

The systematic uncertainties affecting the search for non-resonant and resonant Higgs boson pair production are discussed in the following. Experimental uncertainties are described in Section 6.6.1 but exclude uncertainties related to the estimation of fake-$\tau_{\text{had-vis}}$ backgrounds in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. These were previously described in Sections 6.4.2 and 6.4.3. Theory uncertainties are described in Section 6.6.2 and include uncertainties on the modelling of physics processes using MC simulations.

The searches for Higgs boson pair production presented in this thesis are generally limited by statistical uncertainties due to the finite number of events observed in the SRs. Theory uncertainties play a lesser role and instrumental uncertainties are only relevant for searches for scalar resonance with low masses ($m_X \lesssim 300\,\text{GeV}$). The impact of statistical and systematic uncertainties on the results are discussed in the context of the statistical interpretation in Section 6.7.

[17] The discrimination power of the PNN as a function of the m_X of the signal hypothesis is shown in Figure A.4 in the appendix.

6.6.1 Experimental Uncertainties

Experimental uncertainties arise from the measurement of the integrated luminosity, the re-weighting of the pile-up conditions in simulation, the reconstruction of physics objects, and the efficiencies of selections applied to these objects. Unless otherwise noted, all experimental uncertainties apply to signal and background processes estimated using simulation.

The uncertainty on the integrated luminosity of 1.7 % [100] is applied to all processes normalised using theoretical cross section predictions. Moreover, an uncertainty is assigned on the re-weighting of the simulated event samples to match the pile-up conditions in the recorded pp collision dataset. Uncertainties related to the reconstruction and selection of physics objects are provided by dedicated calibration measurements performed by the ATLAS collaboration. In this search, these calibrations are used for electrons [122, 267], muons [124], $\tau_{\text{had-vis}}$ [135], jets [128], flavour-tagging [268–270], and the $p_{\text{T}}^{\text{miss}}$ reconstruction [271]. The calibration measurements yield uncertainties on the momentum scale, momentum resolution, and selection efficiency of reconstructed objects. The major categories of instrumental uncertainties are summarised in Table 6.19. Uncertainties on the selection efficiencies of electron, muon, and $\tau_{\text{had-vis}}$ triggers are only considered in channels where these triggers are used. All uncertainties affecting the four-momentum of reconstructed and selected objects are also propagated to the object-based $p_{\text{T}}^{\text{miss}}$ reconstruction.

Table 6.19: Summary of instrumental uncertainties. The number of independent NPs describing the uncertainty is given in the right-most column.

Category	Affected quantities	N_{NPs}
Electrons	Momentum scale and resolution; Reconstruction, identification, isolation, and trigger efficiencies.	7
Muons	Momentum scale and resolution; Reconstruction, track-to-vertex-association, isolation, and trigger efficiencies.	15
$\tau_{\text{had-vis}}$	Momentum scale; Reconstruction, identification, e-veto, and trigger efficiencies; $e \rightarrow \tau_{\text{had-vis}}$ mistag rates for e-veto.	38
Jets	Momentum scale and resolution; Jet vertex tagging efficiency.	48
b-tagging	Tagging efficiencies for b-jets and mistag rates for c- and light-quark jets.	13
$p_{\text{T}}^{\text{miss}}$	Momentum scale and resolution.	3

Experimental uncertainties on the fake-$\tau_{\text{had-vis}}$ background estimates in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels account for the statistical uncertainties of the $\text{FF}_{\text{multi-jet}}$, $\text{FF}_{t\bar{t}}$, and $r_{\text{multi-jet}}$ estimates. Moreover, uncertainties on the subtraction of $t\bar{t}$ and non-$t\bar{t}$ processes in the combined FF method are taken into account. An additional uncertainty is assigned on $r_{\text{multi-jet}}$ due to it being derived using a simulation-based estimate of the $t\bar{t}$ + fake-$\tau_{\text{had-vis}}$ contribution in the Anti-ID region. In the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channel, the total uncertainty on the normalisation of the fake-$\tau_{\text{had-vis}}$ background are $^{+18\,\%}_{-23\,\%}$ and $^{+28\,\%}_{-30\,\%}$, respectively. The effect of these uncertainties on the shape of the final discriminants are considered in

the background model.

6.6.2 Theory Uncertainties

A number of theoretical uncertainties need to be considered for signal and background processes estimated using simulation. For a given process, these uncertainties are split into uncertainties on the cross section and uncertainties on the acceptance due to analysis selections. Generally, acceptance uncertainties for a given signal or background process are assumed to be correlated across regions if the uncertainty originates from the same source.

The description of the theory uncertainties is structured as follows: First, the uncertainties on the major Z + HF and $t\bar{t}$ backgrounds are described. Second, uncertainties on minor background processes are estimated. Lastly, uncertainties on the modelling of the signal processes are presented.

Acceptance Uncertainties on Z + HF and $t\bar{t}$ Backgrounds

The normalisation of the Z + HF and $t\bar{t}$ backgrounds are measured in the simultaneous fit to observed data in all regions. Constraints on the normalisation of the Z + HF background can be obtained from data in the Z + HF CR. Similarly, the $t\bar{t}$ normalisation can be constrained in the SR of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel and the Z + HF CR. Since the normalisation of these processes is determined in the fit to data, any uncertainties on the normalisation (e.g. cross section uncertainties) are omitted. Instead, uncertainties on the relative acceptance of Z + HF and $t\bar{t}$ between analysis regions have to be considered. These uncertainties are hereafter referred to as *relative acceptance uncertainties*.

The derivation of relative acceptance uncertainties is described using an example of a background that is estimated using simulation and normalised by a fit to data in two regions A and B, with region A being declared as the *reference region*. The probability of an event being selected into region R is given by the product of acceptance and efficiency, $(\mathcal{A} \times \varepsilon)_{\text{R}}$, for this region. The simulation of the background process predicts how $\mathcal{A} \times \varepsilon$ relates between regions and one can define the ratio

$$\mathcal{R} = \frac{(\mathcal{A} \times \varepsilon)_{\text{B}}}{(\mathcal{A} \times \varepsilon)_{\text{A}}},$$

which is referred to as the *relative acceptance* between regions A and B. Uncertainties on the modelling of $\mathcal{R}$ in simulation are assigned as uncertainties on the normalisation of the background process in region B and no additional uncertainties are assigned in the reference region. Uncertainties on $\mathcal{R}$ are estimated by performing variations of the simulation and estimating a relative change in $\mathcal{R}$ according to

$$\frac{\Delta\mathcal{R}}{\mathcal{R}} = \frac{\mathcal{R}(\text{variation}) - \mathcal{R}(\text{nominal})}{\mathcal{R}(\text{nominal})}, \tag{6.3}$$

where $\mathcal{R}(\text{nominal})$ and $\mathcal{R}(\text{variation})$ is the relative acceptance predicted by the nominal and varied simulation setup, respectively. Using relative acceptances to define modelling uncertainties leads to cancellations of variations that lead to the same relative change in $\mathcal{A} \times \varepsilon$ in both regions. This corresponds to an overall change in normalisation of the process in both regions that can be absorbed into the free normalisation factor.

Relative acceptance uncertainties are defined for the Z + HF and $t\bar{t}$ backgrounds using the Z + HF CR as a reference region. The uncertainties are estimated separately for the SRs of the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$

SLT, and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel.

For the Z + HF background, the uncertainties are derived by performing variations of the simulation setup following recommendations by the ATLAS collaboration, which are summarised in Appendix A.3.2. The following aspects of the simulation are varied: the matrix element generator and parton shower program, the factorisation and renormalisation scales, the resummation scale, the multi-jet merging scale, the PDFs and α_{s}, and the PDF set. The relative acceptance uncertainties on the Z + HF background are summarised in Table 6.20(a) for the three SRs.

A similar approach is taken for the $t\bar{t}$ background. Variations of the simulation setup are performed according to prescriptions developed by the ATLAS collaboration, which are documented in Appendix A.3.1. The following elements of the simulation setup are varied: the matrix element generator, the simulation of the parton shower and hadronisation model, the factorisation and renormalisation scales, the simulation of initial- and final-state radiation, the PDFs and α_{s}. The relative acceptance uncertainties resulting from these prescriptions are summarised in Table 6.20(b).

Relative acceptance uncertainties are implemented as uncertainties on the normalisation of backgrounds in the SRs. Variations of the modelling in simulation can also change the shapes of discriminants used for signal extraction, including $m_{\ell\ell}$ in the Z + HF CR, which need to be considered. Therefore, the impact of modelling uncertainties on the distributions of the MVA discriminants in the SRs and the $m_{\ell\ell}$ distribution in the Z + HF CR are investigated. In cases where the shapes do not differ significantly, no additional uncertainties are assigned. When deviations are observed, the shape uncertainties are propagated to the discriminants and correlated with the corresponding normalisation uncertainties.

Shape uncertainties on the Z + HF background are considered for the variation of the matrix element generator and parton shower program ($\tau_{\text{had}}\tau_{\text{had}}$ channel) and for variations of the factorisation and renormalisation scales ($\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ SLT/LTT channels). All other variations are found to have negligible impact on the shape of the discriminating variables, including $m_{\ell\ell}$ in the Z + HF CR.

Shape uncertainties on the $t\bar{t}$ background are considered for the comparison with an alternative matrix element generator and parton shower program ($\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channels) and the uncertainty on the modelling of initial- and final-state radiation ($\tau_{\text{had}}\tau_{\text{had}}$ channel). All other variations have negligible impact on the shapes of the relevant distributions.

Uncertainties on Minor Backgrounds

The minor backgrounds considered in the analysis are normalised using theoretical cross section predictions and thus both cross section and acceptance uncertainties are considered. Except for the tW acceptance uncertainties, only uncertainties on the normalisation of minor backgrounds are considered. A brief description of the uncertainties on minor backgrounds is given in the following:

Z + LF A normalisation uncertainty of 5 % is assigned to account for the uncertainty on the predicted cross section of Z + jets production at NNLO [219]. Additionally, an acceptance uncertainty of 23 % is adopted from Ref. [238].

W + jets A 5 % cross section uncertainty is assigned to the NNLO cross section prediction [219]. A normalisation uncertainty of 37 % is assigned in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, which is adopted from Ref. [238]. This uncertainty is inflated to 50 % in the $\tau_{\text{had}}\tau_{\text{had}}$ channel since the small W + jets contribution is not part of the data-driven fake-$\tau_{\text{had-vis}}$ estimation.

Table 6.20: Relative acceptance uncertainties on the Z + HF (a) and $t\bar{t}$ (b) background in the three SRs. All uncertainties are estimated using the Z + HF CR as the reference region. The relative sign of the effect of variations is indicated by the ± and ∓ prefixes. The total uncertainty is given for illustration purposes only.

(a) Z + HF background

	Uncertainty / %		
Uncertainty source	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
Matrix element and parton shower generator	±7.0	∓2.1	∓11
Factorisation and renormalisation scale	$^{+12}_{-9.7}$	$^{+5.4}_{-3.0}$	$^{+8.5}_{-5.4}$
Resummation scale	∓6.0	±1.7	±1.6
Multi-jet merging scale	±5.4	±7.0	±7.2
PDF+α_s	±0.77	±0.27	±0.35
PDF set choice	±1.0	±1.0	±1.1
Total	$^{+16}_{-15}$	$^{+9.3}_{-8.1}$	$^{+16}_{-14}$

(b) $t\bar{t}$ background

	Uncertainty / %		
Uncertainty source	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
Matrix element generator	±3.8	±0.3	∓0.9
Parton shower and hadronisation model	±2.2	±7.2	±8.8
Initial-state radiation (incl. scale variations)	±0.3	±0.9	∓1.3
Final-state radiation	$^{+2.0}_{-4.5}$	$^{+1.5}_{-1.0}$	$^{+1.0}_{-3.2}$
PDF+α_s	±0.2	±0.6	±0.8
Total	$^{+4.8}_{-6.3}$	$^{+7.5}_{-7.4}$	$^{+9.0}_{-9.5}$

Diboson Normalisation uncertainties of 20 %, 26 %, and 25 % are applied to ZZ, WZ, and WW production, respectively. These uncertainties are adopted from Ref. [238].

Single top quark Uncertainties on the cross section used to normalise the predictions are taken from Ref. [212]. An acceptance uncertainty of 20 % is assigned to the minor contribution of single-top-quark production via s- and t-channel diagrams, which is adopted from Ref. [238].

Acceptance uncertainties in the phase space selected by the analysis are derived for tW production, which is the dominant source of single-top-quark background in this analysis. The acceptance uncertainties from the NLO+PS matching and the choice of parton shower program are estimated by comparison with alternative simulation setups using MadGraph5_aMC@NLO and Herwig 7, respectively. Uncertainties from PDF+α_s and the simulation of initial- and final-state radiation are obtained by re-weighting of the nominal simulation result. Finally, an uncertainty on the treatment of the interference between tW production at NLO and $t\bar{t}$ production is estimated by performing MC-to-MC comparisons of the nominal *diagram removal* and *diagram subtraction* schemes [272].

The total acceptance uncertainty on tW production is 34 % in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, 14 % in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel, and 23 % in the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel. Shape effects of the final-state radiation and tW-$t\bar{t}$ interference uncertainties on the MVA discriminants are taken into account in the background model.

Single SM H Uncertainties on the total cross section of the SM H production modes considered as backgrounds are assigned according to the recommendations in Ref. [40] for $m_H = 125.0\,\text{GeV}$. Uncertainties on the branching ratios of $H \to \tau^+\tau^-$ and $H \to b\bar{b}$ are assigned to the relevant processes and are taken from Ref. [40].

An acceptance uncertainty of 100 % is assigned to $H \to \tau^+\tau^-$ backgrounds produced via ggF, VBF, and WH to account for difficulties in the modelling of associated production of SM H with heavy-flavour quarks. This uncertainty is a convention adopted by the ATLAS collaboration for searches for Higgs boson pair production using the pp collision dataset collected during Run 2 of the LHC.

Acceptance uncertainties in the phase space selected by the analysis are derived for single-Higgs-boson backgrounds produced via ZH and $t\bar{t}H$ for both $H \to \tau^+\tau^-$ and $H \to b\bar{b}$. They are derived by varying the parton shower program, the factorisation and renormalisation scale, the PDFs and α_s. Additionally, uncertainties on the NLO+PS matching and the modelling of initial- and final-state radiation are assigned to the $t\bar{t}H$ background.

Signal Modelling Uncertainties: SM HH Production

Uncertainties on the predicted SM HH production cross section are taken into account when the signal strength is used as the POI in the statistical analysis. In contrast, when the SM HH production cross section is the POI, cross section uncertainties on the signal processes are omitted. In either case, uncertainties on the branching ratios of $H \to \tau^+\tau^-$ and $H \to b\bar{b}$ are taken from Ref. [40] and assigned as uncertainties on the signal normalisation.

The uncertainty on the SM HH production cross section via ggF is $^{+6}_{-23}$ % for the scale variations and the treatment of the finite top-quark mass [47], and ±3.0 % for variations of PDF+α_s [48]. Similarly,

the uncertainty on the SM HH production cross section via VBF is $^{+0.03}_{-0.04}$ % for the scale variations and ±2.1 % for variations of PDF+α_s [48].

Signal acceptance uncertainties are derived in all SRs by performing variations of the generator setup. An uncertainty on the choice of parton shower and hadronisation model is estimated by comparing the nominal setup using Pythia 8 with an alternative setup using Herwig 7 for parton showering and hadronisation. An uncertainty from missing higher orders of the truncated perturbative expansion in α_s is estimated by performing six variations of the factorisation and renormalisation scales. For SM HH production via ggF (VBF), uncertainties on the PDFs are taken into account following the prescriptions of the PDF4LHC15nlo [197] (NNPDF3.0nlo [199]) set of PDFs. Lastly, the value of $\alpha_s(Q^2 = m_Z^2)$ is varied up and down by 0.0015 (0.001) about the central value of 0.118 in the PDF4LHC15nlo (NNPDF3.0nlo) PDF set. The SM HH acceptance uncertainties are summarised in Table 6.21 for the ggF and VBF production modes.

Table 6.21: Uncertainties on the acceptance of SM HH events produced via ggF and VBF in the three SRs. Uncertainties marked as "–" are found to be negligible.

Uncertainty source	Acceptance uncertainty / %		
	$\tau_{had}\tau_{had}$	$\tau_{lep}\tau_{had}$ SLT	$\tau_{lep}\tau_{had}$ LTT
SM HH (ggF)			
Parton shower and hadronisation model	±4.3	±7.6	±7.5
Factorisation and renormalisation scale	±1.4	±1.2	±1.0
PDF+α_s	–	–	–
SM HH (VBF)			
Parton shower and hadronisation model	±3.0	±6.3	±2.1
Factorisation and renormalisation scale	±0.1	±1.0	±1.0
PDF+α_s	±1.0	–	–

Signal Modelling Uncertainties: Resonant HH Production

Cross section uncertainties are not considered for signals from resonant HH production because the cross section $\sigma(pp \to X \to HH)$ is considered to be the POI. However, uncertainties on the Higgs boson branching ratios are taken into account.

Signal acceptance uncertainties are estimated for the choice of parton shower and hadronisation model, the factorisation and renormalisation scales, and PDF+α_s. These uncertainties are derived for a subset of resonance masses considered in the analysis which are then extrapolated to the full set.

The acceptance uncertainty from the choice of parton shower program and hadronisation model is estimated by comparing the default configuration using Herwig 7 with Pythia 8 for signals with $m_X/\text{GeV} \in \{251, 260, 280, 400, 500, 1000\}$. The uncertainties evaluated at these points are linearly interpolated (extrapolated) for $m_X \leq 1\,000\,\text{GeV}$ ($m_X > 1\,000\,\text{GeV}$) to provide estimates of the uncertainty for other mass points. This procedure is illustrated in Figure 6.29 for all three channels. The extrapolation to $m_X = 1\,600\,\text{GeV}$ can lead to an uncertainty of up to 20 %.

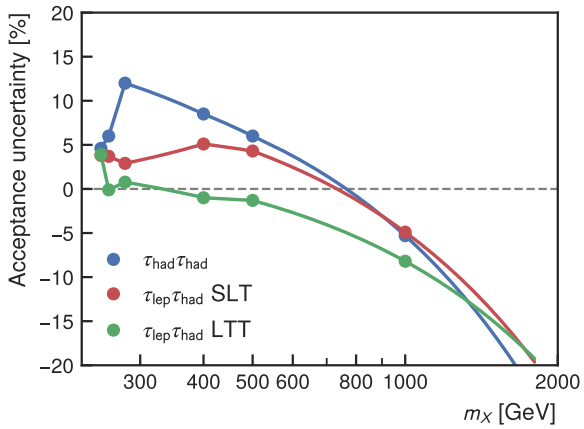


Figure 6.29: Uncertainties on the acceptance of $pp \to X \to HH$ events in the SRs due to the choice of parton shower and hadronisation model. A positive (negative) sign of the uncertainty indicates that the alternative configuration (Pythia) predicts a larger (smaller) $\mathcal{A} \times \varepsilon$ than the nominal one (Herwig). The lines indicate the linear inter-/extrapolation used to obtain the uncertainties for other mass points considered in the analysis.

Acceptance uncertainties from factorisation and renormalisation scales as well as PDF+α_s are evaluated at generator-level for two mass points with $m_X = 500\,\text{GeV}$ and $1\,000\,\text{GeV}$ after approximating the selections applied in the analysis. The uncertainties are found to be negligible and are therefore omitted.

An additional uncertainty is assigned in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ LTT channels to signal samples using fast simulation of the ATLAS detector. The efficiency of $\tau_{\text{had-vis}}$-triggers is found to deviate between full and fast simulation, without dedicated calibrations of $\tau_{\text{had-vis}}$-trigger efficiencies in fast simulation being available. An uncertainty is estimated by comparing the acceptance between full and fast simulation for a benchmark signal with $m_X = 400\,\text{GeV}$. The acceptance predicted using fast simulation is 6.5 % (3.6 %) larger than in full simulation for the $\tau_{had}\tau_{had}$ ($\tau_{lep}\tau_{had}$ LTT) SR for the benchmark point. Distributions of kinematic and MVA input variables are compared between fast and full simulation showing no significant deviations in their shapes. Therefore, the difference in acceptance is assigned as an additional normalisation uncertainty in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ LTT channel for all signal samples using fast detector simulation, i.e. all signals with $m_X \leq 1\,000\,\text{GeV}$.

6.7 Statistical Interpretation

The statistical interpretation of the search for resonant and non-resonant HH production proceeds using the methods introduced in Section 4.1. The main results of the analysis are upper limits on the signal strength and cross section of SM HH production via ggF and VBF and upper limits on the cross section of resonant HH production via scalar, narrow-width resonances produced in ggF.

This section is structured as follows: First, the statistical model is introduced in Section 6.7.1. Second, the results of the search for SM HH and resonant HH production are presented in Section 6.7.2 and Section 6.7.3, respectively. Finally, the statistical interpretation is concluded with an estimation of the statistical significance of the largest excess observed in the search for resonant HH production in Section 6.7.4.

6.7.1 The Statistical Models

A separate statistical model is constructed for every signal hypothesis to be probed. These models describe the distribution of event counts in bins of the MVA discriminants in the SRs and $m_{\ell\ell}$ in the Z + HF CR. They are built with HistFactory (cf. Section 4.1.1) and include the signal and background estimates, including the associated uncertainties, described in previous sections. The discriminants used in all analysis channels are summarised in Table 6.22.

Table 6.22: Summary of the discriminants used in the SRs and the Z + HF CR. The search for resonant HH production uses the PNN discriminant evaluated with a mass parameter set to the resonance mass under test.

	Discriminant used in channel			
Search	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT	Z + HF CR
SM HH production	BDT	NN	NN	$m_{\ell\ell}$
Resonant HH production	PNN(m_X)	PNN(m_X)	PNN(m_X)	$m_{\ell\ell}$

The POIs of the search for non-resonant HH production are the total SM HH production cross section via ggF and VBF, $\sigma_{gg\text{F+VBF}}$, and the corresponding signal strength

$$\mu = \frac{\sigma_{gg\text{F+VBF}}}{\sigma^{\text{SM}}_{gg\text{F+VBF}}} \qquad \text{with} \qquad \sigma^{\text{SM}}_{gg\text{F+VBF}} = 32.78\,\text{fb}\,,$$

which measures the cross section relative to the SM prediction from Refs. [46, 49]. Notably, the interpretation of the SM HH search does not distinguish between the ggF and VBF production modes. In the search for resonant HH production, the cross section $\sigma(pp \to X \to HH)$ is used as the POI. Moreover, SM HH production is not considered as a background in the search for resonant HH production. In either case, free normalisation factors that scale the contributions of Z + HF and $t\bar{t}$ backgrounds in all channels are included in the model. Lastly, statistical uncertainties on the background rates estimated using finite samples of simulated events or CR data are implemented according to the simplified Barlow–Beeston method [141, 142].

The statistical models were scrutinised during their development by performing fits to Asimov data, CR data, CR and SR data but blinding the signal-like regions at high MVA score, and finally fits to observed data in all regions. Alternative models that exclude certain SRs were investigated to understand the effect of individual channels on the results. At every step, the MLE of the model parameters, including the parameter errors and correlations, were inspected to ensure that the model behaves as expected. During early stages of the analysis, these checks informed the construction of the statistical model. After unblinding of the SRs the model remained fixed to avoid the introduction of biases.

Binning of MVA Discriminants

The signal sensitivity of the search depends on the binning used for the MVA discriminants in the statistical interpretation. The binning has to be chosen such that regions with high signal-to-background ratio (high MVA score) are well-separated from regions with low signal-to-background ratio (low/intermediate MVA score).

An iterative re-binning algorithm is used to determine the binning of the MVA discriminants. The aim of the algorithm is to maximise the expected sensitivity to a given signal while ensuring that the background prediction obeys constraints on the statistical uncertainty and expected number of events in each bin. These constraints ensure that asymptotic approximations can be used for the statistical interpretation. The algorithm described in the following was previously used in Ref. [184] and is continued to be used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel of this search. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, a minor modification is applied to improve the signal sensitivity.

The algorithm is provided with MVA score histograms with fine, equidistant binning separately for the signal and total background expectation at the nominal values of all NPs. It proceeds by iteratively merging bins, starting from the most signal-like MVA score bins, until the bin fulfils a set of requirements:

1. The relative statistical uncertainty of the background prediction in the bin must be smaller than $50\,\% \times f_{\text{s}} + 1\,\%$, where f_{s} is the fraction of signal events in the bin with respect to the total number of signal events selected in the channel. This requirement limits the statistical uncertainty of the background estimate in the most signal-like bins to be in a range of 10 to 20 % after re-binning.

   In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, a requirement of $10 f_{\text{s}} + 5 f_{\text{b}} > 1$ is used instead, where f_{b} is the fraction of background events defined in analogy to f_{s}.

2. The expected number of background events in the bin must be larger than five.

When a bin fulfilling all requirements is found, the process is repeated starting from the next bin that is not yet merged. The algorithm terminates with a final bin at low MVA score. If the final bin does not fulfil the criteria, it is merged with the preceding bin. The size of bins resulting from the re-binning procedure can vary by multiple orders of magnitude. For visualisation purposes, the variable-width MVA score bins are therefore displayed as equidistant bins.

Symmetrisation of Systematic Uncertainties

Uncertainties derived from comparisons of the nominal with an alternative prediction only provide a single systematic variation and cannot be readily incorporated in the statistical model. In this case, variations are symmetrised by mirroring their effect with respect to the nominal prediction.

Uncertainties with up- and down-variations that change the prediction in one or multiple bins in the same direction are also subject to symmetrisation. Such variations usually result from statistical fluctuations and can lead to artificial over- or underconstraints of the associated NPs after the fit. In this case, uncertainties are symmetrised by assigning half of the difference between the up- and down-variation as a symmetric uncertainty. This symmetrisation method is selectively applied to affected uncertainties, such as the jet energy scale and resolution uncertainties.

Smoothing and Pruning of Systematic Uncertainties

All uncertainties are split into separate normalisation and shape uncertainties that are correlated in the statistical model. The templates used to define shape uncertainties are susceptible to statistical fluctuations, which can introduce spurious pulls or constraints of NPs after the fit. This is often the case when deriving shape uncertainties from a two-sample comparison or from variations that change

the four-momentum of reconstructed physics objects. In these cases, a smoothing procedure is applied that is adopted from Ref. [273]. Shape uncertainties based on a re-weighting of the nominal prediction are less susceptible to statistical fluctuations and thus no smoothing is applied by default. An exception is made for the variations of the final-state radiation in single-top-quark and $t\bar{t}$ production. These are subject to large statistical fluctuations and are therefore smoothed.

After symmetrisation and smoothing, the model is simplified by removing (pruning) small systematic uncertainties. This procedure is applied separately to all samples and channels, and separately for the normalisation and shape components of uncertainties. The normalisation component of an uncertainty is removed if both the up- and down-variation change the normalisation by less than 0.5 %. The shape component of an uncertainty is removed if the relative change of the up- and down-variations in all bins of a given channel is less than 0.5 %.

Nuisance Parameter Correlation Scheme

All normalisation factors are correlated between channels. NPs related to instrumental uncertainties are correlated between all channels and physics processes. Theory uncertainties on cross sections and acceptances for a given process are assumed to be correlated between channels provided they originate from the same source. NPs related to the fake-$\tau_{\text{had-vis}}$ background estimation are not correlated between the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels since different estimation methods are used. However, they are correlated between the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channel, which use the same estimation techniques.

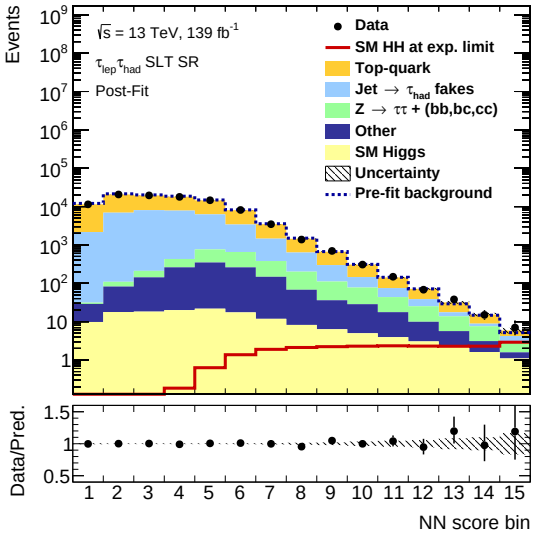
An exception to this scheme are the parton shower uncertainties for $t\bar{t}$ which were decorrelated between the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ SLT, and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel. This decision is based on observed tensions between the best-fit NP values when performing fits of the individual channels.

6.7.2 Results of the Search for SM HH Production

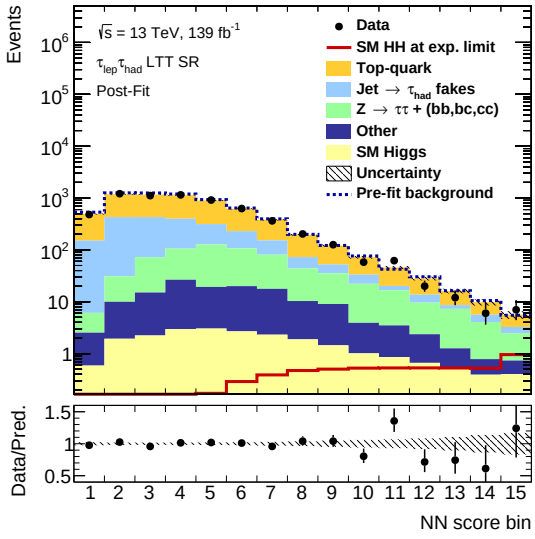
The results of the search for SM HH production are presented for the combination of all channels. Results are also presented prior to the combination of the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels to illustrate the signal sensitivity of individual channels and differences in the impact of systematic uncertainties. Selected background processes are merged for illustration purposes only. The contributions from $t\bar{t}$ and single-top-quark production are combined into a "top-quark" background category. Contributions from minor backgrounds are combined into the "other" background category, which includes $Z \to \tau^+\tau^-$+LF, $Z \to e^+e^-$, $Z \to \mu^+\mu^-$, W + jets, diboson, and $t\bar{t}V$.

The background-only model is fitted to observed data in all channels. Figure 6.30 compares the post-fit MVA score distributions in the three SRs to data. The post-fit prediction describes the observed data in the SRs well. In addition, the expected event yields per process in the SRs after the maximum likelihood fit are summarised in Table 6.23(a). Finally, the normalisation factors of $t\bar{t}$ and Z + HF resulting from the fit are 0.97 ± 0.04 and 1.40 ± 0.11, respectively, which are consistent with the results obtained from fits restricted to the Z + HF CR in Section 6.4.1.

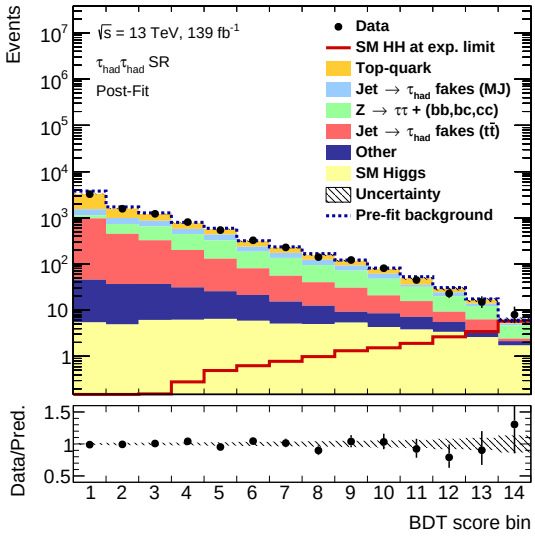
The expected event yields in the two most signal-like bins are summarised in Table 6.23(b) to illustrate the background composition in a kinematic region similar to the one occupied by the signal process. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the dominant background in the two most signal-like bins of the BDT is the production of Z + HF with an expectation of about 8 events. The contribution of top-quark, single-Higgs-boson, and fake-$\tau_{\text{had-vis}}$ backgrounds is similar with an expectation of about 4 events each. The fraction of signal events populating the two most signal-like bins is close to 50 % with respect to



(a) $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel



(b) $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel



(c) $\tau_{\text{had}}\tau_{\text{had}}$ channel

Figure 6.30: Distributions of the MVA discriminants used for the SM *HH* search in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT (a), $\tau_{\text{lep}}\tau_{\text{had}}$ LTT (b), and $\tau_{\text{had}}\tau_{\text{had}}$ (c) channel after the fit of the background-only model to observed data in all regions. The signal overlay is scaled to the expected upper limit on the signal strength of 3.9 from the combination of all channels. The fake-$\tau_{\text{had-vis}}$ background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel is shown separately for fake $\tau_{\text{had-vis}}$ from multi-jet (MJ) and $t\bar{t}$.

Table 6.23: Event yields in the three SRs (a) and in the two most signal-like bins of the MVA discriminants (b) after the background-only fit to the observed data in all regions. The category "other backgrounds" combines minor contributions from $Z \to \tau^+\tau^-$ + LF, $Z \to e^+e^-$, $Z \to \mu^+\mu^-$, W + jets, diboson, and $t\bar{t}V$. The expected SM HH signal yield is shown with $\mu = 1$ and NPs at their best-fit values except for those only affecting the signal processes, which are kept at their nominal values.

(a) Event yields in the SRs.

Process	Event yield		
	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
SM HH (ggF + VBF)	5.16 ± 0.84	5.9 ± 1.0	1.42 ± 0.24
Top-quark	3 250 ± 160	61 000 ± 1 400	4 040 ± 200
$Z \to \tau^+\tau^-$ + HF	1 550 ± 160	1 620 ± 130	529 ± 57
Single Higgs boson	66 ± 13	148 ± 18	23.0 ± 4.3
Jet → fake $\tau_{\text{had-vis}}$ (combined)	–	34 300 ± 1 500	1 640 ± 170
Jet → fake $\tau_{\text{had-vis}}$ (multi-jet)	1 270 ± 130	–	–
Jet → fake $\tau_{\text{had-vis}}$ ($t\bar{t}$)	2 080 ± 200	–	–
Other backgrounds	196 ± 33	1 308 ± 86	121 ± 14
Total background	8 414 ± 90	98 430 ± 390	6 357 ± 79
Observed data	8 380	98 456	6 351

(b) Event yields in the two most signal-like bins of the BDT ($\tau_{\text{had}}\tau_{\text{had}}$) and NN ($\tau_{\text{lep}}\tau_{\text{had}}$) discriminants.

Process	Event yield (two most signal-like bins)		
	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
SM HH (ggF + VBF)	2.37 ± 0.39	1.34 ± 0.23	0.381 ± 0.066
Top-quark	3.80 ± 0.64	8.2 ± 1.8	6.55 ± 0.89
$Z \to \tau^+\tau^-$ + HF	8.3 ± 1.2	6.0 ± 1.0	5.02 ± 0.88
Single Higgs boson	4.3 ± 1.1	2.71 ± 0.51	0.79 ± 0.20
Jet → fake $\tau_{\text{had-vis}}$ (combined)	–	2.35 ± 0.56	2.36 ± 0.84
Jet → fake $\tau_{\text{had-vis}}$ (multi-jet)	1.94 ± 0.51	–	–
Jet → fake $\tau_{\text{had-vis}}$ ($t\bar{t}$)	2.87 ± 0.46	–	–
Other backgrounds	1.54 ± 0.38	1.98 ± 0.24	0.72 ± 0.11
Total background	22.8 ± 1.9	21.2 ± 2.1	15.4 ± 1.7
Observed data	23	22	13

the expected signal yield in the channel. The $\tau_{\text{had}}\tau_{\text{had}}$ channel provides the largest signal-to-background ratio of any individual channel in this search.

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, the dominant backgrounds in the two most signal-like bins originate from the production of top quarks and Z + HF. Compared to the $\tau_{\text{had}}\tau_{\text{had}}$ channel, top-quark backgrounds are more abundant with a large contribution of about 50 % (15 %) from single-top production in the SLT (LTT) channel. As a result, the signal-to-background ratio in the most signal-like bins is reduced, in part also due to the decrease in $m_{\tau\tau}^{\text{MMC}}$ and m_{HH} resolution resulting from an additional neutrino in the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ decay chain.

Post-fit plots of the MVA input variables in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are depicted in Figure 6.31. The background model describes the observed data well with minor discrepancies in the angular observables. For example, at low values of $\Delta R(\tau,\tau)$ the background prediction exceeds the observed data by 10 to 15 %, which is not fully covered by uncertainties. This mismodelling can be partially explained by a dependency of the Z + HF normalisation factor on the transverse momentum of the Z boson, $p_{\text{T}}(Z)$, that is not accounted for. As a cross-check, the determination of the normalisation factors in the Z + HF CR is repeated in bins of $p_{\text{T}}(Z)$ as estimated by the reconstructed transverse momentum of the lepton pair. This test shows that the Z + HF normalisation factors tend to decrease with increasing $p_{\text{T}}(Z)$. For typical values of $p_{\text{T}}(Z)$ after the $\tau_{\text{had}}\tau_{\text{had}}$ SR selection, this dependency can lead to differences of up to 10 % compared to the $p_{\text{T}}(Z)$-inclusive normalisation factor. Due to the anticorrelation of $\Delta R(\tau,\tau)$ and $p_{\text{T}}(Z)$, this effect might be further enhanced in regions of low $\Delta R(\tau,\tau)$. The $p_{\text{T}}(Z)$-dependency of the normalisation factor was only discovered after unblinding and could therefore not be included in the model. In future analyses it might be beneficial to control for this effect.

A comparison of the background-only and signal-plus-background hypothesis is performed using a test for the discovery of a positive signal (cf. Section 4.1.2). This test yields an observed p-value of 27 %; thus, the background-only hypothesis cannot be rejected. Moreover, the signal strength obtained from the unconditional fit is $\hat{\mu} = 0.9^{+1.8}_{-1.5}$ for the combination of all channels, which is compatible with non-resonant HH production predicted by the SM but also with its absence.[18]

The dominant uncertainties affecting the measurement of the SM HH signal strength are summarised in Table 6.24 for the combination of all channels. The measurement is mainly limited by the statistical uncertainty originating from the small number of events observed at high values of the MVA discriminants, which explains about two-thirds of the variance on $\hat{\mu}$. Systematic uncertainties play a lesser role, explaining about one-third of the variance on $\hat{\mu}$. The largest source of systematic uncertainty are due to uncertainties on the modelling of backgrounds and the statistical precision of the background estimate.

The effect of uncertainties on $\hat{\mu}$ is examined on the level of individual NPs in Figure 6.32, separately for the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels as well as their combination. A ranking of NPs with the largest impact on the estimated signal strength is shown in the figure, including the MLE of the NPs and their 68 % confidence intervals. Generally, the NPs are compatible with their pre-fit values. For few NPs the fit provides more stringent constraints than suggested by their prior measurement, which is consistent with results from fits to Asimov dataset.

The largest constraint observed in the $\tau_{\text{had}}\tau_{\text{had}}$ channel is on the NP related to the Z + HF acceptance uncertainty determined by an MC-to-MC comparison of Sherpa (NLO) and Mad-

[18] The best-fit signal strength of the combination of $\tau_{\text{had}}\tau_{\text{had}}$ channel and Z + HF CR is $\hat{\mu} = 0.7^{+1.9}_{-1.6}$ and for the combination of the $\tau_{\text{lep}}\tau_{\text{had}}$ channels and Z + HF CR $\hat{\mu} = 1.9^{+3.7}_{-3.2}$.

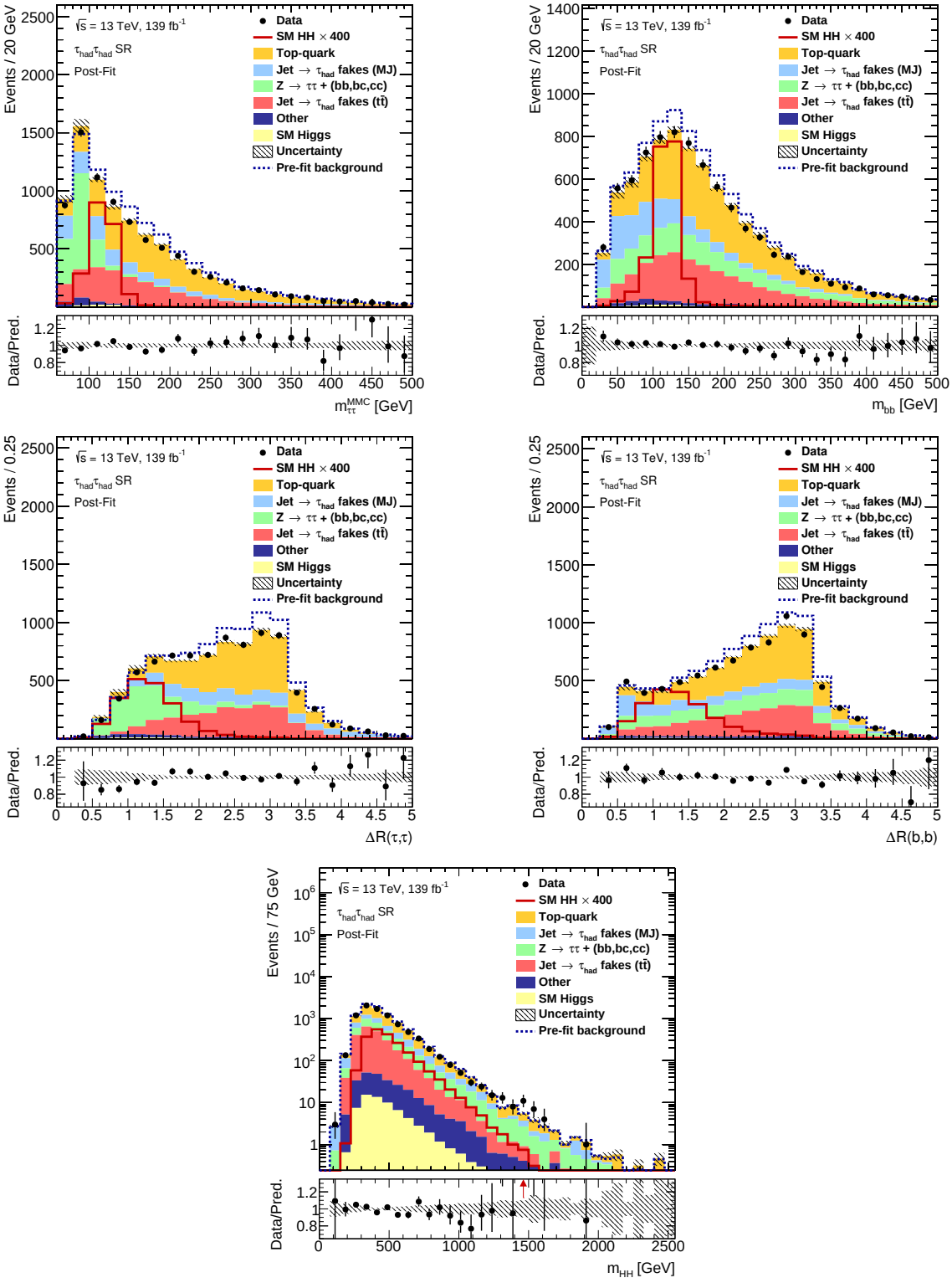


Figure 6.31: Distributions of the BDT input variables in the $\tau_{\text{had}}\tau_{\text{had}}$ SR after the fit of the background-only model to observed data in all regions. The expected SM HH signal is overlayed with a normalisation corresponding to $\mu = 400$.

Table 6.24: Breakdown of the variance of $\hat{\mu}$ by uncertainty category for the unconditional fit to observed data in all regions. The fraction of the variance on $\hat{\mu}$ from a category is approximated using $(\Delta\hat{\mu}^2_{\text{tot}} - \Delta\hat{\mu}^2_{\text{w/o cat}})/\Delta\hat{\mu}^2_{\text{tot}}$, where $\Delta\hat{\mu}^2_{\text{tot}}$ is the estimate of the total variance of $\hat{\mu}$ and $\Delta\hat{\mu}^2_{\text{w/o cat}}$ its variance after fixing the NPs of a given category to their best-fit values. The variance of $\hat{\mu}$ from data statistical uncertainties is determined from the model with all NPs fixed to their best-fit values. The fractions of subcategories do not necessarily sum to the fraction of the parent category due to correlations between NPs.

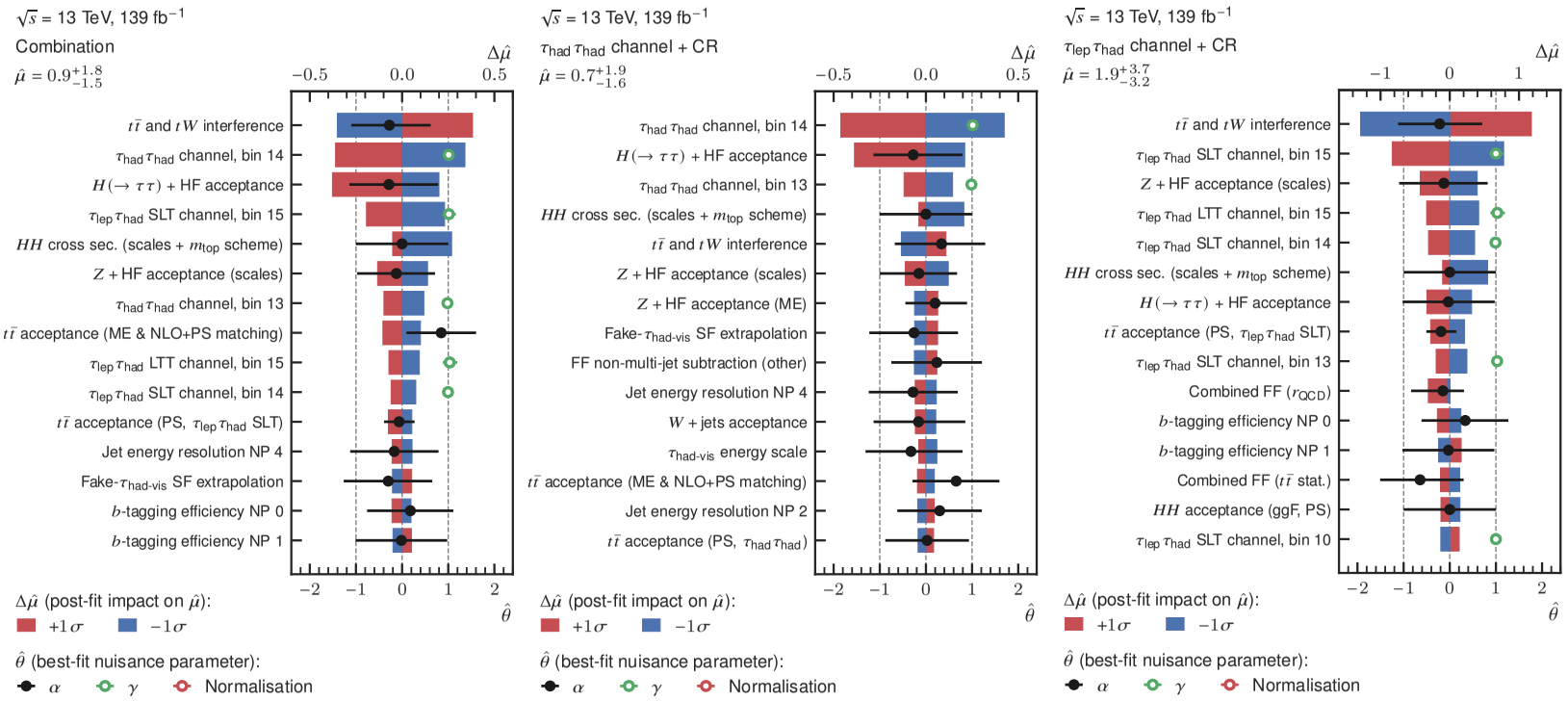
Source	Explained fraction of variance on $\hat{\mu}$
Data statistical uncertainty	66 %
Systematic uncertainties	34 %
Instrumental uncertainties	1 %
Signal modelling uncertainties	4 %
Background statistical uncertainties	8 %
Background modelling uncertainties	18 %
– Top-quark (incl. free normalisation)	6 %
– Z + HF (incl. free normalisation)	1 %
– SM Higgs boson	8 %
– Fake-$\tau_{\text{had-vis}}$	< 1 %
– Other	< 1 %

Graph5+Pythia (LO). This uncertainty is conservative since it compares the nominal matrix element generator at NLO with a lower order prediction. Therefore, the associated NP is expected to be constrained in the fit. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, large constraints are observed on the NP associated with the uncertainty on $r_{\text{multi-jet}}$ and the $t\bar{t}$ acceptance uncertainty from the comparison of parton shower programs. The uncertainty on $r_{\text{multi-jet}}$ is derived to be conservative, varying $r_{\text{multi-jet}}$ between 0 and 100 %, explaining the large constraints on this parameter. The constraints on the $t\bar{t}$ acceptance uncertainty based on the parton shower comparison originates in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel. Since the constraint is large, this uncertainty source is decorrelated between channels to prevent underestimating the acceptance uncertainty in other channels.

Although the analysis is mostly limited by the size of the recorded pp collision dataset, a few of the leading sources of uncertainty are discussed in the following:

- Statistical uncertainties on the expected background rates in high MVA score bins are among the uncertainties with the largest impact on the analysis sensitivity. The MC simulations used for background estimation only populate the high MVA score regions sparsely, resulting in large statistical uncertainties. In addition, the fake-$\tau_{\text{had-vis}}$ background estimation in the $\tau_{\text{had}}\tau_{\text{had}}$ channel requires a large subtraction of non-multi-jet events that further reduces the statistical precision of the background rate predictions. As a result, the associated uncertainties have non-negligible effect on the sensitivity of the analysis.

- The tW acceptance uncertainty targeting the tW and $t\bar{t}$ interference is the systematic uncertainty with the largest impact on $\hat{\mu}$ in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels and the combination. The large impact originates from the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel, where the uncertainty can reach up to 80 % at high NN score. Moreover, tW production makes up half of the top-quark background at high NN



(a) Combination of all channels

(b) $\tau_{had}\tau_{had}$ channel and Z + HF CR

(c) $\tau_{lep}\tau_{had}$ channels and Z + HF CR

Figure 6.32: Ranking of NPs based on the post-fit impact on $\hat{\mu}$ as measured by the change in the estimated signal strength $\Delta\hat{\mu}$ when performing the fit after fixing a given NP to $\hat{\theta} \pm \Delta\hat{\theta}$, where $\hat{\theta}$ is the MLE and $\Delta\hat{\theta}$ the post-fit uncertainty on the NP. The impact on $\hat{\mu}$ is shown on the top axis and the NPs are given in descending order of the mean $|\Delta\hat{\mu}|$. The bottom axis shows the MLE of the NPs and the post-fit uncertainty, with marker colours indicating the type of NP: NPs with Gaussian constraints (α, black), NPs with Poisson constraints from the Barlow–Beeston method (γ, green), and normalisation factors (red). The rankings are obtained from fits of the signal-plus-background model to observed data in all regions (a), in the $\tau_{had}\tau_{had}$ SR and Z + HF CR (b), and in the $\tau_{lep}\tau_{had}$ SRs and Z + HF CR (c).

score in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel. This uncertainty is less relevant in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channels due to smaller fractions of tW events in the most signal-like bins of the MVA discriminants.

- The uncertainty of 100 % on the acceptance of $pp \to H \to \tau^+\tau^-$ production in association with quarks of heavy flavour has a large impact on the fitted μ. The most signal-like bins of the BDT discriminant in the $\tau_{\text{had}}\tau_{\text{had}}$ channel select a considerable amount of single-Higgs-boson events (cf. Table 6.23(b)). About a fourth of these events are expected to be from $H \to \tau^+\tau^-$ production via ggF, which are subject to the heavy-flavour uncertainty. As a result, this uncertainty is among the leading uncertainties affecting the background prediction in the most signal-like bins.

Given the absence of a statistically significant signal, upper limits are set on the SM HH signal strength and cross section using the CL_s method at 95 % CL. Table 6.25 summarises the exclusion limits separately for the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, the $\tau_{\text{had}}\tau_{\text{had}}$ channel, and their combination. The observed (expected) upper limit on the signal strength is 4.7 (3.9) for the combination of all channels. The upper limits are largely driven by the high sensitivity of the $\tau_{\text{had}}\tau_{\text{had}}$ channel to SM HH production, yielding observed (expected) upper limits on μ of 5.0 (4.4). Further discussion of these results is given in Section 6.8.

Table 6.25: Upper limits on the SM HH production cross section via ggF and VBF, $\sigma_{gg\text{F+VBF}}$, and the SM HH signal strength, μ, at 95 % CL. The expected limits are obtained under the assumption of the background-only hypothesis. The table is adapted from Ref. [192].

		Upper limit on POI at 95 % CL					
	POI	Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
$\tau_{\text{had}}\tau_{\text{had}}$ channel	$\sigma_{gg\text{F+VBF}}$ / fb	150	70	95	130	180	240
	μ	5.0	2.4	3.2	4.4	6.1	8.2
$\tau_{\text{lep}}\tau_{\text{had}}$ channel	$\sigma_{gg\text{F+VBF}}$ / fb	280	120	170	230	320	430
	μ	9.7	4.2	5.6	7.8	11	15
Combination	$\sigma_{gg\text{F+VBF}}$ / fb	140	62	83	110	160	210
	μ	4.7	2.1	2.8	3.9	5.4	7.2

6.7.3 Results of the Search for Resonant HH Production

In the search for resonant HH production, the cross section $\sigma(pp \to X \to HH)$ is used as the POI. Otherwise, the statistical interpretation proceeds in analogy to the SM HH case after replacing the BDT/NN discriminants by PNN discriminants evaluated with mass parameters set to the m_X of the signal hypothesis of interest.

The PNN discriminants in the $\tau_{\text{had}}\tau_{\text{had}}$ SR are shown for four exemplary mass points in Figure 6.33 after the background-only fit to observed data in all regions. Moreover, Table 6.26 summarises the expected number of events in the most signal-like bins of the PNN discriminant after the fit. Figures of the PNN discriminants in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels are summarised in Appendix A.5.2. The background processes relevant to the search vary with the m_X of the considered signal hypothesis. For low-mass

resonances, the dominant backgrounds are top-quark production and fake-$\tau_{\text{had-vis}}$ backgrounds. These background processes become less important for larger m_X, at which point the production of Z + jets becomes the dominant background.

Table 6.26: Expected and observed number of events in the $\tau_{\text{had}}\tau_{\text{had}}$ SR for signal-like bins of the PNN discriminant after a fit of the background-only fit model to observed data in all regions. The two most signal-like bins are shown for $m_X = 300\,\text{GeV}$ and 500 GeV. †: Only the most signal-like bin is shown for $m_X = 1\,000\,\text{GeV}$ and 1 600 GeV.

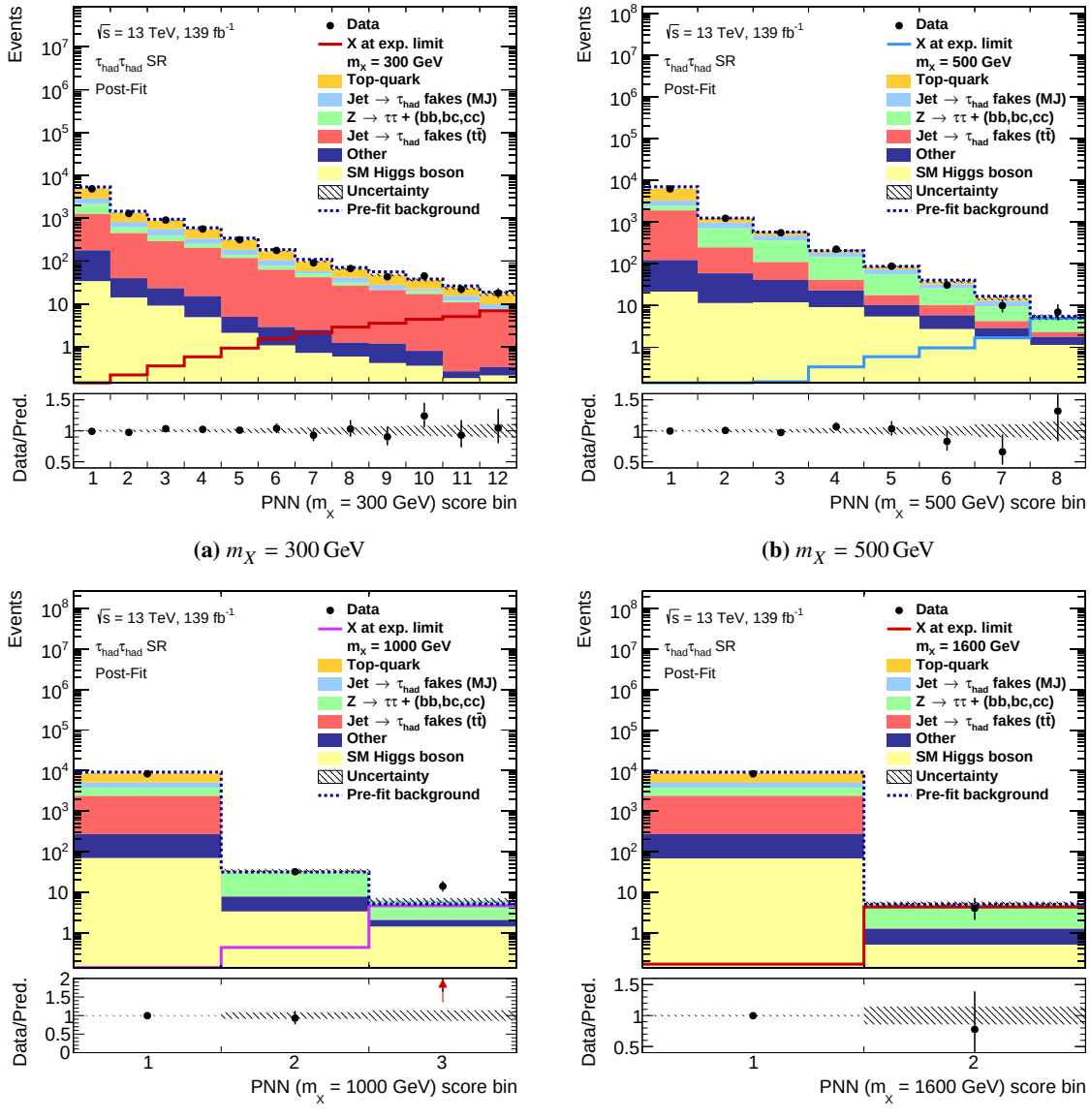
	Event yield in the most signal-like PNN bin(s)			
Process	m_X 300 GeV	500 GeV	1 000 GeV (†)	1 600 GeV (†)
$X \to HH$ ($\sigma = 1\,\text{pb}$)	18.2 ± 2.9	156 ± 16	379 ± 38	139 ± 28
Top-quark	15.6 ± 2.0	2.70 ± 0.44	0.12 ± 0.03	0.37 ± 0.05
$Z \to \tau^+\tau^- + (bb, bc, cc)$	1.38 ± 0.26	7.7 ± 1.1	3.50 ± 0.65	3.12 ± 0.54
Single Higgs boson	0.40 ± 0.07	2.91 ± 0.61	1.40 ± 0.42	0.50 ± 0.27
Jet → fake $\tau_{\text{had-vis}}$ (multi-jet)	5.89 ± 0.92	3.48 ± 0.64	0.63 ± 0.12	0.42 ± 0.07
Jet → fake $\tau_{\text{had-vis}}$ ($t\bar{t}$)	17.4 ± 2.3	1.89 ± 0.28	0	0
Other backgrounds	0.21 ± 0.03	1.75 ± 0.34	0.65 ± 0.13	0.74 ± 0.15
Total background	40.9 ± 3.1	20.4 ± 1.9	6.29 ± 0.88	5.15 ± 0.74
Observed data	40	17	14	4

The comparison of the PNN distributions after the background-only fit with the observed data show decent agreement with exception of the PNN discriminant for resonances with $m_X = 1\,000\,\text{GeV}$ shown in Figure 6.33(c). In the most signal-like bin of the corresponding PNN distribution 14 events are observed, while the background-only model predicts a total of 6.29 ± 0.88 events. This represents a large excess in the observed number of events over the expectation.

The background-only and the signal-plus-background models are compared using tests for the discovery of a signal. The combination of all channels yields an excess at $m_X = 1\,000\,\text{GeV}$ with an observed p-value of 1.3×10^{-3}, which corresponds to a discovery significance of 3.0σ. The excess is localised in the $\tau_{\text{had}}\tau_{\text{had}}$ channel with a significance of 2.8σ when restricting the test to the $\tau_{\text{had}}\tau_{\text{had}}$ SR and the Z + HF CR. The p-values and significances for tests of all considered signal hypotheses are shown in Figure 6.34, separately for the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels and their combination. The significances quoted thus far do not account for multiple hypothesis testing and are therefore referred to as *local significances*. The effect of multiple testing is discussed in Section 6.7.4.

A broad excess can be observed in Figure 6.34 with tests of resonances with masses in the range of 800 to 1 200 GeV being significant at the 2σ-level. At mass scales of 1 000 GeV a broad significance response is expected for a true signal primarily for two reasons: First, the absolute resolution of the m_{HH} reconstruction degrades with increasing m_X. Second, the PNNs are not optimised to distinguish between different signal hypotheses but rather between signal and background, which leads to additional broadening. Figure 6.35 compares the observed significances with the expectation after injecting signals with cross sections at their best-fit values. These injection tests show that the width of the observed significance response is in decent agreement with the expectation for a signal with a mass of about 1 000 GeV.

The best-fit cross section of $pp \to X(m_X = 1\,000\,\text{GeV}) \to HH$ obtained from the fit of the PNN



(a) $m_X = 300\,\text{GeV}$ **(b)** $m_X = 500\,\text{GeV}$

(c) $m_X = 1\,000\,\text{GeV}$ **(d)** $m_X = 1\,600\,\text{GeV}$

Figure 6.33: Distributions of selected PNN discriminants in the $\tau_{\text{had}}\tau_{\text{had}}$ channel after the fit of the background-only model to observed data in all regions. The distributions are shown for four different mass parameter values of the PNN ranging from 300 GeV to 1 600 GeV. The signal overlay is scaled to the expected upper limit on $\sigma(pp \to X \to HH)$ for a given m_X.

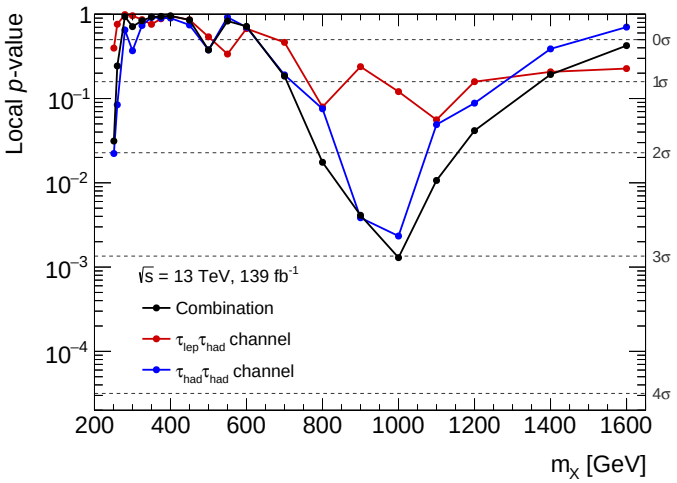


Figure 6.34: Observed local p-values for the comparison of the background-only and the signal-plus-background model as a function of the mass of the resonance. The asymptotic approximation is used for the p-value computation.

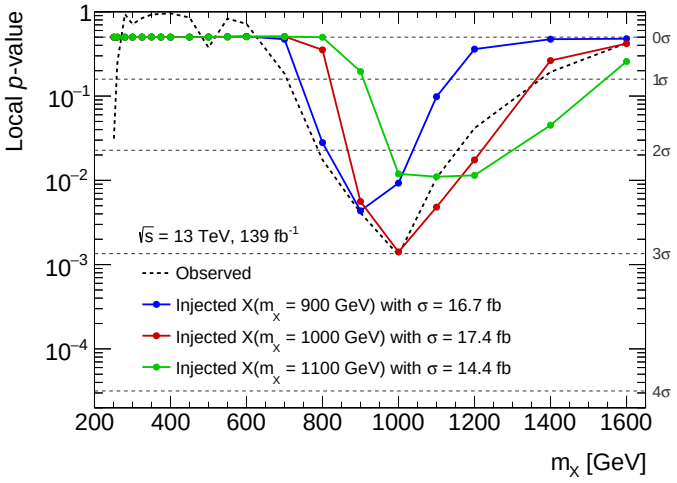


Figure 6.35: Expected local p-values of discovery tests performed on Asimov datasets with injected signals with m_X ranging from 900 to 1 100 GeV. The injected signals are normalised to their best-fit cross section from the unconditional fit to observed data in all regions. The observed p-values (dashed line) are shown for comparison.

discriminants to observed data in all analysis channels is $(17.4\,^{+7.5}_{-6.5})$ fb. The compatibility of the cross section measurement in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels is compared under the assumption of negligible correlation between the measurements in both channels. While a slight tension at the level of 1σ is observed, the measurement in both channels are generally compatible. Further discussion of the excess at $m_X = 1\,000$ GeV is given in Sections 6.7.4 and 6.8.

Upper limits are set on $\sigma(pp \to X \to HH)$ as a function of m_X using the CL_s method at 95 % CL. The observed and expected exclusion limits on the cross section are shown in Figure 6.36. Scalar resonances with a production cross section ranging from 20 to 900 fb, depending on m_X, are excluded given the observed data in all regions. The expected and observed limits are tabulated in Appendix A.5.1 for all m_X and separately for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and their combination.

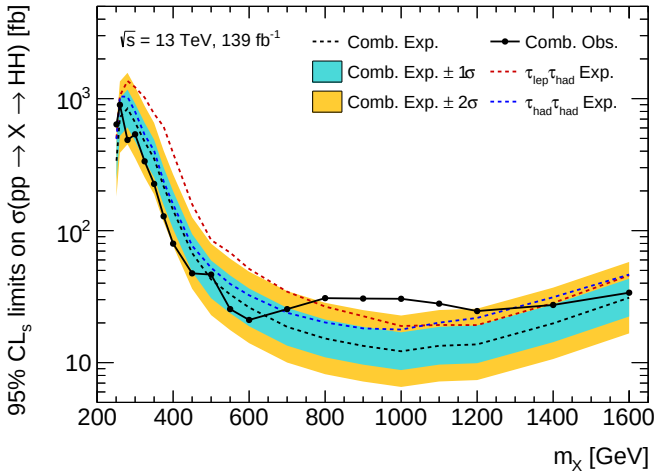


Figure 6.36: Upper limits on $\sigma(pp \to X \to HH)$ as a function of the mass of the scalar resonance. The exclusion limits are obtained using the CL_s method at 95 % CL. The expected upper limits from the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels are overlayed for comparison. The figure is adapted from Ref. [192].

The upper limits on the cross section improve quickly in the m_X range from 300 to 500 GeV as a consequence of the increasing signal acceptance and increasing signal-background separation of the PNNs with m_X. In particular, the ability to distinguish between signal and top-quark backgrounds, which have a large contribution for $m_X \lesssim 500$ GeV, improves quickly as larger resonance masses are considered. For $m_X \geq 1\,000$ GeV, the upper limits start to degrade due to the inability to resolve the constituents of the $H \to \tau^+\tau^-$ and $H \to b\bar{b}$ candidates.

The importance of individual analysis channels to the expected upper limits varies with the considered resonance mass. Except for the smallest m_X, where the $\tau_{\text{had}}\tau_{\text{had}}$ channel has only limited signal acceptance, the $\tau_{\text{had}}\tau_{\text{had}}$ channel gives the most stringent limits in the range of low to intermediate m_X. This is due to the $\tau_{\text{lep}}\tau_{\text{had}}$ channels being dominated by the large, irreducible top-quark background in this regime. For signals with $m_X \gtrsim 1\,000$ GeV both channels are dominated by Z + jets backgrounds in the high PNN score regions and yield similar exclusion limits.

The search for resonant HH production is primarily limited by the data statistical uncertainty, particularly for signals with intermediate to high m_X. This is illustrated in Table 6.27, where the variance on the best-fit cross section from the unconditional fit is decomposed into categories for four

exemplary signal hypotheses. Systematic uncertainties have a sizeable effect only for searches for low-mass resonances. Instrumental uncertainties, most significantly uncertainties affecting jets, $p_{\mathrm{T}}^{\mathrm{miss}}$, and $\tau_{\text{had-vis}}$, have the largest impact on the search when the reconstructed Higgs boson candidates have low transverse momenta, as is the case for signals with low m_X. Similarly, fake-$\tau_{\text{had-vis}}$ backgrounds and uncertainties related to their data-driven estimation play a more important role at low m_X.

Table 6.27: Breakdown of the variance of $\hat{\sigma}$, the MLE of the cross section $\sigma(pp \to X \to HH)$, by uncertainty category for the fit to observed data in all regions. The decomposition is determined in analogy to Table 6.24.

	Explained fraction of variance on $\hat{\sigma}$			
Source	300 GeV	500 GeV	1 000 GeV	1 600 GeV
Data statistical uncertainty	58 %	81 %	86 %	82 %
Systematic uncertainties	42 %	19 %	14 %	18 %
Instrumental uncertainties	10 %	1 %	2 %	< 1 %
Signal modelling uncertainties	2 %	1 %	2 %	3 %
Background statistical uncertainties	19 %	11 %	3 %	11 %
Background modelling uncertainties	14 %	6 %	7 %	4 %
– Top-quark (incl. free normalisation)	3 %	2 %	1 %	< 1 %
– Z + HF (incl. free normalisation)	4 %	1 %	2 %	1 %
– SM Higgs boson	< 1 %	2 %	2 %	2 %
– Fake-$\tau_{\text{had-vis}}$	4 %	< 1 %	1 %	< 1 %
– Other	< 1 %	< 1 %	< 1 %	< 1 %

6.7.4 Global Significance Estimation in the Search for Resonant HH Production

In the search for resonant HH production, multiple hypothesis tests are performed, probing a total of 20 different signal hypotheses. When performing multiple tests, it needs to be considered that under the background-only hypothesis any test could yield a statistically significant result by chance, which would constitute a false discovery (type I error). As the number of tests increases, so does the probability of making one or more false discoveries when not controlling for this effect. The previously quoted significance of 3.0σ of the test at $m_X = 1\,000\,\text{GeV}$ does not account for multiple testing; therefore, the significance was referred to as the local significance. In this section, an alternative test statistic referred to as the *global significance* is introduced that controls for the false discovery rate over a set of hypothesis tests.

The probability of making one or more false discoveries over a set (family) of hypothesis tests is referred to as the family-wise error rate (FWER) [274]. The FWER can be controlled by setting the critical thresholds of individual hypothesis tests such that the FWER remains at an acceptable level. In HEP this is conventionally done by defining the global p-value

$$p_{\text{global}} = \mathbb{P}(Z_{\text{local}}^{\text{max}} > z_{\text{local, obs}}^{\text{max}} \mid H_0)\,,$$

where $Z_{\text{local}}^{\text{max}}$ is the random variable denoting the maximum local significance over all tests, $z_{\text{local, obs}}^{\text{max}}$ is the value of $Z_{\text{local}}^{\text{max}}$ observed in data, and H_0 refers to the background-only hypothesis. The global

significance is then defined as

$$z_{\text{global}} = \Phi^{-1}(1 - p_{\text{global}}) \,,$$

where Φ^{-1} is the quantile function of the Standard Normal distribution. By convention, discovery is claimed if p_{global} (z_{global}) is less (greater) than 2.87×10^{-7} (5σ), which ensures that the FWER is at most 2.87×10^{-7}.

The distribution of $Z^{\text{max}}_{\text{local}}$ under the background-only hypothesis is required to estimate the global significance. In general, this distribution is unknown and needs to be determined by simulation. This is done as follows: First, a toy experiment is drawn from the background-only model. Second, the set of hypothesis tests is performed for a given toy experiment and the maximum local significance over all tests is determined. These steps are repeated N times yielding a sample of realisations of $Z^{\text{max}}_{\text{local}}$ under the background-only hypothesis. For a given value of $z^{\text{max}}_{\text{local, obs}}$, the global p-value can then be estimated according to

$$p_{\text{global}} \approx \frac{1}{N} \times \text{Number of toys with } z^{\text{max}}_{\text{local}} > z^{\text{max}}_{\text{local, obs}} \,,$$

where $z^{\text{max}}_{\text{local}}$ refers to the realisation of $Z^{\text{max}}_{\text{local}}$ for a given toy experiment. Other methods of calculating global significances exist, see for example Ref. [275]; however, they also rely on the ability of drawing toy experiments from the background-only model.

In the search for resonant HH production, the primary difficulty in estimating the global significance lies in the generation of toy experiments from the background-only model. While the likelihood functions used for the statistical interpretation are built based on probabilistic models of the binned PNN discriminants, these models only provide a description of the discriminants for a single mass point. However, a background model that jointly describes all discriminants is required to generate toy experiments for the global significance estimation. The need for such a model was already indicated by the large width of the significance response in the signal injection tests (cf. Figure 6.35), which is caused by a partial overlap in events selected by the most signal-like bins of the PNN discriminants for adjacent mass points.

A substitute background-only model is constructed that accounts for dependencies between all observables entering the statistical interpretation. The model is constructed such that it closely approximates (after marginalising out observables not relevant for a given mass point) the background-only models used for the statistical interpretation in Section 6.7.3. This ensures that the same statistical analysis can be applied to toy experiments without introducing biases due to a mismatch between models. A detailed description of the substitute model developed for this thesis is given in Appendix A.4.2; however, the key parts are highlighted here.

The model is constructed from the pre-fit expectation for all backgrounds except for the $t\bar{t}$ and Z + HF normalisation factors, which are set to approximate post-fit values of 0.97 and 1.35, respectively. The model is divided into three parts, each part describing a certain type of random variable:

Observables Observables are random variables representing the number of events that are observed in a given bin (i.e. n_{cb} in Equation (4.1)). The marginal distributions of the observables are Poisson distributions with known expected value under the background-only hypothesis. When only considering a single PNN discriminant, these observables are mutually independent due to all bins being pairwise disjoint. However, this is not the case when considering the PNN

discriminants for all 20 values of m_X. In this case, the joint distribution of all observables is a multivariate Poisson distribution with non-trivial dependencies between observables.

The joint distribution of observables is estimated as follows: First, the marginal distributions of the observables are extracted from the nominal background-only models. Second, the linear correlation coefficients for all pairs of observables are estimated using MC simulation and CR data, yielding correlation matrices for the observables. Lastly, the marginal distributions of the observables are linked using a Gaussian copula to form the joint distribution of all observables,[19] where the Gaussian copula is defined by the correlation matrix estimated in the second step. Methods of drawing random variates from the joint distribution described by the marginals and the Gaussian copula exist [277] and are used for the generation of toy experiments.

Global observables (Barlow–Beeston method) The global observables related to the simplified Barlow–Beeston method are random variables that describe the effective number of events from simulation or CR data in a given bin (i.e. m_{cb} in Equation (4.2)). Similar to the observables, the effective number of events has non-trivial dependencies between bins of different discriminants. Resampling techniques are applied to MC simulations and CR data to produce alternative datasets for background estimation. Subsequently, the effective number of events is recalculated for every bin using the resampled datasets. The resulting values of the global observables are then used as part of the toy experiments.

Global observables (other systematic uncertainties) The global observables related to all other systematic uncertainties (i.e. a_p in Equation (4.2)) are assumed to be fully correlated for all hypothesis tests and are drawn from the Standard Normal distribution.

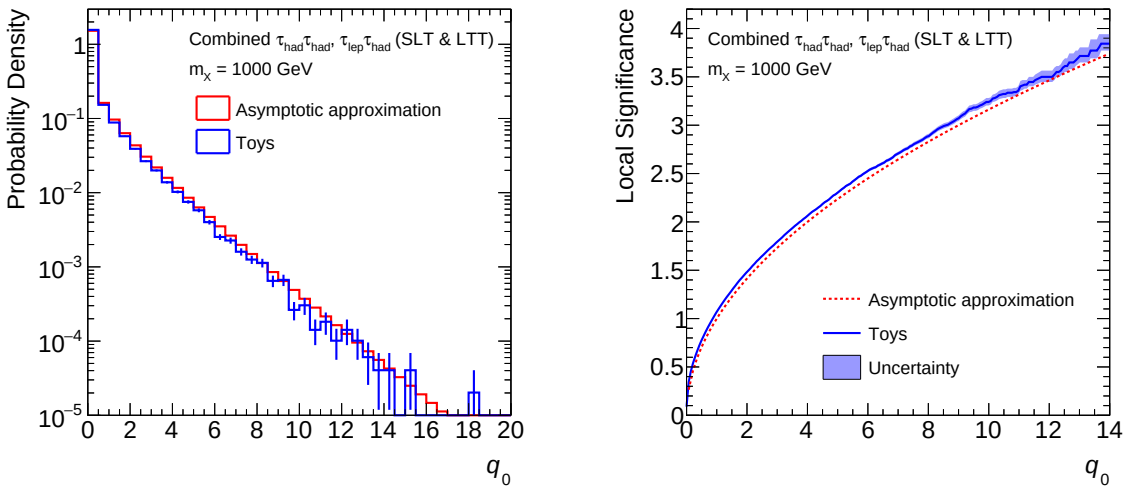
The statistical model used in the Z + HF CR is the same for all hypothesis tests; therefore, standard toy generation methods are used for the observables and the global observables related to the Barlow–Beeston method in the Z + HF CR.

Sampling Distribution of the Discovery Test Statistic

Before proceeding with the estimation of the global significance, the validity of the asymptotic approximation used to determine local significances from observed values of the q_0 test statistic is examined. For this purpose, separate sets of toy experiments are generated using standard toy generation methods. The quality of the approximation is investigated for all 20 hypothesis tests using 100 000 toy experiments per test. A representative example of the q_0 sampling distribution under the background-only hypothesis and the resulting relationship between q_0 and the local significance is shown in Figure 6.37 and compared to the asymptotic approximation.

The asymptotic approximation provides a good description of the relationship between q_0 and z_{local} for all 20 hypothesis tests considered in the search for resonant HH production. Both methods of estimating the local significance agree within a few percent for all hypothesis tests. Nevertheless, toy-based estimates are used for the global significance estimation. In this case, the local significance of the excess at $m_X = 1\,000\,\text{GeV}$ with $q_0 = 9.08$ is estimated to be 3.09 ± 0.03, which is obtained from Figure 6.37(b).

[19] This approach is motivated by Sklar's theorem [276], which states that any n-dimensional joint distribution function can be factorised into n one-dimensional marginal distribution functions and an n-dimensional copula [277]. The copula describes the dependencies between the random variables.



(a) **(b)**

Figure 6.37: Comparison of the q_0 sampling distribution under the background-only hypothesis for the asymptotic approximation and an estimate using 100 000 toy experiments (a). The resulting relationship between the local significance and the observed value of q_0 (b). Both are shown for the test of the $m_X = 1\,000\,\text{GeV}$ signal hypothesis combining all analysis channels.

Global Significance Estimation

A total of 10 000 toy experiments are performed that use observables and global observables drawn from the substitute background-only model outlined previously. The following steps performed for every toy experiment:

1. The values of q_0 are determined for all 20 hypothesis tests considered in the analysis. This step involves performing a conditional (background-only) and an unconditional maximum likelihood fit of the model parameters after replacing observables and global observables in the likelihood functions with values from the toy experiment.

2. The observed values of q_0 are translated into local significances using the toy-based estimates of the relationship between q_0 and z_{local}.

3. The maximum local significance over all tests is determined.

The fits required to obtain the discovery test statistic do not always succeed. A total of 74 fits[20] did not converge even after retrying with altered optimiser settings. The failures are restricted to unconditional fits in cases where a large deficit is observed in a bin with large signal sensitivity. In these cases, the value of q_0 is set to 0 since a deficit does not constitute evidence in favour of the signal-plus-background hypothesis.

The result of the toy experiments is shown in Figure 6.38, which illustrates the effect of multiple hypothesis testing leading to an expectation of $Z_{\text{local}}^{\text{max}}$ of about 1.8 even in the absence of a signal. The

[20] Out of 400 000 fits, i.e. 20 (number of tests) × 2 (conditional & unconditional fit) × 10 000 (number of toys).

toy experiments yield an estimate of the global p-value of

$$p_{\text{global}} = 0.0222 \pm 0.0015\,(\text{toy stat.})$$

and an equivalent global significance of

$$z_{\text{global}} = 2.01 \pm 0.03\,(\text{toy stat.})\,,$$

where only statistical uncertainties from the finite number of toy experiments for the global significance estimation are considered. In addition, the statistical uncertainty of the toy-based estimate of the relationship between q_0 and z_{local} is propagated to the global significance using the bootstrap method [278, 279]. The resulting uncertainty on z_{global} is 0.04 yielding a final result of

$$z_{\text{global}} = 2.01 \pm 0.05\,(\text{total})\,.$$

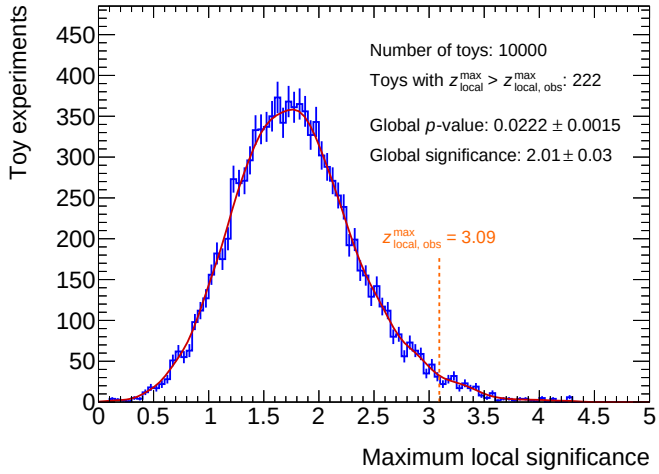


Figure 6.38: Toy-based estimate of the distribution of the maximum local significance, $Z^{\text{max}}_{\text{local}}$, under the background-only hypothesis. The local significance of the largest excess observed in data is indicated in orange. The distribution after kernel smoothing is overlayed in red for illustration purposes. The uncertainty on the global p-value/significance only accounts for the finite number of toy experiments.

Lastly, the toy experiments can be used to estimate the correlations between local significances for all tests under the background-only hypothesis. For this purpose, a signed definition of the local significance given by

$$z^{\text{signed}}_{\text{local}} = \text{sgn}(\hat{\sigma})\sqrt{q_0}$$

is used, where $\text{sgn}(\hat{\sigma})$ refers to the sign of the best-fit cross section. This choice is made for easier interpretation since $z^{\text{signed}}_{\text{local}}$ follows a Standard Normal distribution under the background-only assumption. The correlation matrix between $z^{\text{signed}}_{\text{local}}$ for all 20 hypothesised values of m_X is shown in Figure 6.39. The figure illustrates the increasing correlation between tests at large m_X that was previously observed in the signal injection tests.

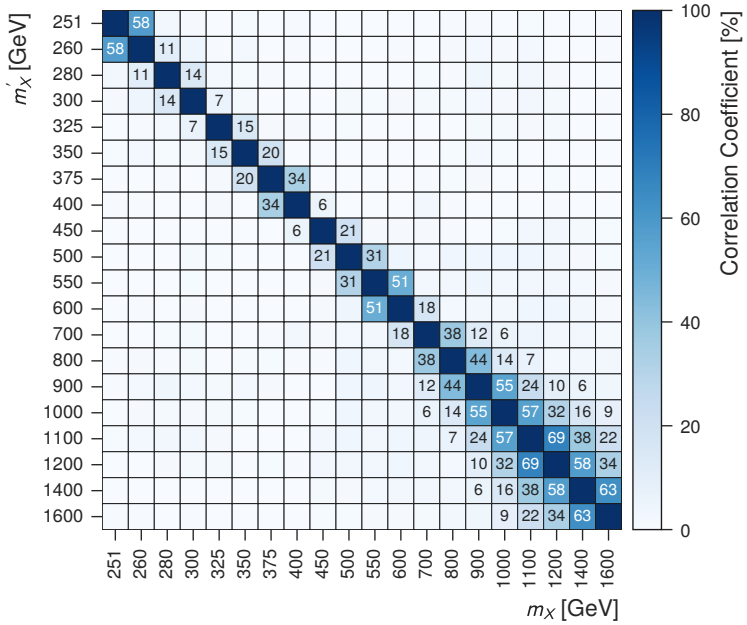


Figure 6.39: Linear correlation coefficients between the signed local significances of hypothesis tests testing the discovery of a signal with mass m_X and m_X' under the background-only hypothesis. The coefficients are estimated using 10 000 toy experiments drawn from the background-only substitute model. Cells are annotated, omitting the diagonal, if the correlation coefficient is greater than or equal to 5 %.

In conclusion, the local (global) significance of the excess observed in data for the test of the $m_X = 1\,000$ GeV signal hypothesis is found to be 3.1σ (2.0σ) using a toy-based estimation method. The uncertainties on the estimated significances are below 0.1 and are omitted in subsequent discussions.

6.8 Conclusion and Outlook

The results of the search for SM HH and resonant HH production in the $b\bar{b}\tau^+\tau^-$ channel are discussed in the following. The search is primarily limited by uncertainties originating from the finite size of the pp collision dataset collected during Run 2 of the LHC, which have the largest impact on the extracted signal strengths and cross sections. Systematic uncertainties are small and only relevant for searches for low-mass resonances. As a result, the sensitivity of this search greatly benefits from the increase in integrated luminosity compared to earlier searches (cf. Section 2.5). Improvements in the analysis, mostly in the reconstruction and selection of physics objects, further increase the sensitivity to resonant and non-resonant HH production. Finally, this chapter is concluded with an outlook on the expected sensitivity of searches for Higgs boson pair production at the end of the High-Luminosity LHC (HL-LHC).

6.8.1 The Search for SM HH Production

To date, the SM HH search presented in this chapter has the highest expected sensitivity to SM HH production of any single analysis channel. This is shown in Figure 6.40, which compares the upper

limits on the SM HH signal strength of searches conducted by the ATLAS and CMS collaborations. In addition, the ATLAS and CMS collaborations performed statistical combinations of the most sensitive channels (cf. Figure 6.40), yielding observed (expected) upper limits on the SM HH signal strength of 2.4 and 3.4 (2.9 and 2.5), respectively [38, 280]. In general, similar results are obtained by both collaborations; however, different analysis strategies are used in the $b\bar{b}b\bar{b}$ channels, which show the largest difference between collaborations. While the ATLAS collaboration only considers event topologies where the jet reconstruction can resolve all four b-jets [187], the CMS collaboration also considers boosted topologies in which the b-jet pairs are reconstructed as large-radius jets [188, 281].

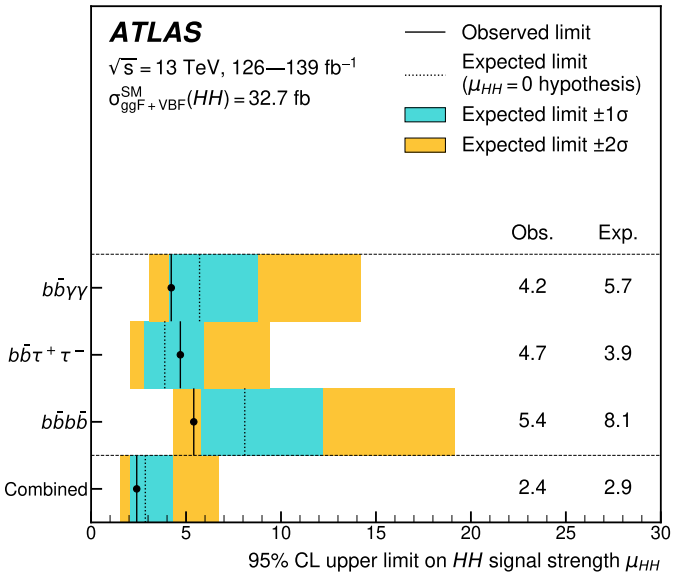
The upper limit on the SM HH signal strength determined in this chapter is compared to the previous result of the ATLAS collaboration in the $b\bar{b}\tau^+\tau^-$ channel (cf. Section 2.5). Using 36.1 fb^{-1} of pp collisions recorded at the beginning of Run 2, the ATLAS collaboration obtained an expected upper limit on the signal strength of SM HH production via ggF of 14.8 for the combination of the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels [184]. With the 139 fb^{-1} pp collision dataset, an expected upper limit of 3.9 is set on the signal strength of SM HH production via ggF and VBF. The improvement in the upper limit exceeds the expectation from a naive extrapolation of the previous result to an integrated luminosity of 139 fb^{-1}, which would yield an expected upper limit of approximately 7.[21] This is primarily for two reasons:

- The signal acceptance in the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{had}}$) channel improved by a factor of about 2 (1.5) due to improvements in b-tagging, $\tau_{\text{had-vis}}$ reconstruction, and $\tau_{\text{had-vis}}$ identification. At the same time, the rates of backgrounds with mistagged jets or fake $\tau_{\text{had-vis}}$ remain comparable to the earlier analysis.

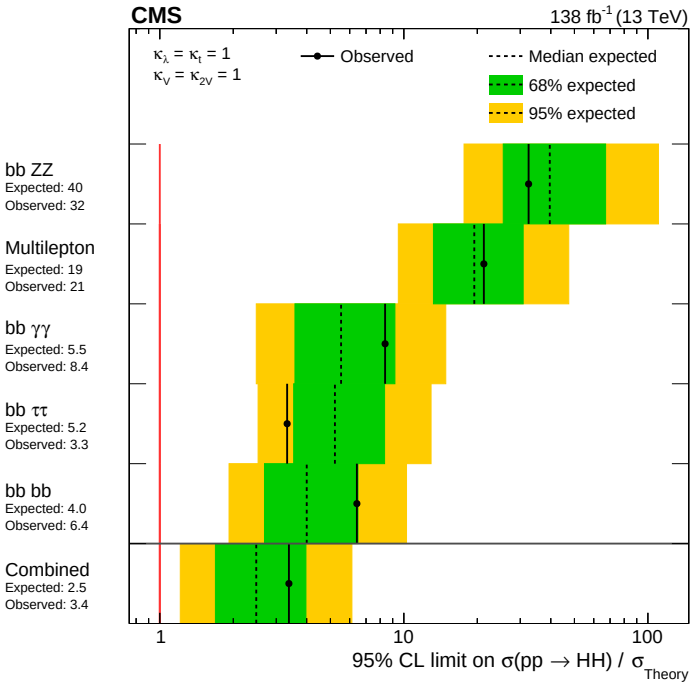
- More events populate the signal-like region of the MVA discriminants as the integrated luminosity increases. This affects the re-binning algorithm due to the constraints on the minimum expected number of background events imposed on all bins. Consequently, as the integrated luminosity increases, the re-binning algorithm produces narrower bins in the high MVA score regions. Since the signal-to-background ratio increases quickly at high MVA score, narrower bins allow for better exploitation of the distinct signature of SM HH production.

An outlook on future searches for Higgs boson pair production is provided by extrapolations of the Run 2 results to the conditions after HL-LHC data-taking. The SM HH search presented in this chapter was extrapolated by the ATLAS collaboration to an integrated luminosity of 3 000 fb^{-1} and $\sqrt{s}$ = 14 TeV in Ref. [283]. This extrapolation is conducted under the assumption that the performance of the ATLAS detector can be maintained at the level of the Run 2 performance in spite of the increase in instantaneous luminosity at the HL-LHC. In addition, the evolution of systematic uncertainties follows the recommendations in Ref. [284]. Given these assumptions the extrapolation of the search for SM HH production in the $b\bar{b}\tau^+\tau^-$ channel yields an expected discovery significance of 2.8σ under the SM hypothesis [283]. A similar extrapolation was performed in the $b\bar{b}\gamma\gamma$ channel resulting in an expected significance of 2.2σ [285]. Furthermore, the prospects of combining the $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ channels at the end of the HL-LHC were investigated in Ref. [45] suggesting that evidence for SM HH production can be obtained with an expected significance of 3.2σ for the combination of both channels. Exceeding the discovery threshold of 5σ at the end of the HL-LHC is realistic but

[21] Neglecting systematic uncertainties and assuming that the upper limit scales with the integrated luminosity as a Poisson counting experiment.



(a) Results of SM HH searches by the ATLAS collaboration. The figure is taken from Ref. [280].



(b) Results of SM HH searches by the CMS collaboration. The multilepton channel targets $WWWW$, $WW\tau\tau$, and $\tau\tau\tau\tau$ final states with two or more reconstructed e, μ, or $\tau_{\text{had-vis}}$ [282]. The figure is taken from Ref. [38].

Figure 6.40: Upper limits on the signal strength of SM HH production at 95 % CL by the ATLAS (a) and CMS (b) collaborations using pp collisions recorded during Run 2 of the LHC. The upper limits are shown for different analysis channels and their combination. The expected upper limits are derived assuming the absence of SM HH production.

would require the combination of the most sensitive channels of the ATLAS and CMS collaborations as well as substantial improvements in the individual search channels.

6.8.2 The Search for Resonant HH Production

The sensitivity of the $b\bar{b}\tau^+\tau^-$ channel to Higgs boson pair production via scalar, narrow-width resonances is complementary to the $b\bar{b}\gamma\gamma$ and $b\bar{b}b\bar{b}$ channels. This is illustrated in Figure 6.41, which compares the upper limits on $\sigma(pp \to X \to HH)$ set by the ATLAS collaboration in the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$ channel. The $b\bar{b}\tau^+\tau^-$ channel provides the highest expected sensitivity of the three channels for resonances in an intermediate mass range from 375 to 800 GeV. For $m_X < 375$ GeV, the $b\bar{b}\gamma\gamma$ channel yields the most stringent upper limits due to its use of di-photon triggers and the excellent background rejection based on the invariant di-photon mass. The highest expected sensitivity to resonances with $m_X > 800$ GeV is provided by the $b\bar{b}b\bar{b}$ channel due to the large $H \to b\bar{b}$ branching ratio, b-jet triggers becoming increasingly efficient at selecting events with large m_{HH}, and the use of special techniques to reconstruct events with highly boosted Higgs bosons for 900 GeV $\leq m_X \leq 5$ TeV.

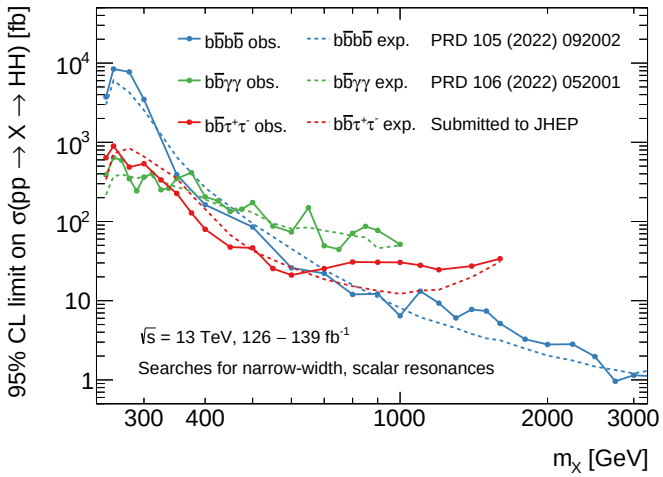


Figure 6.41: Expected (exp.) and observed (obs.) upper limits on $\sigma(pp \to X \to HH)$ at 95 % CL, where X is a scalar, narrow-width resonance with mass m_X. Upper limits are shown separately for searches by the ATLAS collaboration in the $b\bar{b}b\bar{b}$ [286, 287], $b\bar{b}\gamma\gamma$ [189, 288], and $b\bar{b}\tau^+\tau^-$ channel [192].

The excess observed for $m_X = 1\,000$ GeV in the $b\bar{b}\tau^+\tau^-$ channel with a best-fit cross section of $(17.4^{+7.5}_{-6.5})$ fb is compared to results of other searches. The search by the ATLAS collaboration in the $b\bar{b}b\bar{b}$ channel yields an observed (expected) upper limit of 6.5 fb (8.1 fb) on $\sigma(pp \to X \to HH)$ at $m_X = 1\,000$ GeV [286, 287]. Therefore, the best-fit cross section of the excess in the $b\bar{b}\tau^+\tau^-$ channel is excluded at 95 % CL by the search in the $b\bar{b}b\bar{b}$ channel. In addition, the CMS collaboration performed a search for resonant HH production in boosted topologies with $b\bar{b}$ and one or two charged leptons (electrons or muons) in Ref. [289]. This search provides a sensitivity similar to the ATLAS search in the $b\bar{b}b\bar{b}$ channel for $m_X \geq 1$ TeV and yields observed (expected) upper limits on $\sigma(pp \to X \to HH)$ of 4.8 fb (9.6 fb) for $m_X = 1\,000$ GeV [289, 290]. Similarly, this search excludes the excess observed in the $b\bar{b}\tau^+\tau^-$ channel.

In conclusion, a search for resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ channel was presented. This search provides the best expected sensitivity of ATLAS searches in the mass range from 375 to 800 GeV, with upper limits on $\sigma(pp \to X \to HH)$ ranging from 130 to 30 fb in this mass range. A broad excess is observed at $m_X = 1\,000$ GeV with a local (global) significance of 3.1σ (2.0σ). The width of the excess is generally compatible with the expectation of a real signal. However, after accounting for multiple hypothesis testing, the excess was not found to be statistically significant. In addition, other searches by the ATLAS and CMS collaborations do not support the observed excess. Therefore, it is assumed that the excess originates from a statistical fluctuation or a systematic issue, for example, a mismodelling of backgrounds.

CHAPTER 7

Constraining the Strength of the Higgs Boson Self-Coupling

With the pp collision datasets collected during Run 2 of the LHC, direct searches for non-resonant Higgs boson pair production constitute the most sensitive probes of the Higgs boson self-coupling constant, λ_{HHH}. This is due to the large sensitivity of the non-resonant HH production cross section to anomalous values of λ_{HHH}. Hereafter, the self-coupling constant is given in terms of the modifier $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{\text{SM}}$ relating an assumed value of the self-coupling constant to the value predicted by the SM. Accessing the Higgs boson self-coupling is compelling to test the predictions of the SM and to search for deviations that can, for example, originate from BSM phenomena appearing at large energy scales that manifest as changes in the (effective) Higgs boson self-coupling constant.

This chapter presents a reinterpretation of the search for SM HH production (i.e. $\kappa_\lambda = 1$) from Chapter 6 in terms of non-resonant HH production with anomalous values of the Higgs boson self-coupling constant. The reinterpretation is performed in order to set upper limits on the non-resonant production cross section of Higgs boson pairs as a function of κ_λ. These upper limits, when compared to the theoretical cross section predictions, allow the exclusion of κ_λ values that are incompatible with the observations made in the SM HH search.

Previous constraints on κ_λ were set by the ATLAS collaboration using up to 36.1 fb^{-1} of pp collision data taken at the beginning of Run 2 of the LHC. An allowed range of $-5.0 < \kappa_\lambda < 12.0$ at 95 % CL was obtained by combining the results of searches for non-resonant HH production in the $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, and $b\bar{b}\gamma\gamma$ channels [88]. The methods used for the reinterpretation performed in this chapter are largely adopted from the earlier result published in Ref. [88]. The following focuses on the reinterpretation of the search in the $b\bar{b}\tau^+\tau^-$ channel with 139 fb^{-1} of pp collision data. Some results of this chapter were published by the ATLAS collaboration in Ref. [291].

This chapter is structured as follows: Section 7.1 describes the phenomenology of a non-resonantly produced HH signal with anomalous values of κ_λ. The reinterpretation of the SM HH search, including the statistical model, assumptions, and limitations, is discussed in Section 7.2. The upper limits on the cross section and the excluded intervals of κ_λ are presented in Section 7.3. A conclusion and outlook is given in Section 7.4.

7.1 Phenomenology of Higgs Boson Pair Production with Anomalous Higgs Boson Self-Coupling Strength

Before describing the reinterpretation of the SM HH search, the experimental signature of non-resonant HH production with anomalous values of κ_λ is discussed. The aim is to understand the sensitivity of direct searches for non-resonant HH production and the sensitivity of individual search channels.

The non-resonant HH production cross section via ggF and VBF is shown in Figure 7.1 as a function of κ_λ. The ggF production mode is the dominant contribution to Higgs boson pair production throughout the considered κ_λ range. For this production mode, the destructive interference between the box and triangle diagram (cf. Figure 2.6) becomes maximal at about $\kappa_\lambda = 2.3$ at which point the cross section reaches a minimum of approximately 13 fb. Similar behaviour is observed for the VBF production mode, although involving different diagrams (cf. Figure 2.7) and resulting in a different location of the minimum.

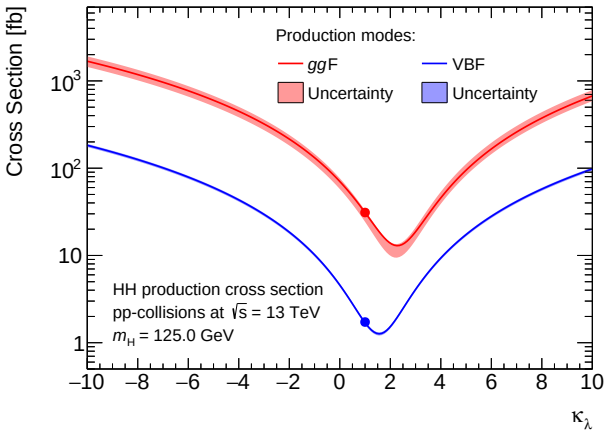
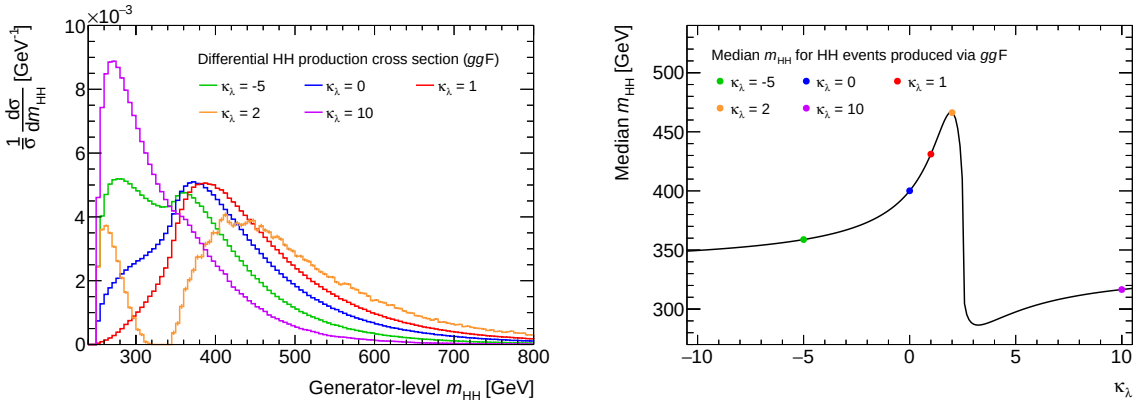


Figure 7.1: Cross section of non-resonant HH production via ggF and VBF as a function of κ_λ. The production cross section for ggF is given at $\text{NNLO}_{\text{NLO-improved}}$ [292] rescaled to $\text{NNLO}_{\text{FTapprox}}$ [46] in the $\kappa_\lambda = 1$ limit [47, 48, 293]. The cross section for the VBF production mode is obtained from simulation with MadGraph5_aMC@NLO at LO after applying an N^3LO k-factor derived for the $\kappa_\lambda = 1$ case [48, 49]. The cross sections are parameterised as quadratic functions of κ_λ. Theoretical uncertainties are shown as coloured bands.

In addition to the change in total cross section, anomalous values of κ_λ alter the differential HH production cross section predominantly in terms of the invariant mass of the pair of Higgs bosons. This is shown in Figure 7.2(a) for the ggF production mode and for five exemplary values of κ_λ. The m_{HH} spectra for different values of κ_λ show large differences in their *hardness* as measured by the median of the m_{HH} distribution in Figure 7.2(b). For κ_λ values just below or at the point of maximum destructive interference between the box and triangle diagram, the m_{HH} spectra are moderately hard and have a pronounced double peak structure. For other values of κ_λ, particularly for $\kappa_\lambda \approx 3$, the cross section at low m_{HH} is enhanced resulting in softer m_{HH} spectra.

The κ_λ-dependency of $\mathrm{d}\sigma/\mathrm{d}m_{HH}$ is an important factor affecting the sensitivity of searches to signals with anomalous κ_λ. This is due to limitations in signal acceptance at low m_{HH} in most search channels. Particularly searches targeting (visible) hadronic final states such as $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ and $b\bar{b}b\bar{b}$



(a) **(b)**

Figure 7.2: Differential cross section of Higgs boson pair production with respect to m_{HH} for the ggF production mode and selected values of κ_λ (a) and the median value of m_{HH} as a function of κ_λ (b). Both are given for pp collisions at $\sqrt{s} = 13\,\text{TeV}$ and assuming $m_H = 125.0\,\text{GeV}$. The cross sections are obtained from simulation with POWHEG BOX v2 at NLO including the top-quark mass dependence [227, 228] for $\kappa_\lambda = 0, 1, 10$. Differential cross sections for other values of κ_λ are estimated using morphing techniques, which are discussed in Section 7.2.1.

are affected. While these channels provide stringent cross section limits for the SM case, scenarios where the cross section at low m_{HH} is enhanced are less favourable due to the strict selections applied by b-jet and $\tau_{\text{had-vis}}$ triggers. In these cases, the $b\bar{b}\gamma\gamma$ channel, due to its use of di-photon triggers, yields larger signal acceptance and the most stringent exclusion limits of a single channel.

7.2 Reinterpretation of the Search for SM HH Production

The SM HH search from Chapter 6 is reinterpreted to set upper limits on the cross section of non-resonant HH production as a function of κ_λ. The reinterpretation adopts the statistical framework presented in Section 6.7 with a few modifications. These modifications are restricted to the signal model used in the statistical interpretation. The background model and final discriminants, including their binning, remain identical to those in the SM HH search.

The signal model used for the statistical interpretation is obtained by replacing the SM HH signal with signals from non-resonant HH production with arbitrary but fixed κ_λ. Similar to the SM HH search, the combination of non-resonant HH production via ggF and VBF is considered as the signal process. The signal is normalised using the total non-resonant HH production cross section via ggF and VBF, $\sigma_{gg\text{F+VBF}}$, which is free to vary and is considered as the POI. The methods for obtaining signal templates in the SRs are described in Section 7.2.1.

The adopted method of reinterpretation makes assumptions that are given in the following: First, except for κ_λ, all coupling strengths are assumed to be at their SM values. Second, the single-Higgs-boson production cross sections and Higgs boson branching ratios are fixed to their SM values and are thus assumed to be independent of κ_λ. This is generally not the case since variations of κ_λ affect both the production cross sections and branching ratios due to higher-order electroweak corrections. These effects and their sensitivity to κ_λ are discussed in Refs. [43, 294, 295]. Since backgrounds from

single-Higgs-boson production are non-negligible in the SM HH search, it is instructive to gauge the quality of this approximation over the allowed interval of $-5.0 < \kappa_\lambda < 12.0$ from previous results of the ATLAS collaboration [88]. In the considered κ_λ interval, the single-Higgs-boson production cross sections deviate from the SM prediction by up to 12 % (25 %) for the ggF, VBF, and VH ($t\bar{t}H$) production modes [294]. The Higgs boson branching ratios to fermions show small relative deviations of up to 3 % from the SM over the relevant κ_λ range [294]. For the purpose of this reinterpretation, signal processes were generated assuming SM Higgs boson branching ratios for consistency with the treatment of single-Higgs-boson backgrounds. The assumptions about single-Higgs-boson production cross sections and branching ratios were dropped in a follow-up analysis performed as part of Ref. [280] by the ATLAS collaboration, which is briefly discussed in the conclusion of this chapter.

7.2.1 Signal Templates and Uncertainties

The distributions of the signal process in the three SRs and for various values of κ_λ are required for the reinterpretation. These are obtained using morphing and re-weighting techniques developed by the ATLAS collaboration that are explained hereafter. An ingredient for both methods are simulated event samples for different assumed values of κ_λ. The event simulation proceeds using the same generator setup described in Section 6.1. Events from non-resonant HH production are generated for $\kappa_\lambda = 1, 10$ for the ggF production mode and $\kappa_\lambda = 0, 1, 10, 20$ for the VBF production mode. In addition, large samples of events from non-resonant HH production via ggF are generated for $\kappa_\lambda = 0, 1, 10, 20$ without simulation of the ATLAS detector. All event samples are normalised using the cross sections previously shown in Figure 7.1.

Established morphing techniques [296] are used to obtain signal templates for arbitrary κ_λ by linearly combining a set of basis templates with fixed κ_λ. This approach is motivated by considering the squared matrix element of non-resonant HH production for either the ggF or VBF production mode at leading order in κ_λ. The squared matrix element can be written as

$$|\mathcal{M}|^2 = \kappa_\lambda^2 |\mathcal{A}|^2 + 2\kappa_\lambda \,\mathrm{Re}(\mathcal{A}\mathcal{B}^*) + |\mathcal{B}|^2 \,,$$

where $\mathcal{A}$ is the (complex) amplitude of diagrams involving the Higgs boson-self coupling after factoring out the value of the anomalous self-coupling constant, and $\mathcal{B}$ the amplitude of diagrams not involving any Higgs boson self-interactions. The knowledge of $|\mathcal{M}|^2$ for three distinct values of κ_λ can be used to calculate the squared matrix element for any value of κ_λ by solving for the squared amplitudes of $\mathcal{A}$ and $\mathcal{B}$, and the interference term. The same considerations can be applied to express the differential cross section for any κ_λ as a linear combination of predictions at $\kappa_\lambda = a, b, c$ according to

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\boldsymbol{x}}(\kappa_\lambda) = w_1(\kappa_\lambda)\, \frac{\mathrm{d}\sigma}{\mathrm{d}\boldsymbol{x}}\bigg|_{\kappa_\lambda=a} + w_2(\kappa_\lambda)\, \frac{\mathrm{d}\sigma}{\mathrm{d}\boldsymbol{x}}\bigg|_{\kappa_\lambda=b} + w_3(\kappa_\lambda)\, \frac{\mathrm{d}\sigma}{\mathrm{d}\boldsymbol{x}}\bigg|_{\kappa_\lambda=c} \,,$$

where the coefficients $w_i(\kappa_\lambda)$ are second degree polynomials in κ_λ and $\boldsymbol{x}$ is a variable characterising the scattering process.[1]

For the ggF production mode, this morphing technique is only used indirectly. Differential distributions at generator-level are obtained by linearly combining the $\kappa_\lambda = 0, 1, 20$ event samples

[1] For non-resonant HH production via ggF, this method was documented by the ATLAS collaboration in Ref. [297].

without detector simulation. These distributions are used to derive a re-weighting in bins of the generator-level m_{HH} that can be used to re-weight the $\kappa_\lambda = 1$ event sample to any value of κ_λ. The re-weighting is then applied to the $\kappa_\lambda = 1$ sample with full detector simulation. This approach avoids having to run the computationally expensive detector simulation for all event samples required for the morphing technique. For the VBF production mode, the morphing technique is used directly by performing linear combinations of the $\kappa_\lambda = 1, 2, 10$ event samples. In this case, the re-weighting method is not used because the effects of κ_λ variations are not easily parameterised as a re-weighting in a single variable. Event samples not used for morphing or re-weighting are used for validation purposes.

Uncertainties on the modelling of the signal processes with anomalous κ_λ are derived according to Section 6.6.2. Variations of the $H \to \tau^+\tau^-$ and $H \to b\bar{b}$ branching ratios, the parton shower simulation, the renormalisation and factorisation scales, and the PDFs and α_s are considered. Additionally, the statistical uncertainties of the κ_λ re-weighting factors are accounted for.

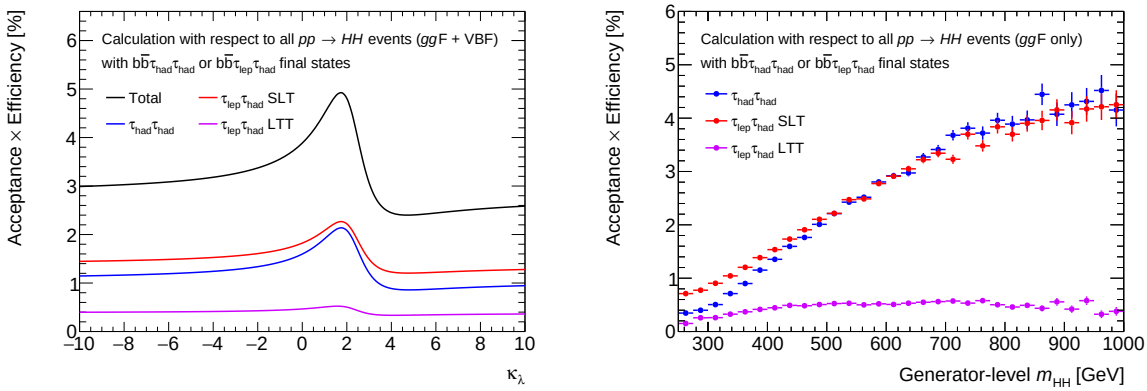
7.2.2 Signal Acceptance in the $b\bar{b}\tau^+\tau^-$ Channel

The strength of the $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ and $b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$ SLT/LTT channels in setting upper limits on the cross section of non-resonant HH production can be qualitatively understood by examining the signal acceptance of the channels as a function of κ_λ. The signal acceptance is shown in Figure 7.3(a) for the SR selections of the three channels and their combination. It reaches a maximum at $\kappa_\lambda \approx 2$, which corresponds to the value of the self-coupling constant with the largest median m_{HH} for signal events produced via ggF (cf. Figure 7.2). In addition, Figure 7.3(b) depicts the signal acceptance in bins of the generator-level m_{HH} for the dominant ggF production mode, illustrating the limitations in selecting signal events with low m_{HH}. While the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channels have similar signal acceptances over a wide range of m_{HH}, the signal acceptance in the low m_{HH} region is dominated by the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel. The inclusion of the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel, which selects events with leptons below the lepton p_{T} threshold of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel, is primarily intended to improve the signal acceptance at low m_{HH}, which is where its relative contribution to the total signal acceptance is largest.

The signal acceptance is not the only factor in determining the sensitivity of the reinterpretation. First, signals with enhanced cross sections at low m_{HH} have a larger overlap with background processes. Second, the BDT used for signal extraction is trained to distinguish between SM HH events and background events. Consequently, the signal-background classification is biased towards events with large m_{HH} due to the moderately hard m_{HH} spectrum of SM HH production. Both factors lead to further reductions in sensitivity to signals with soft m_{HH} spectra.

7.3 Results

Upper limits are set on $\sigma_{gg\text{F+VBF}}$ at 95 % CL by combining the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ SLT, and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channels. The exclusion limits obtained by the ATLAS collaboration are shown in Figure 7.4 as a function of κ_λ. They are compared to the prediction of the combined non-resonant HH production cross section via ggF and VBF from theory, previously shown in Figure 7.1. The most stringent limits are set for $\kappa_\lambda \approx 2$, which follows from signal acceptance considerations discussed previously. The theory prediction for the non-resonant HH production cross section exceeds the observed upper limit outside



(a) Signal acceptance (ggF + VBF) as a function of κ_λ. The combination of all channels is shown in black.

(b) Signal acceptance (ggF only) in bins of the generator-level m_{HH}.

Figure 7.3: Acceptance of events from non-resonant HH production in the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ SLT, and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channels. The signal acceptance is calculated as the fraction of events with $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ or $b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$ final states passing the SR selections of a given channel.

the interval of $\kappa_\lambda \in [-2.4, 9.2]$ thus excluding non-resonant HH production with $\kappa_\lambda \notin [-2.4, 9.2]$ based on the upper limits on the cross sections.[2]

This result represents a significant improvement over earlier searches by the ATLAS collaboration using 36 fb^{-1} of pp collision data collected at the start of Run 2 of the LHC. These searches yielded allowed κ_λ intervals of $\kappa_\lambda \in [-7.4, 15.7]$ and $\kappa_\lambda \in [-5.0, 12.0]$ for the $b\bar{b}\tau^+\tau^-$ channel and the combination of the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels [88], respectively. Due to the increased size of the pp collision dataset and the analysis improvements in the $b\bar{b}\tau^+\tau^-$ channel, the reinterpretation presented in this chapter puts more stringent constraints on κ_λ than the earlier combination of the most sensitive channels.

A comparison between the allowed κ_λ intervals of searches in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels performed at the end of Run 2 by the ATLAS collaboration is given in Table 7.1(a). The results of searches in the $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ channels are complementary. While the $b\bar{b}\gamma\gamma$ channel sets more stringent upper bounds on the allowed κ_λ interval due to its superior acceptance of signal events with low m_{HH}, the $b\bar{b}\tau^+\tau^-$ channel provides a lower bound that is competitive with the result of the $b\bar{b}\gamma\gamma$ channel. The $b\bar{b}b\bar{b}$ channel is the third most sensitive channel due to its limited signal acceptance and large multi-jet backgrounds; however, these disadvantages are partially compensated by the large $H \to b\bar{b}$ branching ratio. Lastly, the corresponding results of the CMS collaboration are summarised in Table 7.1(b) showing similar findings to those of the ATLAS collaboration.

7.4 Conclusion and Outlook

An interpretation of the search for non-resonant HH production in the $b\bar{b}\tau^+\tau^-$ channel in terms of anomalous Higgs boson self-coupling strengths is presented. Upper limits at 95 % CL are set on the

[2] The constraints on κ_λ are largely driven by the $\tau_{\text{had}}\tau_{\text{had}}$ channel. An analysis restricted to the $\tau_{\text{had}}\tau_{\text{had}}$ channel yields observed (expected) κ_λ intervals of $\kappa_\lambda \in [-2.5, 9.3]$ ($\kappa_\lambda \in [-2.3, 9.2]$), while the combination of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channel yields $\kappa_\lambda \in [-3.8, 12.2]$ ($\kappa_\lambda \in [-3.8, 11.9]$) [298].

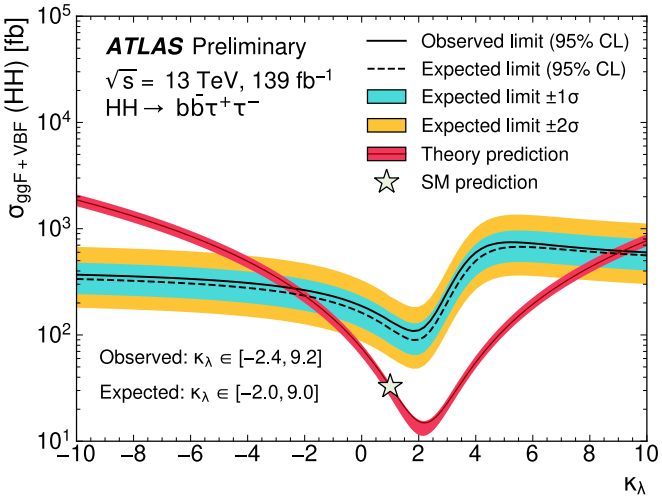


Figure 7.4: Upper limits at 95 % CL on the Higgs boson pair production cross section via ggF and VBF, $\sigma_{gg\text{F+VBF}}$, as a function of κ_λ. The limits are set for the combination of all analysis channels. The expected upper limits assume the absence of non-resonant HH production ($\sigma_{gg\text{F+VBF}} = 0$). The κ_λ intervals defined by the intersection of the theory prediction with the observed/expected limits are given in the lower left of the figure. The figure is taken from Ref. [291].

Table 7.1: Allowed κ_λ intervals from searches for non-resonant HH production in $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels by the ATLAS and CMS collaborations using pp collision datasets collected during Run 2 of the LHC. The expected κ_λ intervals are derived from the expected upper limits on $\sigma_{gg\text{F+VBF}}$ under the background-only ($\sigma_{gg\text{F+VBF}} = 0$) hypothesis.

(a) Results of the ATLAS collaboration with an integrated luminosity of up to 139 fb^{-1}. †: The integrated luminosity of the search in the $b\bar{b}b\bar{b}$ channel amounts to 126 fb^{-1}.

ATLAS	Allowed κ_λ interval		
Channel	Observed	Expected	Reference
$b\bar{b}b\bar{b}^{\dagger}$	$[-3.9, 11.1]$	$[-4.6, 10.8]$	[187]
$b\bar{b}\tau^+\tau^-$	$[-2.4, 9.2]$	$[-2.0, 9.0]$	[291]
$b\bar{b}\gamma\gamma$	$[-1.5, 6.7]$	$[-2.4, 7.7]$	[189]

(b) Results of the CMS collaboration with an integrated luminosity of 138 fb^{-1}. ‡: The search in the $b\bar{b}\gamma\gamma$ channel accounts for the change in single-Higgs-boson production cross section and branching ratio with κ_λ.

CMS	Allowed κ_λ interval		
Channel	Observed	Expected	Reference
$b\bar{b}b\bar{b}$	$[-2.3, 9.4]$	$[-5.0, 12.0]$	[188]
$b\bar{b}\tau^+\tau^-$	$[-1.7, 8.7]$	$[-2.9, 9.8]$	[186]
$b\bar{b}\gamma\gamma^{\ddagger}$	$[-3.3, 8.5]$	$[-2.5, 8.2]$	[190]

total cross section of Higgs boson pair production via ggF and VBF as a function of the self-coupling constant. An allowed range of

$$-2.4 \leq \kappa_\lambda \leq 9.2 \text{ (observed)} \qquad -2.0 \leq \kappa_\lambda \leq 9.0 \text{ (expected for } \sigma_{gg\text{F+VBF}} = 0)$$

is obtained by comparing the cross section exclusion limits at 95 % CL with cross section predictions from theory. The results are based on the assumption that other coupling constants are at their SM values. In addition, changes in single-Higgs-boson production cross sections and Higgs boson branching ratios due variations of κ_λ are neglected.

A follow-up analysis is performed by the ATLAS collaboration in Ref. [280] that combines the results of the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels to perform a direct measurement of κ_λ. Similar to the analysis presented in this chapter, the follow-up analysis includes the results of the SM HH search from Chapter 6 as part of the reinterpretation. The main differences with respect to this chapter are as follows: First, the statistical model uses κ_λ as the POI, allowing the estimation of κ_λ confidence intervals using likelihood ratio tests. Second, the κ_λ-dependencies of single-Higgs-boson production cross sections and Higgs boson branching ratios are accounted for. The follow-up analysis estimates κ_λ confidence intervals at 95 % CL for the combination of the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels to be $\kappa_\lambda \in [-0.6, 6.6]$ for the observed data and $\kappa_\lambda \in [-2.1, 7.8]$ for the expectation under the $\kappa_\lambda = 1$ hypothesis [280]. When considering only the search in the $b\bar{b}\tau^+\tau^-$ channel, the κ_λ confidence intervals are

$$-2.7 \leq \kappa_\lambda \leq 9.5 \text{ (observed)} \qquad -3.1 \leq \kappa_\lambda \leq 10.2 \text{ (expected for } \kappa_\lambda = 1)$$

at 95 % CL [299]. The observed κ_λ confidence interval shows decent agreement with the result obtained in this chapter that neglected κ_λ-dependent electroweak corrections. The expected κ_λ intervals are derived under different assumptions, therefore, larger differences are expected.

Probing the nature of the Higgs boson self-coupling is an important physics goal of the HL-LHC. Prospects of the sensitivity to the self-coupling constant with $3\,000\,\text{fb}^{-1}$ pp collision data at $\sqrt{s} = 14\,\text{GeV}$ are derived in Ref. [283] by projecting the results of the SM HH search in the $b\bar{b}\tau^+\tau^-$ channel. This extrapolation yields an expected confidence interval of $\kappa_\lambda \in [-0.3, 7.4]$ at 95 % CL assuming the $\kappa_\lambda = 1$ hypothesis [283]. With the pp collision dataset collected during the HL-LHC, a combination of searches in the $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ channels is expected to exclude the $\kappa_\lambda = 0$ hypothesis at 95 % CL under the assumption that $\kappa_\lambda = 1$ [45]. This would represent a first milestone in experimentally supporting the Higgs boson self-interaction predicted by the SM, showing that a model without Higgs boson self-coupling would be disfavoured.

CHAPTER 8

Conclusion

This thesis presents a search for Higgs boson pair production in final states with two b quarks and two τ leptons using 139 fb^{-1} of pp collision data recorded by the ATLAS experiment during Run 2 of the LHC. Searches for Higgs boson pair production are instrumental in testing our understanding of electroweak symmetry breaking in the Standard Model, allowing the shape of the Higgs potential—a fundamental component of the theory proposed by Brout, Englert, and Higgs—to be probed. These searches are challenging, however, since Higgs boson pair production is three orders of magnitude less abundant than single-Higgs-boson production at the LHC. Nevertheless, searches for Higgs boson pair production are already of interest for probing physics beyond the Standard Model. This is because the presence of new physics could alter the non-resonant HH production cross section or introduce resonant enhancements of the cross section due to new particles decaying into pairs of Higgs bosons.

The search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ channel considers final states with one or two hadronic τ-lepton decays. A particular focus is put on the sub-channel with two hadronic τ-lepton decays in this thesis. This channel relies on the ability to efficiently identify $\tau_{\text{had-vis}}$ while suppressing backgrounds from quark- or gluon-initiated jets. For this purpose, a novel tau identification method is introduced that uses sequences of reconstructed charged-particle tracks and topo-clusters in the calorimeters as well as high-level information about $\tau_{\text{had-vis}}$ candidates as inputs. This is made possible by the use of a recurrent neural network (RNN) architecture that allows for the processing of variable-length sequences. The new method significantly improves the tau identification performance over the algorithm previously used at the ATLAS experiment. At fixed $\tau_{\text{had-vis}}$ efficiency working points, the RNN-based tau identification improves the rejection of fake $\tau_{\text{had-vis}}$ from quark- or gluon-initiated jets by 80 % (40 to 80 %) for 1-prong (3-prong) $\tau_{\text{had-vis}}$ candidates. The RNN-based tau identification was adopted by the ATLAS collaboration as the recommended tau identification algorithm for analyses of the Run 2 pp collision dataset. Furthermore, it was adapted for the use in $\tau_{\text{had-vis}}$-triggers for data-taking at the end of Run 2 and the beginning of Run 3 of the LHC.

In this thesis, three different modes of Higgs boson pair production are considered: non-resonant HH production predicted by the Standard Model (SM HH production); resonant HH production via massive, scalar particles with small decay widths; and non-resonant HH production with anomalous values of the Higgs boson self-coupling constant.

The search for SM HH production is optimised for the dominant gluon–gluon fusion production mode; however, the vector boson fusion production mode is included in the interpretation of the results. Improved object reconstruction and identification techniques are employed, among them the

RNN-based tau identification, leading to an increase in signal acceptance ranging from 50 to 100 % with respect to earlier searches by the ATLAS collaboration in the $b\bar{b}\tau^+\tau^-$ channel. No statistically significant signal is observed in the search; therefore, upper limits are set on the SM HH signal strength and production cross section. The observed (expected) upper limit at 95 % CL on the SM HH signal strength is 4.7 (3.9). The observed (expected) upper limit at 95 % CL on the SM HH production cross section is 140 fb (110 fb). To date, this search provides the highest expected sensitivity to SM HH production of any individual channel.

The search for resonant HH production targets scalar resonances with masses between 251 and 1 600 GeV decaying into pairs of Higgs bosons. The $b\bar{b}\tau^+\tau^-$ channel provides upper limits that are competitive with other searches in a mass range from 375 to 800 GeV. Over this range, the upper limits on the $pp \to X \to HH$ cross section range from 130 to 30 fb at 95 % CL. The largest excess is observed for a resonance mass of 1 000 GeV with a local (global) significance of 3.1σ (2.0σ). This excess is not statistically significant given the conventional significance thresholds for the discovery of new physics. In addition, it is not supported by other search channels of the ATLAS and CMS collaborations, which set more stringent upper limits for $m_X = 1\,000$ GeV.

Lastly, the search for SM HH production is reinterpreted in the context of non-resonant HH production with anomalous values of the Higgs boson self-coupling constant. Upper limits at 95 % CL are set on the non-resonant HH production cross section for different hypothesised values of the Higgs boson self-coupling modifier, κ_λ. Based on the observed (expected) upper limits on the cross section, κ_λ is constrained to be within $-2.4 \leq \kappa_\lambda \leq 9.2$ ($-2.0 \leq \kappa_\lambda \leq 9.0$).

At present, no golden channel for searches for non-resonant HH production exists. Instead, the searches in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ channels have similar sensitivities and are complementary. As a result, combinations of these and other channels are important to obtain the most stringent constraints on the SM HH signal strength and the Higgs boson self-coupling constant. Such combinations were performed by the ATLAS and CMS collaborations after Run 2 of the LHC, yielding upper limits at 95 % CL on the SM HH signal strength of 2.4 [280] and 3.4 [38], respectively, significantly outperforming any individual search channel. Similarly, the best experimental constraints on the Higgs boson self-coupling are obtained by combining multiple channels. The Higgs boson self-coupling modifier is constrained (at 95 % CL) to be within $-0.6 \leq \kappa_\lambda \leq 6.6$ by the ATLAS collaboration [280] and $-1.2 \leq \kappa_\lambda \leq 6.5$ by the CMS collaboration [38] for the combination of the most sensitive channels.

Searches for Higgs boson pair production remain an important part of the physics programmes at the LHC and future colliders. Under the Standard Model assumption, first evidence for SM HH production is likely to be obtained at the HL-LHC. When combining the most sensitive searches of the ATLAS and CMS collaborations, discovery of SM HH production might even be achieved using the full HL-LHC pp collision dataset. Furthermore, the Higgs boson self-coupling constant will be measured with a precision of the order of 50 % at the HL-LHC [80]. Among the main goals of post-LHC colliders—for example, the FCC [300], ILC [301], or CLIC [302]—is the measurement of the Higgs boson self-coupling with uncertainties of 10 % and below, which would greatly increase the sensitivity to physics beyond the Standard Model.

APPENDIX A

Supplementary Material on the Search for Higgs Boson Pair Production

A.1 Event Displays

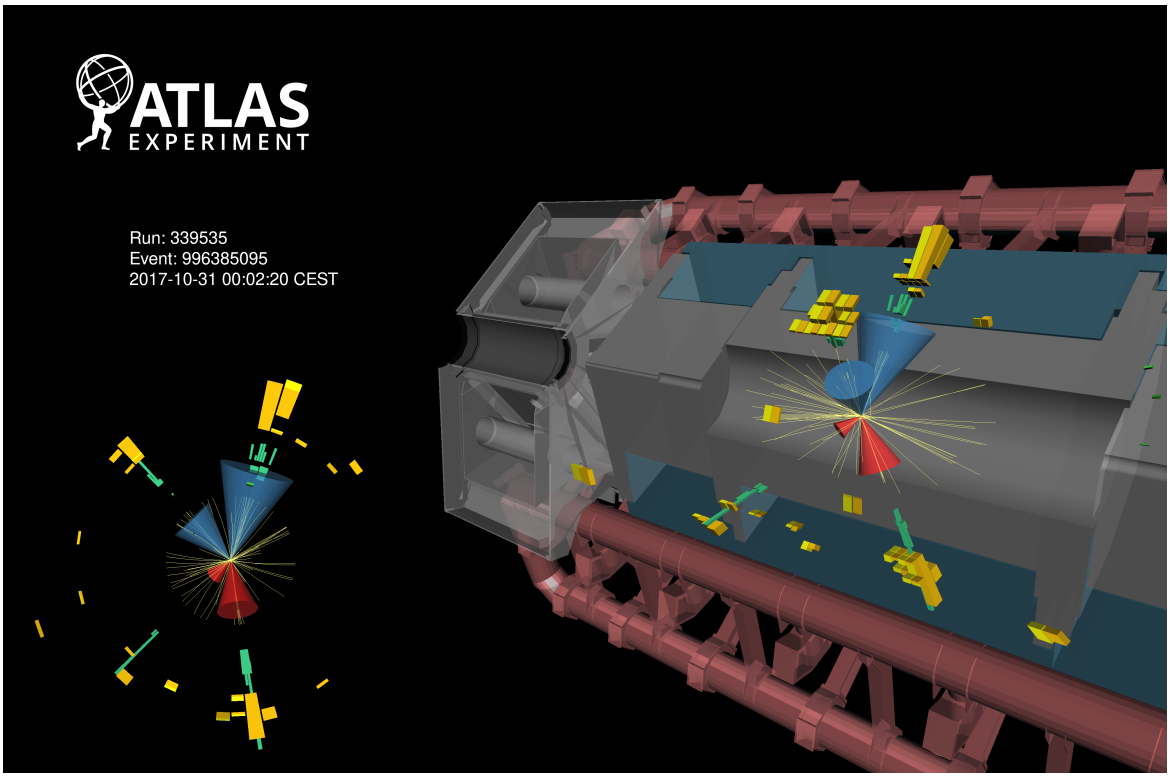


Figure A.1: Visualisation of the SM HH candidate event with the largest BDT score observed in data in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The two b-tagged jets (blue cones) have transverse momenta of 160 GeV and 100 GeV. The two $\tau_{\text{had-vis}}$ candidates (red cones) have transverse momenta of 100 GeV and 40 GeV. Energy deposited in cells of the electromagnetic (hadronic) calorimeters is visualised as green (yellow) towers. Reconstructed charged-particle tracks are depicted as yellow lines. The event has $m_{\tau\tau}^{\text{MMC}} = 130$ GeV, $m_{bb} = 130$ GeV, and $m_{HH} = 510$ GeV. The image is taken from Ref. [192].

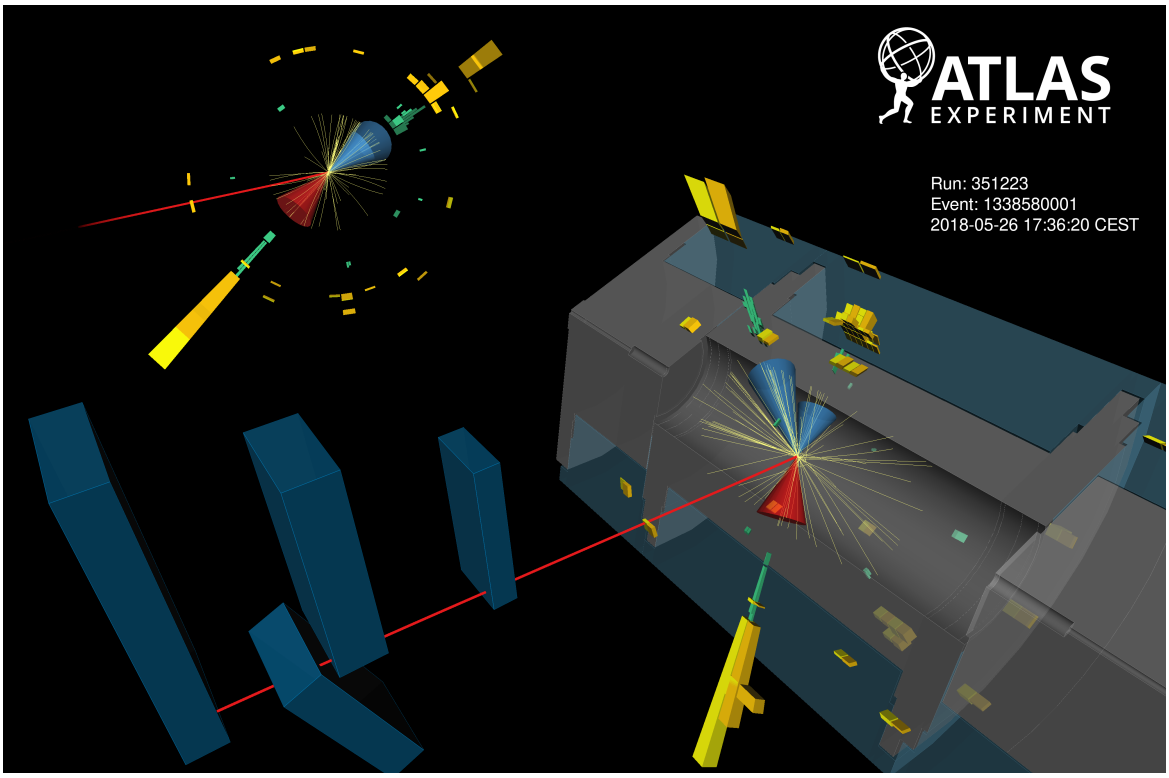


Figure A.2: Visualisation of the SM HH candidate event with the largest NN score observed in data in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel. The two b-tagged jets (blue cones) have transverse momenta of 190 GeV and 90 GeV. The $\tau_{\text{had-vis}}$ candidate (red cone) has a transverse momentum of 80 GeV. The muon (red line) has a transverse momentum of 30 GeV. Energy deposited in cells of the electromagnetic (hadronic) calorimeters is visualised as green (yellow) towers. Reconstructed charged-particle tracks are depicted as yellow lines. The event has $m_{\tau\tau}^{\text{MMC}} = 120\,\text{GeV}$, $m_{bb} = 120\,\text{GeV}$, and $m_{HH} = 680\,\text{GeV}$. The image is taken from Ref. [192].

A.2 Multivariate Analysis

A.2.1 Description of Discriminating Variables used in the $\tau_{\text{lep}}\tau_{\text{had}}$ Channels

A description of the discriminating variables used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel is given (see also Section 6.5.1 and Table 6.14). Reconstructed electrons and muons are collectively referred to as leptons (ℓ).

$\boldsymbol{\Delta p_{\text{T}}(\ell, \tau_{\text{had-vis}})}$ The transverse momentum difference between lepton and $\tau_{\text{had-vis}}$.

$\boldsymbol{m_{\text{T}}^{W}}$ The transverse mass of the lepton and $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ defined as

$$m_{\text{T}}^{W} = \sqrt{2|\boldsymbol{p}_{\text{T}}^{\ell}||\boldsymbol{p}_{\text{T}}^{\text{miss}}|\cos(1-\Delta\phi)}\,,$$

where $\Delta\phi$ is the angle between $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ and $\boldsymbol{p}_{\text{T}}^{\ell}$.

$\boldsymbol{p_{\text{T}}^{\text{miss}}\ \phi}$ **centrality** A measure of the relative angular position of $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ and visible τ-lepton decay products (electrons, muons, or $\tau_{\text{had-vis}}$) in the transverse plane. It is defined as

$$\boldsymbol{p}_{\text{T}}^{\text{miss}}\ \phi \text{ centrality} = \frac{A+B}{\sqrt{A^2+B^2}},$$

where

$$A = \frac{\sin(\phi_{\boldsymbol{p}_{\text{T}}^{\text{miss}}} - \phi_{\tau_1})}{\sin(\phi_{\tau_2} - \phi_{\tau_1})} \qquad B = \frac{\sin(\phi_{\tau_2} - \phi_{\boldsymbol{p}_{\text{T}}^{\text{miss}}})}{\sin(\phi_{\tau_2} - \phi_{\tau_1})},$$

with $\phi_{\boldsymbol{p}_{\text{T}}^{\text{miss}}}$ and $\phi_{\tau_1}/\phi_{\tau_2}$ denoting the azimuthal angle of $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ and visible τ-lepton decay products, respectively [184, 303].

The $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ ϕ centrality is defined relative to the line bisecting the azimuthal angle spanned by the visible τ-lepton decay products. It reaches a maximum of $\sqrt{2}$ (minimum of $-\sqrt{2}$) when $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ is aligned with the bisecting line and pointing into the smaller (larger) angle defined by the decay products. In configurations where $\boldsymbol{p}_{\text{T}}^{\text{miss}}$ is collinear with one of the visible τ-lepton decay products, it takes a value of 1.

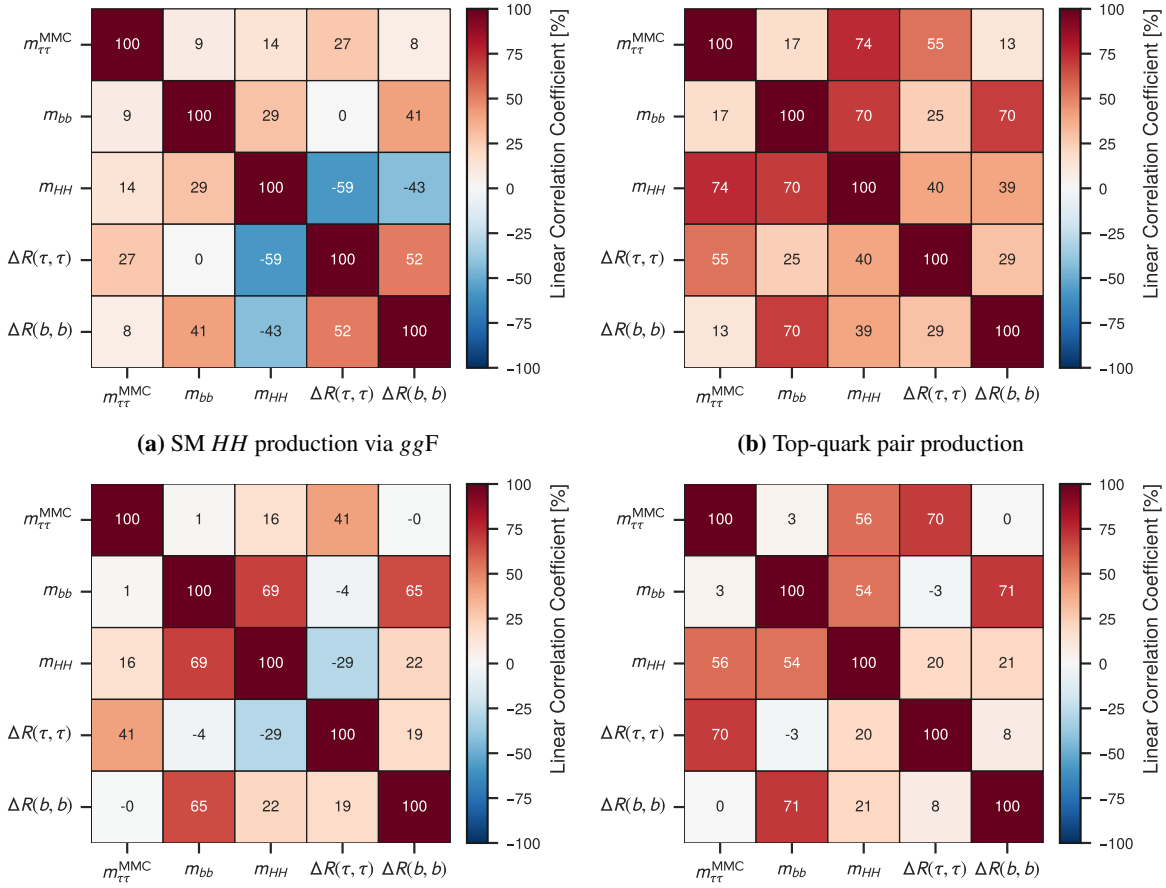
$\boldsymbol{\Delta\phi(\ell\tau_{\text{had-vis}}, bb)}$ Azimuthal angle between the $\ell + \tau_{\text{had-vis}}$ system and the system consisting of the two b-jet candidates.

$\boldsymbol{\Delta\phi(\ell, p_{\text{T}}^{\text{miss}})}$ Azimuthal angle between the lepton and $\boldsymbol{p}_{\text{T}}^{\text{miss}}$.

$\boldsymbol{\Delta\phi(p_{\tau\tau}^{\text{MMC}}, p_{\text{T}}^{\text{miss}})}$ Azimuthal angle between $\tau\tau$-system reconstructed using the MMC and $\boldsymbol{p}_{\text{T}}^{\text{miss}}$.

$\boldsymbol{s_{\text{T}}}$ The scalar sum of transverse momenta of all selected central jets, $\tau_{\text{had-vis}}$, leptons, and $p_{\text{T}}^{\text{miss}}$.

A.2.2 Correlation Matrices of the MVA Input Variables in the $\tau_{\text{had}}\tau_{\text{had}}$ Channel



(a) SM HH production via ggF

(b) Top-quark pair production

(c) $Z \to \tau^+\tau^-$ + jets

(d) Multi-jet

Figure A.3: Correlation coefficients between the MVA input variables used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The correlation matrices are shown separately for the SM HH signal (a) and the three largest backgrounds in the $\tau_{\text{had}}\tau_{\text{had}}$ SR (b-d).

A.2.3 Discrimination Power of the PNN as a Function of m_X in the $\tau_{had}\tau_{had}$ Channel

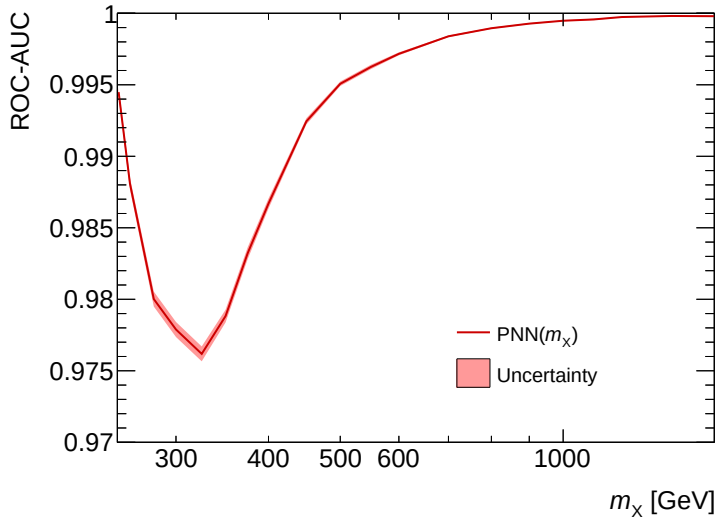


Figure A.4: Area under the receiver operating characteristic curve (ROC-AUC) for the PNN discriminant in the $\tau_{had}\tau_{had}$ channel. The ROC-AUC is calculated for the binary classification task of distinguishing signal events (i.e. resonant HH production via resonances of mass m_X) from background events in the $\tau_{had}\tau_{had}$ SR. Only statistical uncertainties from the finite number of simulated and CR events is shown.

A.3 Systematic Uncertainties

The recommendations of the ATLAS collaboration for the derivation of uncertainties on the modelling of selected physics processes using simulation are summarised. Appendix A.3.1 describes the prescriptions adopted to derive uncertainties for $t\bar{t}$ backgrounds. Appendix A.3.2 summarises the prescriptions for Z + jets backgrounds.

A.3.1 Uncertainties on the Modelling of $t\bar{t}$ Production using Simulation

The nominal event simulation for the $t\bar{t}$ background uses Powheg Box v2 [202] as a matrix element (ME) generator interfaced to Pythia 8.230 [166] for the parton shower (PS) and hadronisation. Uncertainties on the modelling of these processes are derived by varying the simulation setup. The following uncertainties are considered:

Hard scatter simulation and NLO+PS matching An uncertainty due to the choice of ME generator is estimated by comparing the nominal setup with an alternative in which Powheg Box v2 is replaced by MadGraph5_aMC@NLO. This comparison also probes the effect of different schemes for matching NLO ME generators to the PS simulation.[1]

Parton shower and hadronisation model An uncertainty on the modelling of the PS and non-perturbative effects is estimated by replacing Pythia 8 by Herwig 7 for the PS simulation.

Missing higher order contributions The renormalisation scale (μ_r) and factorisation scale (μ_f) is doubled (halved) to probe the effect of truncating the perturbative expansion in α_s when simulating the hard scatter process. Perturbative QCD calculations to sufficiently high order should be approximately independent of the choice of scale.

Initial and final state radiation (ISR / FSR) An uncertainty on the emission of ISR is estimated by varying α_s^{ISR} in the A14 set of tuned parameters for Pythia 8 [167]. An estimate of the uncertainty from the modelling of FSR emissions is estimated by doubling (halving) the renormalisation scale used for FSR branchings in Pythia 8 [166, 305, 306].

Damping factor for additional emissions The damping parameter h_{damp} in Powheg Box v2 is increased to $3m_{\mathrm{top}}$ (from $1.5m_{\mathrm{top}}$) and compared with the nominal simulation setup. The h_{damp} parameter controls the transverse momentum of additional radiation when matching Powheg Box v2 to Pythia 8 using the Powheg-method [307, 308].

These prescriptions are a revised version of the methodology outlined in Ref. [308]. Uncertainties on the NNPDF3.0nlo set of PDFs and the value of α_s were found to be negligible in the search for Higgs boson pair production presented in this thesis.

Variations of the renormalisation and factorisation scales are provided by internal re-weighting in Powheg Box v2. Similarly, Pythia 8 provides variations of initial and final state emissions by varying the renormalisation scales in the PS by re-weighting [305, 306]. This approach allows estimating uncertainties without changing the particle-level predictions of the simulation program thus avoiding the need to re-run the detector simulation.

[1] Powheg Box v2 uses the Powheg method [194–196] and MadGraph5_aMC@NLO the MC@NLO method [304] for NLO+PS matching.

A.3.2 Uncertainties on the Modelling of Z + jets Production in Simulation

The nominal Z + jets event simulation uses Sherpa 2.2.1 for the simulation of the hard scatter event and parton showering. The following uncertainties are considered:

Factorisation and renormalisation scales Six variations of the factorisation and renormalisation scales are performed using internal re-weighting implemented in Sherpa 2.2.1 [216], altering the scales by factors of $\frac{1}{2}$ and 2. The following variations are performed:

$$\left(\frac{\mu_\mathrm{f}}{\mu_\mathrm{f}^\mathrm{nom.}}, \frac{\mu_\mathrm{r}}{\mu_\mathrm{r}^\mathrm{nom.}}\right) \in \left\{(\tfrac{1}{2}, \tfrac{1}{2}), (\tfrac{1}{2}, 1), (1, \tfrac{1}{2}), (1, 2), (2, 1), (2, 2)\right\} ,$$

where $\mu_\mathrm{f}^\mathrm{nom.}$ and $\mu_\mathrm{r}^\mathrm{nom.}$ are the nominal scale values.

Resummation scale The scale of the resummation of soft gluon emissions in the Sherpa parton shower is varied by factors of $\frac{1}{2}$ and 2. Variations of the resummation scale are provided in parameterised form with respect to the default Sherpa configuration in Ref. [309].

Multi-jet merging scale The simulation of Z + jets events with Sherpa 2.2.1 uses matrix elements of NLO accuracy for up to two and LO for up to four partons. These multi-parton matrix elements are merged with the parton shower using an extension of the CKKW algorithm [310–312]. The characteristic scale Q_cut of the multi-jet merging algorithm is varied from its nominal value of $Q_\mathrm{cut} = 20\,\mathrm{GeV}$ to 15 GeV and 30 GeV [309]. These variations are provided, following the approach for the resummation scale, in parameterised form in Ref. [309].

PDF+α_s and PDF choice Uncertainties on the NNPDF3.0nnlo set of PDFs [199] are evaluated using 100 replica sets provided through the LHAPDF6 library [313] and implemented using internal re-weighting in Sherpa. The uncertainty on α_s is estimated by comparing NNPDF3.0nnlo PDF sets with $\alpha_\mathrm{s}(m_Z^2) = 0.117$ and 0.119 with the nominal set using a value of 0.118. Finally, an uncertainty on the choice of PDF set is estimated by comparing with two alternative PDF sets MMHT2014nnlo68cl [314] and CT14nnlo [315].

Alternative generator and parton shower The prediction of Z + jets with the default configuration of Sherpa 2.2.1 is compared to an alternative setup using MadGraph5_aMC@NLO 2.2.2 [198] for the calculation of the hard interaction at LO interfaced to Pythia 8.186 [316] for parton showering.

A.4 Statistical Interpretation

A.4.1 Statistical Uncertainties on Background Predictions

The expected number of background events in bin b of a channel c, ν_{cb}, is often estimated using a finite sample of events (e.g. from MC simulation); therefore, ν_{cb} does not correspond to the true expected number of background events. The background predictions are subject to statistical uncertainties that have to be considered when performing inference, particularly when bins are only sparsely populated by events.

This uncertainty is included in the likelihood function, employing the method proposed by Barlow and Beeston [141], by replacing the expected number of background events estimated using the finite sample with its true value, which is unknown and has to be inferred from data. In practice, this is done by performing the substitution $\nu_{cb} \to \gamma_{cb}\nu_{cb}$, introducing new NPs γ_{cb}. These NPs are constrained by auxiliary measurements that contribute terms of the form

$$\text{Pois}(m_{cb}; \gamma_{cb}\tau_{cb}) \qquad \text{with} \qquad \tau_{cb} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} = \text{const.}$$

to the likelihood function [140], where the sums go over all events in bin b of channel c with event weights w_i. This corresponds to a measurement of the effective number of events ($\gamma_{cb}\tau_{cb}$) based on the observed value m_{cb}, which is nominally equal to τ_{cb},[2] for the finite sample of events.

This approach is based on the approximation of the *compound Poisson distribution* (CPD), which describes the distribution of the sum of a Poisson number of random weights, by a *scaled Poisson distribution* (SPD) [317]. Formally, the CPD can be defined as

$$X = \sum_{i=1}^{N} W_i \,,$$

with $N \sim \text{Pois}(\lambda)$ and i.i.d. W_i that are independent of N. The CPD can be approximated using the SPD defined by

$$\tilde{X} = s \cdot \tilde{N} \quad \text{with} \quad \tilde{N} \sim \text{Pois}(\tilde{\lambda})$$

and

$$s = \frac{\text{E}(W^2)}{\text{E}(W)} \qquad \tilde{\lambda} = \frac{\lambda\,\text{E}(W)^2}{\text{E}(W^2)}$$

where $\text{E}(W)$ and $\text{E}(W^2)$ are the first and second moments of the weight distribution, respectively [317]. The Barlow–Beeston method makes the assumption that the expectation values can be approximated

[2] Generally, m_{cb} is not integer-valued and thus not covered by the support of the Poisson distribution; therefore, the factorial term in the Poisson PMF is replaced by the gamma function to generalise the distribution to $\mathbb{R}^+$.

by sample averages such that

$$s = \frac{\sum_i w_i^2}{\sum_i w_i} \qquad \tilde{\lambda} = \frac{\lambda}{n} \frac{(\sum_i w_i)^2}{\sum_i w_i^2}$$

with sample size n. When defining m_{cb} to be the observed value of $\tilde{N}$ and $\gamma_{cb} = \lambda/n$, then the $\text{Pois}(m_{cb}; \gamma_{cb}\tau_{cb})$ terms in the likelihood function are reproduced.

A.4.2 Generation of Toys for the Global Significance Estimation

The toy generation methods used for the global significance estimation in Section 6.7.4 is described in the following.

Generation of Observables

The correlations between observables (i.e. n_{cb} in Equation (4.1)) have to be quantified before proceeding with the generation of toy experiments. Let A and B be the number of events selected by two bins. Additionally, the marginal distributions of A and B are Poisson. The correlation between A and B is given by the overlap in the kinematic region selected by both bins. The kinematic region selected by either bin can be partitioned into three parts:

- The region selected by bin A but not B with number of events X_1.
- The region selected by bin A and B with number of events X_2.
- The region selected by bin B but not A with number of events X_3.

The X_i are distributed according to $\text{Pois}(\lambda_i)$ for $i = 1, 2, 3$ and are mutually independent. Consequently, A and B can be written as

$$A = X_1 + X_2 \sim \text{Pois}(\lambda_1 + \lambda_2)$$
$$B = X_2 + X_3 \sim \text{Pois}(\lambda_2 + \lambda_3)$$

and the linear correlation coefficient between A and B is given by[3]

$$\rho_{AB} = \frac{\text{Cov}(A, B)}{\sqrt{\text{Var}(A)\,\text{Var}(B)}} = \frac{\lambda_2}{\sqrt{(\lambda_1 + \lambda_2) \cdot (\lambda_2 + \lambda_3)}}\,.$$

This equation reflects the intuition that the overlap between bins, described by λ_2, introduces a correlation between both observables. In the following, the model described by A and B is referred to as the bivariate Poisson model [318].

An approximation of the correlation coefficients for all bin pairs in a given channel can be obtained by estimating the parameters λ_i using MC simulation and CR data. Excerpts of the correlation matrix obtained in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are shown in Figure A.5. In total, 144 bins are considered in the $\tau_{\text{had}}\tau_{\text{had}}$ channel; therefore, the full correlation matrix is of dimension 144×144. Similar matrices are obtained in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT SRs with dimension 193×193 and 182×182, respectively.

A few features of the correlation matrix shown in Figure A.5 are noted in the following:

- Off-diagonal elements of submatrices describing the correlation between observables belonging to the same PNN discriminant are zero due to bins being pairwise disjoint by construction.
- The most background-like bins have correlation coefficients of about 90 % or larger. This is due to events that are easily rejected by the PNN discriminants, which predominately populate the first bins.

[3] $\text{Cov}(A, B) = \text{Cov}(X_1 + X_2, X_2 + X_3) = \text{Cov}(X_2, X_2) = \text{Var}(X_2) = \lambda_2$ from the independence of X_1, X_2, and X_3.

- The observables of the most signal-like bins in the intermediate mass range (cf. Figure A.5(a)) have little correlation. In contrast, in the high mass regime (cf. Figure A.5(b)) the observables of the most signal-like bins can have correlations of up to 75 %. This is due to the inability of the PNN to distinguish between different signal hypotheses with large m_X.

The correlation matrices in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channels show similar features.

The toy generation for the global significance estimation needs to fulfil certain requirements. First, the model needs to reproduce the marginal distributions of the observables, i.e. Poisson distributions with mean parameters equal to the number of events predicted by the nominal background model. Second, it needs to model the dependencies between observables due to the overlap in the kinematic regions selected by bins.

A multivariate extension of the bivariate Poisson distribution would fulfil these requirements; however, such a model becomes intractable due to the large number of observables considered and the presence of negatively weighted events in the background estimate. Instead, a model based on the mathematical framework provided by Sklar's theorem [276] is adopted, which can be used to factorise the marginal distributions from the dependency structure of the observables. A review of similar approaches of modelling multivariate count data is given in Ref. [319].

Sklar's theorem states, see for example Ref. [277], that any n-dimensional joint distribution function H can be decomposed into n marginal distribution functions $F_1, \ldots, F_n$ and an n-dimensional copula C. The copula C is an n-dimensional distribution function, $C : [0,1]^n \to [0,1]$, with uniform marginal densities that describes the dependencies between random variables. The joint distribution can be expressed as

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

according to the theorem.

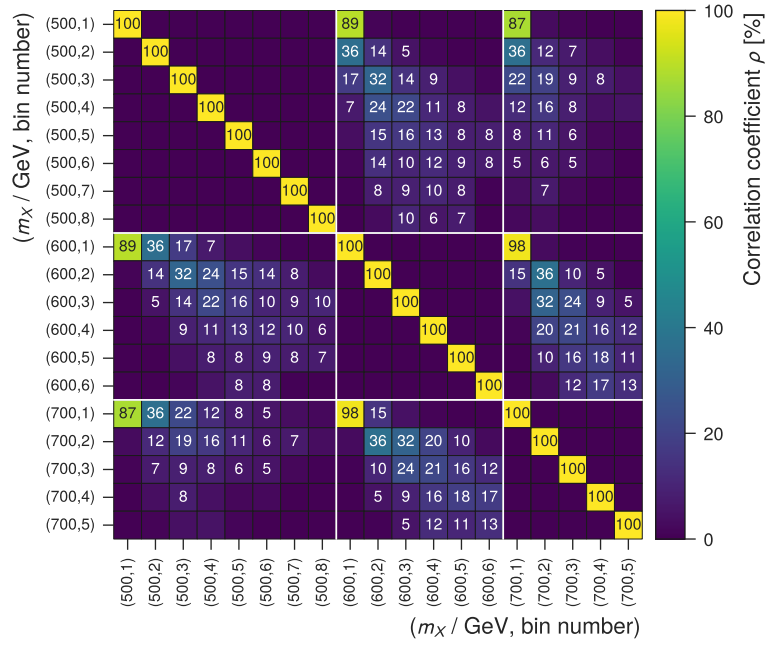
For the task of generating toy experiments, the only missing piece is the functional form of the copula since the marginal distributions are known from the nominal background model. The bivariate Poisson distribution can be approximated by a bivariate Normal distribution provided the means, λ_i, are not too small. This suggests that the copula of a multivariate Normal distribution might be a suitable approximation to model the dependencies between observables. This copula, referred to as the Gaussian copula, can be derived using the "inversion method" described in Ref. [277] and is given by

$$C_{\boldsymbol{R}}(u_1, \ldots, u_n) = \Phi_{\boldsymbol{R}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)) \,, \tag{A.1}$$

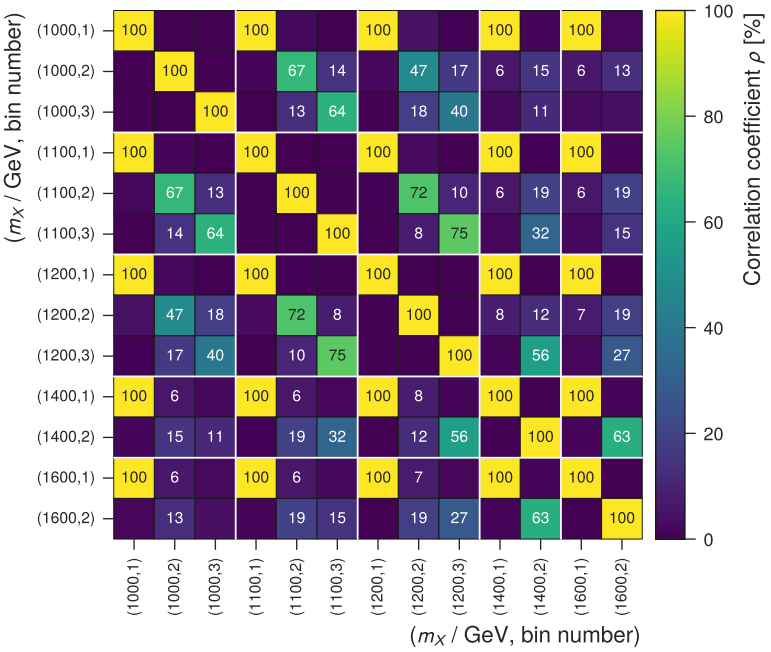
where $\Phi_{\boldsymbol{R}}$ is the CDF of the multivariate Normal distribution with zero mean and covariance equal to the correlation matrix $\boldsymbol{R}$ and Φ^{-1} is the quantile function of the univariate Standard Normal distribution.

The parameters $\boldsymbol{R}$ of the Gaussian copula are estimated by the correlation matrices derived with the bivariate Poisson model. This approximation is investigated empirically in two dimensions by comparing simulated results of the bivariate Poisson model and a copula-based model. The dependence of two observables A and B is illustrated in Figure A.6 for both models in terms of the conditional mean and variance of B given $A = a$. This comparison is performed, instead of a direct comparison of joint distribution functions, to better illustrate the subtle differences between both models. The comparison yields the following findings:

- The conditional mean of B given $A = a$ agrees well between models, even when considering



(a) Submatrix for observables of the $m_X = 500, 600, 700$ GeV PNN discriminants



(b) Submatrix for observables of the $m_X = 1000, 1100, 1200, 1400, 1600$ GeV PNN discriminants

Figure A.5: Estimate of the correlation matrix between observables in $\tau_{\text{had}}\tau_{\text{had}}$ channel. Two submatrices of the full matrix (144×144) are shown. Bins are enumerated in increasing order of the signal probability according to the PNN discriminant (i.e. bins numbered with 1 contain the most background-like events). Cells are annotated if the correlation coefficient is larger than or equal to 5 %. Off-diagonal elements with a correlation coefficient of 100 % are due to rounding.

bin pairs with the lowest λ_i relevant for the analysis.

- The conditional variance of B given $A = a$ illustrates that both models are not equivalent. For bin pairs with small λ_i, the variance of the distribution of B for fixed $A = a$ differs by up 15 % to between both models.
- In the case of disjoint bins (not shown in the figure), the bivariate Poisson model and copula-based model are identical.

With these simplifying assumptions, observables can be randomly generated using the copula-based model. The following algorithm is employed to draw a vector of random variates $\boldsymbol{x} = (x_1, \ldots, x_n)$ from the joint distribution described by the Gaussian copula and marginal distributions [277]:

1. Draw a vector of random variates $\boldsymbol{u} = (u_1, \ldots, u_n)$ from the distribution described by $C_{\boldsymbol{R}}$.

   The vector $\boldsymbol{u}$ is obtained by generating random variates $\boldsymbol{n} = (n_1, \ldots, n_n)$ from the multivariate Normal distribution with zero mean and covariance matrix equal to $\boldsymbol{R}$, followed by an element-wise application of the univariate Standard Normal CDF to $\boldsymbol{n}$, that is, $u_i = \Phi(n_i)$ for $i = 1, \ldots, n$. This procedure follows from the form of the Gaussian copula in Equation (A.1).

   In practice, the matrices $\boldsymbol{R}$ that are considered here are singular due to multicollinearity between observables. This arises from 19 implicit constraints due to the requirement that the sum of observables is the same for all 20 discriminants in a given analysis channel, thus leading to 19 vanishing eigenvalues of $\boldsymbol{R}$. Therefore, multivariate Normal random variates are generated in a lower dimensional space that removes the collinearity, followed by a back-transformation to the n-dimensional space. The required transformations are provided by an eigendecomposition of $\boldsymbol{R}$.

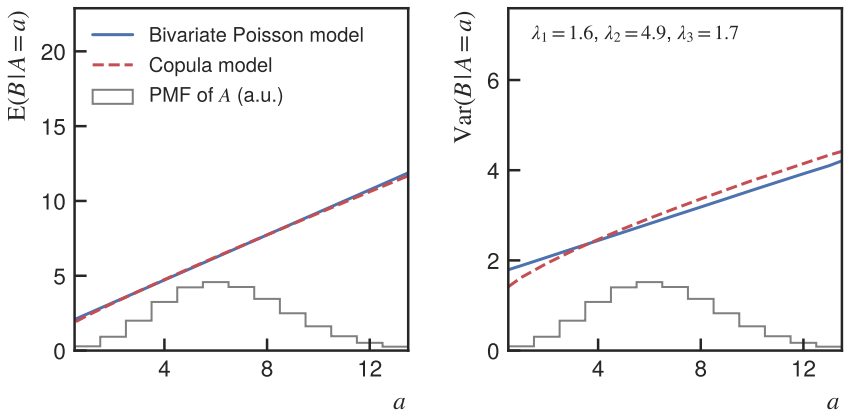
2. Obtain $\boldsymbol{x}$ by evaluating the quantile functions of the n marginal distributions at the values of $\boldsymbol{u}$ from the previous step such that $x_i = F_i^{-1}(u_i)$ for $i = 1, \ldots, n$. The marginal distribution functions are given by Poisson distributions with mean parameters corresponding to the nominal background prediction in a given bin.

The multivariate count data generated with this method was inspected by generating large samples of size $O(10^6)$ showing agreement in terms of the expected marginal distributions of the observables and closely reproducing the pair-wise correlation coefficients estimated with the bivariate Poisson model.

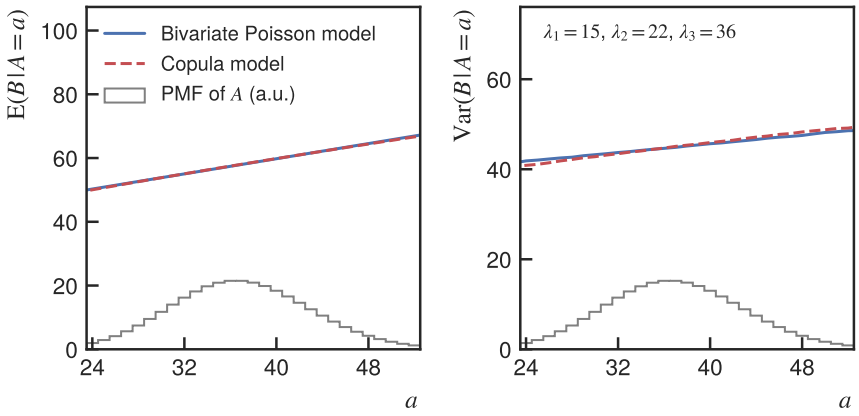
Generation of Global Observables for the Barlow–Beeston Method

This section describes the generation of global observables related to the Barlow–Beeston method (i.e. m_{cb} in Equation (4.2)). The sample of simulated and CR events used for background estimation is resampled using a non-parametric bootstrap method [278, 279] to obtain new samples that capture the statistical uncertainties and correlations of the background predictions. Specifically, random weights $w_i^{\text{bootstrap}}$ drawn from Pois(1) are assigned to all events [320]. These weights define new background estimates $\sum_i w_i^{\text{bootstrap}} w_i$ for all bins, where w_i is the nominal weight of event i and the sum goes over all events in a given bin. The resulting background estimate is translated, separately for every bin, into an effective number of events according to

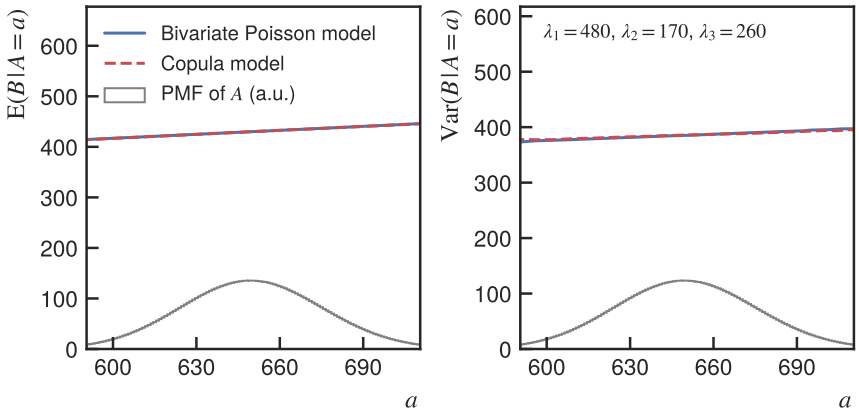
$$m_{cb} = \frac{1}{s_{cb}} \sum_i w_i^{\text{bootstrap}} w_i \,,$$



(a) Most signal-like bins of the $m_X = 1\,100\,\text{GeV}$ and $m_X = 1\,200\,\text{GeV}$ PNN discriminants ($\rho_{AB} = 75\,\%$).



(b) Second bin of the $m_X = 1\,000\,\text{GeV}$ and second bin of the $m_X = 1\,200\,\text{GeV}$ PNN discriminant ($\rho_{AB} = 47\,\%$).

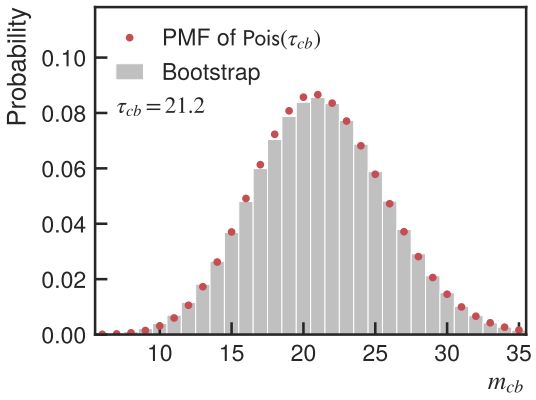


(c) Third bin of the $m_X = 500\,\text{GeV}$ and second bin of the $m_X = 600\,\text{GeV}$ PNN discriminant ($\rho_{AB} = 32\,\%$).

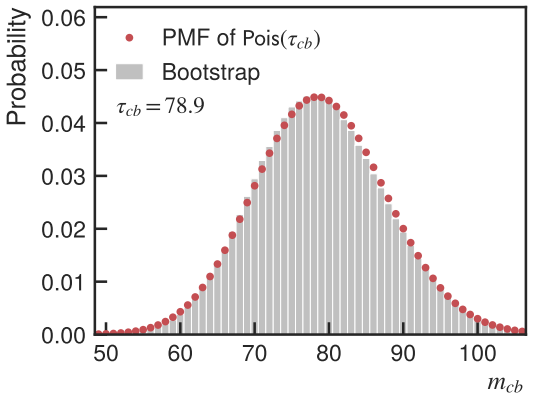
Figure A.6: Comparison of the bivariate Poisson model and a model using the Gaussian copula defined by $\rho_{AB} = \lambda_2/\sqrt{(\lambda_1+\lambda_2)\cdot(\lambda_2+\lambda_3)}$ and marginal distributions given by $A \sim \text{Pois}(\lambda_1+\lambda_2)$ and $B \sim \text{Pois}(\lambda_2+\lambda_3)$. The conditional mean (left) and variance (right) of B given $A = a$ is shown to illustrate the dependence of both random variables. Three different scenarios with varying λ_i and ρ_{AB} are shown, which are chosen from bin pairs in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The probability mass function (PMF) of A is overlayed in arbitrary units (a.u.).

with a scaling factor $s_{cb} = \sum_i w_i^2 / \sum_i w_i$ (cf. Appendix A.4.1) and the sums going over all events in bin b of channel c. Multiple toy experiments are generated by repeating the bootstrap procedure.

In Figure A.7, exemplary distributions of m_{cb} in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are shown. The figure compares the distribution obtained from the bootstrap procedure with the distribution predicted by the Barlow–Beeston method, i.e. the distribution of $m_{cb} \sim \text{Pois}(\tau_{cb})$. In general, both distributions are not expected to agree perfectly since the Barlow–Beeston method is itself an approximation. Nevertheless, both approaches show decent agreement for scenarios relevant to this analysis. Even in the case depicted in Figure A.7(a), which corresponds to the bin with the smallest effective number of events in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the agreement is acceptable.



(a) Most signal-like bin of the $m_X = 251\,\text{GeV}$ PNN discriminant in the $\tau_{\text{had}}\tau_{\text{had}}$-channel.



(b) Most signal-like bin of the $m_X = 1\,000\,\text{GeV}$ PNN discriminant in the $\tau_{\text{had}}\tau_{\text{had}}$-channel.

Figure A.7: Distributions of m_{cb} derived using a bootstrap procedure (grey histogram) and predicted by the Barlow–Beeston method (red points). Two scenarios with different τ_{cb} are depicted, which are taken from the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

A.5 Results

A.5.1 Tables of Upper Limits for the Search for Resonant HH Production

Table A.1: Expected and observed upper limits on the cross section of $pp \to X \to HH$ at 95 % CL for the combination of all channels.

	Upper limit on $\sigma(pp \to X \to HH)$ at 95 % CL / fb					
m_X/GeV	Observed	-2σ	-1σ	Expected	+1σ	+2σ
251	640	180	240	340	470	630
260	900	390	520	720	1 000	1 400
280	490	450	610	840	1 200	1 600
300	540	350	480	660	920	1 200
325	340	250	340	470	660	880
350	230	190	250	350	490	660
375	130	120	160	220	300	400
400	80	77	100	140	200	270
450	47	36	49	68	94	130
500	46	23	31	43	59	80
550	25	18	24	33	46	61
600	21	14	19	26	36	49
700	25	10	13	19	26	35
800	31	8.2	11	15	21	28
900	31	7.2	9.7	13	19	25
1 000	30	6.5	8.8	12	17	23
1 100	28	7.2	9.7	13	19	25
1 200	25	7.4	9.9	14	19	26
1 400	27	11	14	20	28	37
1 600	34	17	22	31	43	58

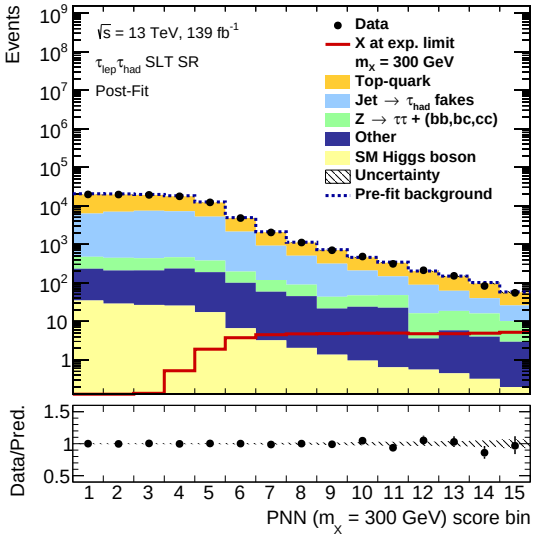
Table A.2: Expected and observed upper limits on the cross section of $pp \to X \to HH$ at 95 % CL for the combination of the $\tau_{\text{had}}\tau_{\text{had}}$ channel and the Z + HF CR.

	Upper limit on $\sigma(pp \to X \to HH)$ at 95 % CL / fb					
m_X/GeV	Observed	-2σ	-1σ	Expected	+1σ	+2σ
251	970	270	360	490	690	920
260	1 600	560	750	1 000	1 400	1 900
280	890	550	740	1 000	1 400	1 900
300	900	430	570	800	1 100	1 500
325	430	290	380	530	740	990
350	260	220	290	400	560	750
375	150	130	170	240	330	440
400	100	85	110	160	220	290
450	62	41	55	77	110	140
500	56	28	38	52	73	98
550	26	21	28	40	55	74
600	27	17	23	32	45	60
700	32	13	17	24	33	45
800	33	11	15	20	28	38
900	42	9.8	13	18	25	34
1 000	44	9.6	13	18	25	33
1 100	36	11	14	20	28	37
1 200	35	12	16	22	30	41
1 400	35	17	23	31	44	58
1 600	38	25	34	47	65	87

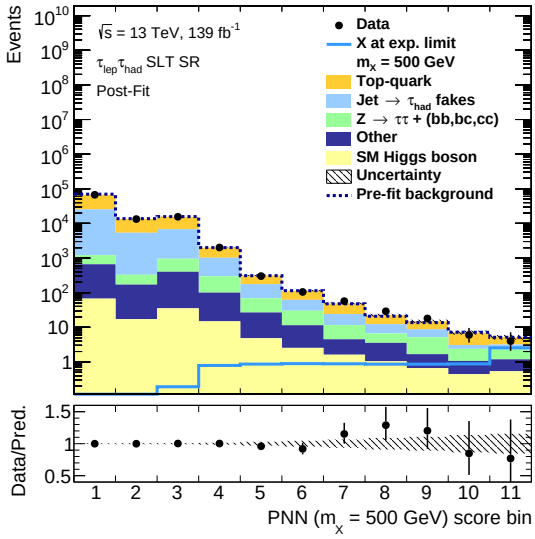
Table A.3: Expected and observed upper limits on the cross section of $pp \to X \to HH$ at 95 % CL for the combination of the $\tau_{\text{lep}}\tau_{\text{had}}$ channels and the Z + HF CR.

	Upper limit on $\sigma(pp \to X \to HH)$ at 95 % CL / fb					
m_X/GeV	Observed	-2σ	-1σ	Expected	+1σ	+2σ
251	590	270	360	500	700	940
260	820	570	760	1 100	1 500	2 000
280	650	730	980	1 400	1 900	2 500
300	700	660	880	1 200	1 700	2 300
325	710	550	740	1 000	1 400	1 900
350	620	410	550	770	1 100	1 400
375	420	330	440	610	860	1 100
400	230	210	280	390	540	720
450	110	85	110	160	220	300
500	86	46	61	85	120	160
550	78	37	49	68	95	130
600	42	28	37	51	71	96
700	36	19	25	34	48	64
800	43	14	19	27	37	50
900	29	12	16	23	31	42
1 000	28	10	14	19	26	35
1 100	33	10	14	19	27	36
1 200	27	10	14	19	27	36
1 400	39	15	20	28	39	53
1 600	61	25	33	46	64	86

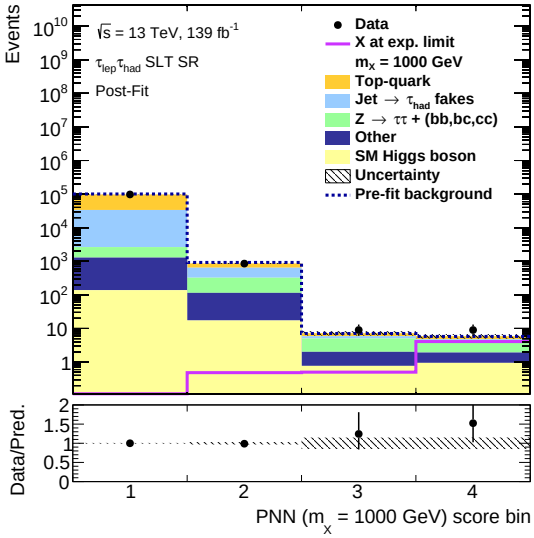
A.5.2 Post-Fit Distributions of PNN Discriminants in the $\tau_{\text{lep}}\tau_{\text{had}}$ Channels



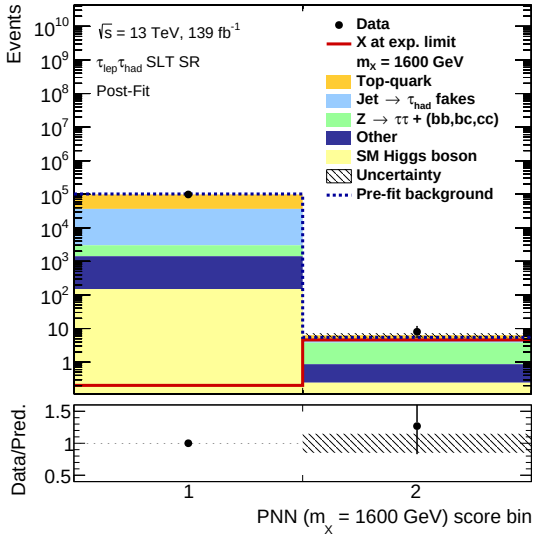
(a) $m_X = 300\,\text{GeV}$



(b) $m_X = 500\,\text{GeV}$

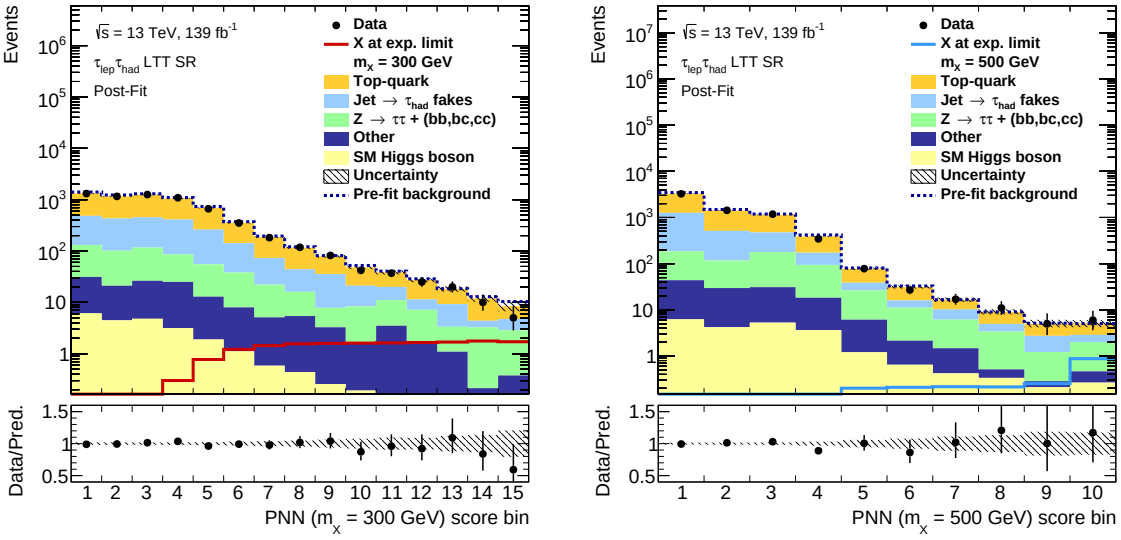


(c) $m_X = 1\,000\,\text{GeV}$



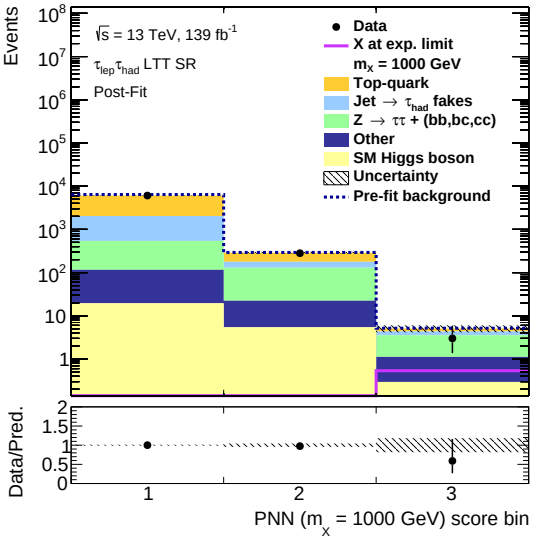
(d) $m_X = 1\,600\,\text{GeV}$

Figure A.8: Distribution of selected PNN discriminants in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel after the background-only fit to data in all channels. The signal overlay is scaled to the expected upper limit on $\sigma(pp \to X \to HH)$.

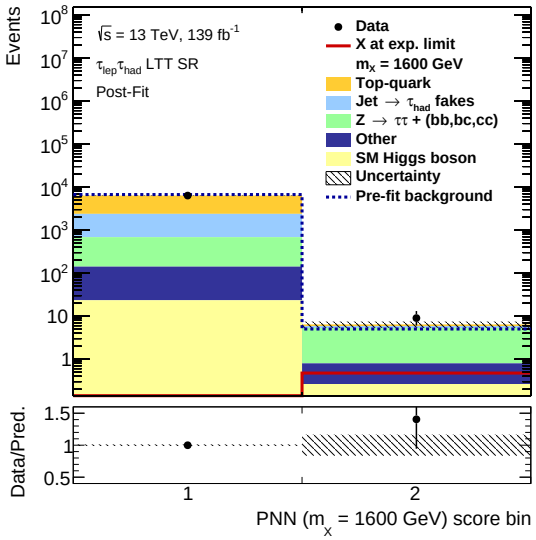


(a) $m_X = 300\,\text{GeV}$

(b) $m_X = 500\,\text{GeV}$



(c) $m_X = 1\,000\,\text{GeV}$



(d) $m_X = 1\,600\,\text{GeV}$

Figure A.9: Distribution of selected PNN discriminants in the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel after the background-only fit to data in all channels. The signal overlay is scaled to the expected upper limit on $\sigma(pp \to X \to HH)$.

A.5.3 Nuisance Parameter Rankings in the Search for Resonant HH Production

The rankings of NPs after the fit to data in all channels are shown in Figures A.10 to A.12 for the serach for resonant HH production with $m_X = 300\,\text{GeV}$, 500 GeV, and 1 000 GeV. The rankings are defined in analogy to Figure 6.32 except that the cross section $\sigma(pp \to X \to HH)$ is used as the POI instead of the signal strength.

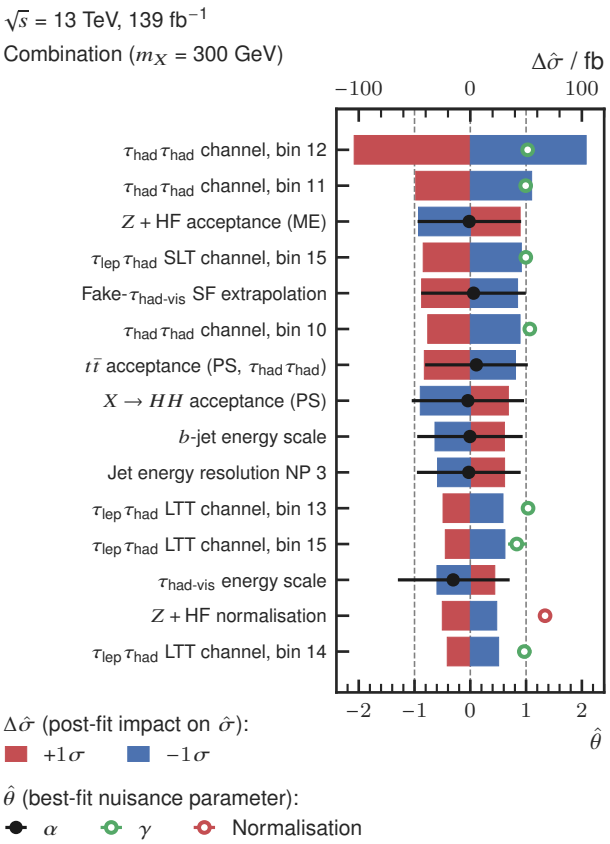


Figure A.10: Rankings of NPs in the search for resonant HH production with $m_X = 300\,\text{GeV}$. The rankings are shown after the fit to observed data in all analysis channels. The best-fit cross section is $\hat{\sigma}(pp \to X \to HH) = (-180\,^{+310}_{-320})$ fb.

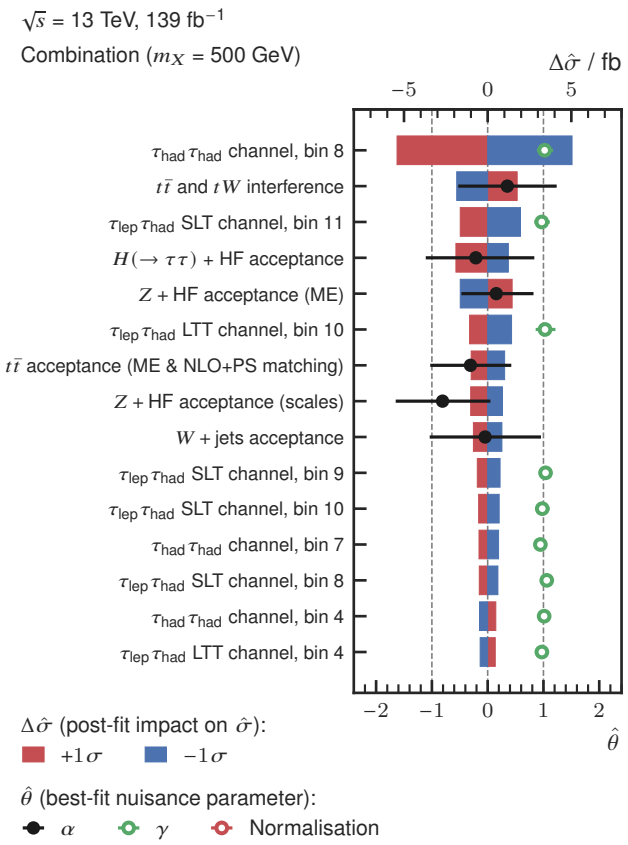


Figure A.11: Rankings of NPs in the search for resonant HH production with $m_X = 500\,\text{GeV}$. The rankings are shown after the fit to observed data in all analysis channels. The best-fit cross section is $\hat{\sigma}(pp \rightarrow X \rightarrow HH) = (6\,^{+20}_{-17})$ fb.

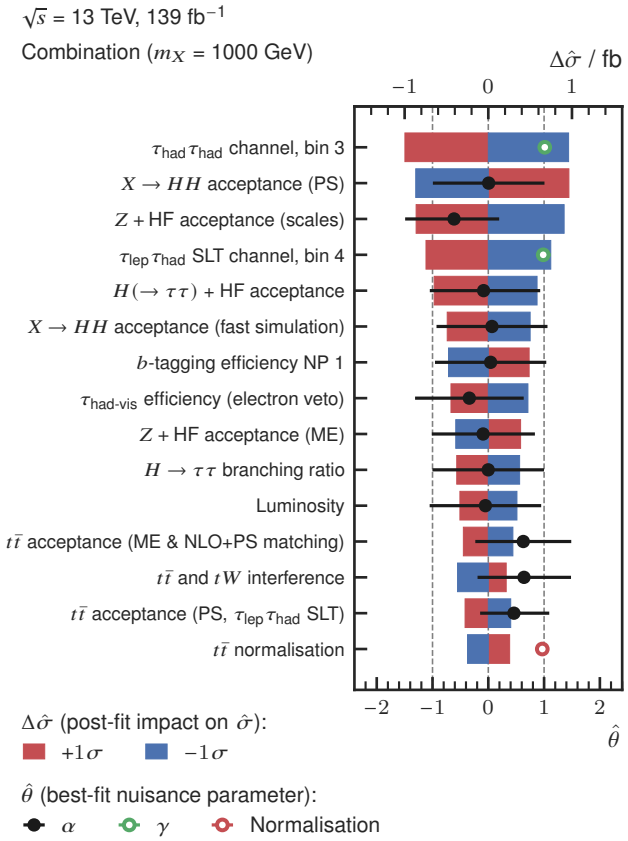


Figure A.12: Rankings of NPs in the search for resonant HH production with $m_X = 1\,000\,\text{GeV}$. The rankings are shown after the fit to observed data in all analysis channels. The best-fit cross section is $\hat{\sigma}(pp \to X \to HH) = (17.4\,^{+7.5}_{-6.5})$ fb.

A.6 Additional Signal, Control, and Validation Region Plots

Transverse Momenta of $\tau_{\text{had-vis}}$ and b-Jet Candidates in the Signal Region

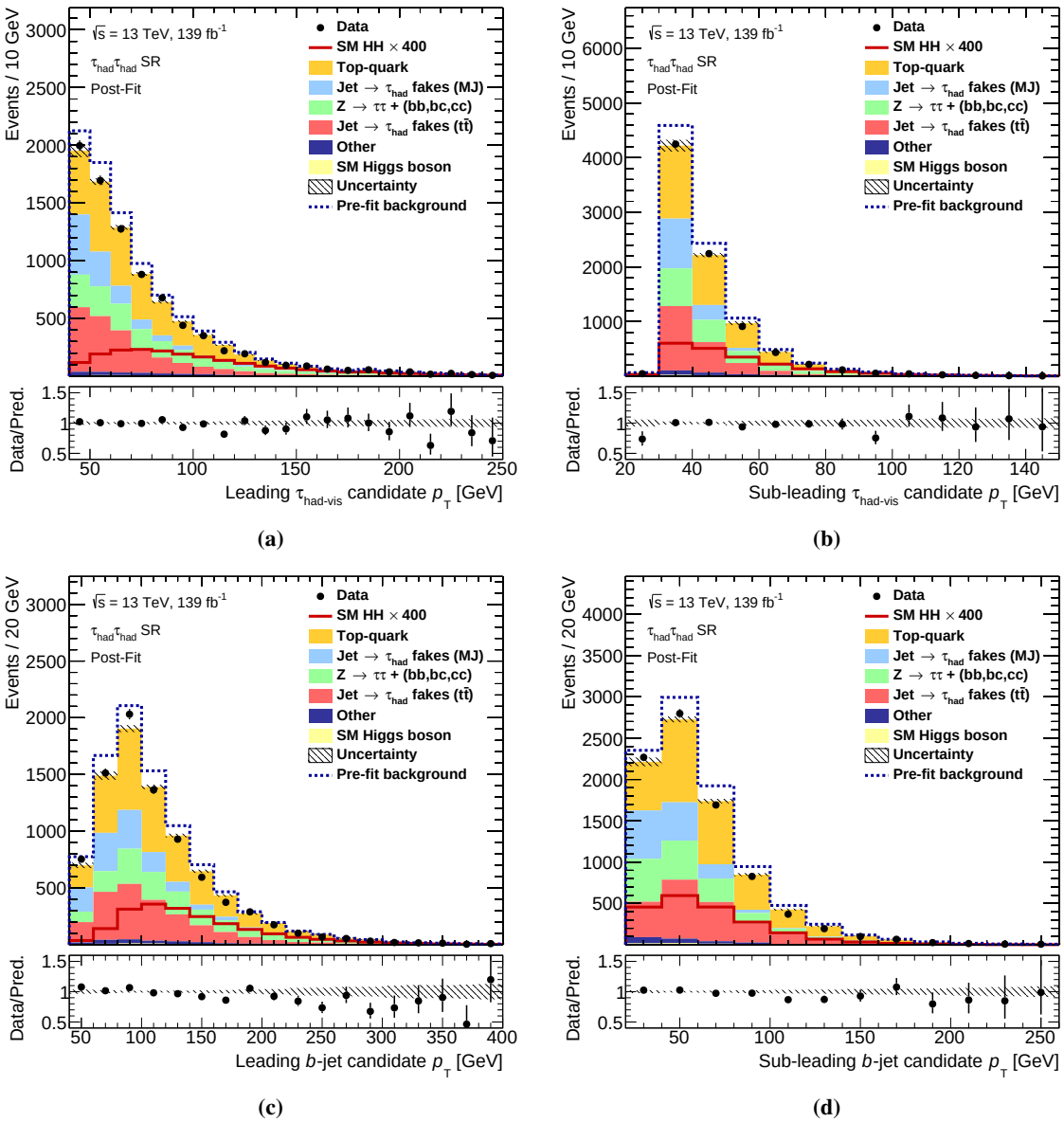


Figure A.13: Distributions of the leading and sub-leading $\tau_{\text{had-vis}}$ candidate p_{T} (a,b) and the leading and sub-leading b-jet candidate p_{T} (c,d) in the SR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel after the background-only fit to data in all regions.

MVA Score Distributions in the Z + Jets Validation Region

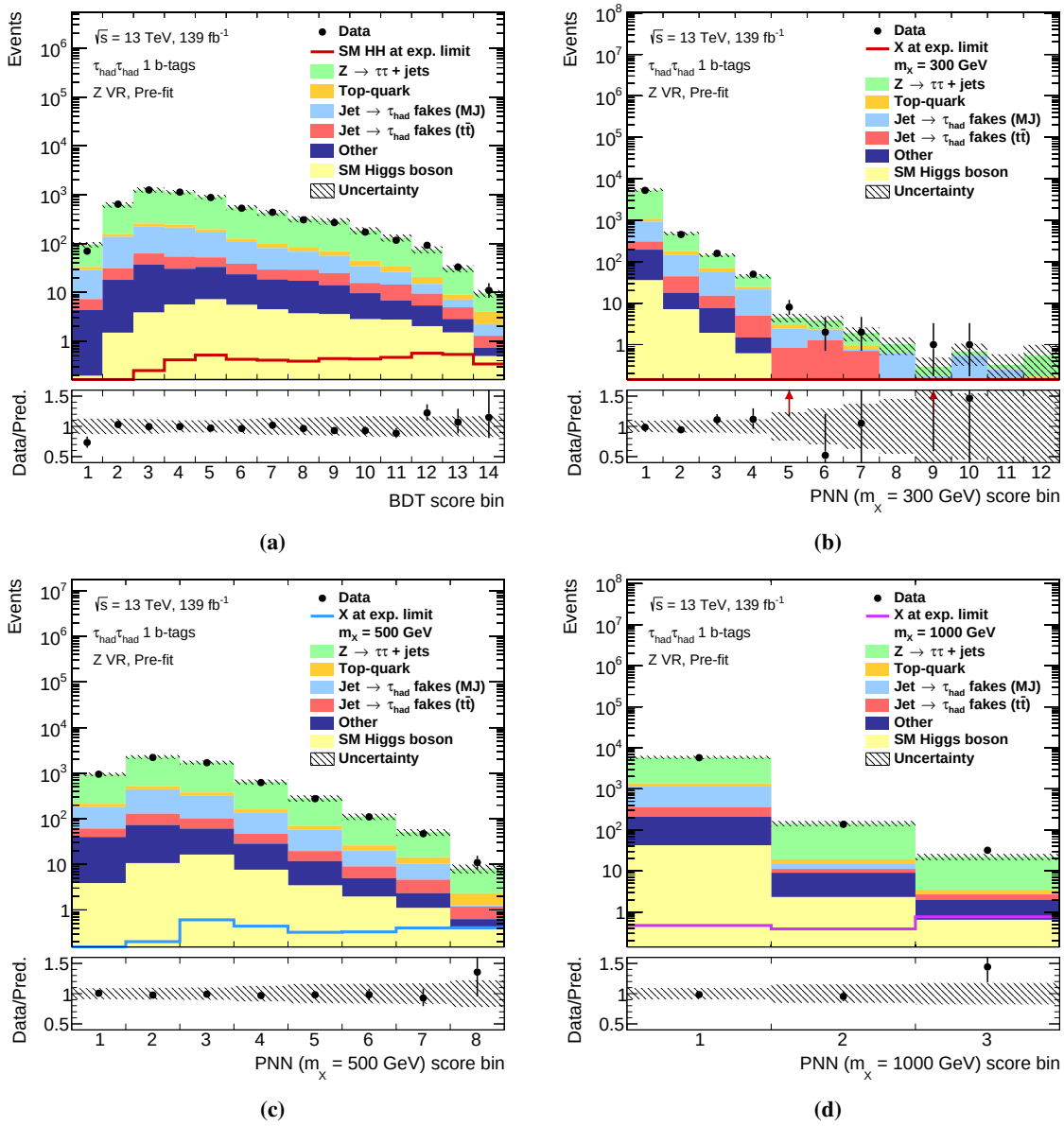


Figure A.15: BDT (a) and PNN (b-d) distributions in the Z + jets VR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel prior to the fit. The Z + jets VR is defined by requiring exactly 1 b-tagged jet, $70\,\text{GeV} < m_{\tau\tau}^{\text{MMC}} < 110\,\text{GeV}$, and $\Delta R(\tau,\tau) < 0.015\,\text{GeV}^{-1} \cdot m_{\tau\tau}^{\text{MMC}} + 0.06$. The signals are normalised to the expected upper limit for the combination of all channels.

MVA Scores in the SS Control Region

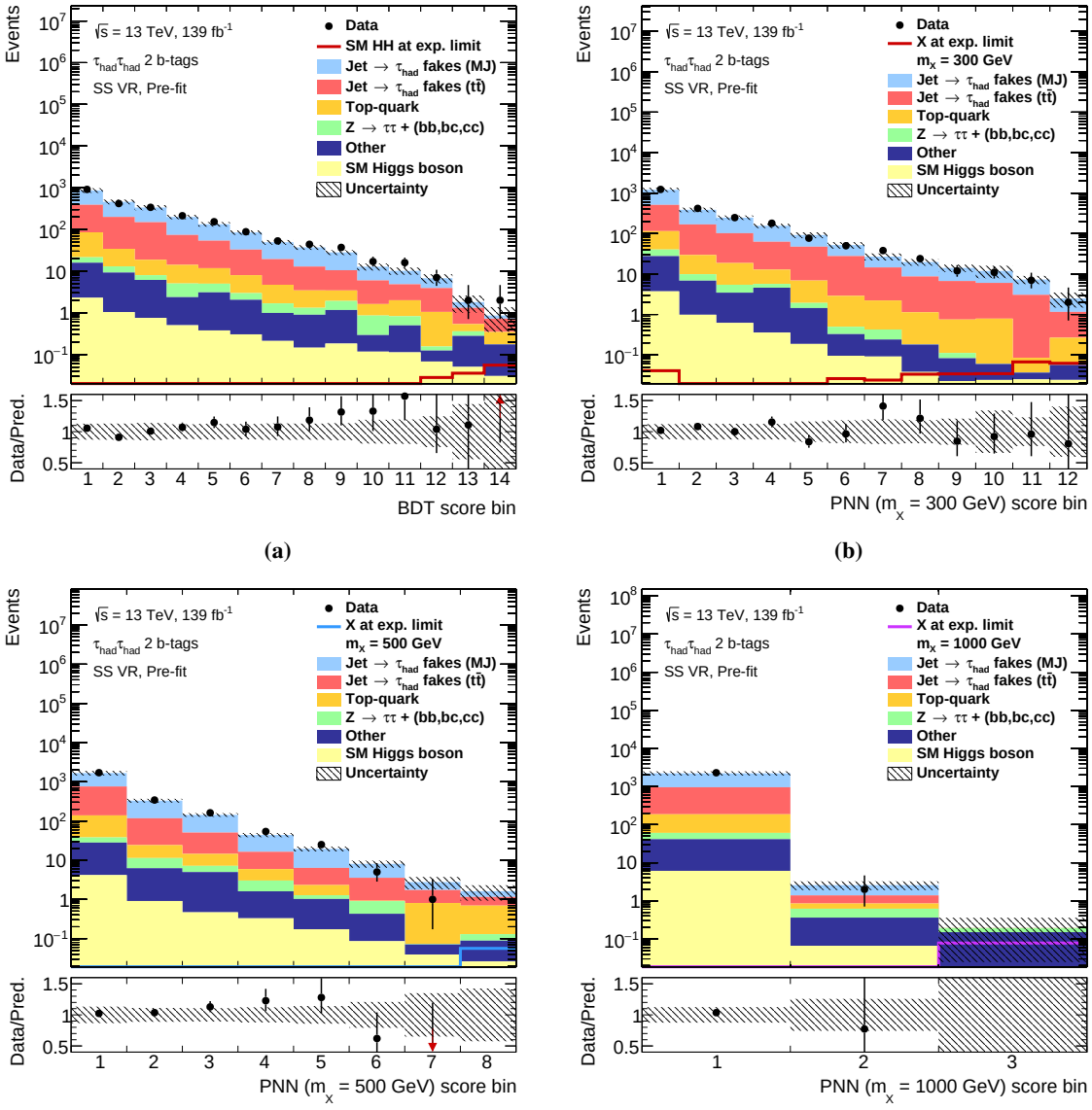


Figure A.17: BDT (a) and PNN (b-d) distributions in the SS CR of the $\tau_{\text{had}}\tau_{\text{had}}$ channel prior to the fit. The SS CR is defined by the SR event selection but requiring $\tau_{\text{had-vis}}$ candidates with electric charges of the same sign. The signals are normalised to the expected upper limit for the combination of all channels.

Bibliography

[1] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. B **47** (1973) 365.

[2] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343.

[3] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346.

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[5] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168.

[6] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[7] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex].

[8] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex].

[9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[11] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: https://cds.cern.ch/record/2688062.

[12] ATLAS Collaboration, *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, ATL-PHYS-PUB-2022-044, 2022, URL: https://cds.cern.ch/record/2827111.

[13] G. 't Hooft, *Renormalizable Lagrangians for Massive Yang-Mills Fields*, Nucl. Phys. B **35** (1971) 167.

[14] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, John Wiley & Sons, 1984.

[15] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013.

[16] A. Djouadi, *The Anatomy of Electro–Weak Symmetry Breaking*, Phys. Rept. **457** (2008) 1, arXiv: hep-ph/0503172 [hep-ph].

[17] C. Burgard, *Diagram of the Standard Model of Physics*,
Work published under the Creative Commons Attribution 2.5 Generic license, URL:
https://texample.net/tikz/examples/model-physics/ (visited on 05/09/2022).

[18] P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01.

[19] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*,
Phys. Rev. Lett. **81** (1998) 1562, arXiv: hep-ex/9807003 [hep-ex].

[20] Q. R. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301,
arXiv: nucl-ex/0204008 [nucl-ex].

[21] C.-N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*,
Phys. Rev. **96** (1954) 191.

[22] K. G. Wilson, *Confinement of Quarks*, Phys. Rev. D **10** (1974) 2445.

[23] H. Fritzsch and M. Gell-Mann, "Current algebra: Quarks and what else?",
Proceedings of the 16th International Conference On High-Energy Physics: Volume 2,
ed. by J. D. Jackson and A. Roberts, Batavia, Illinois, 1972,
arXiv: hep-ph/0208010 [hep-ph].

[24] H. Fritzsch and P. Minkowski, *Psi Resonances, Gluons and the Zweig Rule*,
Nuovo Cim. A **30** (1975) 393.

[25] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214.

[26] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson,
Experimental Test of Parity Conservation in β Decay, Phys. Rev. **105** (1957) 1413.

[27] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*,
Addison-Wesley, 1995.

[28] J. Riebesell, *Higgs Potential*, Work published under the MIT license,
URL: https://tikz.net/higgs-potential/ (visited on 14/09/2022).

[29] Y. Nambu, *Quasiparticles and Gauge Invariance in the Theory of Superconductivity*,
Phys. Rev. **117** (1960) 648.

[30] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19** (1961) 154.

[31] H. Yukawa, *On the Interaction of Elementary Particles I*,
Proc. Phys. Math. Soc. Jap. **17** (1935) 48.

[32] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.

[33] M. Kobayashi and T. Maskawa,
CP Violation in the Renormalizable Theory of Weak Interaction,
Prog. Theor. Phys. **49** (1973) 652.

[34] ATLAS Collaboration,
Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ decay channel using $139\,fb^{-1}$ of $\sqrt{s} = 13$ TeV pp collisions recorded by the ATLAS detector at the LHC,
Submitted to Phys. Lett. B, 2022, arXiv: 2207.00320 [hep-ex].

[35] ATLAS Collaboration,
Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector,
Eur. Phys. J. C **75** (2015) 476, arXiv: 1506.05669 [hep-ex],
Erratum: Eur. Phys. J. C **76** (2016) 152.

[36] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at* 7 *and* 8 *TeV*, Phys. Rev. D **92** (2015) 012004,
arXiv: 1411.3441 [hep-ex].

[37] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022) 52, arXiv: 2207.00092 [hep-ex].

[38] CMS Collaboration,
A portrait of the Higgs boson by the CMS experiment ten years after the discovery,
Nature **607** (2022) 60, arXiv: 2207.00043 [hep-ex].

[39] 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** (2011) 041803,
arXiv: 1010.0991 [hep-ex].

[40] LHC Higgs Cross Section Working Group,
Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,
ed. by D. de Florian et al., CERN, 2017, arXiv: 1610.07922 [hep-ph].

[41] CMS Collaboration, *Evidence for Higgs boson decay to a pair of muons*, JHEP **01** (2021) 148,
arXiv: 2009.04363 [hep-ex].

[42] F. Maltoni, E. Vryonidou and M. Zaro,
Top-quark mass effects in double and triple Higgs production in gluon-gluon fusion at NLO,
JHEP **11** (2014) 079, arXiv: 1408.6542 [hep-ph].

[43] G. Degrassi, P. P. Giardino, F. Maltoni and D. Pagani,
Probing the Higgs self coupling via single Higgs production at the LHC, JHEP **12** (2016) 080,
arXiv: 1607.04251 [hep-ph].

[44] ATLAS Collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV*,
ATLAS-CONF-2022-050, 2022, URL: https://cds.cern.ch/record/2816332.

[45] ATLAS Collaboration, *Projected sensitivity of Higgs boson pair production combining the $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states with the ATLAS detector at the HL-LHC*,
ATL-PHYS-PUB-2022-005, 2022, URL: https://cds.cern.ch/record/2802127.

[46] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*,
JHEP **05** (2018) 059, arXiv: 1803.02463 [hep-ph].

[47] J. Baglio et al., $gg \to HH$ *: Combined uncertainties*, Phys. Rev. D **103** (2021) 056002,
arXiv: 2008.11626 [hep-ph].

[48] LHC Higgs Working Group, *Current recommendations for HH cross-sections*,
URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGHH (visited on 21/03/2022).

[49] F. A. Dreyer and A. Karlberg, *Vector-Boson Fusion Higgs Pair Production at N^3LO*,
Phys. Rev. D **98** (2018) 114016, arXiv: 1811.07906 [hep-ph].

[50] J. Veatch, *Searches for Resonant Scalar Boson Pair Production Using Run 2 LHC Proton-Proton Collision Data*, Symmetry **14** (2022) 260.

[51] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[52] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[53] A. D. Sakharov,
Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe,
JETP Lett. **5** (1967) 24.

[54] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay,
Evidence for the 2π Decay of the K_2^0 Meson, Phys. Rev. Lett. **13** (1964) 138.

[55] The T2K Collaboration,
Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations,
Nature **580** (2020) 339, arXiv: 1910.03887 [hep-ex], Erratum: Nature **583** (2020) E16.

[56] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helv. Phys. Acta **6** (1933) 110.

[57] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*,
Astrophys. J. **86** (1937) 217.

[58] V. C. Rubin and W. K. Ford Jr.,
Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions,
Astrophys. J. **159** (1970) 379.

[59] V. C. Rubin, N. Thonnard and W. K. Ford Jr., *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 $R = 4$ kpc to UGC 2885 $R = 122$ kpc*,
Astrophys. J. **238** (1980) 471.

[60] D. Clowe et al., *A direct empirical proof of the existence of dark matter*,
Astrophys. J. **648** (2006) L109, arXiv: astro-ph/0608407 [astro-ph].

[61] A. G. Riess et al., *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, Astron. J. **116** (1998) 1009,
arXiv: astro-ph/9805201 [astro-ph].

[62] S. Perlmutter et al., *Measurements of Ω and Λ from 42 High Redshift Supernovae*,
Astrophys. J. **517** (1999) 565, arXiv: astro-ph/9812133 [astro-ph].

[63] G. F. Giudice, "Naturally Speaking: The Naturalness Criterion and Physics at the LHC",
Perspectives on LHC Physics, ed. by G. Kane and A. Pierce, World Scientific, 2008,
arXiv: 0801.2562 [hep-ph].

[64] E. Gildener, *Gauge Symmetry Hierarchies*, Phys. Rev. D **14** (1976) 1667.

[65] S. Weinberg, *Gauge Hierarchies*, Phys. Lett. B **82** (1979) 387.

[66] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*,
Phys. Rev. D **20** (1979) 2619.

[67] E. Majorana, *Teoria simmetrica dell'elettrone e del positrone*, Nuovo Cim. **14** (1937) 171.

[68] M. Aker et al.,
Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN,
Phys. Rev. Lett. **123** (2019) 221802, arXiv: 1909.06048 [hep-ex].

[69] L. Di Luzio, R. Gröber and M. Spannowsky,
Maxi-sizing the trilinear Higgs self-coupling: how large could it be?,
Eur. Phys. J. C **77** (2017) 788, arXiv: 1704.02311 [hep-ph].

[70] B. W. Lee, C. Quigg and H. B. Thacker,
Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass,
Phys. Rev. D **16** (1977) 1519.

[71] J. F. Gunion, H. E. Haber, G. L. Kane and S. Dawson, *The Higgs Hunter's Guide*,
Perseus Publishing, 2000.

[72] B. Patt and F. Wilczek, *Higgs-field portal into hidden sectors*, 2006,
arXiv: hep-ph/0605188 [hep-ph].

[73] R. M. Schabinger and J. D. Wells, *A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the large hadron collider*, Phys. Rev. D **72** (2005) 093007,
arXiv: hep-ph/0509209 [hep-ph].

[74] M. Bowen, Y. Cui and J. D. Wells, *Narrow trans-TeV Higgs bosons and $H \to hh$ decays: Two LHC search paths for a hidden sector Higgs boson*, JHEP **03** (2007) 036,
arXiv: hep-ph/0701035 [hep-ph].

[75] V. Barger, P. Langacker, M. McCaskey, M. J. Ramsey-Musolf and G. Shaughnessy,
LHC Phenomenology of an Extended Standard Model with a Real Scalar Singlet,
Phys. Rev. D **77** (2008) 035005, arXiv: 0706.4311 [hep-ph].

[76] M. J. Dolan, C. Englert and M. Spannowsky,
New Physics in LHC Higgs boson pair production, Phys. Rev. D **87** (2013) 055002,
arXiv: 1210.8166 [hep-ph].

[77] J. M. No and M. Ramsey-Musolf,
Probing the Higgs Portal at the LHC Through Resonant di-Higgs Production,
Phys. Rev. D **89** (2014) 095031, arXiv: 1310.6035 [hep-ph].

[78] C.-Y. Chen, S. Dawson and I. M. Lewis,
Exploring resonant di-Higgs boson production in the Higgs singlet model,
Phys. Rev. D **91** (2015) 035015, arXiv: 1410.5488 [hep-ph].

[79] T. Robens and T. Stefaniak,
LHC Benchmark Scenarios for the Real Higgs Singlet Extension of the Standard Model,
Eur. Phys. J. C **76** (2016) 268, arXiv: 1601.07880 [hep-ph].

[80] J. Alison et al., *Higgs boson potential at colliders: Status and perspectives*,
Rev. Phys. **5** (2020) 100045, ed. by B. Di Micco, M. Gouzevitch, J. Mazzitelli and C. Vernieri,
arXiv: 1910.00012 [hep-ph].

[81] D. O'Connell, M. J. Ramsey-Musolf and M. B. Wise,
Minimal Extension of the Standard Model Scalar Sector, Phys. Rev. D **75** (2007) 037701,
arXiv: hep-ph/0611014 [hep-ph].

[82] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*,
Phys. Rept. **516** (2012) 1, arXiv: 1106.0034 [hep-ph].

[83] H. E. Haber and G. L. Kane,
The Search for Supersymmetry: Probing Physics Beyond the Standard Model,
Phys. Rept. **117** (1985) 75.

[84] M. Trodden, *Electroweak baryogenesis*, Rev. Mod. Phys. **71** (1999) 1463,
arXiv: hep-ph/9803479 [hep-ph].

[85] D. Fontes et al., *The C2HDM revisited*, JHEP **02** (2018) 073, arXiv: 1711.09419 [hep-ph].

[86] A. Djouadi and J. Quevillon, *The MSSM Higgs sector at a high M_{SUSY}: reopening the low tanβ regime and heavy Higgs searches*, JHEP **10** (2013) 028, arXiv: 1304.1787 [hep-ph].

[87] A. Djouadi et al., *The post-Higgs MSSM scenario: Habemus MSSM?*,
Eur. Phys. J. C **73** (2013) 2650, arXiv: 1307.5205 [hep-ph].

[88] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103,
arXiv: 1906.02025 [hep-ex].

[89] CMS Collaboration, *Combination of Searches for Higgs Boson Pair Production in Proton–Proton Collisions at $\sqrt{s}$ = 13 TeV*, Phys. Rev. Lett. **122** (2019) 121803,
arXiv: 1811.09689 [hep-ex].

[90] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[91] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[92] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[93] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*,
JINST **3** (2008) S08006.

[94] TOTEM Collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*,
JINST **3** (2008) S08007.

[95] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*,
MoEDAL-TDR-001; CERN-LHCC-2009-006, 2009,
URL: https://cds.cern.ch/record/1181486.

[96] FASER Collaboration, *FASER: ForwArd Search ExpeRiment at the LHC*,
UCI-TR-2019-01; KYUSHU-RCAPP-2018-08, 2019, arXiv: 1901.04468 [hep-ex].

[97] A. Boyarsky, O. Mikulenko, M. Ovchynnikov and L. Shchutska,
Searches for new physics at SND@LHC, JHEP **03** (2022) 006,
arXiv: 2104.09688 [hep-ph].

[98] E. Mobs, *The CERN accelerator complex in 2019*, 2019,
URL: https://cds.cern.ch/record/2684277.

[99] LHC Commissioning, *LHC long term schedule*,
URL: https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm
(visited on 18/08/2022).

[100] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019,
URL: https://cds.cern.ch/record/2677054.

[101] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: 1606.02625 [hep-ex].

[102] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det].

[103] ATLAS Collaboration, *Online Luminosity Summary Plots over Multiple Years including full Run 2*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 15/08/2022).

[104] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633, Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[105] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det].

[106] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, Eur. Phys. J. C **80** (2020) 1194, arXiv: 2007.07624 [hep-ex].

[107] C. Grupen and B. Schwartz, *Particle Detectors*, Cambridge University Press, 2008.

[108] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, JINST **12** (2017) P05002, arXiv: 1702.06473 [hep-ex].

[109] H. Kolanoski and N. Wermes, *Teilchendetektoren: Grundlagen und Anwendungen*, Springer, 2016.

[110] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, 1996, URL: https://cds.cern.ch/record/331061.

[111] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, ATLAS-TDR-3; CERN-LHCC-96-042, 1996, URL: https://cds.cern.ch/record/331062.

[112] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601, arXiv: 1603.09222 [hep-ex].

[113] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, JINST **15** (2020) P10004, arXiv: 2007.12539 [hep-ex].

[114] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex].

[115] F. Akesson et al., *ATLAS Tracking Event Data Model*, ATL-SOFT-PUB-2006-004, 2006, URL: https://cds.cern.ch/record/973401.

[116] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, ed. by A. Salzburger, ATL-SOFT-PUB-2007-007, 2007, URL: https://cds.cern.ch/record/1020106.

[117] A. Salzburger, *Optimisation of the ATLAS Track Reconstruction Software for Run-2*, J. Phys. Conf. Ser. **664** (2015) 072042.

[118] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, arXiv: 1611.10235 [hep-ex].

[119] R. Fruhwirth, W. Waltenberger and P. Vanlaer, *Adaptive vertex fitting*, J. Phys. G **34** (2007) N343.

[120] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex].

[121] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, ATL-PHYS-PUB-2017-022, 2017, URL: https://cds.cern.ch/record/2298955.

[122] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex].

[123] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017, arXiv: 1812.03848 [hep-ex].

[124] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex].

[125] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[126] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph].

[127] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex].

[128] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: 2007.02645 [hep-ex].

[129] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Submitted to Eur. Phys. J. C, 2022, arXiv: 2211.16345 [hep-ex].

[130] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281.

[131] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL: https://cds.cern.ch/record/2255226.

[132] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: https://cds.cern.ch/record/2270366.

[133] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-PUB-2018-025, 2018, URL: https://cds.cern.ch/record/2645405.

[134] C. Deutsch, *Identification and Classification of Hadronic Tau Lepton Decays in the ATLAS Experiment for Run 2 of the LHC*, Master's thesis: University of Bonn, 2017.

[135] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}$ = 13 TeV*, ATLAS-CONF-2017-029, 2017, URL: https://cds.cern.ch/record/2261772.

[136] D. U. Duschinger, *Search for neutral bosons decaying into the fully hadronic di-tau final state with the ATLAS detector at the LHC*, PhD Thesis: TU Dresden, 2018, URL: https://cds.cern.ch/record/2673361.

[137] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv: 0802.1188 [hep-ph].

[138] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, Eur. Phys. J. C **76** (2016) 295, arXiv: 1512.05955 [hep-ex].

[139] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex].

[140] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, 2012, URL: https://cds.cern.ch/record/1456844.

[141] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219.

[142] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, ed. by H. B. Prosper and L. Lyons, Geneva, Switzerland, 2011, arXiv: 1103.0354 [physics.data-an].

[143] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv: 1007.1727 [physics.data-an], Erratum: Eur. Phys. J. C **73** (2013) 2501.

[144] G. Casella and R. L. Berger, *Statistical Inference*, Cengage Learning, 2001.

[145] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. A **434** (1999) 435, arXiv: hep-ex/9902006 [hep-ex].

[146] A. L. Read, "Modified frequentist analysis of search results (The CL_s method)", *Proceedings of the Workshop on Confidence Limits*, ed. by L. Lyons, Y. Perrin and F. E. James, Geneva, Switzerland, 2000, URL: https://cds.cern.ch/record/411537.

[147] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693.

[148] L. Breiman, J. Friedman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees*, Chapman & Hall/CRC, 1984.

[149] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.

[150] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.

[151] J. Friedman, T. Hastie and R. Tibshirani, *Additive logistic regression: a statistical view of boosting*, Ann. Stat. **28** (2000) 337.

[152] J. H. Friedman, *Greedy function approximation: A gradient boosting machine*, Ann. Stat. **29** (2001) 1189.

[153] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an].

[154] G. Cybenko, *Approximation by Superpositions of a Sigmoidal Function*, Math. Control Signals Syst. **2** (1989) 303.

[155] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are universal approximators*, Neural Netw. **2** (1989) 359.

[156] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

[157] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys. **4** (1964) 1.

[158] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533.

[159] I. Sutskever, J. Martens, G. Dahl and G. Hinton, "On the importance of initialization and momentum in deep learning", *Proceedings of the 30th International Conference on Machine Learning*, ed. by S. Dasgupta and D. McAllester, Atlanta, USA, 2013.

[160] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Comput. **9** (1997) 1735.

[161] F. A. Gers, J. Schmidhuber and F. Cummins, *Learning to forget: Continual prediction with LSTM*, Neural Comput. **12** (2000) 2451.

[162] F. Chollet et al., *Keras*, 2015, URL: https://keras.io.

[163] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and Collider Physics*, Cambridge University Press, 2011.

[164] ATLAS Collaboration, *Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 322, arXiv: 1602.00988 [hep-ex].

[165] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187.

[166] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph].

[167] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[168] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[169] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[170] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[171] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, URL: https://cds.cern.ch/record/2064383.

[172] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, URL: https://www.tensorflow.org.

[173] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines", *Proceedings of the 27th International Conference on Machine Learning*, ed. by J. Fürnkranz and T. Joachims, Haifa, Israel, 2010.

[174] D. H. Guest et al., *lwtnn: Version 2.8.1*, 2019, URL: https://doi.org/10.5281/zenodo.2583131.

[175] ATLAS Collaboration, *Trigger Menu in 2018*, ATL-DAQ-PUB-2019-001, 2019, URL: https://cds.cern.ch/record/2693402.

[176] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV*, JINST **13** (2018) P10005, arXiv: 1809.02816 [hep-ex].

[177] CMS Collaboration, *Identification of hadronic tau lepton decays using a deep neural network*, JINST **17** (2022) P07023, arXiv: 2201.08458 [hep-ex].

[178] M. Maerker, *Search for heavy neutral Higgs bosons decaying into the fully hadronic di-tau final state with ATLAS*, PhD Thesis: TU Dresden, 2021, URL: https://cds.cern.ch/record/2800824.

[179] M. Zaheer et al., "Deep Sets", *Advances in Neural Information Processing Systems 30*, ed. by I. Guyon et al., 2017.

[180] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014, 2020, URL: https://cds.cern.ch/record/2718948.

[181] Particle Data Group, P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01.

[182] ATLAS Collaboration, *Searches for Higgs boson pair production in the $hh \to bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector*, Phys. Rev. D **92** (2015) 092004, arXiv: 1509.04670 [hep-ex].

[183] CMS Collaboration, *Search for Higgs boson pair production in the $bb\tau\tau$ final state in proton–proton collisions at $\sqrt{(}s) = 8$ TeV*, Phys. Rev. D **96** (2017) 072004, arXiv: 1707.00350 [hep-ex].

[184] ATLAS Collaboration, *A search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. Lett. **121** (2018) 191801, arXiv: 1808.00336 [hep-ex], Erratum: Phys. Rev. Lett. **122** (2019) 089901.

[185] CMS Collaboration, *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*, Phys. Lett. B **778** (2018) 101, arXiv: 1707.02909 [hep-ex].

[186] CMS Collaboration, *Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV*, Submitted to Phys. Lett. B, 2022, arXiv: 2206.09401 [hep-ex].

[187] ATLAS Collaboration, *Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Submitted to Phys. Rev. D, 2023, arXiv: 2301.03212 [hep-ex].

[188] CMS Collaboration, *Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **129** (2022) 081802, arXiv: 2202.09617 [hep-ex].

[189] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **106** (2021) 052001, arXiv: 2112.11876 [hep-ex].

[190] CMS Collaboration, *Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton–proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **03** (2021) 257, arXiv: 2011.12373 [hep-ex].

[191] ATLAS Collaboration, *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector*, ATLAS-CONF-2021-030, 2021, URL: https://cds.cern.ch/record/2777236.

[192] ATLAS Collaboration, *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector*, Submitted to JHEP, 2022, arXiv: 2209.10910 [hep-ex].

[193] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[194] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 [hep-ph].

[195] S. Frixione, P. Nason and C. Oleari,
Matching NLO QCD computations with parton shower simulations: the POWHEG method,
JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[196] 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** (2010) 043,
arXiv: 1002.2581 [hep-ph].

[197] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*,
J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph].

[198] 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*, JHEP **07** (2014) 079,
arXiv: 1405.0301 [hep-ph].

[199] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040,
arXiv: 1410.8849 [hep-ph].

[200] S. Gieseke, C. Rohr and A. Siodmok, *Colour reconnections in Herwig++*,
Eur. Phys. J. C **72** (2012) 2225, arXiv: 1206.0041 [hep-ph].

[201] J. Bellm et al., *Herwig 7.1 Release Note*, 2017, arXiv: 1705.06919 [hep-ph].

[202] S. Frixione, G. Ridolfi and P. Nason,
A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction,
JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[203] M. Beneke, P. Falgari, S. Klein and C. Schwinn,
Hadronic top-quark pair production with NNLL threshold resummation,
Nucl. Phys. B **855** (2012) 695, arXiv: 1109.1536 [hep-ph].

[204] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*,
Phys. Lett. B **710** (2012) 612, arXiv: 1111.5869 [hep-ph].

[205] P. Bärnreuther, M. Czakon and A. Mitov, *Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$*,
Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph].

[206] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv: 1207.0236 [hep-ph].

[207] M. Czakon and A. Mitov,
NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction,
JHEP **01** (2013) 080, arXiv: 1210.6832 [hep-ph].

[208] M. Czakon, P. Fiedler and A. Mitov,
Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$,
Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph].

[209] M. Czakon and A. Mitov,
Top++: A program for the calculation of the top-pair cross-section at hadron colliders,
Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph].

[210] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130, arXiv: 1207.5391 [hep-ph].

[211] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, arXiv: 0907.4076 [hep-ph], Erratum: JHEP **02** (2010) 011.

[212] LHC Top Physics Working Group, *ATLAS-CMS recommended predictions for single-top cross sections using the Hathor v2.1 program*, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec (visited on 13/01/2022).

[213] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: 1009.2450 [hep-ph].

[214] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D **82** (2010) 054018, arXiv: 1005.4451 [hep-ph].

[215] N. Kidonakis, "Top Quark Production", *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)* (15th–28th July 2013), Dubna, Russia, arXiv: 1311.0283 [hep-ph].

[216] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph].

[217] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph].

[218] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, ATL-PHYS-PUB-2017-006, 2017, URL: https://cds.cern.ch/record/2261937.

[219] C. Anastasiou, L. Dixon, K. Melnikov and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266 [hep-ph].

[220] K. Hamilton, P. Nason, E. Re and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, arXiv: 1309.0017 [hep-ph].

[221] K. Hamilton, P. Nason and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140, arXiv: 1501.04637 [hep-ph].

[222] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex].

[223] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 037, arXiv: 0911.5299 [hep-ph].

[224] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence*, Phys. Rev. Lett. **117** (2016) 012001, arXiv: 1604.06447 [hep-ph], Erratum: Phys. Rev. Lett. **117** (2016) 079901.

[225] J. Baglio et al., *Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme*, Eur. Phys. J. C **79** (2019) 459, arXiv: 1811.05692 [hep-ph].

[226] G. Heinrich, S. P. Jones, M. Kerner, G. Luisoni and E. Vryonidou, *NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers*, JHEP **08** (2017) 088, arXiv: 1703.09252 [hep-ph].

[227] G. Heinrich, S. P. Jones, M. Kerner, G. Luisoni and L. Scyboz, *Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects*, JHEP **06** (2019) 066, arXiv: 1903.08137 [hep-ph].

[228] G. Heinrich, S. P. Jones, M. Kerner and L. Scyboz, *A non-linear EFT description of $gg \to HH$ at NLO interfaced to POWHEG*, JHEP **10** (2020) 021, arXiv: 2006.16877 [hep-ph].

[229] S. Dawson, S. Dittmaier and M. Spira, *Neutral Higgs boson pair production at hadron colliders: QCD corrections*, Phys. Rev. D **58** (1998) 115012, arXiv: hep-ph/9805244 [hep-ph].

[230] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett. **111** (2013) 201801, arXiv: 1309.6594 [hep-ph].

[231] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex].

[232] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex].

[233] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex], Erratum: Eur. Phys. J. C **77** (2017) 712.

[234] ATLAS Collaboration, *Forward jet vertex tagging using the particle flow algorithm*, ATL-PHYS-PUB-2019-026, 2019, URL: https://cds.cern.ch/record/2683100.

[235] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B **659** (2008) 119, arXiv: 0707.1378 [hep-ph].

[236] ATLAS Collaboration, *Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector*, JHEP **12** (2017) 024, arXiv: 1708.03299 [hep-ex].

[237] ATLAS Collaboration, *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, Phys. Lett. B **786** (2018) 59, arXiv: 1808.08238 [hep-ex].

[238] ATLAS Collaboration, *Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at* 13 *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 178, arXiv: 2007.02873 [hep-ex].

[239] A. D. Bukin, *Fitting function for asymmetric peaks*, 2007, arXiv: 0711.4449 [physics.data-an].

[240] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. A **654** (2011) 481, arXiv: 1012.4686 [hep-ex].

[241] ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2022) 175, arXiv: 2201.08269 [hep-ex].

[242] M. Hübner, *Higgs Boson Production Cross-Section Measurements in the Di-Tau Final State at ATLAS*, PhD Thesis: University of Bonn, 2022, URL: https://hdl.handle.net/20.500.11811/9734.

[243] ATLAS Collaboration, *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: https://cds.cern.ch/record/2037702.

[244] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, ATL-DAQ-PUB-2016-001, 2016, URL: https://cds.cern.ch/record/2136007.

[245] ATLAS Collaboration, *Trigger Menu in 2016*, ATL-DAQ-PUB-2017-001, 2017, URL: https://cds.cern.ch/record/2242069.

[246] ATLAS Collaboration, *Trigger Menu in 2017*, ATL-DAQ-PUB-2018-002, 2018, URL: https://cds.cern.ch/record/2625986.

[247] ATLAS Collaboration, *The ATLAS Tau Trigger in Run 2*, ATLAS-CONF-2017-061, 2017, URL: https://cds.cern.ch/record/2274201.

[248] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[249] ATLAS Collaboration, *The ATLAS Inner Detector Trigger performance in pp collisions at* 13 *TeV during LHC Run 2*, Eur. Phys. J. C **82** (2021) 206, arXiv: 2107.02485 [hep-ex].

[250] ATLAS Collaboration, *Performance of the ATLAS Level-1 topological trigger in Run 2*, Eur. Phys. J. C **82** (2021) 7, arXiv: 2105.01416 [hep-ex].

[251] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731.

[252] ATLAS Collaboration, *Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using* 13 *TeV proton–proton collision data in ATLAS*, JHEP **11** (2020) 163, arXiv: 2007.14811 [hep-ex].

[253] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph].

[254] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph].

[255] A. Denner, S. Dittmaier and L. Hofer, *Collier: A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph].

[256] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph].

[257] ATLAS Collaboration, *Measurements of the production cross-section for a Z boson in association with b-jets in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **07** (2020) 044, arXiv: 2003.11960 [hep-ex].

[258] P. Bokan, *Pair production of Higgs bosons in the final state with bottom quarks and τ leptons in the ATLAS experiment: Search results using LHC Run 2 data and prospect studies at the HL-LHC*, PhD Thesis: Uppsala University, 2020, URL: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-419002.

[259] ATLAS Collaboration, *Object-based missing transverse momentum significance in the ATLAS Detector*, ATLAS-CONF-2018-038, 2018, URL: https://cds.cern.ch/record/2630948.

[260] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, Eur. Phys. J. C **76** (2016) 235, arXiv: 1601.07913 [hep-ex].

[261] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[262] G. C. Cawley and N. L. Talbot, *On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation*, J. Mach. Learn. Res. **11** (2010) 2079.

[263] M. Stone, *Cross-Validatory Choice and Assessment of Statistical Predictions*, J. R. Stat. Soc. B **36** (1974) 111, URL: http://www.jstor.org/stable/2984809.

[264] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[265] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.

[266] L. Breiman, *Random Forests*, Machine Learning **45** (2001) 5.

[267] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex].

[268] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex].

[269] ATLAS Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS detector*, Eur. Phys. J. C **82** (2021) 95, arXiv: 2109.10627 [hep-ex].

[270] ATLAS Collaboration, *Light flavour mistag calibration of the 2019 ATLAS particle-flow jets b-taggers using the Z+jets Negative Tag method*, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-002/ (visited on 29/07/2022).

[271] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, URL: https://cds.cern.ch/record/2625233.

[272] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv: 0805.3067 [hep-ph].

[273] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **01** (2015) 069, arXiv: 1409.6212 [hep-ex].

[274] E. L. Lehmann, J. P. Romano and G. Casella, *Testing Statistical Hypotheses*, Springer, 2005.

[275] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, Eur. Phys. J. C **70** (2010) 525, arXiv: 1005.1891 [physics.data-an].

[276] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8** (1959) 229.

[277] R. B. Nelsen, *An Introduction to Copulas*, Springer, 2006.

[278] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Stat. **7** (1979) 1.

[279] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, CRC Press, 1994.

[280] ATLAS Collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s}$ = 13 TeV*, (2022), arXiv: 2211.01216 [hep-ex].

[281] CMS Collaboration, *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*, (2022), arXiv: 2205.06667 [hep-ex].

[282] CMS Collaboration, *Search for Higgs boson pairs decaying to WWWW, WW$\tau\tau$, and $\tau\tau\tau\tau$ in proton–proton collisions at $\sqrt{s}$ = 13 TeV*, (2022), arXiv: 2206.10268 [hep-ex].

[283] ATLAS Collaboration, *Projected sensitivity of Higgs boson pair production in the $bb\tau\tau$ final state using proton–proton collisions at HL-LHC with the ATLAS detector*, ATL-PHYS-PUB-2021-044, 2021, URL: https://cds.cern.ch/record/2798448.

[284] ATLAS Collaboration, *Expected performance of the ATLAS detector at the High-Luminosity LHC*, ATL-PHYS-PUB-2019-005, 2019, URL: https://cds.cern.ch/record/2655304.

[285] ATLAS Collaboration, *Measurement prospects of Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state with the ATLAS experiment at the HL-LHC*, ATL-PHYS-PUB-2022-001, 2022, URL: https://cds.cern.ch/record/2799146.

[286] ATLAS Collaboration, *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **105** (2022) 092002, arXiv: 2202.07288 [hep-ex].

[287] ATLAS Collaboration, *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, HEPData, https://doi.org/10.17182/hepdata.111124, 2022.

[288] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, HEPData, https://doi.org/10.17182/hepdata.105864, 2022.

[289] CMS Collaboration, *Search for heavy resonances decaying to a pair of Lorentz-boosted Higgs bosons in final states with leptons and a bottom quark pair at $\sqrt{s}$ = 13 TeV*, JHEP **05** (2021) 005, arXiv: 2112.03161 [hep-ex].

[290] CMS Collaboration, *Search for heavy resonances decaying to a pair of Lorentz-boosted Higgs bosons in final states with leptons and a bottom quark pair at $\sqrt{s}$= 13 TeV*, HEPData, https://doi.org/10.17182/hepdata.115024, 2021.

[291] ATLAS Collaboration, *Combination of searches for non-resonant and resonant Higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS-CONF-2021-052, 2021, URL: https://cds.cern.ch/record/2786865.

[292] S. Borowka et al., *Full top quark mass dependence in Higgs boson pair production at NLO*, JHEP **10** (2016) 107, arXiv: 1608.04798 [hep-ph].

[293] S. Amoroso et al., "Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report", *Proceedings of the 11th Les Houches Workshop on Physics at TeV Colliders*, Les Houches, France, 2020, arXiv: 2003.01700 [hep-ph].

[294] ATLAS Collaboration, *Constraint of the Higgs boson self-coupling from Higgs boson differential production and decay measurements*, ATL-PHYS-PUB-2019-009, 2019, URL: https://cds.cern.ch/record/2667570.

[295] F. Maltoni, D. Pagani, A. Shivaji and X. Zhao, *Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC*, Eur. Phys. J. C **77** (2017) 887, arXiv: 1709.08649 [hep-ph].

[296] ATLAS Collaboration, *A morphing technique for signal modelling in a multidimensional space of coupling parameters*, ATL-PHYS-PUB-2015-047, 2015, URL: https://cds.cern.ch/record/2066980.

[297] ATLAS Collaboration, *Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector*, ATL-PHYS-PUB-2019-007, 2019, URL: http://cds.cern.ch/record/2665057.

[298] C. Dimitriadi, personal communication, 2022.

[299] ATLAS Collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s}$ = 13 TeV*, HEPData, https://doi.org/10.17182/hepdata.135471, 2022.

[300] FCC Collaboration,
FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1,
Eur. Phys. J. C **79** (2019) 474.

[301] P. Bambade et al., *The International Linear Collider: A Global Project*, 2019,
arXiv: 1903.01629 [hep-ex].

[302] CLIC Collaboration, *The CLIC Potential for New Physics*, ed. by J. de Blas et al.,
CERN Yellow Reports: Monographs, CERN, 2018, arXiv: 1812.02093 [hep-ph],
URL: http://cds.cern.ch/record/2650541.

[303] ATLAS Collaboration,
Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector,
JHEP **04** (2015) 117, arXiv: 1501.04943 [hep-ex].

[304] S. Frixione and B. R. Webber,
Matching NLO QCD computations and parton shower simulations, JHEP **06** (2002) 029,
arXiv: hep-ph/0204244 [hep-ph].

[305] S. Mrenna and P. Skands, *Automated parton-shower variations in PYTHIA 8*,
Phys. Rev. D **94** (2016) 074005, arXiv: 1605.08352 [hep-ph].

[306] C. Bierlich et al., *Pythia 8.230 Documentation: Automated Variations of Shower Parameters for Uncertainty Bands*,
URL: https://pythia.org/manuals/pythia8230/Variations.html (visited on 30/05/2022).

[307] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168.

[308] ATLAS Collaboration, *Study of top-quark pair modelling and uncertainties using ATLAS measurements at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2020-023, 2020,
URL: https://cds.cern.ch/record/2730443.

[309] J. Anders, *Searches for direct pair production of third generation squarks, and dark matter, in final states containing b-jets and* E_T^{miss} *using the ATLAS detector at the LHC*,
PhD Thesis: University of Liverpool, 2017,
URL: https://livrepository.liverpool.ac.uk/id/eprint/3009386.

[310] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*,
JHEP **11** (2001) 063, arXiv: hep-ph/0109231 [hep-ph].

[311] S. Höche, F. Krauss, S. Schumann and F. Siegert,
QCD matrix elements and truncated showers, JHEP **05** (2009) 053,
arXiv: 0903.1219 [hep-ph].

[312] S. Höche, F. Krauss, M. Schönherr and F. Siegert,
QCD matrix elements + parton showers. The NLO case, JHEP **04** (2013) 027,
arXiv: 1207.5030 [hep-ph].

[313] A. Buckley et al., *LHAPDF6: parton density access in the LHC precision era*,
Eur. Phys. J. C **75** (2015) 132, arXiv: 1412.7420 [hep-ph].

[314] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne,
Parton distributions in the LHC era: MMHT 2014 PDFs, Eur. Phys. J. C **75** (2015) 204,
arXiv: 1412.3989 [hep-ph].

[315] S. Dulat et al.,
New parton distribution functions from a global analysis of quantum chromodynamics,
Phys. Rev. D **93** (2016) 033006, arXiv: 1506.07443 [hep-ph].

[316] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*,
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[317] G. Bohm and G. Zech, *Statistics of weighted Poisson events and its applications*,
Nucl. Instrum. Meth. A **748** (2014) 1, arXiv: 1309.1287 [physics.data-an].

[318] H. Teicher, *On the Multivariate Poisson Distribution*, Scand. Actuar. J. **1954** (1954) 1.

[319] D. I. Inouye, E. Yang, G. I. Allen and P. Ravikumar,
A review of multivariate distributions for count data derived from the Poisson distribution,
Wiley Interdiscip. Rev. Comput. Stat. **9** (2017) e1398.

[320] ATLAS Collaboration,
Evaluating statistical uncertainties and correlations using the bootstrap method,
ATL-PHYS-PUB-2021-011, 2021, URL: https://cds.cern.ch/record/2759945.

List of Figures

List of Tables

Acknowledgements

The completion of this dissertation would not have been possible without the support and encouragement of many people. I am deeply grateful for the contributions of those who have been a part of my journey. Although it is not possible to mention each person by name, I would like to extend my gratitude to all those who have helped to make this thesis a reality.

First of all, I am deeply indebted to Prof. Dr. Jochen Dingfelder for accepting me as part of his research group and giving me the freedom to pursue many interesting topics during my PhD. I am grateful for your unwavering guidance, support, and patience throughout my MSc and PhD studies in your group. *Thank you!*

I am extremely grateful to my co-supervisor Tatjana Lenz. Thank you for supporting me and for the many occasions where you proofread the things I wrote, be it abstracts, internal notes, or chapters of this thesis. Thank you also for coordinating the $HH \to b\bar{b}\tau^+\tau^-$ effort in ATLAS, which, I am sure, must have felt like herding cats at times.

Many thanks to Prof. Dr. Ian Brock for accepting to be a referee for my dissertation. Moreover, I am grateful to PD Dr. Bastian Kubis and Prof. Dr. Elena Demidova for agreeing to be members of my doctoral committee.

My appreciation goes out to the members and alumni of the HEP1 group. Thank you, Shubham Bansal, Florian Beisiegel, Christina Dimitriadi, Stephan Duell, Peter Falke, Stephan Hageböck, Ruth Jacobs, Christian Nass, Oğul Öncel, Jonathan Pampel, Christos Vergis, and Eckhard von Törne for the lunches, coffee breaks, and fruitful discussion. It was a pleasure to be able to work with all of you. In addition, I am thankful to Shubham, Christina, Peter, Christian, and Eckhard for giving me feedback on parts of this thesis.

Special thanks go to Benedict Winter, William Davey, Michel Janus, and Pier-Olivier DeViveiros for taking me under their wings during the early stages of my career in the ATLAS collaboration. I appreciate your valuable advice and infectious enthusiasm, which was a big part in motivating me to continue on the academic track after finishing my Master's thesis.

I would like to acknowledge my collaborators at the ATLAS experiment: Alessandra Betti, Christian Grefe, Zhiyuan "Jordan" Li, Yanlin Liu, Max Märker, Bertrand Martin Dit Latour, Tulin Mete, Giacinto Piacquadio, Elliot Reynolds, Soshi Tsuno, Song-Ming Wang, Daniele Zanzi, and Bowen Zhang. Many thanks also to my $b\bar{b}\tau^+\tau^-$ friends Petar Bokan, Christina Dimitriadi, and Serhat Ördek for the post-work Zoom hangouts during the pandemic and introducing me to the art of crab walking.

Lastly, I want to thank my parents. None of this would have been possible without your unparalleled support. Thanks should also go to Hilda for occasionally taking my mind off the day-to-day struggles of doing a PhD.

Christopher Deutsch
Bonn, February 2023