# Search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state with the ATLAS detector

Florian Haslbeck

The Queen's College

University of Oxford

A thesis submitted for the degree of
*Doctor of Philosophy*
Trinity Term 2025

# Abstract

The self-interaction of the Higgs boson is a key prediction of the Standard Model (SM), and its experimental verification through Higgs boson pair ($HH$) production remains one of the most important goals in particle physics, thirteen years after the Higgs boson's discovery.

This thesis presents the search for non-resonant $HH$ production via gluon fusion (ggF) and vector boson fusion (VBF), in the final state with two $b$-jets and two $\tau$-leptons. The analysis is based on the full ATLAS Run 2 dataset of proton-proton collisions, corresponding to an integrated luminosity of $140\,\mathrm{fb}^{-1}$ at a centre-of-mass energy of $13\,\mathrm{TeV}$. An observed (expected) upper limit on the signal strength of $\mu_{HH} < 5.9\,(3.3)$ times the SM prediction is set at the $95\,\%$ confidence level, ranking this analysis among the three most sensitive $HH$ searches to date. The expected limit is improved by $15\,\%$, corresponding to approximately $60\,\mathrm{fb}^{-1}$ of additional data in this statistically limited analysis. Moreover, the optimised analysis strategy enables the first independent measurements of the ggF and VBF $HH$ production signal strengths. This leads to enhanced sensitivity to the Higgs boson self-coupling modifier $\kappa_\lambda$ and the quartic coupling modifier $\kappa_{2V}$ ($V = W, Z$) with observed (expected) $95\,\%$ confidence intervals of $\kappa_\lambda \in [-3.1, 9.0]\,([-2.5, 9.3])$ and $\kappa_{2V} \in [-0.5, 2.7]\,([-0.2, 2.4])$, corresponding to expected improvements of $11\,\%$ and $19\,\%$, respectively.

A dedicated extrapolation study conducted as part of this thesis demonstrates that the observation of $HH$ production is within realistic reach at the High-Luminosity LHC (HL-LHC). In preparation for the HL-LHC, this thesis also investigates the Detector Safety System (DSS), focusing on alarm behaviour during upgrade and maintenance activities. A detailed study reveals that alarms are significantly more frequent during these periods. To improve operational safety and diagnostics, a new tool - the Alarm Helper - was developed. It facilitates alarm follow-up and provides a structured database that will support future safety improvements during ATLAS operations.

# Contents

**1 Introduction** **1**

**2 Theoretical foundations** **4**

- 2.1 The Standard Model of particle physics . . . 4
  - 2.1.1 Particles and forces . . . 5
  - 2.1.2 Quantum electrodynamics . . . 5
  - 2.1.3 Quantum chromodynamics . . . 6
  - 2.1.4 Electroweak unification . . . 7
  - 2.1.5 The Higgs mechanism and particle masses . . . 8
  - 2.1.6 Shortcomings of the SM . . . 12
- 2.2 Simulation of proton-proton collisions . . . 13
  - 2.2.1 Hard scatter cross-section . . . 13
  - 2.2.2 Event generation . . . 15

**3 Higgs boson physics** **17**

- 3.1 The Higgs boson . . . 17
- 3.2 Production and decay of Higgs bosons at the LHC . . . 18
  - 3.2.1 Single Higgs boson production and decay . . . 19
  - 3.2.2 Higgs boson pair production and decay . . . 21
- 3.3 Effective Higgs boson coupling modifiers . . . 23
  - 3.3.1 Phenomenology of $HH$ production with anomalous couplings . . . 24
- 3.4 Experimental status . . . 28

**4 The ATLAS Experiment** **33**

- 4.1 The Large Hadron Collider . . . 33
- 4.2 The ATLAS detector . . . 35
  - 4.2.1 The ATLAS coordinate system . . . 36
  - 4.2.2 Particle identification with the sub-detectors . . . 37
  - 4.2.3 The Inner Detector . . . 38
  - 4.2.4 The calorimeter system . . . 40
  - 4.2.5 The Muon Spectrometer . . . 41
  - 4.2.6 The luminosity measurement system . . . 43
  - 4.2.7 The trigger and data acquisition system . . . 43
  - 4.2.8 Computing and data processing . . . 43
  - 4.2.9 Run 2 performance . . . 45
- 4.3 Outlook to the High Luminosity LHC . . . 45

4.4 Detector services . . . . . . . . 46
4.5 Safety systems . . . . . . . . 47
4.5.1 Environmental monitoring systems . . . . . . . . 48
4.6 Detector Safety System . . . . . . . . 49
4.7 Detector Control System . . . . . . . . 52
4.8 Expert System . . . . . . . . 53
4.8.1 Database . . . . . . . . 53
4.8.2 User interfaces . . . . . . . . 55

**5 Improvements of the detector safety 58**
5.1 DSS alarm analysis . . . . . . . . 58
5.1.1 Dataset . . . . . . . . 60
5.1.2 Evolution of errors and interventions in the last six years . . . . . . . . 62
5.1.3 Differences across operation modes . . . . . . . . 63
5.1.4 Time distribution of alarms . . . . . . . . 65
5.1.5 Distribution of alarms across sub-systems . . . . . . . . 67
5.2 The Alarm Helper tool . . . . . . . . 69
5.2.1 Interfaces . . . . . . . . 70
5.3 Conclusion and outlook . . . . . . . . 73

**6 Reconstruction and identification of physics objects 75**
6.1 Charged particle tracks and vertex reconstruction . . . . . . . . 76
6.2 Calorimeter clusters . . . . . . . . 77
6.3 Electrons . . . . . . . . 77
6.4 Photons . . . . . . . . 78
6.5 Muons . . . . . . . . 79
6.6 Jets . . . . . . . . 80
6.6.1 Flavour tagging . . . . . . . . 82
6.7 $\tau$-leptons . . . . . . . . 84
6.7.1 anti-$\tau_{\text{had-vis}}$ . . . . . . . . 87
6.8 Missing transverse energy . . . . . . . . 88
6.9 Overlap removal . . . . . . . . 88

**7 Search for non-resonant Higgs boson pair production 90**
7.1 Introduction . . . . . . . . 90
7.2 Analysis Overview . . . . . . . . 91
7.2.1 Event selection and categorisation . . . . . . . . 91
7.2.2 Main backgrounds . . . . . . . . 93
7.2.3 Signal extraction . . . . . . . . 95
7.3 Data and simulated event samples . . . . . . . . 96
7.3.1 Data . . . . . . . . 96
7.3.2 Simulated Samples . . . . . . . . 96
7.4 Object reconstruction and selection . . . . . . . . 104
7.5 Trigger and event selection . . . . . . . . 104
7.5.1 Trigger selection . . . . . . . . 104
7.5.2 Event selection . . . . . . . . 107
7.6 Boosted decision tree strategy . . . . . . . . 109
7.6.1 Decision trees and Boosted decision trees . . . . . . . . 109
7.6.2 Training and hyper-parameter optimisation . . . . . . . . 110

7.6.3 Input variable selection . . . . . . . . . . 111
7.7 Event categorisation . . . . . . . . . . 112
7.8 Signal acceptance times efficiency . . . . . . . . . . 114
7.9 Background composition and estimation . . . . . . . . . . 117
7.9.1 Top quark . . . . . . . . . . 117
7.9.2 Z+HF . . . . . . . . . . 119
7.9.3 Fake $\tau_{\text{had-vis}}$ . . . . . . . . . . 120
7.9.4 Single Higgs and other backgrounds . . . . . . . . . . 123
7.10 Multivariate signal extraction . . . . . . . . . . 123
7.11 Systematic uncertainties . . . . . . . . . . 126
7.11.1 Experimental uncertainties . . . . . . . . . . 126
7.11.2 Modelling Uncertainties . . . . . . . . . . 127
7.11.3 Signal modelling . . . . . . . . . . 131
7.11.4 Background modelling . . . . . . . . . . 134
7.12 Statistical Analysis . . . . . . . . . . 138
7.12.1 Statistical model . . . . . . . . . . 139
7.12.2 Fit setup . . . . . . . . . . 144
7.13 Results . . . . . . . . . . 146
7.13.1 Inclusive signal strength results . . . . . . . . . . 146
7.13.2 Signal strengths for ggF and VBF HH production . . . . . . . . . . 156
7.13.3 Constraints on the coupling modifiers $\kappa_\lambda$ and $\kappa_{2V}$ . . . . . . . . . . 157
7.14 VZ verification measurement . . . . . . . . . . 161
7.15 Conclusion and outlook . . . . . . . . . . 166

**8 Projection of the Run 2 result to the High Luminosity LHC period 168**
8.1 Extrapolation procedure . . . . . . . . . . 169
8.2 Results . . . . . . . . . . 172
8.2.1 HH production signal strength . . . . . . . . . . 172
8.2.2 Constraints on Higgs boson coupling modifiers . . . . . . . . . . 177
8.3 Improvements to the identification of $b$-jets and $\tau$-leptons . . . . . . . . . . 180
8.4 Combination with other final states . . . . . . . . . . 182

**9 Conclusion and outlook 183**

**A Appendix 187**
A.1 VBF signal parametrisation . . . . . . . . . . 187
A.2 Estimation of the fake-$\tau_{\text{had-vis}}$ background . . . . . . . . . . 188

**Bibliography 191**

**Acknowledgements 207**

# 1 Introduction

The Standard Model (SM) of particle physics describes the smallest constituents of the universe, subatomic particles, and three out of the four known fundamental forces. The theory has been developed and tested through the interplay of experimental observations and theoretical predictions. The SM has not only successfully explained all recorded collider data, but has also predicted multiple new particles that were subsequently discovered. For example, the ATLAS and CMS collaborations at the CERN Large Hadron Collider (LHC) discovered the Higgs boson in 2012 [1, 2], after the Higgs mechanism and its associated particle were hypothesised in the 1960s to explain the masses of the fundamental particles [3–5].

Notwithstanding its impressive success, substantial shortcomings suggest that the SM is not yet the final theory of nature. Theoretical issues and inexplicable observations, such as that the SM can only describe approximately 15 % of the universe's known matter content [6], underscore the need for new physics beyond the SM. In spite of exhaustive searches and measurements, no significant deviations have been observed thus far, fuelling the search for any deviation from SM predictions that might hint at a more comprehensive understanding of the underlying principles governing particle physics.

The Higgs boson presents an ideal laboratory for new physics searches at the LHC. The particle is not only intrinsically linked to various theoretical issues of the SM [7], but has yet to be fully characterised experimentally. This limited experimental precision leaves room for deviations from SM predictions, which might point to physics beyond the SM. One particularly important and remaining test is the measurement of Higgs boson self-coupling. In the SM, the Brout-Englert-

Higgs mechanism [3–5, 8] predicts the coupling of the Higgs boson to itself. The strength of this coupling is directly linked to the shape of the Higgs potential, another core ingredient of the SM. Today, the most performant measurement of the self-coupling is the measurement of Higgs boson pair production ($HH$) at high energy collisions, which can only be achieved by the LHC at CERN. However, $HH$ production is extremely rare and experimental evidence for it has yet to be established. Therefore, analyses only have set loose bounds on the interaction strength, and thus the shape of the potential. Notably, some alternative theories that seek to address the shortcomings of the SM, such as extended Higgs sectors, propose diverse shapes for the Higgs potential [9, 10] that are not yet ruled out with our current knowledge of the self-coupling. Therefore, establishing experimental evidence for $HH$ production provides an important test of a vital ingredient of the SM.

This thesis presents the search for Higgs boson pair production in final states with two $b$-quarks and two $\tau$-leptons with the ATLAS detector ($HH \to b\bar{b}\tau^+\tau^-$), utilising the LHC Run 2 proton-proton collision dataset of $140\,\mathrm{fb}^{-1}$ that has been recorded at a centre-of-mass energy of $13\,\mathrm{TeV}$.

The search is primarily limited by the available sample size of the Run 2 dataset. As part of the High Luminosity (HL) LHC project, the ATLAS detector will be upgraded, and is expected to record an order of magnitude more data until the 2030s. Thus, this thesis also presents an extrapolation of the expected sensitivity of future analyses that might consider up to $3000\,\mathrm{fb}^{-1}$ of data.

The safety of the ATLAS detector is essential not only during its operation, but especially during upgrade periods. Therefore, improvements to the alarm management of the Detector Safety System (DSS) are discussed, too. This thesis presents the first statistical analysis of DSS alarms and the new Alarm Helper tool, which helps distinguish alarms stemming from error from alarms that result from interventions.

## Outline of the thesis and personal contributions

The research presented herein has been conducted in close collaboration with colleagues within the ATLAS Collaboration. The structure of this thesis is outlined below. In each chapter, the author's individual contributions are discussed with greater detail.

The theoretical foundations of searches for Higgs boson pair production, including the SM and the Higgs mechanism, and the simulation of proton proton collisions are summarised in Chapter 2.

Chapter 3 provides an overview of Higgs boson physics at the LHC. This includes introducing effective Higgs boson coupling modifiers, and describing how non-SM couplings alter the phenomenology, as well as a summary of the current experimental status of $HH$ searches.

The ATLAS experiment and the LHC are introduced in Chapter 4. Special emphasis is placed on DSS operations.

Chapter 5 describes improvements to detector safety. It presents the first analysis of DSS alarms and the novel Alarm Helper tool, both of which were were devised by the author and made public in Reference [11].

The algorithms used by the ATLAS collaboration to reconstruct physics objects from raw detector signals are explained in Chapter 6.

Chapter 7 presents the search for $HH \to b\bar{b}\tau^+\tau^-$, published in Reference [12]. The author's contributions, as one of the search's main analysts, are highlighted, notably the derivation of systematic uncertainties, performing the statistical analysis, as well as research towards a verification measurement of the analysis strategy.

Chapter 8 describes the sensitivity estimate of the search in HL-LHC conditions, which is made public in Reference [13]. The author served as the principal analyst and studied the impact of advancements in theory, event generation, and detector performance.

Chapter 9 concludes this thesis and provides an outlook.

# 2 Theoretical foundations

This chapter describes the theoretical foundations of the Standard Model of particle physics (SM) and the simulation of proton-proton collisions at the LHC.

## 2.1 The Standard Model of particle physics

The SM is at present the most successful theoretical framework describing the fundamental particles of the visible universe and their interactions through fundamental forces with the exception of gravity. This theory has emerged from the close interplay between experimental observations and theoretical predictions, culminating in a Lorentz invariant, relativistic quantum field theory that describes reality through quantum fields. Excitations of the underlying quantum fields are described as particles, which are categorised by their quantum numbers. Interactions between particles arise from local symmetry requirements with respect to the three gauge groups

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \ , \tag{2.1}$$

which are discussed in the following. The groups provide a theoretical description of the electromagnetic, the weak and the strong nuclear forces, which are mediated between the particles of the SM by the gauge bosons associated with the gauge groups. Gravity is not described by the SM, but instead by Einstein's theory of general relativity on a macroscopic level. This section follows in large parts Reference [14], and makes use of the Einstein summation convention and Lorentz-Heaviside units, where $\hbar = c = 1$.

### 2.1.1 Particles and forces

An overview of all SM particles, their quantum numbers, and their masses is given in Figure 2.1. Particles are categorised by their spin into fermions with half-integer spin, and bosons with integer spin. Fermions are further divided into quarks that carry the colour charge and leptons that do not. Bosons are divided into gauge bosons with spin 1, and the Higgs boson with spin 0.

**Fermions** are matter particles and exist in three generations that differ by the mass of the particles. The exception are neutrinos that are massless in the SM. Anti-fermions have the exact opposite quantum charges, but the same mass as the corresponding fermions. All stable matter is exclusively made of the first generation, since particles of the second and third generation decay to lower generations after a characteristic lifetime.

**Gauge bosons** mediate the three forces and arise from symmetry requirements. The strong force is mediated by eight massless gluons. They couple to the colour charge which is carried by quarks and gluons. One massless photon mediates the electromagnetic force between electrically charged fermions. Two charged $W^{\pm}$ bosons, and one neutral $Z$ boson mediate the weak force. Their masses arise during electroweak symmetry breaking (EWSB), described in Subsection 2.1.5.

**The Higgs boson** arises from the need to include masses for the $W$ and $Z$ bosons as well as fermions in the theory without violating local gauge invariance.

### 2.1.2 Quantum electrodynamics

Quantum electrodynamics (QED) describes the electromagnetic interactions between particles. The QED Lagrangian consists of a Dirac term, describing the motion of free fermions, and a kinetic term, describing the dynamics of the electromagnetic field

$$\mathcal{L}_{\text{QED}} = \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{kin.}} = \sum_{f} \bar{\psi}_f (i\gamma_\mu D^\mu - m_f)\psi_f - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad , \tag{2.2}$$

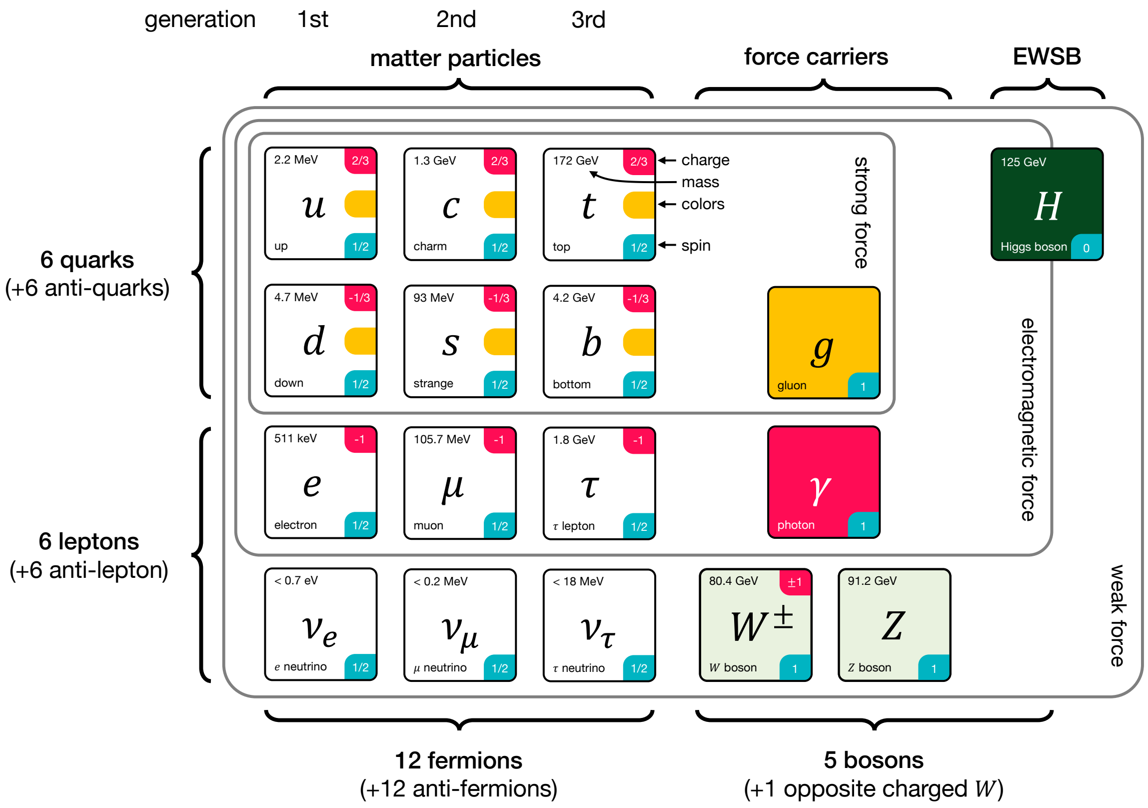


Figure 2.1: The SM particles. Adapted from Reference [15], values from Reference [16].

where $\psi_f$ is the Dirac spinor of a charged fermion $f$ with mass $m_f$ and the electric charge $q$. The covariant derivative is given by $D_\mu = \partial_\mu + iqA_\mu$, where $A_\mu$ is the photon field and the electromagnetic field strength tensor is $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The QED Lagrangian is invariant under local gauge transformations of the $\mathrm{U}(1)_{\mathrm{QED}}$ group, as $A_\mu$ transforms as $A_\mu \to A_\mu - \frac{1}{q}\partial_\mu \alpha$, where $\alpha$ is a local, space-time dependent phase. The associated conserved quantum number is the electric charge. The $\mathrm{U}(1)_{\mathrm{QED}}$ group is incorporated into the SM during EWSB, during which the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry is spontaneously broken to the $\mathrm{U}(1)_{\mathrm{QED}}$ subgroup (see Subsection 2.1.4).

### 2.1.3 Quantum chromodynamics

Quantum chromodynamics (QCD) describes the strong interaction between gluons and quarks through the non-Abelian symmetry group $\mathrm{SU}(3)_C$. Gluons and quarks carry the associated colour charge that exists as red, green and blue. Anti-quarks carry the respective anti-colours, and gluons carry a non-zero combination of colour and anti-colour. The QCD Lagrangian is given by

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{\psi}_q (i\gamma_\mu D^\mu - m_q)\psi_q - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \ , \tag{2.3}$$

where the quark spinors $\psi_q$ with mass $m_q$ are $\mathrm{SU}(3)_C$ triplets. The covariant derivative is given by $D_\mu = \partial_\mu + i g_s \frac{\lambda^a}{2} G^a_\mu$, where $g_s$ is the strong coupling constant, $\lambda^a$ are the eight Gell-Mann matrices, and $G^a_\mu$ are the eight gluon fields. Instead of $g_s$, often $\alpha_s = \frac{g_s}{4\pi}$ is used. The field strength tensors $G^a_{\mu\nu}$ are defined as $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu$, where $f^{abc}$ are the $\mathrm{SU}(3)_C$ structure constants. Since the $\mathrm{SU}(3)_C$ group is non-Abelian, the $f^{abc}$ are non-zero, which leads to triple or quartic self-interactions between gluons.

These self-interactions lead to "colour confinement" of gluons and quarks as they can only exist in colour neutral states. Therefore, gluons do not propagate over macroscopic distances and quarks hadronise into colour neutral bound states, such as mesons (colour- anti-colour pairs) or baryons (red-green-blue triplets). Furthermore, the value of coupling constant $g_s$ depends strongly on the transferred momentum $Q$ of the interaction. For large $Q$, $g_s$ decreases significantly, leading to the phenomenon of "asymptotic freedom" and allowing perturbative calculations for $Q > O(1\,\mathrm{GeV})$. Infinities arising from higher-order corrections can be absorbed and renormalised by introducing an arbitrary renormalisation scale $\mu_{\mathrm{R}}$.

### 2.1.4 Electroweak unification

The electromagnetic and weak interactions are described jointly by the symmetry group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$.

Given the experimental observation that $W$ bosons only couple to left-handed fermions, the fermion fields $\psi$ are split into left- and right-handed components ($\psi_L$, $\psi_R$), and the symmetry group describing the weak charged interaction is labelled $\mathrm{SU}(2)_L$. Weak interactions conserve the component of the isospin $I$ around an arbitrary axis $I_3$. The symmetry group $U(1)_Y$ is added to incorporate the observed interactions of $Z$ bosons with both left- and right-handed particles, as well as also to describe electromagnetic interactions. Two electrically neutral and one charged field $W_\mu^1, W_\mu^2, W_\mu^3$ are introduced to ensure local gauge invariance with respect to the $\mathrm{SU}(2)_L$ symmetry group, and one $B_\mu$ field with respect to the $\mathrm{U}(1)_Y$ group. The associated hypercharge $Y$ is related to the electrical charge $Q$ by the Gell-Mann-Nishijima relation $Q = I_3 + \frac{Y}{2}$. The interaction strengths are expressed by the parameters $g$ and $g'$, respectively. The Glashow-Salam-Weinberg (GSW) model [17–19] relates the physical photon field $A_\mu$ and the $Z$ boson field $Z_\mu$ by mixing the $W_\mu^3$ and $B_\mu$ fields as they carry the same quantum charges. This mixing is parametrised by the weak mixing angle $\theta_W = \arctan(g'/g)$. The non-Abelian nature of the $\mathrm{SU}(2)_L$ group leads to triple and quartic interactions of the $W$ and $Z$ bosons.

### 2.1.5 The Higgs mechanism and particle masses

The combined description of the electroweak and strong interactions by the $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group cannot provide a theoretical description of the experimentally observed masses of the fermions and $W$ and $Z$ bosons without breaking the principle of local gauge invariance. The Brout-Englert-Higgs-mechanism, or EWSB [3–5, 8], introduces mass terms for $W$ and $Z$ bosons while preserving local gauge invariance.

**Brout-Englert-Higgs mechanism** The SM Lagrangian is extended by a kinematic and a potential term

$$\mathcal{L}_{\text{Higgs}} = \mathcal{L}_{\text{kin.}} - \mathcal{L}_{\text{potential}} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \ , \tag{2.4}$$

where $\phi$ is a complex scalar $\mathrm{SU}(2)_L$ doublet field with hypercharge $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad . \tag{2.5}$$

The electroweak covariant derivative is given by $D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu + ig\frac{\sigma^a}{2}W_\mu^a$.

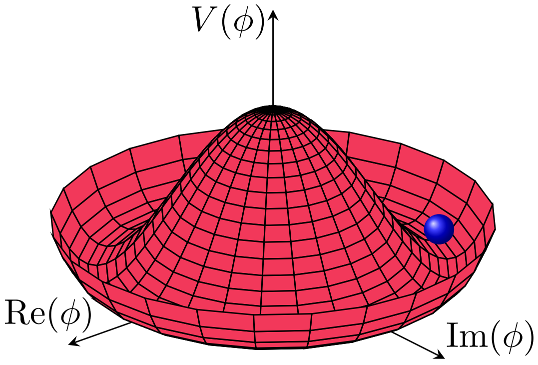


Figure 2.2: Illustration of the Higgs potential $V(\phi)$ as a function of the complex field $\phi$. The ball indicates the minimum. Adapted from Reference [20].

The Higgs potential $V(\phi)$ has two free parameters $\mu$ and $\lambda$ and is defined as

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2 \ . \tag{2.6}$$

The parameters are chosen such that a minimum exists for a non-zero value of $\phi$. This is the case for $\mu^2 < 0$ and $\lambda > 0$. An illustration of such a potential is given in Figure 2.2. Without loss of generality, one minimum is chosen at $\phi_1 = \phi_2 = \phi_4 = 0$, and $\phi_3 = \nu = \sqrt{-2\mu^2/\lambda}$. The chosen minium corresponds to the vacuum expectation value of the Higgs field $\nu$. This choice of a specific vacuum breaks the symmetry of the electroweak SM Lagrangian. As the vacuum is electrically neutral, it leaves the $\mathrm{U(1)_{QED}}$ symmetry unbroken, which results in a photon mass of zero.

Expanding the Higgs doublet $\phi$ around its vacuum expectation value $\nu$, with small fluctuations $\{H, \theta_1, \theta_2, \theta_4\}$ introduces three unphysical Goldstone fields $\theta_i$. The Goldstone fields can be absorbed into an overall phase factor, which in turn can be removed by adopting the unitary gauge

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1(x) + i\theta_2(x) \\ v + H(x) + i\theta_4(x) \end{pmatrix} \approx \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} e^{i\frac{\sigma^j}{2}\frac{\theta_j(x)}{v}} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \ . \tag{2.7}$$

Inserting the simplified expression for $\phi$ back into Equation 2.6, and ignoring all terms that do not rely on $H$ or are linear in $H$ gives

$$V = \lambda\nu^2 H^2 + \lambda\nu H^3 + \frac{\lambda}{4} H^4 \ . \tag{2.8}$$

The first term corresponds to a mass term for the scalar Higgs field $H$ with $m_H = \sqrt{2\lambda\nu^2}$. The second and third terms are the tri-linear and the quartic self-interactions with the coupling strength $\lambda$ defined as

$$\lambda = \frac{m_H^2}{2\nu^2} \ . \tag{2.9}$$

The kinematic term of the Higgs Lagrangian generates mass terms for the gauge bosons once the Higgs field acquires its vacuum expectation value by the transition into the minimum

$$\mathcal{L}_{\text{kin.}} = \frac{1}{2}\partial_\mu H\,\partial^\mu H + \frac{g^2}{4}(v+H)^2 W_\mu^+ W^{-\,\mu} + \frac{g^2}{8\cos(\theta_W)^2}(v+H)^2 Z_\mu Z^\mu + 0\,A_\mu A^\mu \ , \tag{2.10}$$

where $\theta_W$ is the weak mixing angle. From this, the masses of the $W^\pm$ and $Z$ boson, as well as the photon are obtained as

$$m_W = \frac{vg}{2} \ , \quad m_Z = \frac{m_W}{\cos(\theta_W)} \ , \text{and} \quad m_\gamma = 0 \ . \tag{2.11}$$

The physical fields $W_\mu^\pm$, $Z_\mu$, and the photon field $A_\mu$ are defined by the linear combination of the gauge fields $W_\mu$ and $Z_\mu$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad \text{and} \quad \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} . \tag{2.12}$$

The excitation of the physical Higgs field $H$ is referred to as the Higgs boson. By convention, the two degrees of freedom of the potential ($\mu$, $\lambda$) are typically expressed in terms of the Higgs boson mass $m_H$ and the vacuum expectation value $\nu$.

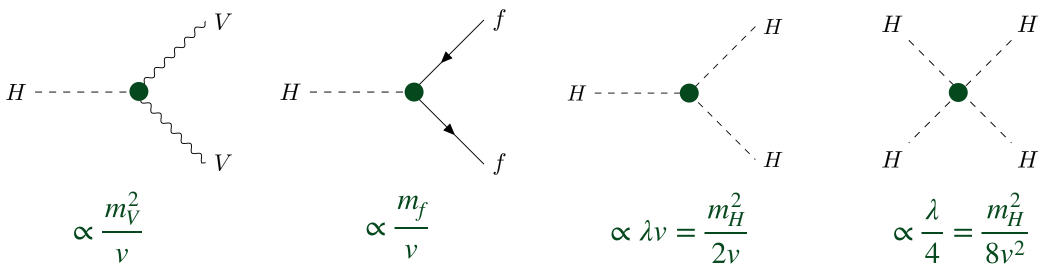


Figure 2.3: Fundamental vertices of the SM involving Higgs bosons.

**Fermion masses** The Higgs field $H$ also enables the theoretical description of fermion masses by adding a "Yukawa term" to the SM Lagrangian

$$\mathcal{L}_{\text{Yukawa}} = \sum_f -Y_f \frac{\nu}{\sqrt{2}} \bar{\psi}_f \psi_f - Y_f \frac{1}{\sqrt{2}} \bar{\psi}_f \psi_f H \ , \tag{2.13}$$

where $\psi_f$ are the fermion mass eigenstate fields, $\nu$ is the Higgs vacuum expectation value, and $Y_f$ is the Yukawa coupling of fermion $f$. The latter can be parametrized in terms of the fermion mass $Y_f = \sqrt{2} m_f / \nu$.

In the quark sector, the difference between the eigenstates in the weak interaction basis and the mass basis is accounted for by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [21, 22]. The CKM matrix empirically describes the quark mixing, and can be parametrised by three mixing angles and one complex phase that allows for Charge-Parity (CP) violation. The matrix is approximately diagonal, meaning that quark mixing primarily occurs within the same generation.

In summary, the SM Lagrangian contains four types of couplings of the Higgs boson to other SM particles and itself. The four diagrams and the proportionality of the coupling to the mass of the involved particles are given in Figure 2.3.

The SM has 19 parameters that cannot be calculated from first principles but must be determined from observations. One possible parametrisation is given by the 3 coupling strengths of the strong and electroweak interactions ($g_s$, $g$, $g'$), the 9 mass parameters of the quarks and charged leptons, the 3 mixing angles and 1 complex phase of the CKM matrix, the 2 free parameters of the Higgs potential ($\mu$, $\lambda$), and 1 parameter related to CP violation in strong interactions.

### 2.1.6 Shortcomings of the SM

The SM provides a well-tested theoretical description of the smallest constituents of the universe and their interactions. The theory has successfully explained all recorded collider data thus far, and has also made multiple predictions for particles that were subsequently discovered, such as the $W$ and $Z$ bosons [23] and the top quark [24, 25]. Most recently, the Higgs boson was discovered by the ATLAS and CMS collaborations at the CERN Large Hadron Collider (LHC) in 2012 [1, 2]. However, despite its impressive success, the SM shows substantial shortcomings, partially stemming from theoretical issues, but also from inexplicable observations.

**Theoretical issues** Several theoretical issues are linked to the Higgs mechanism itself. For example, the SM provides no explanation for the shape of the potential or the values of $\mu$ and $\lambda$. Any even polynomial fulfils the symmetry requirements, and any $\lambda > 0$ would provide a stable vacuum. Furthermore, the Higgs boson is affected by virtual corrections from fermion loops by $\Delta m_H \sim \Delta m_f^2$. If the Higgs boson couples to heavy non-SM particles, this would break the principle of separation of scales. Other theoretical shortcomings include the lacking description of gravity within the SM, as well as a lack of understanding of the relative weakness of the gravitational interaction of elementary particles with respect to the other three fundamental forces and whether the SM forces can be unified to a larger symmetry. Likewise, it remains unclear why there are three generations of leptons and three generations of quarks, as well as the origin of the mass values. Moreover, no explanation is given for the difference in CP violation between weak and strong interactions.

**Experimental issues** Cosmological observations show that the SM can only describe approximately 15 % of the known matter content of the universe [6]. Furthermore, the SM cannot explain the phenomena of dark matter, a form of matter that does not interact electromagnetically or strongly. Another experimental issues is, for instance, that the SM provides no explanation for the observed imbalance between matter and anti-matter in the universe due to a lack of a viable baryogenesis mechanism [6]. Furthermore, neither the observed neutrino masses [26] or the observed accelerated expansion of the universe [27, 28] are described by the SM.

## 2.2 Simulation of proton-proton collisions

The theoretical description and simulation of proton-proton ($pp$) collisions at the LHC is very complex as it requires a good understanding of protons, the scattering of their constituents at large momenta transfers, and the behaviour of the proton remnants. This section describes the simulation of LHC $pp$ collisions with Monte Carlo (MC) generators.

The proton is a composite particle and is comprised of three valence quarks in a bound state and a sea of virtual gluons and quarks. The strong interactions among these constituents account for most of the proton's mass of approximately 1 GeV. At low energies, the strong interaction is non-perturbative and cannot be treated analytically. However, at very high energies, or equivalently, at large momentum transfers, asymptotic freedom of QCD allows a perturbative, parton-level description of proton scattering to be applied reliably.

### 2.2.1 Hard scatter cross-section

The scattering with the largest momentum transfer during the crossing of the LHC beams is referred to as the hard scatter of the collision. The scattering of two protons at the LHC can be factorised into the probability of finding two partons $i$ and $j$ in the colliding protons, the hard scatter interaction between those quasi-free partons, and the fragmentation and hadronisation of the final state partons $f$. By splitting the interaction at the factorisation scale $\mu_\mathrm{F}$ into hard and soft components, the cross-section is given by

$$\sigma_{pp\to f} = \sum_{i,j} \int dx_1 \int dx_2 \; \mathrm{PDF}_i(x_1, \mu_F^2)\; \mathrm{PDF}_j(x_2, \mu_F^2)\; \hat{\sigma}_{ij\to f}(x_1, x_2, \alpha_S(\mu_\mathrm{R}^2), \frac{Q^2}{\mu_\mathrm{F}^2}, \frac{Q^2}{\mu_R^2}) \; . \tag{2.14}$$

Partons emitted below $\mu_\mathrm{F}$ are described by parton distribution functions (PDFs) and partons emitted above $\mu_\mathrm{F}$ by the hard scatter. At leading order, the $\mathrm{PDF}_i(x_i, \mu_\mathrm{F}^2)$ describes the probability of parton $i$ to carry the fraction $x_i$ of total proton momentum along the beam axis. The PDFs

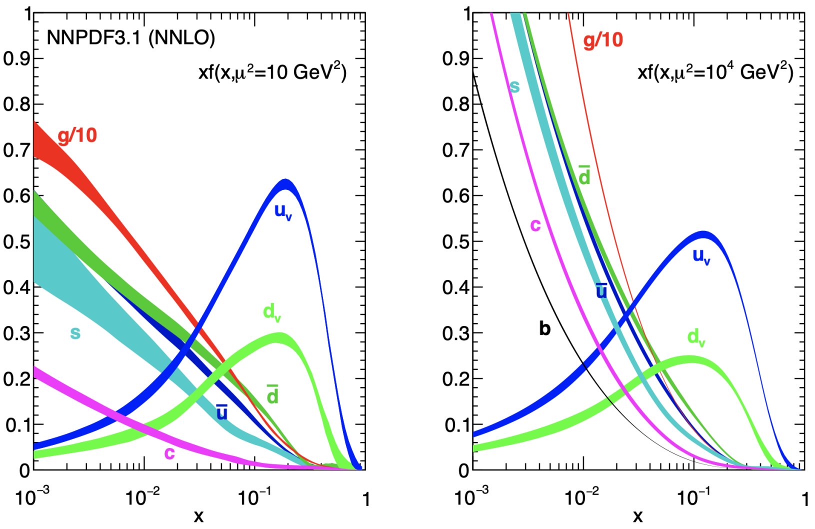


Figure 2.4: Parton distribution functions as a function of $x$ for two different $\mu_{\mathrm{F}}^2$ scales as determined by NNPDF collaboration. Taken from Reference [29].

are universal and do not depend on the final state, but only on $\mu_{\mathrm{F}}$. Moreover, they cannot be calculated from first principles, and are thus typically measured in deep inelastic scattering experiments. The PDFs shown in Figure 2.4 demonstrate that for low values of $x$, the PDFs are greatly enhanced for gluons. The processes studied by the $HH \to b\bar{b}\tau^+\tau^-$ analyses typically occur at an energy scale of $Q^2 \approx \mathcal{O}((100\,\mathrm{GeV})^2)$. With a beam energy of 6.5 TeV, this is reached for partons carrying $x \approx 10^{-2}$ of the proton's momentum along the beam axis. Thus, production occurs predominantly via gluons.

The hard scatter cross-section $\hat{\sigma}_{ij}$ can be calculated from the matrix element (ME). The ME can be calculated perturbatively in QCD up to fixed order using Feynman rules and considering the available final state phase space. The renormalisation scale $\mu_{\mathrm{R}}$ is used to treat non-physical infinities in fixed beyond leading order calculations, and determines the value of the strong coupling constant $\alpha_S$. The cross-section also depends on $\mu_{\mathrm{F}}$ as it is related to the momentum transfer of the hard scatter. In general, the cross-section dependence on the choice of $\mu_{\mathrm{F}}$ and $\mu_{\mathrm{R}}$ decreases as more orders in perturbation theory are considered.

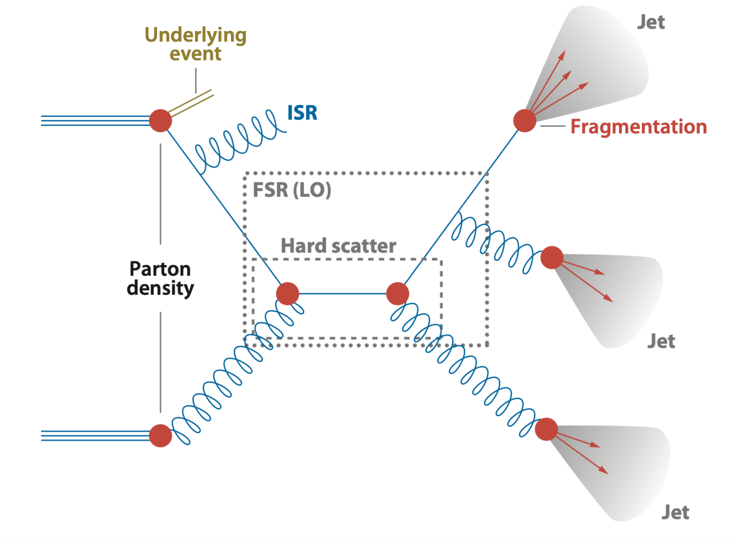


Figure 2.5: Sketch of a proton-proton collision, illustrating some of the physics processes. Adapted from References [30, 31].

### 2.2.2 Event generation

An “event” refers to the collision of two proton bunches that make up the LHC beams. The simulation of an event involves many different processes that occur at a wide range of energy scales and cannot all be described with perturbative methods. A schematic illustration of a *pp* collision is shown in Figure 2.5. Two partons scatter with a large momentum transfer. Depending on the perturbative order of the calculation, additional real and virtual emissions are accounted for using parton shower (PS) algorithms. PS algorithms simulate the initial and final state radiation (ISR, FSR) in a perturbative formalism until the emissions reach the soft limit of about 1 GeV, where perturbation theory breaks down. At this point, the partons fragment into colour-neutral hadrons, which then form jets. This process is described by phenomenological hadronisation models. The simulation also accounts for the interaction with the proton remnants, known as the underlying event, as well as “pileup”, which refers to softer interactions occurring in the same or nearby bunch crossings.

The number of events $N$ for a specific process with cross-section $\sigma$ depends on the instantaneous luminosity $\mathcal{L}_{\text{inst}}$ integrated over the run time $t$

$$N = \sigma \times \int dt \; \mathcal{L}_{\text{inst}} \; . \tag{2.15}$$

The instantaneous luminosity is determined by properties of the collider and the beam.

Due to the complexity of the processes and the stochastic nature of phenomena such as fragmentation, the processes are simulated using MC generators. Differences across algorithms and models, and in the assumptions and input parameters, are evaluated as systematic uncertainties in the statistical interpretation. These are described in Subsection 7.11.2.

The comparison of simulated and recorded collisions also requires the simulation of the detector response. A detailed model of the ATLAS detector is implemented in Geant4 [32, 33]. It simulates the interaction of final state particles from the hard scatter generator ("truth-level" particles) with the detector material, such as their respective energy deposits and their corresponding analogue and digital signals. The detector response is simulated up to the level of the readout electronics. From then onwards, the same processing steps are used as for the data. The computationally most expensive step is the simulation of the ATLAS electromagnetic calorimeter due to its complicated, accordion-shaped geometry. Systematic uncertainties are partially derived using an alternative, faster simulation, which replaces this step with a parametrised response model [34].

# 3 Higgs boson physics

This chapter describes the current state-of-the-art theoretical and experimental knowledge on the Higgs boson and in particular the Higgs boson self-interaction. First, the production and decay of Higgs bosons and Higgs boson pairs at the LHC are discussed. Then, the indirect measurement of contributions from beyond the Standard Model contributions to Higgs boson pair production in terms of effective coupling modifiers are explained. Finally, the experimental status at the time of writing this thesis is summarized.

## 3.1 The Higgs boson

The Higgs boson was discovered in 2012 by the ATLAS and CMS collaborations [1, 2], marking the beginning of an extensive research programme into its properties and couplings to other elementary particles, including itself. Although the Higgs boson is related to many inexplicable observations and shortcomings of the Standard Model (SM), all properties measured so far are consistent with SM predictions. This poses many questions on the nature of the Higgs mechanism and the boson itself [35, 36]. For example, the SM Higgs mechanism is based on the mathematically simplest potential that can explain boson and fermion masses without violating local gauge invariance. However, alternative potentials are compatible with current data and some could potentially explain phenomena such as the observed baryon asymmetry or the "hierarchy problem" [37], which might, for example, be explained by a composite Higgs boson. As the shape of the Higgs potential is directly linked to the Higgs boson self-interaction, its measurement is key to improve our understanding of our universe. Therefore, the experimental study of the Higgs boson, particular its self-interaction, remains a main priority of modern particle physics.

## 3.2 Production and decay of Higgs bosons at the LHC

The Higgs boson couples to all SM particles that have a non-zero mass. The couplings to fermions, gauge bosons and itself arise from different parts of the SM Lagrangian. In general, the interaction strength depends on the mass of the particles. For fermions, the strength is proportional to the fermion mass ($\propto m_f$), for massive vector-bosons it is proportional to the squared mass of the vector boson ($\propto {m_V}^2$). As gluons and photons are massless, the Higgs boson can only couple to them indirectly via loops that either involve heavy fermions or massive bosons.

The Higgs boson self-interaction vertices are proportional to the parameter $\lambda$, which is related to the Higgs boson mass $m_H$ and the vacuum expectation value $\nu$ as given in Equation 2.9. The Higgs boson mass has been measured by ATLAS to be $m_H = 125.09 \pm 0.11\,\text{GeV}$ [38], and the vacuum expectation value has also experimentally measured with sub-percent precision [39]. Therefore, the SM coupling strength is predicted to be $\lambda_{\text{SM}} = \frac{1}{2}(\frac{m_H}{\nu})^2 \approx \frac{1}{2}(\frac{125\,\text{GeV}}{246\,\text{GeV}})^2 \approx \frac{1}{8}$.

Once the Higgs boson mass has been measured, all free parameters of the Higgs sector are known and thus production cross-sections and decay branching ratios can be calculated. In the SM the Higgs boson total decay width for $m_H = 125\,\text{GeV}$ is predicted to be $\Gamma_H^{\text{total}} \approx 4\,\text{MeV}$ with a proper lifetime of $10^{-22}\,\text{s}$ [40]. Since the decay width is more than five orders of magnitude smaller than the mass, the Narrow Width Approximation allows the factorisation of the cross-section $\sigma$ of any process $pp \to H \to X$ into the production cross-section and the decay branching fraction $\mathcal{B}$ as

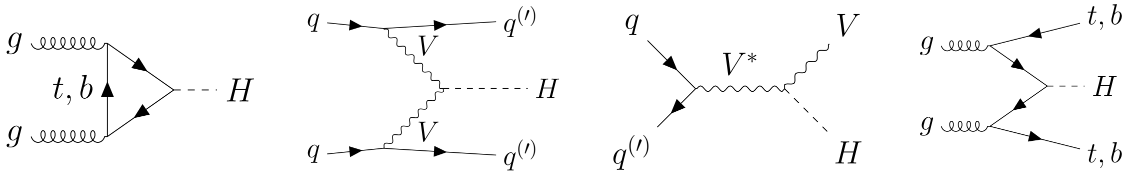
$$\sigma(pp \to H \to X) = \sigma(pp \to H) \times \mathcal{B}(H \to X)\ . \tag{3.1}$$

The branching fraction $\mathcal{B}(H \to X)$ is defined as the ratio of the partial width $\Gamma_H^X$ of the decay $H \to X$ and the total Higgs boson decay width $\Gamma_H^{\text{total}}$, i.e. the sum over all partial widths

$$\mathcal{B}(H \to X) = \frac{\Gamma_H^X}{\sum_i \Gamma_H^i} = \frac{\Gamma_H^X}{\Gamma_H^{\text{total}}}\ . \tag{3.2}$$

Henceforth, the production and decay of Higgs bosons and Higgs boson pairs can be treated separately.

### 3.2.1 Single Higgs boson production and decay



(a) ggF $H$ (b) VBF $H$ (c) $VH$ (d) $t\bar{t}H$

Figure 3.1: Feynman diagrams for different single Higgs production mechanisms.

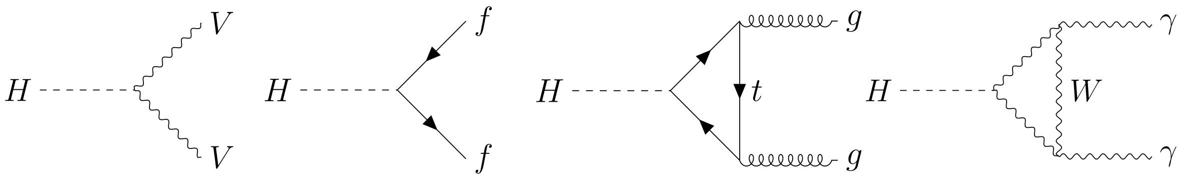
Single Higgs boson ($H$) production occurs via multiple production modes in proton-proton ($pp$) collisions. Example Feynman diagrams are shown in Figure 3.1 and the production cross-sections are given in Table 3.1.

The probability of producing a Higgs boson in a $pp$ collision is largest if the transferred momenta is approximately equal to the Higgs boson mass ($\sqrt{\hat{s}} \sim m_H$). At the LHC, the gluon PDFs are enhanced for small fractional momenta $x$ of the proton. Therefore, and because of the large Yukawa top-quark coupling, Higgs bosons are mostly produced via gluon fusion (ggF $H$), where two gluons fuse via a quark loop to a Higgs boson. The loop is dominated by top-quarks contributions as the coupling to fermions is proportional to the fermion mass as explained above.

The second most common production mode is vector-boson fusion (VBF $H$). Two quarks scatter and each radiate one vector-boson which then fuse to produce a Higgs boson. The VBF Higgs production cross-section is about ten times lower than for ggF. The production of a Higgs boson

| | **Production cross-section [pb]** | | | | | |
|---|---|---|---|---|---|---|
| $\sqrt{s}$ [TeV] | ggF $H$ | VBF $H$ | $WH$ | $ZH$ | $t\bar{t}H$ | |
| 13.0 | 48.6 | 3.78 | 1.37 | 0.88 | 0.51 | Calculation [40] |
| 13.6 | 52.2 | 4.08 | 1.46 | 0.94 | 0.57 | Interpolation [41] |
| 14.0 | 54.7 | 4.28 | 1.51 | 0.99 | 0.61 | Calculation [40] |

Table 3.1: Higgs boson production cross-sections assuming $m_H = 125\,\text{GeV}$.



(a) $H \to VV^*$ (b) $H \to ff$ (c) $H \to gg$ (d) $H \to \gamma\gamma$

Figure 3.2: Example Feynman diagrams for different Higgs boson decays.

in association with a heavy vector boson ($VH$) is the third largest production mode. In the main diagram, a Higgs boson is radiated off a vector boson. The fourth most likely production mode is the production of a Higgs boson together with quark- anti-quark pair, most often a top-quark pair ($t\bar{t}H$). The cross-section is more than hundred times smaller than of ggF. Additional production modes exist but they are omitted in the context of this thesis because of their even smaller cross-sections.

The Higgs boson decays directly into pairs of fermions or massive bosons. The respective Feynman diagrams are shown in Figure 3.2. Even though the Higgs boson coupling to other particles depends on their mass, the branching fractions for the decays to vector bosons ($H \to WW^*$ and $H \to ZZ^*$) are significantly reduced despite the large coupling strength ($\propto {m_V}^2$), because the Higgs boson mass is less than twice the vector boson mass ($m_H < 2m_V$). Therefore, a decay is only possible if one of the two vector bosons is off-shell. Similarly, the decay to a pair of top-quarks is maximally suppressed as the combined mass is significantly heavier than the Higgs boson ($m_H \lll 2m_t$), and the coupling is weaker than for vector bosons ($\propto m_f$). As a result, the majority of Higgs bosons decay into a pair of $b$-quark with a branching fraction of 58 %. The Higgs boson can also decay into a pair of gluons via a quark loop, or a pair of photons either via a loop consisting of either quarks or $W$ bosons. The calculated branching fractions assuming $m_H = 125\,\text{GeV}$ are listed in Table 3.2.

| $H \to X$ | $b\bar{b}$ | $WW^*$ | $gg$ | $\tau^+\tau^-$ | $c\bar{c}$ | $ZZ^*$ | $\gamma\gamma$ | $Z\gamma$ | $\mu^+\mu^-$ | ... |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{B}$ [%] | 58 | 21 | 8.2 | 6.3 | 2.9 | 2.6 | 0.23 | 0.15 | 0.022 | ... |

Table 3.2: Higgs boson branching fractions assuming $m_H = 125\,\text{GeV}$. Values taken from Reference [40].

(a) Triangle diagram.

(b) Box diagram.

Figure 3.3: Leading order diagrams for ggF $HH$ production.

### 3.2.2 Higgs boson pair production and decay

At the LHC, Higgs boson pairs ($HH$) are predominantly produced via gluon fusion (ggF $HH$). At leading order, the two diagrams depicted in Figure 3.3 contribute. The first diagram is referred to as the "triangle diagram". It contains one gluon fusion vertex that is proportional to the top-quark Yukawa coupling $Y_t$, and one that is proportional to $\lambda$, at which two Higgs bosons are emitted from one Higgs boson, and allows the measurement of the Higgs boson self-interaction strength. The second diagram, referred to as the "box-diagram", does not contain a self-interaction vertex. Instead, the two Higgs bosons are produced individually from a quark loop. The interference of the two diagrams approximately halves the total ggF $HH$ production cross-section to about 31 fb, which is about one thousand times smaller than the ggF $H$ cross-section [42].

Vector-boson fusion (VBF) $HH$ production has the second largest cross-section, which is one order of magnitude smaller than the ggF $HH$ cross-section [43, 44]. The three leading order diagrams, depicted in Figure 3.4, interfere with each other. The first diagram is not only sensitive to $\lambda$, but also to the coupling of the Higgs boson to vector bosons ($g_{HVV}$). The second diagram is sensitive to the coupling of two Higgs bosons to two vector bosons ($g_{HHVV}$). The third diagram features two vertices that are sensitive to the Higgs boson coupling to two vector bosons ($g_{HVV}$). In the SM, $g_{HHVV}$ and $g_{HVV}$ are related via $g_{HHVV} = g_{HVV}/(2\nu)$.

Table 3.3 lists the calculated ggF and VBF $HH$ cross-sections for different values of the centre-of-mass energy $\sqrt{s}$. Additional production modes are not considered in this thesis given their smaller cross-sections. The same applies to triple Higgs boson production ($HHH$) given its small

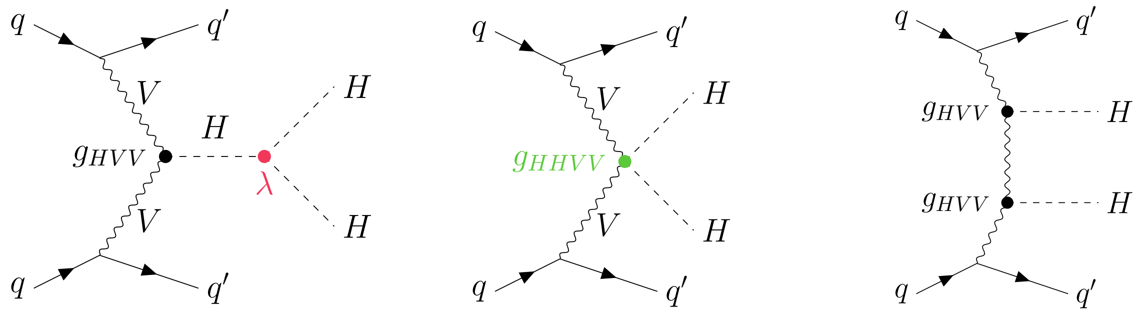


Figure 3.4: Leading order Feynman diagrams of VBF $HH$ production.

cross-section of about 0.08 fb [45]. At higher order in perturbation theory, diagrams of single Higgs boson production and decays also feature self-interaction vertices.

In $HH$ production, each Higgs boson decays individually with the single Higgs boson branching fractions as listed in Table 3.2. The combined $HH$ branching fraction is calculated as

$$\mathcal{B}(HH \to XX) = \mathcal{B}(H \to X) \times \mathcal{B}(H \to X) \ , \tag{3.3}$$

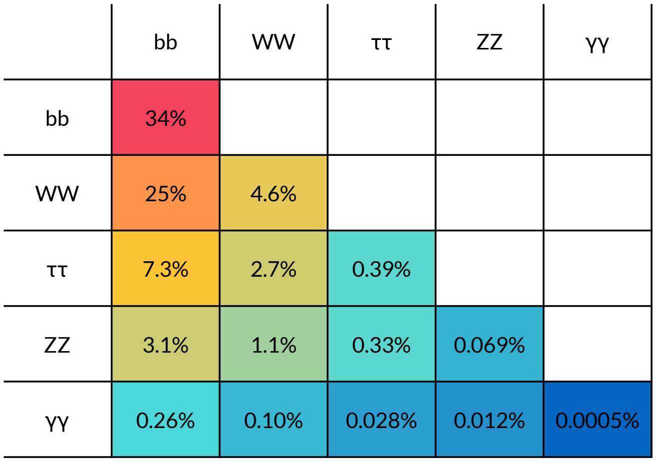
if both decay to the same final state $X$, else it is given by

$$\mathcal{B}(HH \to XY) = 2 \times \mathcal{B}(H \to X) \times \mathcal{B}(H \to Y) \ . \tag{3.4}$$

Figure 3.5 illustrates the calculated $HH$ branching fractions for selected final states that are targeted by the ATLAS collaboration with the Run 2 dataset. About one third of Higgs boson pairs decay into two pairs of $b$-quarks, about one quarter decay into a pair $W$-bosons. Approximately one third of the $W$ bosons subsequently decay into a lepton and a neutrino, the rest decays hadronically.

| | **Production cross-section [fb]** | |
|---|---|---|
| $\sqrt{s}$ [TeV] | ggF $HH$ [42] | VBF $HH$ [43, 44] |
| 13.0 | 30.77 | 1.687 |
| 13.6 | 34.13 | 1.874 |
| 14.0 | 36.37 | 2.005 |

Table 3.3: Higgs boson pair production cross-sections assuming $m_H = 125\,\text{GeV}$.



| | bb | WW | ττ | ZZ | γγ |
|---|---|---|---|---|---|
| bb | 34% | | | | |
| WW | 25% | 4.6% | | | |
| ττ | 7.3% | 2.7% | 0.39% | | |
| ZZ | 3.1% | 1.1% | 0.33% | 0.069% | |
| γγ | 0.26% | 0.10% | 0.028% | 0.012% | 0.0005% |

Figure 3.5: Branching fractions for selected final states of the decay of a system of two Higgs bosons assuming $m_H = 125\,\text{GeV}$. Modified from Reference [46].

The focus of this thesis is the $HH \to b\bar{b}\tau^+\tau^-$ decay, where the $\tau$-leptons decay either hadronically or leptonically. The final state has a branching fraction of 7.3 %, which is smaller than that of the $b\bar{b}b\bar{b}$ final state but larger than that of $b\bar{b}\gamma\gamma$. Importantly, the final state benefits from a favourable signal-to-background ratio.

## 3.3 Effective Higgs boson coupling modifiers

Events with Higgs boson pairs are sensitive to indirect contributions from beyond the Standard Model phenomena (BSM), that can alter the self-interaction strength $\lambda$. BSM contributions can be expressed as an effective vertex, illustrated in Figure 3.6. The SM coupling strength $\lambda$ is multiplied by an effective coupling modifier $\kappa_\lambda$. Theoretical considerations limit the possible range of the coupling modifier to $\mid \kappa_\lambda \mid \lesssim 6$ [47].

Additional effective coupling modifiers are defined for other Higgs boson interactions, impacting both production and decay processes. The $\kappa$-framework has been developed during LHC Run 1 to combine modifications to the Higgs boson interaction across different analyses [48]. The SM interaction strength corresponds to the case in which all coupling modifiers are equal to unity ($\kappa_i = 1$). The $\kappa_\lambda$ interpretation assumes that any BSM particle's decay widths is sufficiently small to uphold the narrow width approximation and that unlike for the single Higgs boson

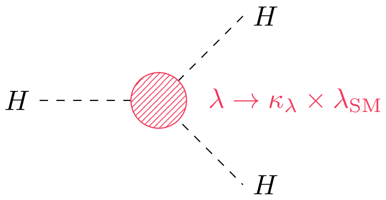


Figure 3.6: Effective vertex

$\kappa$-framework, the kinematics of the two Higgs bosons can be altered with respect to the SM. Furthermore, the coupling modifiers are real numbers and are independent of the decay width of the BSM particles.

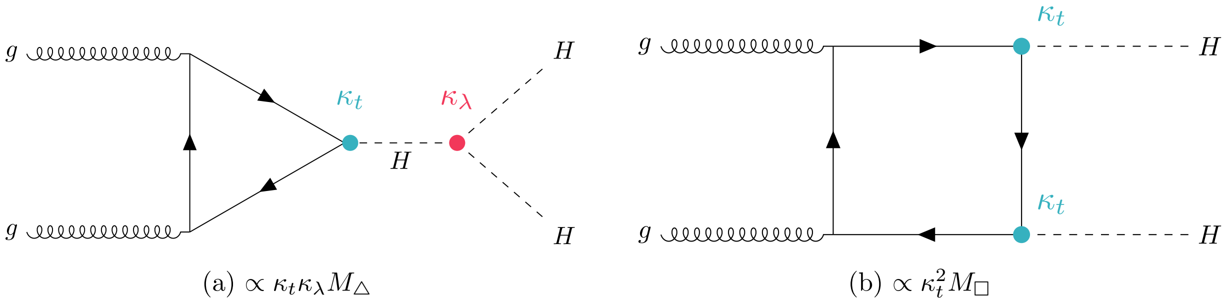
In this thesis the following effective coupling modifiers are considered

$$\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{\mathrm{SM}}$$
$$\kappa_{2V} = g_{HHVV}/g_{HHVV}^{\mathrm{SM}}$$
$$\kappa_V = g_{HVV}/g_{HVV}^{\mathrm{SM}}$$
$$\kappa_t = Y_t/Y_t^{\mathrm{SM}}$$

### 3.3.1 Phenomenology of *HH* production with anomalous couplings

Anomalous couplings alter both the cross-section and kinematics with respect to the SM predictions. In particular, the $HH$ cross-section and kinematics are especially sensitive to the value of $\kappa_\lambda$ as it changes the interference between the leading order diagrams for both ggF and VBF $HH$ production.

Figure 3.7 shows the affected vertices for ggF $HH$ production. The triangle diagram comprises of one vertex proportional to $\kappa_\lambda$, and one proportional to $\kappa_t$. The box diagram features two vertices proportional to $\kappa_t$. The total cross-section is proportional from the squared sum of all ggF $HH$ matrix elements, namely the matrix elements of the two diagrams elements $(M_\triangle, M_\square)$, multiplied



(a) $\propto \kappa_t \kappa_\lambda M_\triangle$ (b) $\propto \kappa_t^2 M_\square$

Figure 3.7: Leading order for Feynman diagrams for ggF $HH$ production. The effective coupling modifiers $\kappa_\lambda$ and $\kappa_t$ are drawn at the corresponding vertices.

by their corresponding effective coupling modifiers

$$\sigma_{\text{ggF } HH}(\kappa_\lambda) \propto | \; \kappa_t \kappa_\lambda M_\triangle + \kappa_t^2 M_\square \; |^2 \quad . \tag{3.5}$$

$\kappa_t$ is constrained to be compatible with unity from measurements of $t\bar{t}H$ production. Therefore, in the context of the $HH \to b\bar{b}\tau^+\tau^-$ analysis, $\kappa_t$ is fixed to unity, whereas $\kappa_t$ is left floating in global combinations of Higgs boson measurements. With $\kappa_t = 1$, the cross-section simplifies to

$$\begin{aligned}\sigma_{ggF\ HH}(\kappa_\lambda) &\propto | \; \kappa_\lambda M_\triangle + M_\square \; |^2 \\ &= \kappa_\lambda^2 \, |M_\triangle|^2 + \kappa_\lambda \, ( \, M_\triangle^* M_\square + M_\square^* M_\triangle \, ) \, + \, |M_\square|^2 \; .\end{aligned} \tag{3.6}$$

Figure 3.8 highlights the quadratic dependence of total cross-section on $\kappa_\lambda$, arising from the scaling of the triangle diagram. Contributions from the box diagram are constant and independent of $\kappa_\lambda$, while the interference term between the two diagrams scales linearly with $\kappa_\lambda$. For $\kappa_\lambda = 0$, there are no contributions from the triangle diagram and all Higgs boson pairs are produced via the box diagram. For $\kappa_\lambda \approx 2.4$, the two diagrams interfere maximally destructive and the cross-section reaches its minimum.

Anomalous $\kappa_\lambda$ values also affect the $HH$ kinematics as the three triangle, box and interference terms correspond to different $m_{HH}$ spectra, as shown in Figure 3.9. The impact of different values of $\kappa_\lambda$ on the $HH$ invariant mass distribution is illustrated in Figure 3.10. The triangle diagram exhibits a soft $m_{HH}$ spectrum that peaks at around twice the Higgs boson mass. Since the

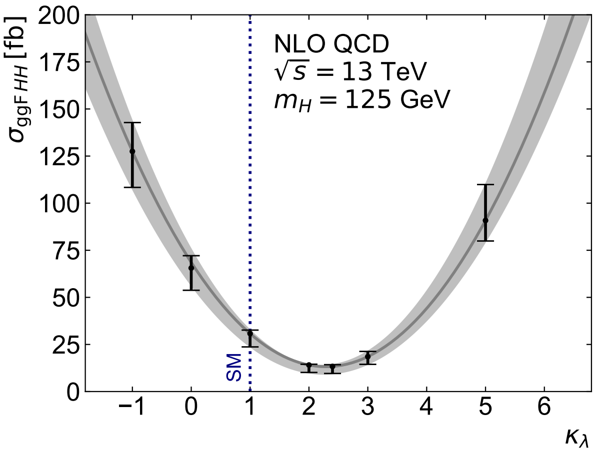


Figure 3.8: The ggF $HH$ cross-section for different values of $\kappa_\lambda$. The points indicate NLO in QCD calculated cross-sections. The error bars indicate the combined QCD scale and top-quark mass uncertainty. The grey line and band are obtained from a fit of the calculated values. Values and fit are taken from References [49, 50].

triangle contributions scale as $\kappa_\lambda^2$, the total $m_{HH}$ spectrum becomes soft for large absolute values of $\kappa_\lambda$. The box diagram has a harder $m_{HH}$ spectrum that peaks at around twice the top-quark mass, and a smaller peak at twice the Higgs boson mass. The box diagram contributions dominate for small values of $\kappa_\lambda$, particularly for $\kappa_\lambda = 0$. The interference depends on the value of $\kappa_\lambda$, and is maximal for $\kappa_\lambda \approx 2.4$ and leads to a minimum at around 350 GeV.

Similarly, VBF $HH$ production is also affected by anomalous couplings. The three leading order

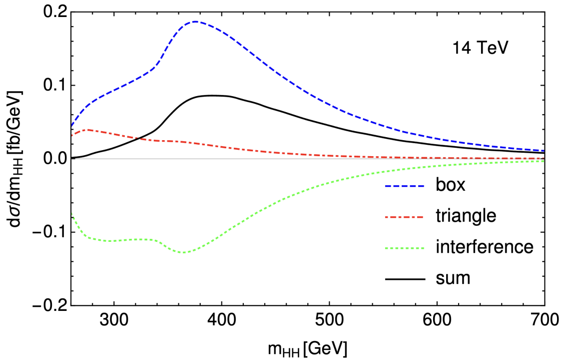


Figure 3.9: Differential $m_{HH}$ distributions for the two ggF $HH$ diagrams and their interference. Taken from Reference [47].

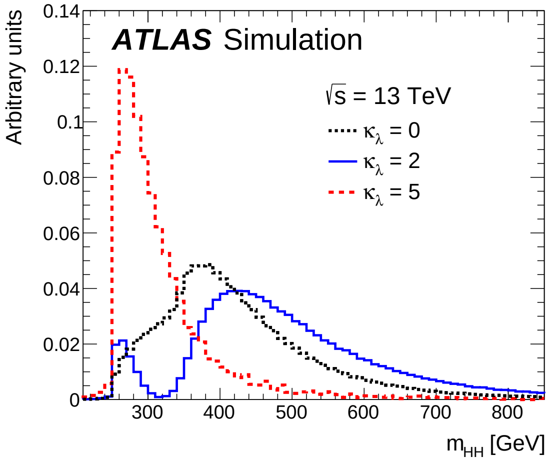
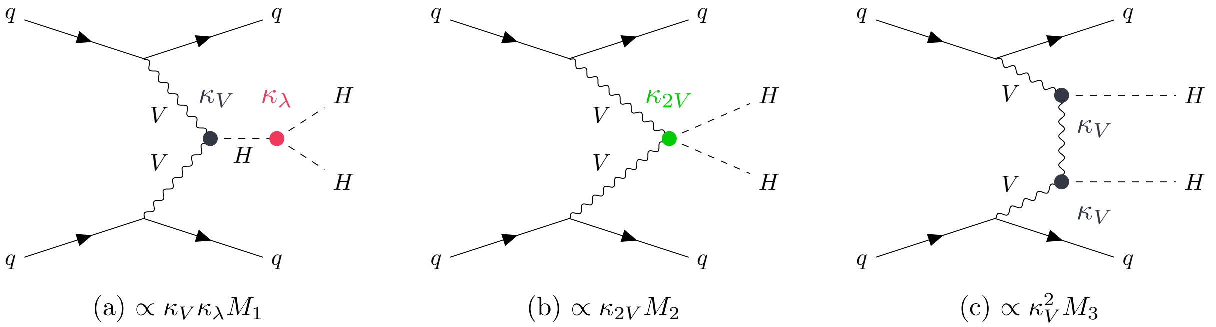


Figure 3.10: Resulting $m_{HH}$ spectra for various values of $\kappa_\lambda$. Taken from Reference [51].



(a) $\propto \kappa_V \kappa_\lambda M_1$ (b) $\propto \kappa_{2V} M_2$ (c) $\propto \kappa_V^2 M_3$

Figure 3.11: Feynman diagrams for VBF $HH$ production. The Figures indicate the vertices that are affected by the effective coupling modifiers $\kappa_\lambda$, $\kappa_{2V}$ and $\kappa_V$.

diagrams are shown in Figure 3.11. The interaction vertices can be modified by $\kappa_\lambda$, $\kappa_V$ and $\kappa_{2V}$. The $HH \to b\bar{b}\tau^+\tau^-$ analysis fixes $\kappa_V$ to unity since no significant deviations from unity have been observed yet. Similar to $\kappa_t$, $\kappa_V$ is included in global combinations, where it is primarily constrained from single Higgs boson measurements. With $\kappa_V = 1$, the VBF $HH$ cross-section depends quadratically on $\kappa_\lambda$ and $\kappa_{2V}$

$$\begin{aligned}\sigma_{\text{VBF } HH}(\kappa_\lambda, \kappa_{2V}) &\propto |\, \kappa_\lambda M_1 + \kappa_{2V} M_2 + M_3 \,|^2 \\ &= \kappa_\lambda^2 \, |M_1|^2 + \kappa_{2V}^2 \, |M_2|^2 + |M_3|^2 \\ &\quad + \kappa_\lambda \kappa_{2V} \, (\, M_1 M_2^* + M_1^* M_2 \,) \\ &\quad + \kappa_\lambda \, (\, M_1 M_3^* + M_1^* M_3 \,) + \kappa_{2V} \, (\, M_2 M_3^* + M_2^* M_3 \,) \; . \end{aligned} \tag{3.7}$$

The three interference terms depend linearly on $\kappa_\lambda$, $\kappa_{2V}$ and their combination. With $\kappa_{2V} = 1$, the minimum is reached for $\kappa_\lambda \approx 1.5$. The anomalous couplings also affect the kinematic distributions. The impact on the $m_{HH}$ spectrum is significantly smaller as the VBF $HH$ cross-section is approximately ten times smaller than the ggF $HH$ cross-section.

In addition, effective coupling modifiers also affect single Higgs boson production. For example, $\kappa_\lambda$ appears in higher order electroweak corrections to single Higgs boson production and decays. Representative diagrams are shown in Figure 3.12. Given the small impact on single Higgs boson production, the effect of $\kappa_\lambda$ is considered by the $HH \to b\bar{b}\tau^+\tau^-$ analysis only with respect to cross-sections, evaluated independently for each production mode. Any modification of the

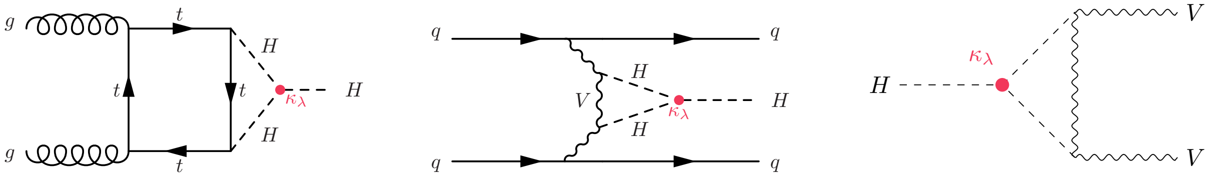


Figure 3.12: Higher order Feynman diagrams for single Higgs boson production and decay that are sensitive to the Higgs boson self-coupling $\kappa_\lambda$.

kinematics of single Higgs boson production are neglected. Furthermore, all other coupling modifiers are assumed to be unity for single Higgs boson production, including $\kappa_{2V}$.

## 3.4 Experimental status

The discovery of the Higgs boson started an extensive research programme into its properties and couplings with other SM particles and itself. The Higgs boson mass has been measured with a precision of 0.09 % to

$$m_H = 125.11 \pm 0.09\,(\text{stat.}) \pm 0.06\,(\text{syst.})\,\text{GeV} = 125.11 \pm 0.11\,\text{GeV} \tag{3.8}$$

by ATLAS [38]. The direct measurement of the Higgs boson width is limited by the detector resolution of a few GeV. The latest indirect measurement of the decay width by ATLAS sets an upper limit on the decay width at the 68 % confidence level (CL) of $\Gamma_H < 4.3^{+2.7}_{-1.9}\,\text{MeV}$ [52] from the ratio of off-shell to on-shell Higgs boson production. The Higgs boson has been measured to be consistent with a CP-even quantum state predicted by the SM, and alternative pure CP quantum state hypotheses have been ruled out at high confidence levels [53]. Most production and several decay modes have been observed. In particular, decays into bosons via the $WW^*$, $ZZ^*$ and $\gamma\gamma$, as well as decays into fermions through the $b\bar{b}$ and $\tau^+\tau^-$ channels, have been observed. There is evidence for the decay to $Z\gamma$. In addition to inclusive production cross-sections and decay branching fractions, also differential cross-sections have been measured [35]. Of particular interest is the measurement of effective coupling modifiers to probe BSM contributions. As of today, there is no evidence for significant deviations with respect to the SM predictions [54].

Thus far, there has been no evidence for Higgs boson pair production or the Higgs boson self-interaction. Current $HH$ searches target ggF and VBF production in a combination of final states to maximize the use of the available dataset. In addition, different final states allow an optimisation across various regions of phase space. Most analyses select final states where one Higgs boson decays into a pair of $b$-quarks taking advantage of the large branching fraction to increase the expected signal yield. The three most sensitive channels are discussed below:

**$b\bar{b}b\bar{b}$:** This channel has the largest branching fraction, but is limited by a large multi-jet background. The all-hadronic final state requires relatively high $p_{\mathrm{T}}$ trigger thresholds to not exceed bandwidth limitations. In addition, the reconstruction of the four $b$-jets originating from two Higgs bosons is ambiguous, leading to significant combinatorial backgrounds. Finally, simulating the multi-jet background is computationally intensive, so it is typically estimated using data-driven techniques.

**$b\bar{b}\tau^+\tau^-$:** While the requirement of two $\tau$-leptons in the final state reduces the branching fraction to 7.3 %, their distinct decay signatures allows the suppression of the multi-jet background. The main backgrounds are the production of $t\bar{t}$, Z + heavy flavour quarks, multi-jet and single Higgs events. The analysis targets $\tau$-lepton pairs that either both decay hadronically ($\tau_{\mathrm{had}}\tau_{\mathrm{had}}$), or one leptonically and the other hadronically ($\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$). Events where both $\tau$-leptons decay leptonically (about 10 %), are included in a different search that considers the $b\bar{b}\ell\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ final state.

**$b\bar{b}\gamma\gamma$:** This channel has a very small branching fraction of only 0.26 %. However, the two photons allows for low trigger thresholds, and provides an order of magnitude better mass resolution compared to the $b\bar{b}b\bar{b}$ and $b\bar{b}\tau^+\tau^-$ channels, enabling effective background rejection. The main backgrounds are $\gamma\gamma$ and single Higgs boson events.

An additional search targets final states where the two Higgs bosons decay fully into leptons.

The latest combination of the ATLAS Run 2 $HH$ searches [55] sets a 95 % confidence level (CL) upper limit on the $HH$ signal strength at 2.9 times the SM prediction, with an expected

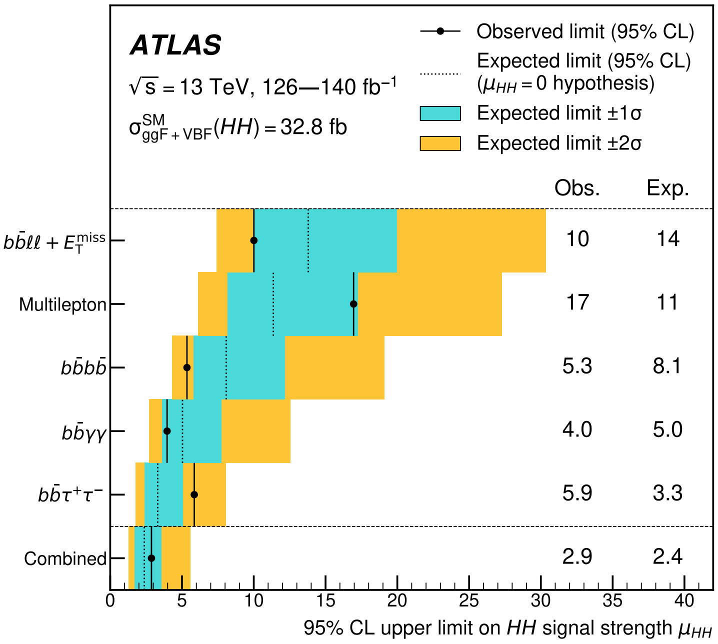
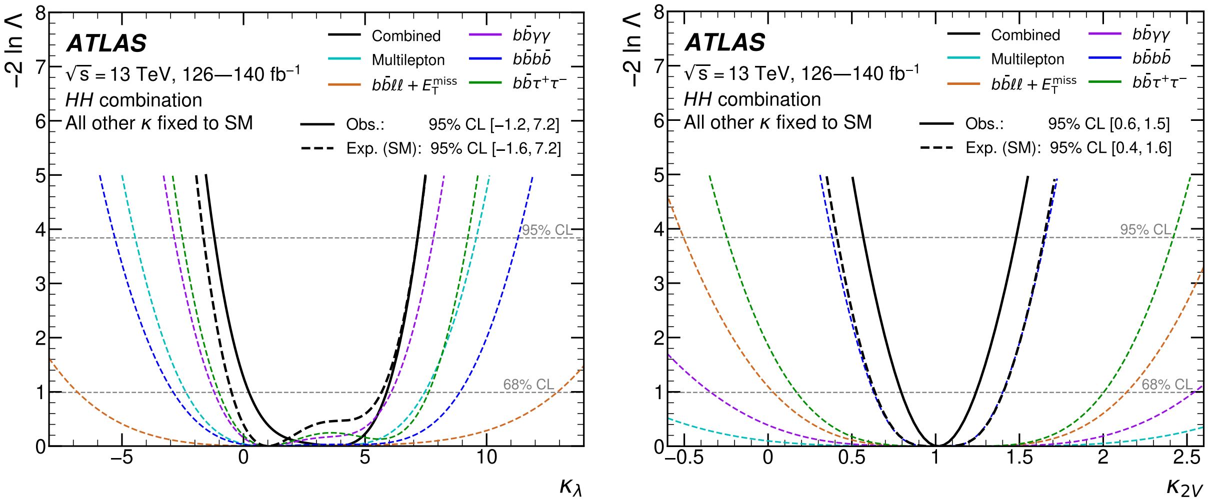


Figure 3.13: Observed and expected upper limits at the 95 % CL on the signal strength for inclusive ggF $HH$ and VBF $HH$ production from various decay channels, and their statistical combination assuming $m_H = 125\,\text{GeV}$. Taken from Reference [55].

limit of 2.4 in the absence of $HH$ production. The combination includes the most recent Run 2 results in the $b\bar{b}b\bar{b}$ [56, 57], $b\bar{b}\tau^+\tau^-$ [12], $b\bar{b}\gamma\gamma$ [58], multilepton [59], and $b\bar{b}\ell^+\ell^- + E_{\mathrm{T}}^{\mathrm{miss}}$ [60] final states. Figure 3.13 shows the observed and expected limits for the combination and the individual channels in the absence of $HH$ production. The $HH \to b\bar{b}\tau^+\tau^-$ analysis provides the best expected limit on $HH$ production at 3.3 times the SM prediction with Run 2, owing to an optimal balance between signal acceptance and background rejection. The observed limit of 5.9 × SM is larger than the expected limit, which is credited to an upwards fluctuation of the data in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

The CMS Collaboration reports an observed 95 % CL upper limit on $HH$ production of 3.4 times the SM prediction, with an expected limit of 2.5 times the SM value under the assumption of HH production [61]. In the latest CMS analysis of the $b\bar{b}\tau^+\tau^-$ final state, the observed and expected limits on the $HH$ signal strength are 3.3 and 5.2 times the SM prediction, respectively [62].

The different $HH$ analyses have different signal to background ratios across the $m_{HH}$ spectrum. Therefore, different analyses have varying sensitivities to different values of $\kappa_\lambda$. The scan of the



(a) $\kappa_\lambda$ (b) $\kappa_{2V}$

Figure 3.14: Scans of the test statistic as a function of (a) $\kappa_\lambda$ and (b) $\kappa_{2V}$ for the different $HH$ analyses and their statistical combination. The observed results are shown as solid lines, the expected results as dashed. The intersections with the horizontal dashed lines indicate the 68 and 95 % confidence limits. The results assume all other Higgs boson couplings to be SM-like. Taken from Reference [55].

test statistics for various values of $\kappa_\lambda$ in the different ATLAS analyses and their combination is shown in Figure 3.14a. The combined result constrains the Higgs boson self-coupling modifier $\kappa_\lambda$ at the 95 % CL to the range $\kappa_\lambda \in [-1.2, 7.2]$ , assuming all other couplings are SM-like, which is approximately consistent with the theoretical limits derived from perturbative arguments. The expected interval assuming SM-like couplings including the self-interaction is slightly larger with $\kappa_\lambda \in [-1.6, 7.2]$. The upper limit is primarily driven by the $b\bar{b}\gamma\gamma$ analysis as the final state allows for good signal to background ratio for events with a lower $m_{HH}$. The $b\bar{b}\tau^+\tau^-$ channel contributes most significantly to the lower bound given its enhanced sensitivity to SM-like $HH$ production. The $b\bar{b}\tau^+\tau^-$ scan of the test statistic assuming SM-like couplings exhibits a second minimum around $\kappa_\lambda \approx 5$ given the quadratic dependence of the cross-section on $\kappa_\lambda$. Figure 3.15 shows a sketch of the SM Higgs potential that indicates the current constraints on $\kappa_\lambda$ obtained from the most recent ATLAS $HH$ combination.

The likelihood scan of the test statistic for various values of $\kappa_{2V}$ is presented in Figure 3.14b. The observed allowed interval for the coupling modifier at the 95 % CL is $\kappa_{2V} \in [0.6, 1.5]$, with

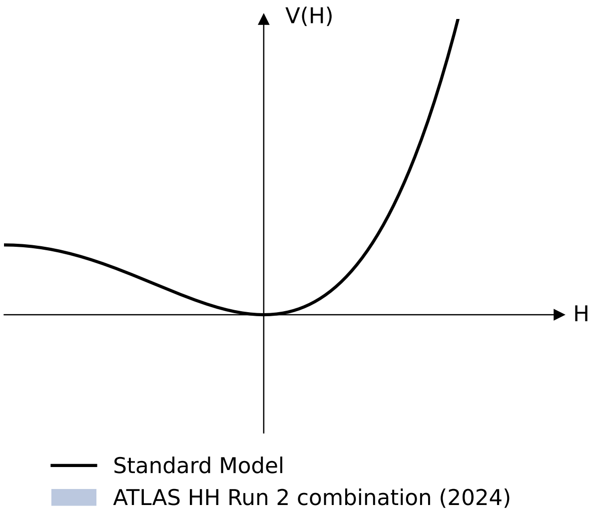


Figure 3.15: Sketch of the the Higgs boson potential after symmetry breaking. The black line illustrates the SM prediction. The blue shaded region illustrates the shape compatible with the observed interval of $1.2 \leq \kappa_\lambda \leq 7.2$ and assuming $\lambda_{HHH} = \lambda_{HHHH}$.

an expected range of $\kappa_{2V} \in [0.4, 1.6]$. These constraints are primarily driven by the $b\bar{b}b\bar{b}$ channel, which, thanks to its large branching fraction, is less affected by the very small VBF $HH$ production cross-section.

The most recent CMS analysis [62] in the $b\bar{b}\tau^+\tau^-$ final state obtains an observed constraint on $\kappa_\lambda$ of $\kappa_\lambda \in [-1.7, 8.7]$, and an expected interval of $\kappa_\lambda \in [-2.9, 9.8]$ at the 95 % CL. $\kappa_{2V}$ is constrained to $\kappa_{2V} \in [-0.4, 2.6]$ in the observed case, and $\kappa_{2V} \in [-0.6, 2.8]$ in the expected.

The most recent combination of $H$ and $HH$ measurements was performed with the first set of Run 2 results [63] in 2022. This combination allows multiple coupling modifiers, such as $\kappa_t$ and $\kappa_V$, to float freely, providing the most model-independent limits to date. It includes the $HH$ analyses in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ final states, and the $H$ analyses in the $\gamma\gamma$, $ZZ^*$, $WW^*$, $\tau^+\tau^-$ and $b\bar{b}$ final states. The resulting constraint on $\kappa_\lambda$ was an observed interval of $\kappa_\lambda \in [-0.4, 6.3]$, with an expected interval of $\kappa_\lambda \in [-1.9, 7.6]$. The most recent combination of $HH$ and $H$ analyses by the CMS collaboration obtained an observed interval at the 95 % CL of $\kappa_\lambda \in [-1.2, 7.5]$ and an expected interval of $\kappa_\lambda \in [-2.0, 7.7]$ [64]. The analysis statistically combined CMS $HH$ analyses in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, leptons and $WWb\bar{b}$ final states, as well as $H$ analyses in the four-lepton, $\gamma\gamma$, $WW$, leptons, $b\bar{b}$, $\tau^+\tau^-$ and $\mu^+\mu^-$ final states.

# 4 The ATLAS Experiment

A model independent test of the Higgs mechanism, particular the Higgs boson self-interaction, requires the production of Higgs boson pairs ($HH$) in a controlled environment such as a particle collider. Today, the Large Hadron Collider (LHC) hosted at CERN, the European Organization for Nuclear Research, is the only facility in the world capable of producing a sufficient number of Higgs bosons to enable a search for $HH$ production. This chapter introduces the LHC and describes the ATLAS detector and its infrastructure, which are used to record and analyse the LHC collisions.

## 4.1 The Large Hadron Collider

The LHC [66] is the world largest particle collider with a circumference of 26.7 km. It is located in the tunnel that formerly housed the LEP collider [67], at a depth ranging from 45 to 170 m beneath CERN, near Geneva, Switzerland. The LHC is designed for the exploration of particles at the TeV scale, and therefore operates at a very high centre-of-mass energy and with the highest possible instantaneous luminosity. The machine is built to collide two opposite directed proton beams at an centre-of-mass energy of 14 TeV, but can also collide heavy ions.

Before the beams enter the LHC, they are accelerated to increasingly higher energies in a chain of smaller accelerators that is sketched in Figure 4.1. The proton beam is created by ionising hydrogen. The beam is focused and the protons are grouped into packets, called “bunches”, that are accelerated to 50 MeV in the LINAC 2. After a series of synchrotrons, the Booster, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) the beam is filled into the LHC in

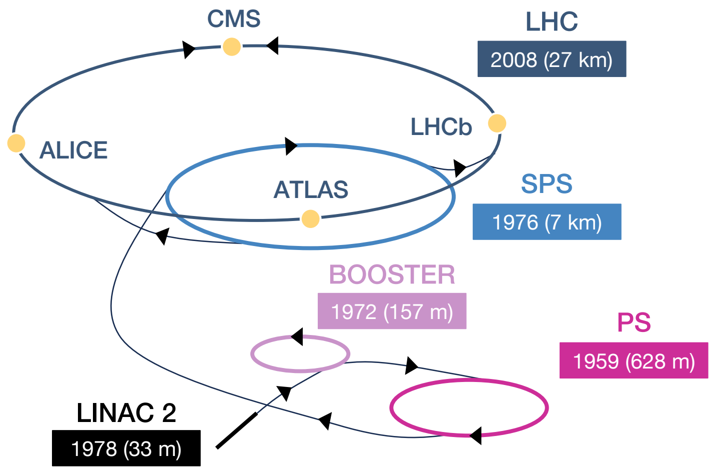


Figure 4.1: Illustration of the LHC proton-proton injector chain. Graphic adapted from Reference [65].

either clockwise or anti-clockwise direction. The two beams are accelerated and focussed in eight straight sections. They are kept on closed loops in eight arc sectors by 1200 superconducting niobium-titanium dipole magnets that generate field strengths up to 8.3 T and are cooled down with liquid Helium to 1.9 K.

The LHC operates in multi-year data-taking periods known as "Runs", separated by extended maintenance and upgrade periods called "Long Shutdowns" (LS). Since the start of Run 1 in 2010, the collision and centre-of-mass energy have been steadily increased after each shutdown. Table 4.1 provides an overview of the operational timeline, showing the centre-of-mass energy and the integrated luminosity delivered to the experiments. This focus of this thesis is Run 2, during which the LHC accelerated each beam from 450 GeV to the collision energy of 6.5 TeV. At the design luminosity of $\mathcal{L} = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, each beam consisted of 2808 bunches of about $10^{11}$ protons separated by 25 ns.

| **Period** | **Years** | $\sqrt{s}$ | **Delivered luminosity to ATLAS and CMS** | |
|---|---|---|---|---|
| Run 1 | 2010 - 2012 | 7 - 8 TeV | $\approx 30\,\text{fb}^{-1}$ | [68] |
| LS1 | 2013 - 2014 | - | - | [69] |
| Run 2 | 2015 - 2018 | 13 TeV | $\approx 160\,\text{fb}^{-1}$ | [70] |
| LS2 | 2019 - 2021 | - | - | [71] |
| Run 3 | since 2022 | 13.6 TeV | $> 300\,\text{fb}^{-1}$ as of October 2025 | [72] |

Table 4.1: Overview of the LHC operation periods.

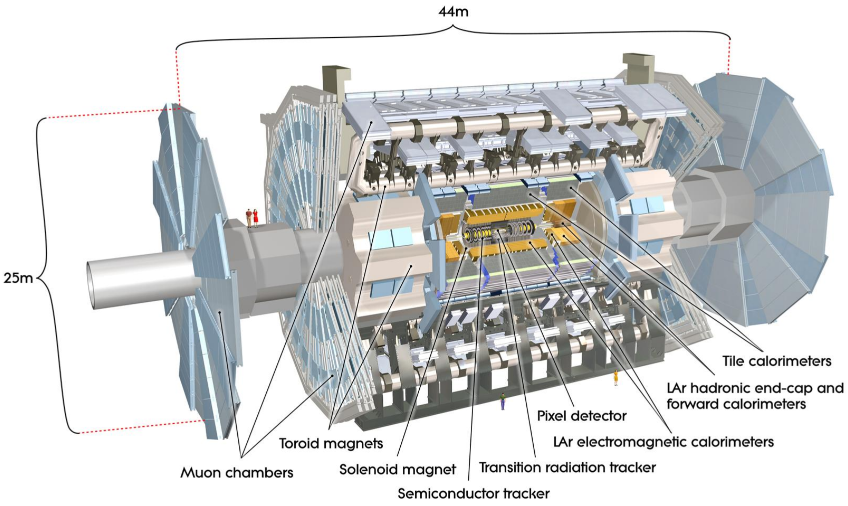


Figure 4.2: Cutaway illustration of the ATLAS detector as operated during LHC Run 2. Taken from Reference [73].

The two LHC beams can be brought to collision at any of the four interaction points, where particle detectors record the collision debris. On opposite sides of the collider, two general purpose detectors, ATLAS [73] and CMS [74] are situated. They are designed independently to cross-confirm each other. The LHCb detector [75] is a forward spectrometer designed to measure heavy-flavour hadron decays. The ALICE detector [76] primarily analyses heavy ions collisions to study very high energy density states namely the quark-gluon plasma. In addition, several smaller experiments are installed around the LHC searching for exotic particles.

## 4.2 The ATLAS detector

The ATLAS[1] detector is one of the largest general-purpose detectors ever built. It is 44 m long, 25 m wide and weights about 7000 t. ATLAS is designed for a broad experimental program ranging from the search for the - until 2012 - undiscovered Higgs boson, SM precision measurements but also searches for new physics phenomena. The detector records the energy deposits of charged and neutral SM particles with the exception of neutrinos. From the detector signals tracks and calorimeter clusters are reconstructed, enabling the identification of particles and the

---

[1] **A** **T**oroida**l** **L**HC **A**pparatu**S** - A Truly Lazy Acronym, Seriously.

reconstruction of their four-momenta.

ATLAS consists of several forward-backward symmetrical sub-detectors. They are placed in a cylindrical barrel region and inside two layered end-cap regions providing an almost symmetric $4\pi$ solid angle coverage of the interaction point (IP). A cutaway illustration is shown in Figure 4.2. Closest to the IP, the Inner Detector (ID) records the tracks of charged particles that are bent by a 2 T magnetic field, generated by a superconducting solenoid magnet. The ID is enclosed by the electromagnetic and hadronic calorimeters. The outermost Muon Spectrometer (MS) is immersed in a magnetic field of up to 3.5 T that is generated by the superconducting toroid magnet system. The MS measures charge deposits that are most likely from muons as they are the only charged particle type escaping the calorimeter. The high particle fluxes necessitate not only a high granularity of the sensors to resolve overlapping signatures but also fast and radiation hard electronics and sensors.

The analysis of $HH \to b\bar{b}\tau^+\tau^-$ events relies in particular on the good track resolution near the IP to resolve the primary, secondary and tertiary interaction vertices as they are vital for the identification of $b$-hadrons. In addition, the fine transverse segmentation of the electromagnetic calorimeters benefits the measurement of electrons, while the measurements of jets and hadronically decaying $\tau$-leptons profit from the segmentation of the hadronic calorimeter system. A short overview of the sub-detectors is given in following sections, further information can be found for example in Reference [73].

### 4.2.1 The ATLAS coordinate system

The ATLAS experiment uses a right-handed coordinate system with its origin at the interaction point. The $x$-axis is defined by the direction from the interaction point to the centre of the LHC ring, $y$ is pointing upwards and the $z$-axis is parallel to the beam. In the transverse plane the cylindrical coordinates radius $r$ and the azimuthal angle $\phi$ around the $z$-axis are used. Quantities in this plane are denoted with the subscript $T$. Given the unknown momentum fractions of the colliding partons in proton-proton collisions, the boost along the beam axis ($z$) is not known.

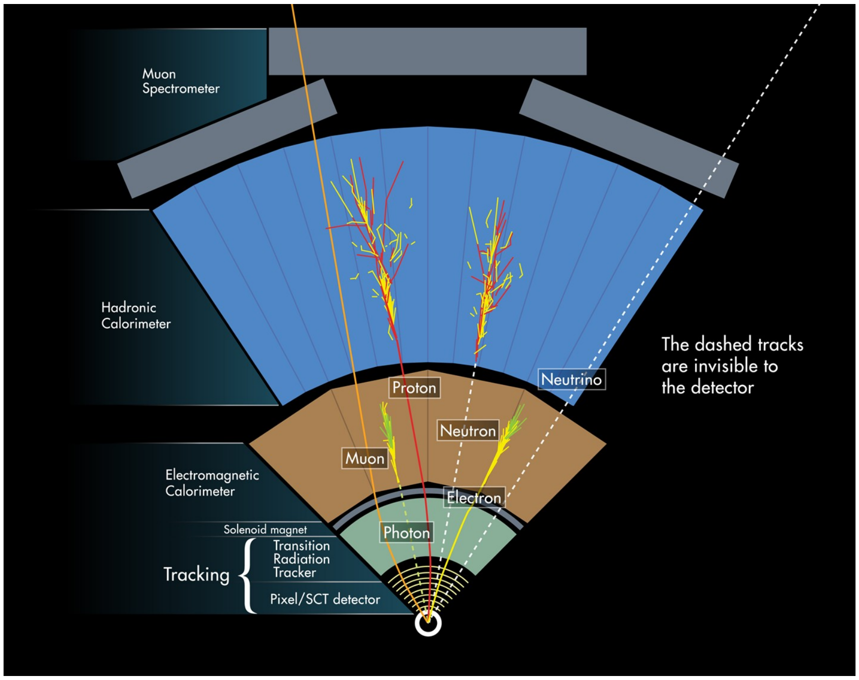


Figure 4.3: Illustration of signatures left by various particles in a cut through of the ATLAS detector. Taken from [77].

Instead of using the polar angle $\theta$, the rapidity $y$ is used as differences in $y$ are invariant under Lorentz boosts along $z$. The rapidity is defined as $y = \frac{1}{2}(\frac{E+p_z}{E-p_z})$, with the particle's energy $E$ and the momentum along the beam axis $p_z$. In the ultra-relativistic limit ($m \approx 0$) the rapidity is approximated by the pseudo-rapidity $\eta = -\ln(\theta/2)$, and the geometric distance between two particles is defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$.

### 4.2.2 Particle identification with the sub-detectors

Figure 4.3 illustrates the signatures left by the interaction of various particles in the sub-detectors that are used to identify them. Due to their charge, electrons leave tracks in the different sub-systems of the ID and their momentum and charge can be inferred from the curvature. Their energy is measured from the signal caused by electromagnetic showers initiated in the electromagnetic calorimeters by electrons. Photons are electrically neutral. Thus, they leave no tracks in the ID. However, they cause similar shower signatures in the electromagnetic calorimeter, and are, as electrons, fully stopped there. Protons leave tracks in the ID and interact little with the electromagnetic calorimeters. They mainly interact via the strong force and are stopped via

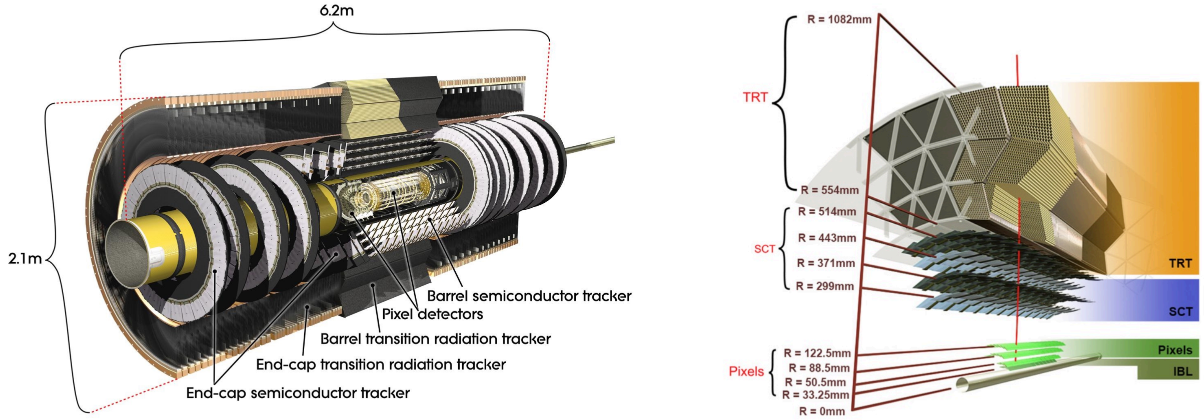


Figure 4.4: Illustration of the ATLAS Inner Detector (left) [73], and of the barrel structure (right) [79].

hadronic showers in the hadronic calorimeters. Charged pions leave tracks in the ID and deposit their energy predominantly in the hadronic calorimeter. Neutral pions, which decay promptly into two photons, leave no tracks and appear as electromagnetic showers in the electromagnetic calorimeter. Neutrons are identified by similar hadronic calorimeter signatures and the absence of tracks in the ID. Muons leave tracks in the ID. Since they are mostly minimum ionising, muons deposit only little energy in the calorimeters and only produce showers if they are above 100 GeV. They are therefore the only particles reaching and leaving tracks in the MS. Lastly, neutrinos leave no traces in the ATLAS detector. Their presence is inferred indirectly from the momentum imbalance in the transverse plane.

### 4.2.3 The Inner Detector

The ID [78] is a precise tracking detector that surrounds the beam pipe and covers $|\eta| < 2.5$. The detector is immersed in a 2 T magnetic field along the beam direction generated by the solenoid magnet and consists of three sub-systems, from inner to outermost: the Pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Each sub-system consists of a barrel region and two end-caps, as illustrated in Figure 4.4.

Charged particles leave local energy deposits ("hits") in the sensor materials. Tracks are reconstructed from the hits, and used to determine the momentum $p$, charge and the primary, secondary and tertiary interaction vertices. The designed momentum resolution is $\sigma_{p_T}/p_T = 0.05\,\%\,p_\mathrm{T}\,[\mathrm{GeV}] \oplus 1\,\%$ for charged particles with $p_T > 0.5\,\mathrm{GeV}$.

**The Pixel Detector with the Insertable B-Layer** consists of four layers of high resolution modules in the barrel region, and three disks in the end-caps. The innermost "Insertable B-Layer" was added for Run 2 and significantly improves the resolution for low-$p_T$ particles with an pixel size of $50\,\mu\mathrm{m} \times 250\,\mu\mathrm{m}$. The other layers have larger pixels ($50\,\mu\mathrm{m} \times 400\,\mu\mathrm{m}$) [80]. On average, a charged particle leaves four hits in the pixel detector. The high resolution of $10\,\mu\mathrm{m}$ in the $r-\phi$ plane and $67\,\mu\mathrm{m}$ in the $z$-direction is important for the reconstruction of jets and the identification of the jet flavour [81].

**The Semiconductor Tracker** consists of silicon microstrip modules. They are arranged in four layers in the barrel and nine disks in each barrel. Each module consists of two pairs of rectangular silicon sensors that are glued back-to-back. Each strip has a pitch of $80\,\mu\mathrm{m}$, and provide a a resolution of $17\,\mu\mathrm{m}$ in $r-\phi$ plane, and $580\,\mu\mathrm{m}$ in the $z$-axis. On average, a charged particle leaves eight hits in the SCT [82, 83].

**The Transition Radiation Tracker** is a straw tube tracker tracking particles up to $|\eta| < 2$ [84]. The $4\,\mathrm{mm}$ wide, gas filled tubes feature a central anode wire. The tubes are placed parallel to the beam pipe in the barrel system and perpendicular to the beam in the end caps. The TRT balances the lesser intrinsic accuracy with a large number of hits in the 350 thousand tubes since a charged particle leaves on average 36 hits in the detector. TRT provides a resolution of $130\,\mu\mathrm{m}$ in the $r-\phi$ plane in the barrel and $z-\phi$ plane in the end-cap. Initially, the tubes were filled with a Xenon gas mixture which allowed the distinction of electrons from pions by the transition radiation of traversing particles. Due to gas leaks the tubes are filled today with an argon based gas mixture [85] and the feature has been lost.

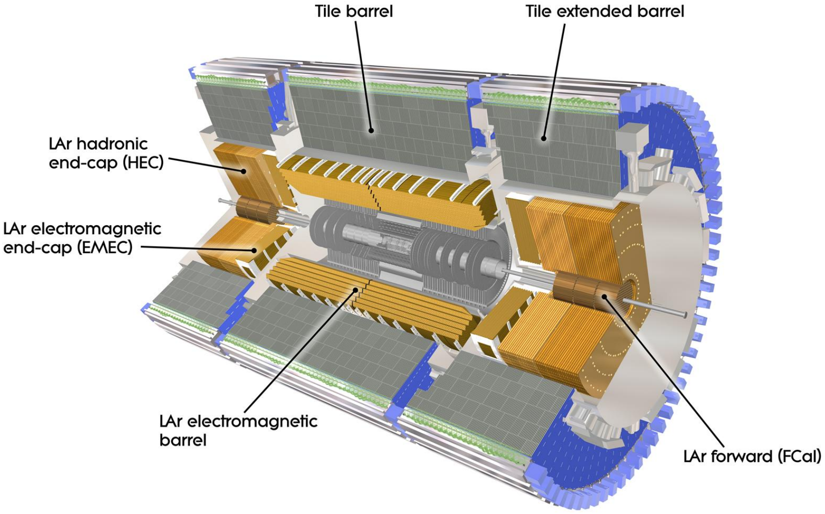


Figure 4.5: Illustration of the ATLAS calorimeter system. Taken from [88].

### 4.2.4 The calorimeter system

The calorimeter system measures the energy of particles. It is split into the electromagnetic (ECAL) [86] and the hadronic calorimeter (HCAL) [87]. They measure the energy of particles that either predominantly shower via the electromagnetic force, such as electrons or photons, or the strong force, such as hadrons. The system is organised into barrel and end-cap structures as shown in Figure 4.5. Both, the ECAL and the HCAL are sampling calorimeters with alternating layers of active detection and passive high-density absorption materials. The calorimeters' segmentation provide directional information, which are used for the reconstruction of electrons, photons, jets or the missing transverse energy. The calorimeters are rotationally symmetric around the beam axis and cover up to $|\eta| < 4.9$. Muons and neutrinos are the only SM particles that traverse the calorimeter.

**The electromagnetic calorimeter** is fully symmetrical in $\phi$ in the barrel region for $|\eta| < 1.475$ given its accordion-shaped alternating lead absorbers and Liquid Argon (LAr) and Kapton electrodes. The ECAL has a depth of at least 22 radiations lengths, which is sufficient to stop electrons and photons up to several TeV. For better spatial resolution, the calorimeter is divided in three layers in $r$ and segmented into “cells” in $\phi \times \eta$. This also improves the photon resolution and the estimation of energy leakage into the HCAL. In each end-cap, an inner wheel covers $1.375 < |\eta| < 2.5$, and an outer wheel $2.5 < |\eta| < 3.2$. Additionally, thin pre-sampler layers correct for energy losses in the material before the ECAL in both barrel and end-caps. The designed relative energy resolution of the ECAL is $\sigma_E/E = 10\,\%/\sqrt{E[\text{GeV}]} \oplus 0.7\,\%$ [89].

**The Hadronic Calorimeter** consists of three subsystems made from different materials, the Tile, the LAr hadronic end-cap and the LAr forward calorimeter. The designed relative energy resolution of the hadronic calorimeter is $\sigma_E/E = 50\,\%/\sqrt{E[\text{GeV}]} \oplus 3\,\%$ in the barrel and end-caps, and $\sigma_E/E = 100\,\%/\sqrt{E[\text{GeV}]} \oplus 10\,\%$ for the forward calorimeter. In total, the HCAL has a nuclear interaction length of about $10\,\lambda$, allowing only muons and neutrinos to pass through to the MS. The Tile calorimeter consists of a barrel ($|\eta| < 1.0$) and an extended barrel ($0.6 < |\eta| < 1.7$) system using a combination of scintillating tiles and steel absorber plates. The calorimeter is segmented in $r$ and $\phi \times \eta$. The LAr end-cap calorimeter consists of LAr and copper absorbers and covers $1.5 < |\eta| < 3.2$. As the Tile calorimeter, its readout is segmented. The LAr forward calorimeter consists of a electromagnetic and a hadronic calorimeter subsystem, using also LAr as the active material and tungsten absorbers.

### 4.2.5 The Muon Spectrometer

Figure 4.6 shows the MS [91], the outmost sub-detector that measures muons with an energy of up to a few TeV before they leave the detector and decay. To increase the momentum resolution, the MS is situated inside an up to 3.5 T magnetic field generated by a toroid magnet system. The muons are deflected in the $\eta$ direction for $|\eta| < 2.7$. Its four sub-detectors are separated into a barrel and two end-caps. The MS consist of three tracking systems that yield a relative momentum resolution of $\sigma/p_T = 10\,\%$ for a 1 TeV muon. Combining the MS measurement with tracking information from the ID enhances the precision particularly for low-$p_T$ muons.

**Monitored Drift Tubes** (MDT) are gas-filled drift tubes that precisely track muons using three layers in the barrel and end-caps with a resolution of 35 µm per tube for $|\eta| < 2.7$. In the

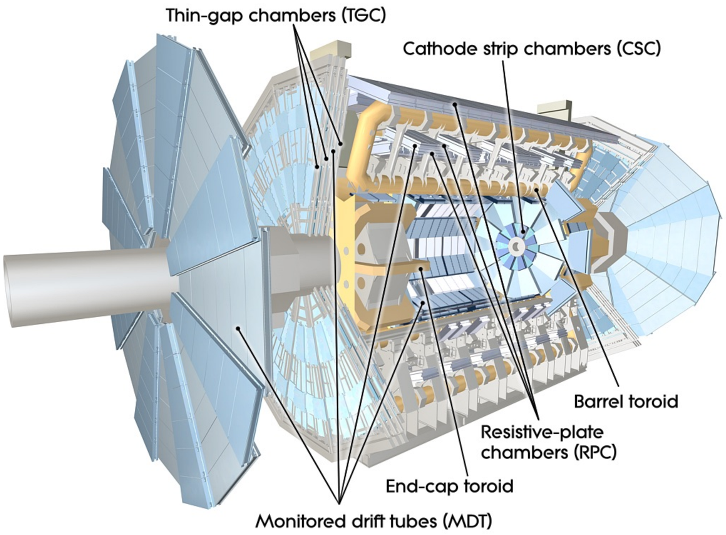


Figure 4.6: Illustration of the ATLAS Muon spectrometer. Taken from [90].

end-caps, the MDTs are mounted on large wheels.

**Cathode Strip Chambers** (CSC) are used for tracking in the forward region ($2.0 < |\eta| < 2.7$). The multi-wire proportional chambers have a resolution of 60 µm and provide a significantly faster readout than the MDTs in the forward region.

**Resistive Plate Chambers** (RPC) provide triggering with a spatial resolution of several millimetres. The chambers consist of two metal plates separated by a thin gas layer, and are installed in the middle and behind the last MDT layer.

**Thin Gap Chambers** enable triggering in the forward region with a similar time resolution of $\mathcal{O}(\text{ns})$ as the RPCs. The multi-wire proportional chambers have a spatial resolution of a few mm.

**The New Small Wheel** was installed after Run 2 in preparation for the high-luminosity LHC. It replaces the CSC and TGC in the small wheel by small-strip Thin Gap Chambers (sTGC) and micro-megas (MMG) chambers [92].

### 4.2.6 The luminosity measurement system

The estimate of the expected number of events for a process relies on a precise measurement of the instantaneous luminosity. Therefore, ATLAS measures the luminosity in various ways. For example, the LUCID-2 [93] detector measures the luminosity from the Cherenkov light produced by particles in its photomultipliers positioned in the forward arms at about 17 m away from the IP. The Beam Conditions Monitor (BCM) [94] provides a luminosity measurement in addition to its supervision of the beam condition. Another estimate is provided from the minimum bias currents in the crack regions of the Tile calorimeter. The different measurements are calibrated in van-der-Meer scans, during which low-luminosity beams are are swept transversely across each other [95].

### 4.2.7 The trigger and data acquisition system

The multi-level trigger and data acquisition system reduces the high event rate of 40 MHz to manageable save rates by only storing events of particular high interest for analysis. During Run 2, a custom hardware-based Level 1 (L1) trigger system was followed by a software based High Level trigger (HLT) [96–98]. The L1 trigger identifies objects with large transverse momenta or large missing transverse energy and defines regions of interest from reduced calorimeter and MS information. “Interesting” events are passed at rate of about 100 kHz from front-end buffers to read-out drivers (ROD). The HLT trigger accesses the full detector information to decide whether the event is kept for offline. The total event save rate is thereby reduced to a maximum of 1.5 kHz at peak times. For each signature of interest, dedicated trigger chains are defined taking into account beam conditions and physics requirements.

### 4.2.8 Computing and data processing

The recorded data is processed with the worldwide Worldwide LHC Computing Grid (WLCG) [99]. The WLCG is a global, distributed computing system designed to provide the storage capacity and processing power required for LHC data, and to ensure efficient access for physicists across participating institutions. It is organised into three hierarchy levels, referred to as “Tiers”, which enables scalable data handling and coordinated resource usage [100]. The Tier 0 centre is located at CERN. It reconstructs the raw detector data, assesses the data quality immediately after data-taking and archives the primary datasets. About ten large national or regional computing facilities serve as Tier 1 centres. There, events are processed by reconstructing and calibrating physics objects from the detector signatures, as described in Chapter 6. Furthermore, the Tier 1 centres store parts of the raw and processed data to ensure redundancy and data integrity across the grid. User-level analysis and Monte Carlo simulation production occurs on the more than hundred Tier 2 centres, which are typically university or laboratory clusters.

The large data volume and complexity of the ATLAS detector output requires a structured computing model to transform the raw detector signals into data formats suitable for physics analysis. A multi-stage data processing model progressively refines the information content by separating detector-level reconstruction from user-level analysis and thus reduces the data size while preserving the relevant physics content.

From the raw data, containing the full detector response information for each event, Event Summary Data (ESD) are produced by reconstructing objects such as tracks. ESD contain detailed information on the reconstructed physics objects and detector calibrations and are intended for re-calibrations and reprocessing. For physics analysis Analysis Object Data (AOD) are derived from ESD. AOD contain only high-level reconstructed physics objects such as electrons, muons, jets, and missing transverse momentum. Subsequent data reduction steps further tailor the datasets to specific analyses. The $HH \to b\bar{b}\tau^+\tau^-$ analysis uses the CxAOD framework [101] that has been developed in the context of the Run 1 $VH, H \to b\bar{b}$ analysis [102]. A preliminary event selection (“skimming”) and variable selection (“slimming”) are applied to produce Derived AODs (DxAODs). In a next step, physics objects are calibrated, and the event and variable selection is refined, yielding Calibrated xAODs (CxAODs). Finally, lightweight datasets containing histograms are produced from the CxAODs for the statistical interpretation of the analysis.

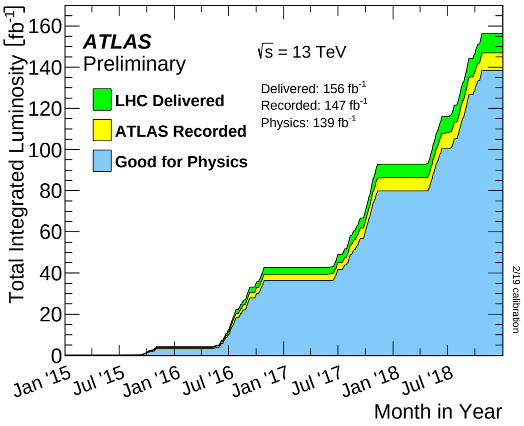


Figure 4.7: The cumulative integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018. Taken from Reference [103].

### 4.2.9 Run 2 performance

During Run 2, the LHC operated at a centre-of-mass energy of 13 TeV from 2015 to 2018. ATLAS recorded in total 140 $\mathrm{fb}^{-1}$ of physics data out of the 159 $\mathrm{fb}^{-1}$ of collisions delivered by the LHC. The cumulative integrated luminosity over time is shown in Figure 4.7. Per bunch crossing there were on average 33.7 collisions. This background of non-interesting collisions is referred to as “pileup”.

## 4.3 Outlook to the High Luminosity LHC

The LHC will be upgraded after Run 3 for the High Luminosity (HL) programme to record up to 3000 $\mathrm{fb}^{-1}$ of proton-proton collisions by the year 2041 [104–106]. The instantaneous luminosity is expected to increase by more than a factor of three with respect to Run 3 [107] and results in more than 200 simultaneous collisions per bunch crossing. The ATLAS detector will be upgraded to remain as least as performant as during Run 2 [108]. Several new detectors will be installed and the safety systems will be upgraded accordingly. A new all-silicon inner tracking detector (ITk) will replace the ID [109, 110]. ITk features a reduced material budget, finer segmentation and an extended coverage up to $|\eta| < 4$. A new High-Granularity Timing Detector (HGTD) will be installed in between ITk and the end-cap calorimeters [111]. HGTD will allow resolving the high pileup backup through timing information from several layers of fast silicon sensors. New muon chambers in the inner barrel region will improve the trigger efficiency and momentum resolution [112]. The electronics of the calorimeter and the muon system will be upgraded and feature a faster readout and a higher resolution for the trigger system. The upgrades will allow a faster trigger and data acquisition system that can use the full detector granularity. The hardware based trigger rate is expected to increase by a factor of ten [113–115].

## 4.4 Detector services

The ATLAS detector is one of the largest and complex particle detectors ever built. The detector is situated 85 m underground and relies on an extensive service infrastructure for its electricity, cooling, and gas supplies. The service infrastructure is distributed across multiple surface and underground buildings. Surface building names begin with the prefix “S”, underground buildings with “U”, such as the experimental cavern UX15, or two service cavern US15 and USA15.

**Electricity** The experimental site receives a primary 66 kV electricity supply from the CERN power network, that is in turn sourced from the French grid. A backup supply is available via the Swiss power grid in the event of service interruptions. Electricity is distributed through a network of switchboards at three voltage levels which are provided by multiple transformers: 18 kV, 400 V, and 230 V. Critical systems are supported by Uninterruptible Power Supply (UPS) battery units. Depending on the criticality of the system UPS provides emergency power between for ten minutes up to several hours. Safety systems are powered by safety switchboards (ESD). They are additionally backed up by a diesel generator on site and diesel generators on the main CERN site. A separate diesel generator backup is available for the magnet system.

**Water and cooling** “Mixed” (15 °C) and “Chilled” (5 °C) water is produced with electrical chillers on the surface in SUX1 that are connected to the Swiss water network. The primary circuits are cooled with evaporative cooling towers in SF1. The mixed and chilled water circuits are closed loops that are thermally coupled via heat exchangers to the secondary circuits of the underground cooling stations in the service caverns. The rack cooling circuits use the mixed water supply. Critical racks feature an additional by-pass to tap water cooling in case of a cooling circuit failure. The Tile and LAr calorimeters, the diffusion pumps, and the Muon system are cooled with water-based mono-phase cooling stations. The Pixel and SCT sub-detectors are cooled by the evaporative two-phase thermosiphon system that uses $C_3F_8$ as the coolant. The TRT sub-detector and the cabling cooling stations use $C_6F_{14}$, and the IBL cooling station $CO_2$ as the coolant.

**Ventilation** Air conditioning units use the chilled water circuits to cool down selected surface buildings and the underground caverns. The experimental cavern is kept at $25 \pm 3\,^\circ\mathrm{C}$ at a relative humidity between 25 and 60 %. The ventilation system is connected to the safety systems. In case of problems, it can be switched to a closed circuit or purge the cavern air within several minutes.

**Cryogenics** The cryogenic cooling system cools the superconducting magnets with liquid helium to 4.5 K. The thermal shield is cooled down to about 60 K with gaseous helium. The helium is cooled with five primary and two secondary helium compressors. The magnet vessel is cooled by a liquid nitrogen refrigeration system.

## 4.5 Safety systems

Different systems supervise the safety of the personnel and the detector with different granularity. They trigger alarms, which are categorised by their severity into four levels that are listed in Table 4.2. Level 0 alarms typically correspond to technical failures of a system. If a fault does not impact other systems significantly, a Level 1 alarm is issued. Alarms related to events with a serious impact on the detector are classified as Level 2. Any situation in which there is a danger to personnel or an immediate threat to the safety of the detector issues a Level 3 and triggering a response by the CERN fire brigade.

| **Severity** | **Description** | **Follow-up** |
|---|---|---|
| Level 3 | Threat to personnel or detector safety | Immediate + fire-brigade |
| Level 2 | Medium impact to detector | Immediate |
| Level 1 | Contained impact to other sub-systems | Daily |
| Level 0 | Technical fault | Weekly |

Table 4.2: ATLAS alarm severity levels.

### 4.5.1 Environmental monitoring systems

**Smoke detection system** The smoke detection system features large area and individual-based smoke detection units. “Sniffing pipes” sample air close to the ceiling to detect smoke in large areas and inside the detector volume. Additionally, each rack cooling turbine is equipped with one smoke detector. The smoke detector signals are connected to “smoke centrals” (SFDIN), from which they are sent forward to the Detector Safety System (DSS) [116], the CERN control centre technical infrastructure (CCC-TI) and the CERN safety alarm monitoring system (CESAM) [117].

**Gas detection system** Flammable gas (FG), oxygen deficiency hazard (ODH) and carbon-dioxide detectors ($CO_2$) are installed at relevant locations in the caverns. The signals are connected to “gas centrals” (SGGAZ), and are forwarded to DSS and CESAM.

**Radiation monitors** Radiation levels are monitored by the CERN Radiation and Environment Monitoring Unified Supervision (REMUS) system [118]. Only the radiation monitors located at the access gates are connected to DSS.

**Water leak detection** Sensing wires on the ground provide large area leak monitoring in the cooling circuitry from cooling stations, manifolds and distribution in all the experimental galleries. At the end of Run 2, a novel detection system, based on a new carbon-nanotube material [119], was installed that provides local detection inside racks.

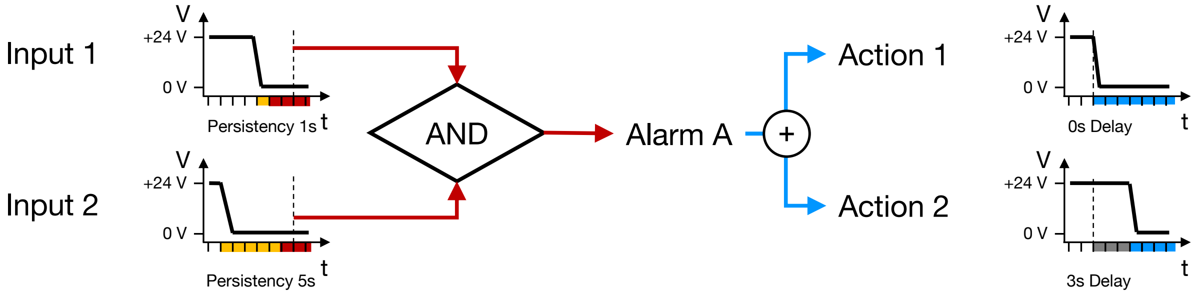


Figure 4.8: Simplified logic flow of DSS. Alarm A is triggered as soon as the two inputs are active for longer their persistencies. The alarm triggers two actions after a specific delay.

## 4.6 Detector Safety System

DSS protects the experimental equipment and services. If any operational problems, abnormal or potentially dangerous situations are detected, DSS returns the affected sub-system(s) into a predefined safe state. DSS is common across all LHC experiments and consist of an autonomous and redundant front-end and a supervising back-end.

The front-end is composed of two redundant Programmable Logic Controllers (PLC) that are configured to trigger an alarm upon the reception of an input signal. A simplified logic flow is illustrated in Figure 4.8. An alarm input is defined from any logical “AND” or “OR” combination of any number of analogue or digital signals. A signal is triggered if an input is active longer than a given amount of time, referred to as persistency. The persistency is measured in seconds. An alarm can trigger any number of “actions” after a custom delay. DSS operates on a positive safety logic. “High” is defined as 24 V. A flip to 0 V is defined as an active signal. This logic allows triggering an input, alarm or action even in the loss of power or connection.

The alarm inputs and actuators are organised by their supervised sub-system in eight Detector Safety Unit racks (DSU). The DSUs are situated across the surface and underground buildings and are connected by redundant optical fibres. The DSS back-end features status displays, such as the active alarms or actions, and limited online reconfiguration capabilities of the PLCs via an industrial Supervisory Control and Data Acquisition (SCADA) system. The PLCs are configured via a configuration database.

In total, DSS comprises 968 analogue and digital inputs, 865 alarms and 433 actuators as of January 2025. Inputs come from flammable gas, smoke, ODH and $CO_2$ sensors, water leak detection, radiation gate monitors, magnet conditions, cooling and electricity systems as well as control systems from sub-detectors. Actions range from interlocking local power or gas supply to switching off entire sub-systems or the entire detector. An overview of the number of alarms and actions per system, and sub-system is provided in Table 4.3. Each alarm is assigned to one system. Some systems are further divided into sub-systems. The assignment of actions per sub-system is not as straightforward since some actions can be triggered by multiple alarms and can be used to ensure the safety of another system. For example, a smoke alarm in the ID will interlock the power in the calorimeters.

70 % of DSS alarms are related to safety systems, such as the highly granular smoke detection system or flammable gas and oxygen deficiency hazards. Others are related to evacuation or emergency stop buttons. 77 of the 176 alarms related to the general infrastructure are water leak alarms. They interlock power if there is a high risk of damage. The rest covers general supplies such as electricity, gas or compressed air. 70 alarms are triggered if cooling power is lost or if a cooling plant is stopped. To prevent overheating, the affected systems are interlocked with varying granularity across sub-systems. It is highest for the Muon and the IBL cooling system, which are both divided into two independently cooled halves. The caverns USA15 and US15 and the surface building SDX1 each feature their own separate rack cooling circuit. A failure of one rack cooling system has wide reaching consequences for the entire detector as many racks are interlocked. Magnet alarms either indicate a loss of the vacuum or the dumping of the magnet current. The loss of vacuum and the resulting warm up can lead to condensation, therefore, all close-by systems are switched of. Due to the large currents during a fast dump, the muon system is interlocked as it contains flammable gas. The beam abort signal from BCM has no effect on the detector and is purely informative to the operators in the control room.

The distribution of the number of actions per alarm is shown in Figure 4.9. About 40 % of alarms, such as water leak or small area smoke occurrences do not trigger any actions. Slightly less than

| **System** | **Alarms** | **Actions** | **Sub-system** | **Alarms** | **Actions** | **Impact** |
|---|---|---|---|---|---|---|
| Safety | 607 | 418 | Smoke | 525 | 414 | 3788 |
| | | | Flammable gas | 36 | 69 | 410 |
| | | | ODH | 14 | 0 | 0 |
| | | | $CO_2$ | 12 | 0 | 0 |
| | | | Evacuation | 6 | 0 | 0 |
| | | | DSU | 4 | 0 | 0 |
| | | | Radiation gate monitors | 3 | 0 | 0 |
| | | | Emergency buttons | 3 | 230 | 231 |
| | | | Other | 4 | 4 | 8 |
| Infrastructure | 176 | 312 | Water leak | 77 | 13 | 21 |
| | | | Flooding | 2 | 291 | 322 |
| | | | UPS | 34 | 30 | 85 |
| | | | Electricity | 7 | 12 | 14 |
| | | | Overheating | 11 | 9 | 11 |
| | | | Gas | 11 | 1 | 4 |
| | | | Ventilation | 11 | 0 | 0 |
| | | | Cryogenics | 8 | 32 | 76 |
| | | | Compressed Air | 2 | 1 | 1 |
| | | | ID air flush | 2 | 1 | 1 |
| | | | Technical fault | 11 | 0 | 0 |
| Cooling | 70 | 205 | Muon | 32 | 17 | 65 |
| | | | IBL | 14 | 3 | 7 |
| | | | ID | 7 | 14 | 14 |
| | | | Beam Pipe | 5 | 3 | 5 |
| | | | Rack cooling | 5 | 172 | 172 |
| | | | Calorimeter | 3 | 15 | 15 |
| | | | Other | 4 | 10 | 10 |
| Magnet | 10 | 46 | | 10 | 46 | 156 |
| BCM | 2 | 0 | | 2 | 0 | 0 |
| **Total** | 865 | 494 | | | | |

Table 4.3: Distribution of DSS alarms and actions per system and sub-system. Each alarm is exclusively assigned to one system and sub-system. Actions are not exclusively assigned per (sub-)system, as they can be triggered by multiple alarms. The column “Actions” therefore refers to the number of unique actions shared by the alarms per sub-system. The overlap of actions is illustrated by the “Impact”, defined as the the total number of actions of the alarms per sub-system.

20 % of alarms interlock other systems with one action. Representative examples are cooling failures interlocking dependent systems or smoke detection alarms inside a rack. 12 % of alarms trigger two actions. Those alarms are similar to alarms with one action, but have a slightly larger impact. Most alarms related to the muon system trigger up to 15 actions as the system is large

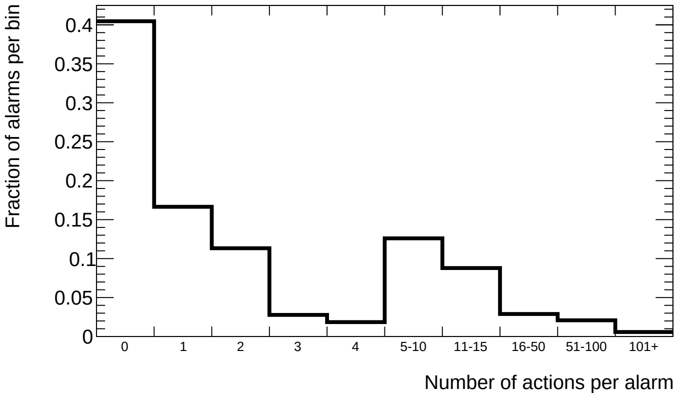


Figure 4.9: Actions per alarm distribution as of January 2025.

and is monitored with a high granularity by DSS. Level-3 alarms trigger up to 100 actions. They indicate an immediate danger to life and the detector, such as large area smoke or flammable gas detection, which requires the shutdown of large areas of detector. Only five alarms trigger more than 100 actions. They correspond to situations where an entire cavern needs to be interlocked, such as significant amount of smoke in USA15, the emergency stop button or flooding of the cavern. The alarm related to the stop of the rack cooling in USA15 triggers the most (127) actions.

## 4.7 Detector Control System

Detector Control System (DCS) [120, 121] controls the individual sub-detector sub-systems and the general infrastructure. It is not a safety system, but allows interacting with the hardware and to change its operational parameters. Critical states trigger DCS alarms alerting the operators albeit at a lower criticality than for DSS alarms. DCS is connected to external control systems such as the LHC control system or the data acquisition system. Compared to DSS, DCS features a more granular resolution up to the single channels of a power supply. The system is organised as a tree-like structure with each leaf defined as a Finite State Machine.

## 4.8 Expert System

The operation and maintenance of the ATLAS detector and its services require in-depth knowledge in many different areas, especially when planning or executing interventions and when unexpected situations arise. Most expert knowledge is documented in the CERN Engineering Data Management Service (EDMS) [122]. Unfortunately, this general purpose engineering documentation framework is often highly compartmentalized and lacks sufficient meta-data categorization. Furthermore, information is stored in many different formats, hindering easy and fast access to the relevant information. Therefore, the Expert System tool [123] has been developed with the goal of improving and facilitating the access to the experiment's knowledge base across a wide range of users. The Expert System is designed to make information easily accessible and highlight interdependencies between systems and describes detector control and safety systems, gas, water, cooling, ventilation, cryogenics, and electricity distribution. It features a simulation of the behaviour and interactions of the described subsystems with each other, for example during an intervention. An new extension of system, the Alarm Helper, is discussed in Chapter 5.

### 4.8.1 Database

The Expert System describes the detector and service subsystems and interdependencies in an object-relationship representation. The abstraction level is chosen such that an appropriate functional description of the reality is achieved. Relations follow a parent-child hierarchy and describe, for example, the power and cooling supply ("poweredBy", "waterFrom"), or DSS actions ("interlockedBy"). The hierarchy is followed when evaluating the state of an object. A child is considered "off" if any parent is "off". Some systems are described as a group of objects. The state of the group is evaluated to "off" if any object belonging to that group is "off".

To illustrate the object-relationship representation, the power and cooling supply of a rack is described in the following. The rack receives power via 230 V supply from a switchboard. A fan and a heat exchanger cool the devices inside rack to avoid overheating. The heat exchanger

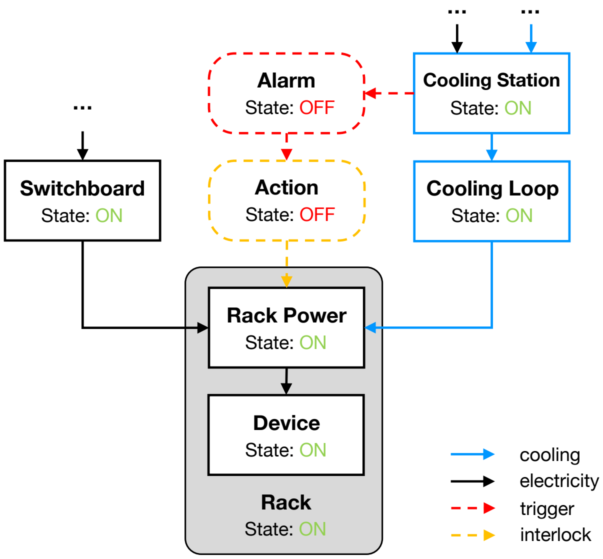


Figure 4.10: Illustration of the Expert System object-relationship description of the electricity and cooling supply of a rack, and the interlock by DSS.

is supplied with mixed water from a cooling loop, that is cooled by a cooling station. In the Expert System, this is described with multiple connected objects as shown in Figure 4.10. A rack power object is connected to a switchboard object via the boolean relation "poweredBy", which is either true or false. Information like the voltage, current or wattage, or the cable number and cable diameter are not described in the Expert System but in dedicated databases. Similarly, the rack cooling is described by the relation "waterFrom" between rack power object and a cooling loop object. Information like the water temperature, flow rate, pressure or the cooling power are omitted. The cooling loop object receives "waterFrom" from a cooling station object. The rack object is "on" if both the rack power and the contained device on. This requires both parents (switchboard, cooling loop) to be "on", too. The impact of a failure of the cooling station to produce enough cooling power is simulated by setting the state of the cooling station object to "off". Following the relations from parent to children, the cooling loop object is simulated as "off", and therefore the rack object, too. In reality, the lack of sufficient cooling power would cause the electronics inside the rack to start overheating and be potentially damaged. Therefore, the rack must be switched off by DSS as soon as the cooling station fails. A DSS alarm "Stop of cooling station" interlocks the rack electricity supply, and powers down the rack and its devices. In the

| **Sub-detector** | **Number of objects** | | | |
|---|---|---|---|---|
| | Total | Sub-divided by | | |
| Pixel | 4 | 4 × 1 | | |
| SCT | 240 | 12 × 9 (Barrel), | 11 × 12 (End-cap) | |
| TRT | 128 | 128 × 1 | | |
| LAr | 180 | 6 × 16 (Barrel), | 6 × 13 (End-cap), | 6 × 1 (Forward) |
| Tile | 256 | 4 × 64 | | |
| TGC | 74 | 74 × 1 | | |
| RPC | 20 | 2 × 10 | | |
| MDT | 1056 | 11 × 96 | | |
| NSW | 64 | 32 × 2 | | |

Table 4.4: Granularity of Expert System description of the ATLAS detectors in terms of objects that require at least one supply of electricity, cooling or gas. The granularity of the description of SCT and LAr is different across the barrel, end-cap and forward region.

Expert System, this is described by a DSS alarm and an action object.

The database comprises of approximately 13000 objects and 89000 relationships that are stored in the ATLAS TDAQ object-oriented configuration database, referred to as Object Kernel Support (OKS) [124]. The state of an object is calculated on a python server. Table 4.4 lists the number of Expert system objects per sub-detector, illustrating the varying levels of abstraction and granularity. For example, the power and cooling services of the Pixel detector are organised into four quadrants. This allows a functional representation by a group of four quadrant objects, where each quadrant is supplied by several power, and one cooling supply. The TRT sub-detector and its services are more complex, and are thus described with a higher granularity. The detector is described as a group of 128 objects, that each have multiple cooling, one power and one gas supply. The granularity of description for SCT and LAr varies between their barrel, end-cap and forward regions. Overall, the MDT has the highest granularity and described by 1056 objects.

### 4.8.2 User interfaces

The Expert System is accessible via a client server based web application providing two interface types. Visual flow chart-like diagrams highlight the interdependencies between high level systems. The diagrams are “hand-made” and only show the relationships needed to illustrate a particular application. An examples is given for the cooling of the muon system in Figure 4.11. Users can click virtual switches to simulate the impact of a failure or switch-off of objects. Visual interfaces exists to varying level of detail. For example, the water distribution is available as an overview of the distribution on the experimental site, and at higher granularity per surface building and underground cavern. A second, table-based interface type provides detailed descriptions of elements including pictures. An example is shown in Figure 4.12. Any change of state is propagated across the different interfaces.

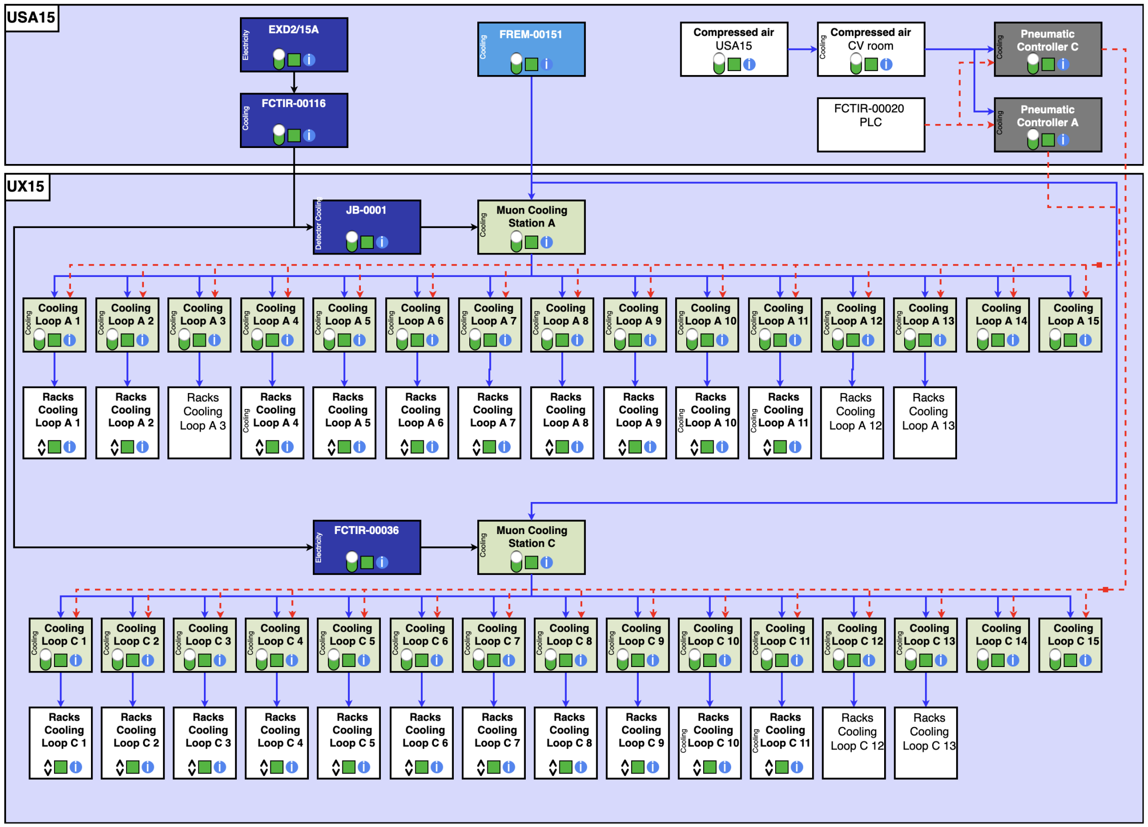
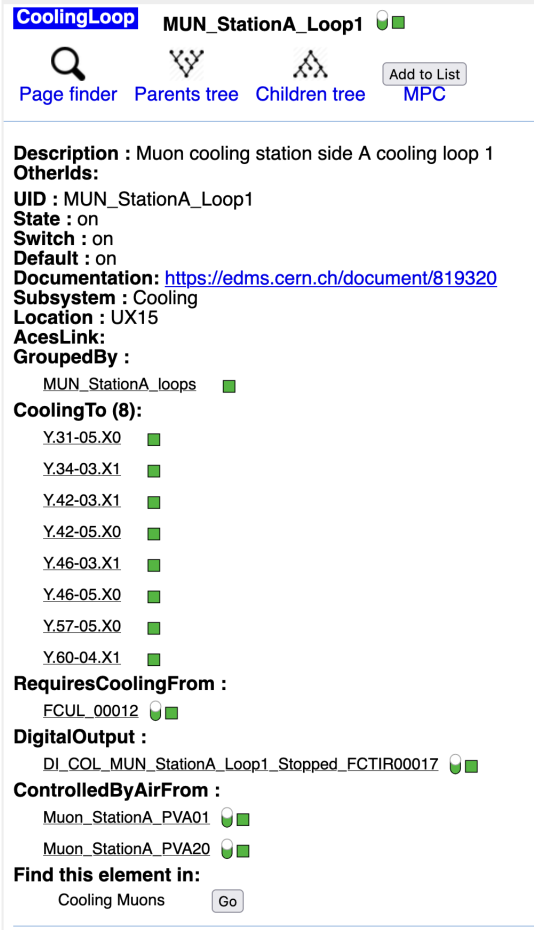


Figure 4.11: Visual interface showing the cooling of the muon system as described in the Expert System.



**MUN_StationA_Loop1** 🟩

**Description :** Muon cooling station side A cooling loop 1
**OtherIds:**
**UID :** MUN_StationA_Loop1
**State :** on
**Switch :** on
**Default :** on
**Documentation:** https://edms.cern.ch/document/819320
**Subsystem :** Cooling
**Location :** UX15
**AcesLink:**
**GroupedBy :**

MUN_StationA_loops 🟩

**CoolingTo (8):**

Y.31-05.X0 🟩
Y.34-03.X1 🟩
Y.42-03.X1 🟩
Y.42-05.X0 🟩
Y.46-03.X1 🟩
Y.46-05.X0 🟩
Y.57-05.X0 🟩
Y.60-04.X1 🟩

**RequiresCoolingFrom :**

FCUL_00012 🟩

**DigitalOutput :**

DI_COL_MUN_StationA_Loop1_Stopped_FCTIR00017 🟩

**ControlledByAirFrom :**

Muon_StationA_PVA01 🟩
Muon_StationA_PVA20 🟩

**Find this element in:**

Cooling Muons

Figure 4.12: Table style interface showing the Muon Cooling Loop 1 as described in the Expert System.

# 5 Improvements of the detector safety

This chapter describes the new Alarm Helper tool that offers a robust database designed to support improved alarm management during future interventions, and provides information on the causes of past DSS alarms, helping to distinguish between those triggered by errors and those resulting from interventions. It is based on the first analysis of DSS alarms that is published in [11], and for which the author was the main analyst. For this thesis, the analysis has been updated to include alarms up to January 2025.

## 5.1 DSS alarm analysis

Any triggered DSS alarm is followed up by the operators to identify its cause, the point of failure and the criticality of the situation is identified. The critically of an alarm depends on a multitude of factors, ranging from which sub-system is affected to the current mode of operation. The alarm recovery demands significant resources and requires the collaboration of multiple stake holders. The Shift Leader in Matters of Safety (SLIMOS) acts as a "first responder". They are responsible for the operational safety of personnel and the detector in the ATLAS control room. Assisted by the DSS and the operations team, their first priority is the determination of the cause of the alarm, and whether it is caused by an error or by an intervention as this influences the follow-up. For example, an alarm indicating the shutdown of a cooling plant can either be caused by an error in the system, or by an intervention on the device. In the case of an error, the relevant experts must establish a recovery procedure. However, if the alarm is triggered during maintenance or repairs, responsible personnel is typically already working on the equipment and the nominal operation condition can be restored more easily in this case. Importantly, alarms caused by intervention distract resources away from real failures, and, over time, lessen the alert readiness for errors.

The Expert System provides multiple tools that assist the operators in understanding DSS alarms and the ATLAS infrastructure. The tools explain and visualize the inputs and the impact of alarms on the detector. They also help in identifying failures given a list of alarms by traversing the database to identify common parents of several affected objects [125]. While the various tools can reliably identify point of failures, they do not provide information on the cause of a failure despite its importance when assessing the criticality of an alarm. Furthermore, the Expert System is unaware of ongoing interventions, despite them causing the majority of alarms as highlighted in the following.

The ATLAS operation procedure demands that any intervention is announced at the weekly meeting during which the impact on the detector is discussed. Nevertheless, despite best efforts, not all alarms and their impact on other sub-systems are communicated effectively beforehand. Members of the DSS team follow up every DSS alarm to understand its cause and to asses the status of the detector. Their first point of contact is often the SLIMOS as they are in contact with the personnel working in the cavern and the operations team. Every week, the DSS team presents a report detailing the previous week's alarms and corresponding actions, grouping those related to the same issue into "events." Since April 2024, events are categorised by three types of causes:

| | |
|---|---|
| **Intervention** | The alarms can be linked to repair or maintenance works. |
| **Error** | The alarms are not associated with any ongoing interventions and include all technical faults related to smoke, gas, and water leak sensors. |
| **Unknown** | The type of cause cannot be determined. |

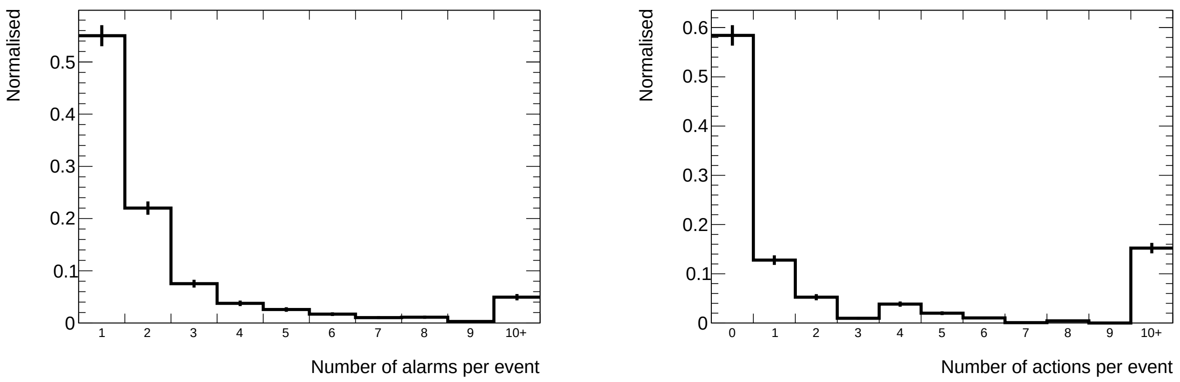
An example for an intervention is a cooling station maintenance that triggered two alarms. One alarm is related to the shutdown of the station. Spilled water triggers also a water leak alarm. Ideally, all expected alarms would be inhibited before the start of the intervention. In the example, this refers to the alarm related to (expected) shutdown of the cooling station. The water leak alarm might have been unexpected. It should only be inhibited if the intervention explicitly tests the corresponding water leak since unnoticed water leaks can have a potentially high impact. In reality, inhibit requests are the responsibility of the team carrying out the intervention. The safety policy requirements of compensatory measures in the case of a reduced safety measures can sometimes result in fewer inhibit requests being made than ideally wanted, and a large number of alarms during interventions despite them being identified and known beforehand.

This section describes the analysis of DSS alarms that occurred between the end of Run 2, the Long Shutdown 2 (LS2) and the first two years of Run 3. While the limited dataset size prevents the identification of definitive error or alarm features, the findings provide valuable insights that enhance the overall transparency of DSS alarms.

### 5.1.1 Dataset

The dataset comprises of all alarms that have been presented in operational management meetings between 1 January 2018 and 19 January 2025. It covers 865 DSS alarms, that have been triggered 5903 times, and assigned to 1673 events. The DSS team manually groups the alarms into events and categorizes them by their type of cause, based on their best understanding of the situation. Since 22 April 2024, the event type has been explicitly presented and discussed in the operational meeting. For events before that date, the type has retroactively been assigned by two members of the DSS team. Sometimes, events are understood better in the weekly operational management meeting and the findings are corrected. Nevertheless, a small uncertainty on the event grouping remains. The uncertainty on the number of events per type $N$ is assumed to be covered by the statistical uncertainty of the number of events $\sqrt{N}$. Each event is considered to start at the time of the first alarm. Due to changes in the reporting style, event times for occurrences after 2022 are recorded with minute-level accuracy. For earlier events, the times of alarms, and therefore the time of the corresponding event, have sometimes been reported rounded by a few minutes. The impact on the validity of results presented in the following is negligible.

The analysis is restricted to alarms and actions that were implemented in January 2025. This



(a) (b)

Figure 5.1: Alarms and actions per event for the dataset. The statistical uncertainty ($\sqrt{N}$) per bin estimates the uncertainty of the manual assignment of alarms to the events, and the resulting number of actions. As described in the text, the number of actions is by definition more likely to be an overestimation, which is considered to not significantly affect the analysis.

limitation arises from the fact that the Expert System database does not feature the description of decommissioned installations. Over the years, the definition of alarms and actions has changed little, and in total five alarms are excluded from the analysis: Two alarms related to the small wheel, which was replaced during Long Shutdown 2 (LS2) by the new small wheel, two alarms related to the decommissioned UPS unit EXS102, and the decommissioned smoke alarm covering the entire SR1 building that has been replaced by a more granular monitoring.

The criticality of an alarm can be partially attributed to the number of associated actions. Since multiple alarms can trigger the same action and DSS does not log which alarm triggered an action, the analysis assumes for simplicity that each alarm has triggered all linked actions. This is assumed to be reasonably accurate when comparing the distributions of actions per alarm and the number of alarms and actions per event shown in Figures 5.1a and 5.1b, respectively. In general, most alarms indicate a local failure, and trigger none or few actions. Only a few highly critical alarms trigger several actions. The majority of events only triggered one or two alarms. For about 40 % of the events, actions have been assigned out of which about 60 % have less than ten linked actions. The number of actions for events with more than ten actions is likely overestimated. The impact is small as those events correspond either to significant errors or interventions that are followed up intensely.

The alarm analysis distinguishes three different detector operation modes:

| | |
|---|---|
| **Physics** | Stable physics operation with either proton-proton or lead-lead collisions. |
| **YETS** | Planned shutdown of at least one week. Includes the End-of-the-year-technical-stops (YETS), and the LS2. |
| **Stop** | Short technical shutdowns and machine development periods. |

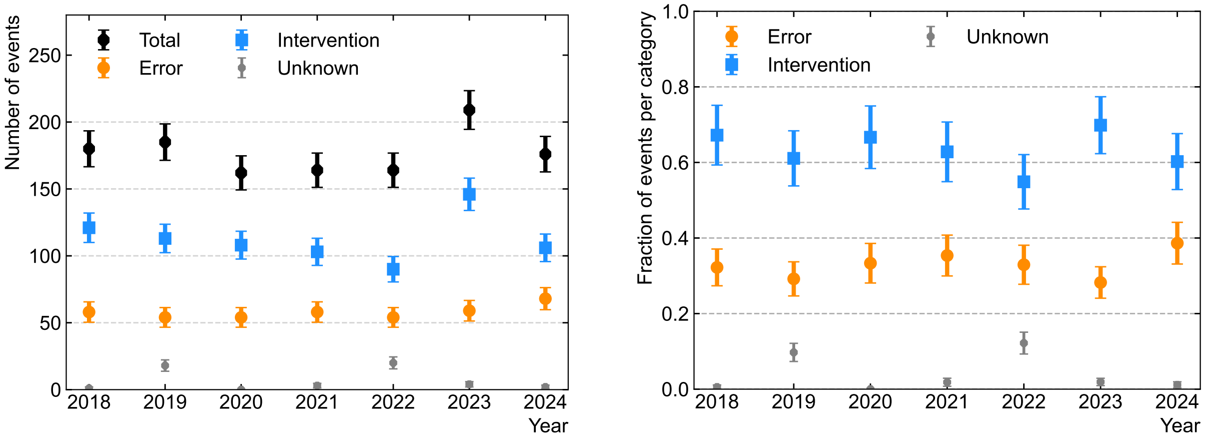
Since access to the experimental cavern is only possible if there is no beam present and safe access conditions for the personnel are guaranteed, the majority of maintenance work is performed during planned shutdown periods, which are defined by the LHC and LS2 schedules [126, 127]. During the last six years, the distribution of days per mode per year changed significantly as shown in Table 5.1. About 60 % of days were spent in YETS mode, mostly due to the three year long LS2. Run 2 ended on 2 December 2018. The LS2 started the next day and lasted until 4 March 2022. After three years in shutdown mode, the LHC commissioning for Run 3 continued for about five months. Data-taking for Run 3 resumed on 4 August 2022. Two large incidents with the LHC machine during the start of Run 3 resulted in several longer periods in Stop mode, reducing the days in Physics mode.

| **Year** | **Physics** | **YETS** | **Stop** | **Total** | Note |
|---|---|---|---|---|---|
| 2018 | 157 | 121 | 87 | 365 | End of Run 2. Start of LS2. |
| 2019 | 0 | 365 | 0 | 365 | LS2. |
| 2020 | 0 | 366 | 0 | 366 | LS2. |
| 2021 | 0 | 365 | 0 | 365 | LS2. |
| 2022 | 74 | 97 | 194 | 365 | End LS2. Start of Run 3. |
| 2023 | 78 | 131 | 156 | 365 | Extended YETS. |
| 2024 | 158 | 58 | 150 | 366 | |
| 2025 | 0 | 18 | 0 | 19 | Dataset includes events up to 19.01. |
| Total | 467 | 1522 | 587 | 2576 | |

Table 5.1: Distribution of days per operation mode.

### 5.1.2 Evolution of errors and interventions in the last six years

1266 events are defined for alarms since 2018, excluding events from interventions that have no effect on the detector such as beam aborts, technical faults and and radiation gate monitor alarms. A majority of 809 events is classified to be a result of interventions, while 409 result from errors. For 48 events the categorisation was not possible. The large number of interventions presents an



(a) Number of events per event type.

(b) Fraction of events per event type.

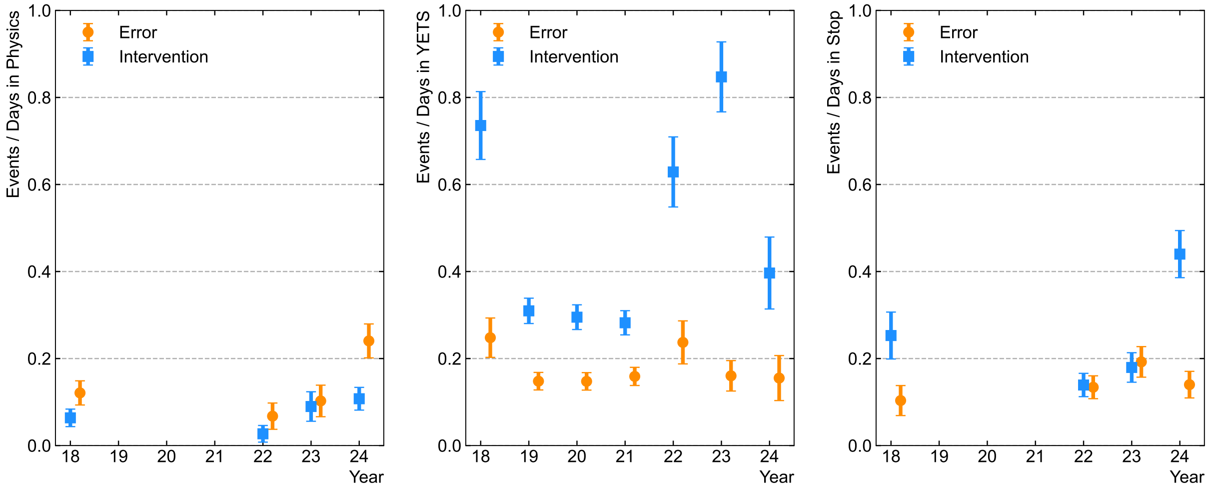
Figure 5.2: The evolution of events between 2018 and 2024, excluding events which no impact to the detector. (a) shows the total number of events and per event type. The indicated uncertainty corresponds to the statistical uncertainty ($\sqrt{N}$). (b) shows the fraction of each event type. The uncertainty is approximated using Gaussian propagation of uncertainty assuming no correlation.

opportunity for improvement. Resources can be potentially saved if more inhibits are requested before interventions start.

Figure 5.2 shows the number of events per year and the fraction of each event type. Each year circa 180 DSS events occur on average. Most of them (ca. 63 %) are classified as interventions. The number of errors is approximately constant at circa 60 events per year. They account for circa a third of events. A small increase is observed for the years 2023 and 2024. The number of interventions varies over the years. They decrease from 121 events in 2018 to circa 110 events per year between 2019 and 2022. In 2023, the number of interventions increased significantly with respect to the previous years to 146 events, and decreased to 106 events the next year.

### 5.1.3 Differences across operation modes

The impact of the operation mode on the occurrence of errors and interventions is illustrated in Figure 5.3, showing the average number events per day classified into physics, YETS and Stop mode per year. Errors and interventions are displayed separately. The number of interventions per day highly depends on the operation mode of the detector. It further shows a large dependence on the year. On the contrary, the rate of errors is stable over the years in both YETS and stop



(a) Physics (b) YETS (c) Stop

Figure 5.3: The number of events per day in operation mode per year. Interventions and errors are shown separately. The plot excludes events that have zero impact to the detector, specifically events with only beam abort, sensor technical faults and radiation gate monitor alarms. The uncertainty is approximated using Gaussian propagation of uncertainty assuming no correlation. The data points are drawn with a slight offset in the $x$-axis.

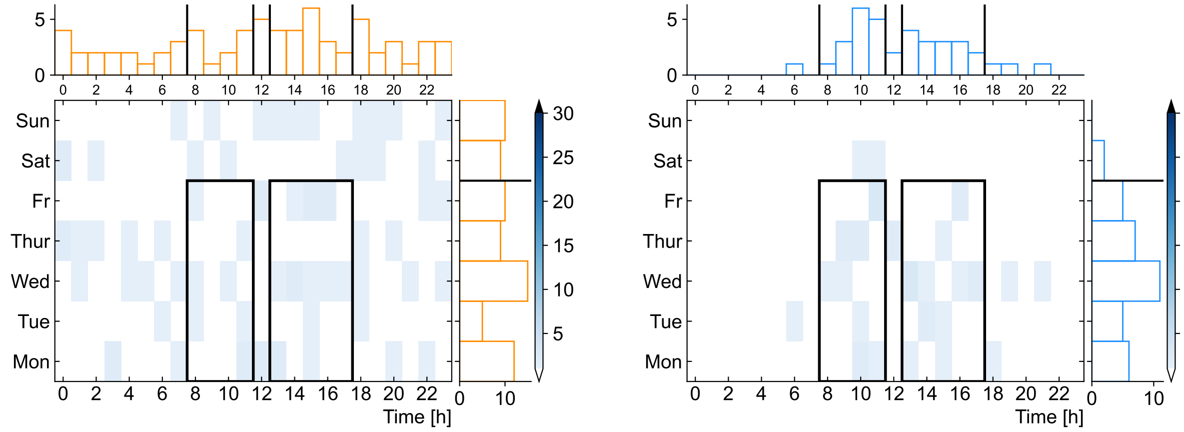
and decreases slightly during physics data-taking.

Fewer interventions and errors occur in physics mode compared to YETS or Stop. Since the start of Run 3, the rate of errors has increased every year. This is probably due to ageing equipment. For 2024, several errors were caused in summer by condensation in the surface rack rooms, that additionally increase the number of events to about one every five days. During YETS, at least twice as many interventions are observed compared to errors. The rate of intervention in YETS decreased since the start of the LS2 in 2018. At the beginning of LS2, the infrastructure, especially the electricity distribution, was upgraded causing several large interventions. This correlates with an increase of errors caused by the operation of new installations. During the LS2, the rate of interventions and errors was stable with about 0.3 interventions per day and 0.15 errors per day. With the end of LS2 and the start of Run 3 in 2022, the number of daily interventions doubled with respect to the previous years. This indicates that the return to physics mode required several adjustments to the services and infrastructure. The high activity in the cavern also correlated to an increase of errors in 2022. In 2023, a maximum 0.8 interventions per day is observed. The high activity is accredited to the extension of the 2023 annual shutdown as a consequence of high electricity prices which allowed for more maintenance activities. The error rate in YETS decreased in 2023 to the same level as during LS2. Following an extended shutdown period the year before, lesser maintenance was required during the YETS in 2024. In stop mode typically only urgent maintenance is conducted. The rate of interventions is lower than in YETS, with the exception of 2024 where several interventions on the cooling systems were required.

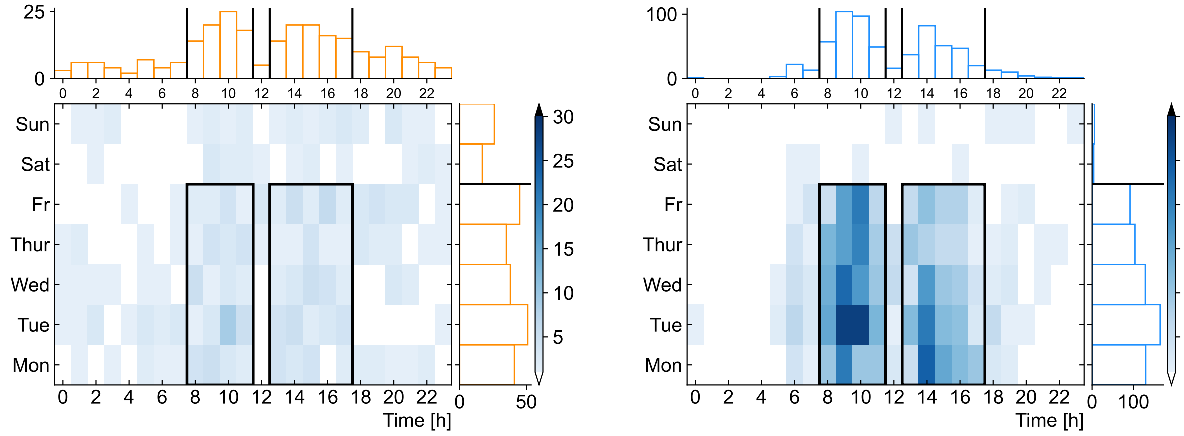
### 5.1.4 Time distribution of alarms

While the LHC and ATLAS are operated in physics mode 24/7, the core working hours for interventions are typically from 8:00 until 17:00 - with a one hour lunch break at 12:00. Nevertheless, urgent interventions can require activities outside of the regular working hours. Figure 5.4 shows the time distribution of errors and interventions per workday and hour for the three operation modes. Several errors seem to occur spontaneously, and are therefore uniformly distributed in time. In addition, errors are also caused by intervention as it can be seen from the increase of errors during working hours. Repairs and maintenance works are mostly performed at the beginning of the week and before the lunch break.

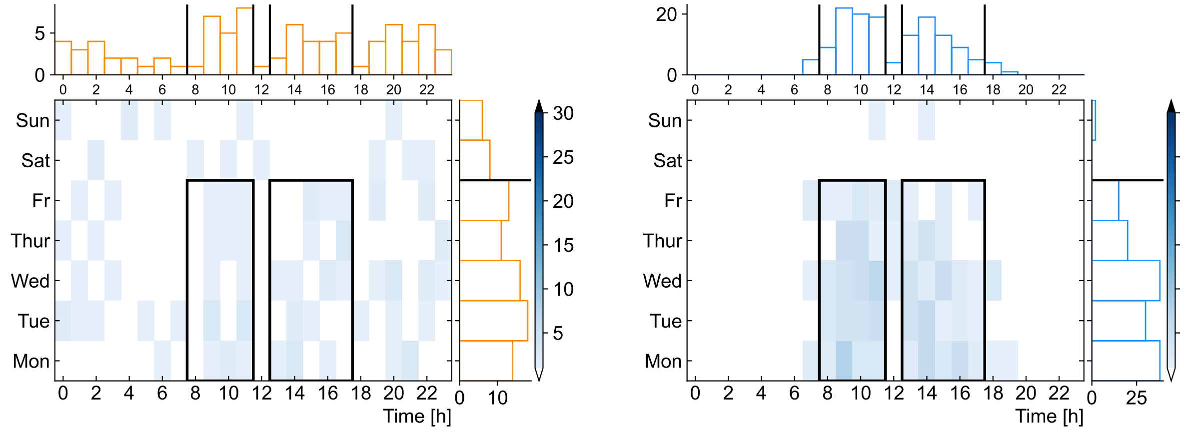
During physics operation, the few scheduled interventions are mainly done during working hours whereas errors occur at any time. During YETS, the higher activity in the cavern leads to more errors occurring during working hours. Interventions are mostly scheduled during regular working hours. Particularly critical or disruptive interventions, such as safety tests or maintenance on general electricity, are usually performed in the very early morning to reduce the impact on other systems. Towards the end of the week the activity decreases. The distribution of errors and interventions during Stop times is similar to the distribution in YETS. Any alarm outside of working hours or during run time is more likely due to an error than an intervention. The high volume of alarms triggered by interventions during working hours creates considerable background noise, potentially obscuring errors that tend to occur more frequently during these periods. Measures reducing the number of alarms caused by interventions are discussed in Section 5.2 and the outlook of this chapter.



(a) Physics - Error (b) Physics - Intervention



(c) YETS - Error (d) YETS - Intervention



(e) Stop - Error (f) Stop - Intervention

Figure 5.4: The time distribution of events since 2018. The occurrence of events is separately shown for errors (left) and interventions (right) for the three operation modes physics (top), YETS (centre) and stop (bottom). The z-axis is the total number of events per hour per day over the last six years. The marginal distributions per hour and per day are shown on the top and right of each Figure. Gray squares and lines show the regular working hours.

### 5.1.5 Distribution of alarms across sub-systems

The number of triggered alarms per sub-system is shown for the three operation modes separately for errors and interventions, in Figure 5.5. All alarms between 2018 and the 19 January are considered with the exception of BCM, technical fault, radiation gate monitor and ID flush alarms. Additionally, one event from 10 October 2024 is excluded as DSS failed during an intervention and every alarm was triggered at least once. Fortunately, this exceptional situation was recovered within half an hour.

In physics mode, only urgent repairs and maintenances are conducted. For almost all sub-systems most alarms are caused by errors. A notable exception are muon cooling alarms as 90% of these alarms were caused by interventions demonstrating that the system is not very stable. During run time, most errors occur due to water leaks followed by issues in the gas infrastructures. Both can be attributed to the ageing of the detector, and are a high priority of the technical management. For instance, new water leak sensors developed by the DSS team were installed before Run 3 to increase the monitoring where repairs of the cooling pipe joints are not possible. Errors in the magnet system triggered several alarms in physics mode, too. This is particularly worrisome as the magnet cannot be exchanged if it were to fail permanently. Evacuation alarms typically occur upon the detection of smoke, flammable gas or ODH that are caused by leaks due to ageing pipes.

In both YETS and Stop mode, significantly more alarms are triggered during interventions than by errors. As in physics mode, most alarms are triggered by muon cooling interventions. During YETS periods, cooling interventions trigger many cooling and water leak alarms. UPS alarms are almost always caused by maintenance works. In stop periods, most alarms are caused by urgent cooling, gas or electricity interventions. Apart from muon cooling, the ratio of errors to intervention is similar as in YETS. The distribution of errors to interventions per subsystem changes with the operation mode. Most alarms are triggered by muon cooling interventions. The findings demonstrate that particularly during maintenance periods, many alarms can be inhibited which heightens the attention for errors.

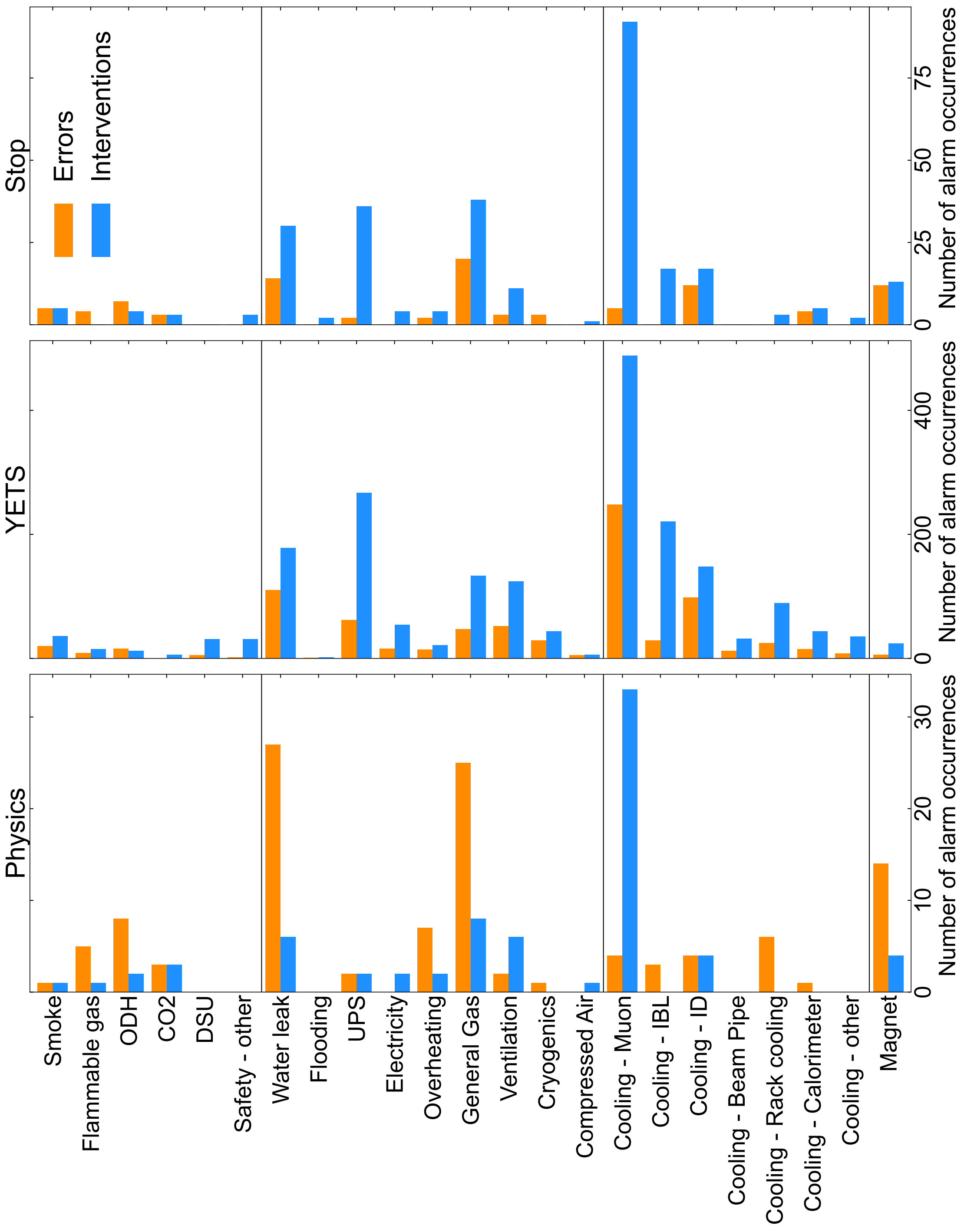
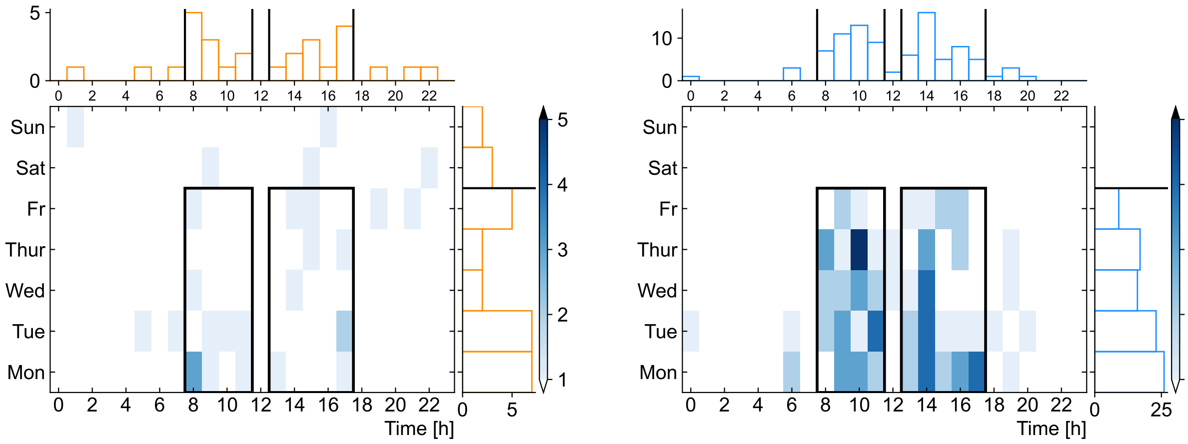


Figure 5.5: The number of triggered alarms grouped by sub-system and separated between errors and intervention for the three operation modes physics (left), YETS (centre) and stop (right). Please note the different x-axis scales. All alarms between 2018 and the 19 January are considered with the exception of BCM, technical fault, radiation gate monitor and ID flush alarms. Additionally, one event on 10 October 2024 is excluded during which all DSS alarms were triggered at least once.



(a)

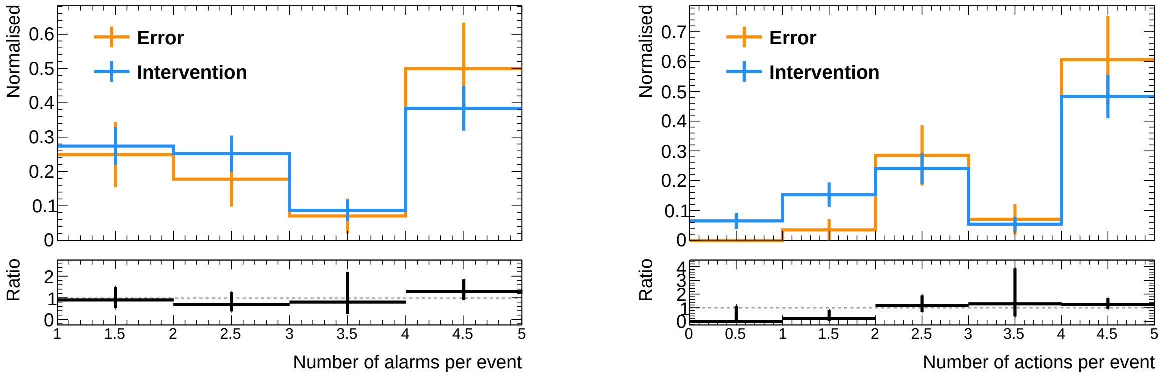
(b)

Figure 5.6: Time distribution of muon cooling errors (a) and interventions (b) for events in YETS mode. The $z$-axis is the total number of events per hour per day over the last six years. Each figure also shows the marginal distributions per hour (top) and per day (right), and the regular working hours in gray.

As discussed before, the muon cooling system is responsible for many alarms and demands significant attention from the operators. Up to one-third of these alarms are triggered by errors during YETS. Errors and intervention occur mostly but not exclusively during during working hours, as shown in Figure 5.6. Unfortunately, the number of alarms or actions per event does not allow a quick assessment on whether the alarm is caused by an error or an intervention since differences between errors and interventions are covered by the statistical uncertainties, see Figure 5.7.

## 5.2 The Alarm Helper tool

The analysis of DSS alarms between January 2018 and January 2025 identifies several trends related to the occurrence of alarms. The majority of alarms is caused by interventions, in particular interventions related to the cooling of the muon system. Improvements to the operational procedure of interventions are advisable since the follow-up on those alarms consumes significant resources and distracts the attention from alarms caused by errors. One improvement is the Alarm Helper tool. At its core, it is a database of all triggered DSS alarms and actions grouped into events which is accessible for several interfaces. The first implementation in 2023 was designed



(a) (b)

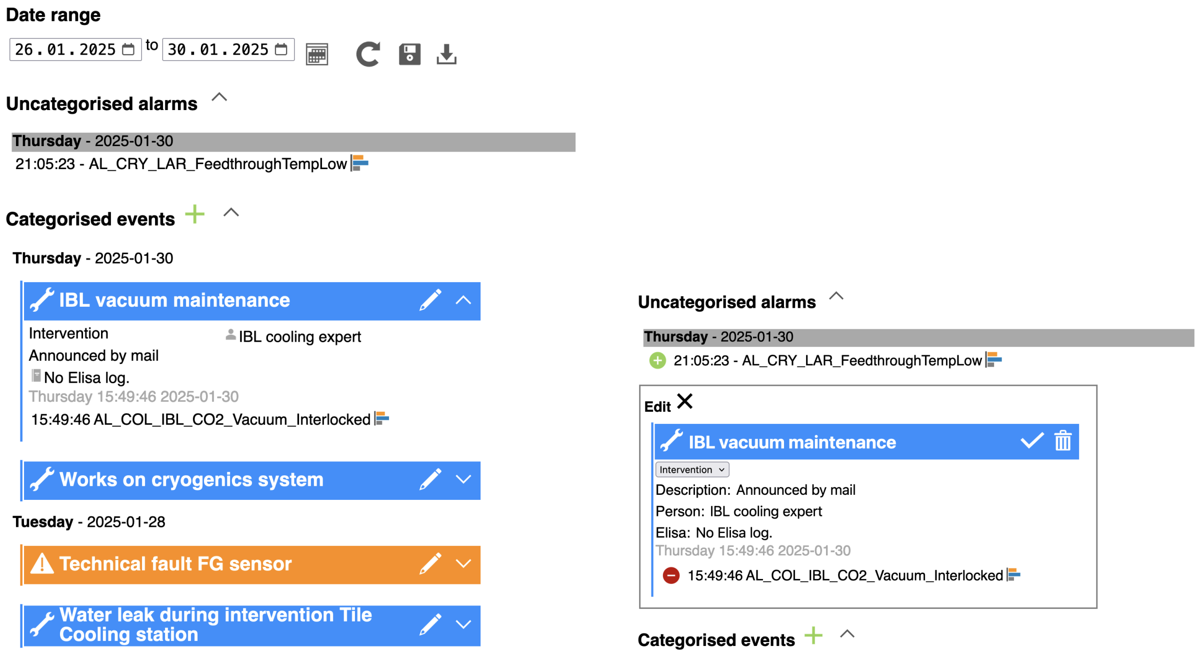
Figure 5.7: Normalised distribution of alarms (a) and actions (b) per muon cooling event. The last bin includes the overflow. The indicated uncertainty corresponds to the statistical uncertainty. The lower panels show the ratio of errors to interventions.

following the first alarm analysis including alarms between 2018 and 2023.

The current Alarm Helper features two main user interfaces that allow the insertion or retrieval of information about alarms and events. The tool is part of the Expert System project and accessible to all ATLAS members via a web browser. The Expert System OKS database schema is extended by two new classes which store and relate triggered alarms and actions to the Alarm and Action classes that are used by the inference engine. A third class is used to store events. The triggered alarms and actions are read from the logbook functionality of DSS that is available outside the ATLAS technical network.

### 5.2.1 Interfaces

**Logbook** Figure 5.8 shows an screenshot of the logbook interface. It enables grouping of DSS alarms and the corresponding triggered actions into events in a web-browser accessible database. Upon loading, the alarms and actions since the last Monday (and thus the last operational management meeting) are displayed. A custom time range is possible, too. Any not yet categorised alarms, and the actions they triggered, are shown on the top. Events can be created and deleted, and alarms can be assigned and removed by clicking the corresponding icons. The selected event type (error, intervention, unknown) is indicated by colour and icons



**Date range**

26.01.2025 to 30.01.2025

**Uncategorised alarms**

**Thursday** - 2025-01-30

21:05:23 - AL_CRY_LAR_FeedthroughTempLow

**Categorised events**

**Thursday** - 2025-01-30

**IBL vacuum maintenance**

Intervention

IBL cooling expert

Announced by mail

No Elisa log.

Thursday 15:49:46 2025-01-30

15:49:46 AL_COL_IBL_CO2_Vacuum_Interlocked

**Works on cryogenics system**

**Tuesday** - 2025-01-28

**Technical fault FG sensor**

**Water leak during intervention Tile Cooling station**

**Uncategorised alarms**

**Thursday** - 2025-01-30

21:05:23 - AL_CRY_LAR_FeedthroughTempLow

**Edit**

**IBL vacuum maintenance**

Intervention

Description: Announced by mail

Person: IBL cooling expert

Elisa: No Elisa log.

Thursday 15:49:46 2025-01-30

15:49:46 AL_COL_IBL_CO2_Vacuum_Interlocked

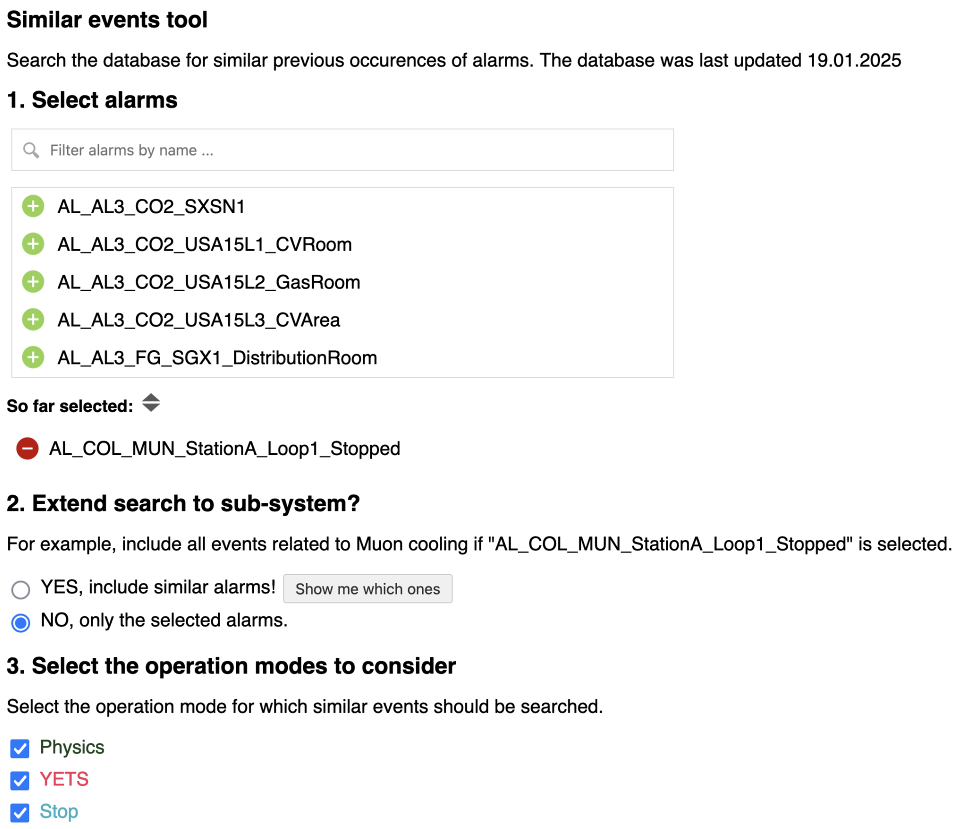
**Categorised events**

(a) View (b) Edit

Figure 5.8: Screenshot of the Alarm Helper logbook interface. (a) shows few events, categorised as interventions (blue, wrench icon) and errors (orange, warning sign). On the top one not assigned alarm is shown. (b) shows the editing view of an event. Icons allow creating and deleting of events, and assigning and removing alarms.

allowing a fast overview of the events. Additional information such as a description, contact persons and links to other databases such as the Elisa logbook [128] can be added as well. The user input is saved on the Expert System server. Once the categorisation has been discussed in the operational meeting, the events are inserted into the Expert System database and are available for other interfaces. Events can also be downloaded formatted as the weekly DSS report for the Operational Management meeting as a PowerPoint file, text file and JSON.

**Similarity of events** To asses the criticality of an alarm, the knowledge of similar situations, as presented by the similar event interface, is useful. Figure 5.9 shows the input section of this interface. Given a list of alarms, similar events and the points of contact defined in the SLIMOS handbook are presented. The search can be extended to alarms of the same sub-system, for example additional loops or the second side system when selecting a system that is divided between sectors A and C. Furthermore, the operation mode can also be selected. Similar events



**Similar events tool**

Search the database for similar previous occurences of alarms. The database was last updated 19.01.2025

**1. Select alarms**

Filter alarms by name ...

- ⊕ AL_AL3_CO2_SXSN1
- ⊕ AL_AL3_CO2_USA15L1_CVRoom
- ⊕ AL_AL3_CO2_USA15L2_GasRoom
- ⊕ AL_AL3_CO2_USA15L3_CVArea
- ⊕ AL_AL3_FG_SGX1_DistributionRoom

**So far selected:** ⇕

- ⊖ AL_COL_MUN_StationA_Loop1_Stopped

**2. Extend search to sub-system?**

For example, include all events related to Muon cooling if "AL_COL_MUN_StationA_Loop1_Stopped" is selected.

○ YES, include similar alarms! Show me which ones
● NO, only the selected alarms.

**3. Select the operation modes to consider**

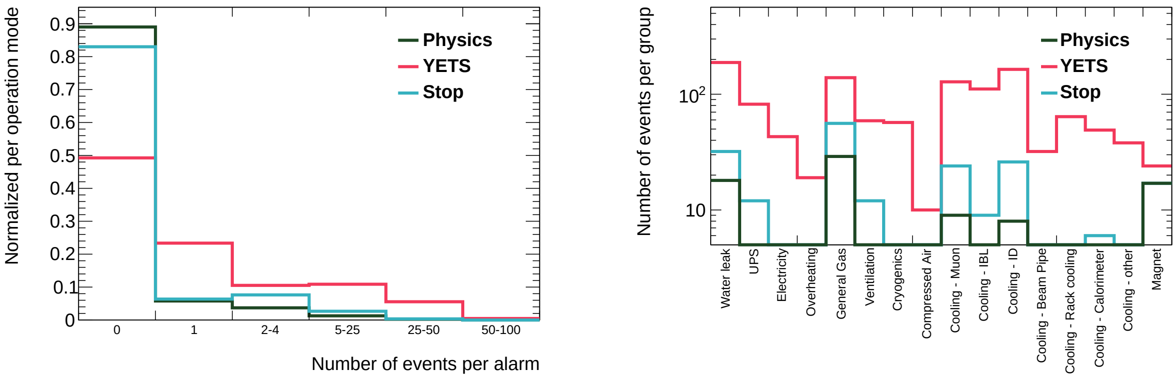
Select the operation mode for which similar events should be searched.

☑ Physics
☑ YETS
☑ Stop

Figure 5.9: Screenshot of the input section of the similar event tool with the alarm “AL_COL_MUN_StationA_Loop1_Stopped” selected as an example.

are presented as a list, similar to Figure 5.8a. Additionally, the ratio of errors to interventions and the list of previous contact persons, is shown as well.

The available statistics of similar events per alarm and sub-system per operation mode as of January 2025 is shown in Figure 5.10. Approximately 90 % of alarms never occurred during physics mode, and there are therefore no similar event in the database. Given that recent events in physics mode are more likely to be caused by errors they should be thus treated as such with the appropriate urgency. Similarly, most alarms never occurred in stop mode, most of the ones that do are related to the infrastructure and cooling. Most alarms are triggered in periods of maintenance and for every second alarm, the Alarm Helper can provide operators with similar events and help identifying true errors from interventions. Similar events are available for all sub-systems, particular water leak alarms, gas and cooling alarms.



(a) (b)

Figure 5.10: (a) events per alarm per operation mode, (b) and events per infrastructure and cooling sub-system. The DSS failure event is excluded.

## 5.3 Conclusion and outlook

This chapter described the analysis of DSS alarm occurrences and the Alarm Helper tool. The number of alarms has remained stable over the last years of operation. Nevertheless, significant upgrade and maintenance works are required to keep the detector in a good operating condition. Recently, ageing has caused an increase in alarms and demand the attention of the operation team, and the recent increase of errors during the last three years of physics operation presents a trend worth noting. The analysis identified that the alarm frequency increases drastically during high activity maintenance periods and that the muon cooling system is responsible for most alarms are related to interventions and errors. These findings will have to be addressed with respect to the planned upgrades of the detector for the High Luminosity LHC period to ensure optimal detector performance.

The Alarm Helper has eased and accelerated follow up of alarms and the identification of possible causes is in daily use. The database has increased the transparency of operational knowledge, especially during repair and maintenance periods. The realisation that the majority of alarms are caused by interventions has highlighted the necessity to change the procedure of announcing interventions and the inhibition of expected alarms. One possible solution is a forecasting tool, that generates the list of expected alarms given a description of an intervention which in turn can then be inhibited, and reducing the number of alarms drastically. The tool would require the connection of the Alarm Helper database containing the description of past interventions with the Intervention Management Planning and Coordination Tool (IMPACT) [129]. Furthermore, in the future, the alarm notification could be adapted to contain an automatically generated overview of past occurrences and their cause, helping the operators events to identify the cause. Moreover, the documentation of alarms and DSS could be enhanced by adapting DSS to also log the triggering of inhibited alarms. This could, for example, be used to generate a weekly review of expected and occurred alarms, and update the database with intervention, for which alarms were inhibited correctly.

# 6 Reconstruction and identification of physics objects

This chapter describes the conversion of raw detector signals to “physics objects”, particularly $b$-jets and $\tau$-leptons, that are used in the $HH \rightarrow b\bar{b}\tau^{+}\tau^{-}$ analysis [12]. While the section names refer to particles, the same procedures are applied to the corresponding anti-particles.

Most particles that result from the LHC collisions produce signals in the various sub-detectors. These signals are used to reconstruct the particles from low-level objects such as tracks of charged particles and vertices in the ID, MS tracks as well as clusters of calorimeter signals. The low-level objects are matched to the detector signatures of particles such as electrons, muons, $\tau$-leptons or jets from the hadronisation of quarks or gluons. They are subsequently used to define high-level physics objects that correspond to the underlying particle. Dedicated algorithms per particle type are applied to the reconstructed object to obtain a calibrated estimate of (parts of) the four-momentum and other properties of the underlying true particle. The same reconstruction and identification algorithms are applied to reconstructed signals from collision data and simulation. The objects’ reconstruction and selection efficiency, energy and momentum scales, and resolutions are calibrated with data, simulations, and differences in performance between data and simulation. Data and simulated events are also used to estimate the rate of mis-identified objects, referred to as “fakes”, by matching the truth-level record of simulated objects to the detector level objects.

## 6.1 Charged particle tracks and vertex reconstruction

The curved trajectory of charged particles inside the magnetic field of the detector is reconstructed from the energy deposits they leave in the tracking sub-detectors, referred to as "hits". Those reconstructed "tracks" are important inputs to the reconstruction of interaction vertices and leptons, but also for the identification of $b$-jets and hadronically decaying $\tau$-leptons.

Pattern recognition algorithms match the hits in the IBL, Pixel, SCT and TRT sub-detectors to tracks that are compatible with the hypothesis of a charged particle traversing the magnetic field [130]. Tracks from particles originating from the hard scatter interaction and from short lived particles are reconstructed by an iterative "inside-out" algorithm [131]. Seeded by three hits in the Pixel and SCT detector, the algorithm extrapolates a track by iteratively adding Pixel and SCT hits. Ambiguities in the assignment of hits to tracks are resolved with track quality criteria such as the number and origin of hits. Tracks are then ranked by their quality by a $\chi^2$-fit, and high quality tracks are extended to the TRT. Tracks of charged particles that are produced in the decay of long-lived particles such as $b$-hadrons are reconstructed with an "outside-in" algorithm starting with TRT measurements. Physics analyses use tracks with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and that are within the acceptance of the ID ($|\eta| < 2.5$) and that fulfil certain quality criteria, such as having a minimum number of hits in the Pixel and SCT detectors [131].

Vertices, i.e. the locations of interactions, are iteratively reconstructed from intersections of tracks [132, 133]. A vertex seed is determined from all good-quality tracks, and the vertex position is identified once all incompatible tracks have been removed. The procedure is repeated until the vertex passes quality criteria, and no more vertices can be built from the rejected tracks. The hard-scatter vertex, also referred to as the primary vertex (PV), is determined as the vertex with the highest sum of $p_{\mathrm{T}}^2$ among all vertices that are compatible with the bunch crossing area and that have at least two tracks with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ associated. All other primary vertex candidates are considered to be pileup interactions. Vertices outside the bunch crossing area are referred to as "secondary" vertices.

## 6.2 Calorimeter clusters

The segmentation of the calorimeter systems into lateral and longitudinal cells in $R \times \eta \times \phi$ (see Chapter 4) provides a directional energy measurement. The cell signals are used to reconstruct electromagnetic and hadronic showers that deposit energy across multiple cells. A topological cluster algorithm groups the energy deposits that originate from a single particle shower [134]. Starting with the locally most significant energy deposit with respect to the cell noise, adjacent cells are added to the cluster if they recorded significant energy deposits.

By default, topological clusters are calibrated to the cells' response to electromagnetic signatures, referred to as the "EM scale". An alternative "local hadronic calibration" accounts for the different cell response to hadronic showers and the local calorimeter density. The energy is estimated considering the energy density of the cluster as well as the shower width and depth. Both calibrations are used for example in the reconstruction of hadronically decaying $\tau$-leptons, which uses topological clusters that are calibrated using a mixture of both methods.

## 6.3 Electrons

Electrons leave tracks in the ID, and deposit their energy in a narrow cluster in $\eta \times \phi$ in the ECAL. Both sub-detectors are used to reconstruct them in the central detector region ($|\eta| < 2.47$). A small transition region between the barrel and the end-caps ($1.37 < |\eta| < 1.52$) is excluded. Electrons in the forward region are not considered by the $HH \to b\bar{b}\tau^+\tau^-$ analysis.

The electron reconstruction [135] starts with a topological cluster that has been calibrated to the EM scale. ID tracks are geometrically matched to this seed cluster. The matching algorithm considers bremsstrahlung effects and interactions with dead detector material. The matched tracks and the topological cluster are used to define "super clusters", which account for additional "satellite clusters" from bremsstrahlung photons [136].

In the context of physics analyses, electrons are typically required to be "prompt", meaning that they originate from the primary vertex. Reconstructed electrons are required to have $p_{\mathrm{T}} > 7\,\mathrm{GeV}$, and be within $|\eta| < 2.47$, excluding the transition region between the calorimeter barrel and end-cap at $1.37 < |\eta| < 1.52$. The electron identification algorithm features loose, medium, and tight working points, each offering a different balance between signal efficiency and the rejection of fake electron signatures. These "fake" electrons can be caused by mis-reconstructed and mis-identified jets or by non-prompt electrons, such as electrons from photon conversion ($\gamma \to e^+e^-$) or electrons from semi-leptonic heavy-flavour hadron decays. The $HH \to b\bar{b}\tau^+\tau^-$ analysis considers electron candidates that pass the loose working point. On average, 93 % of true hard-scatter electrons are selected by those selection and quality criteria. Additional isolation criteria rely on the typically limited nearby activity in terms of tracks and energy deposit in a cone around a prompt electron. The $HH \to b\bar{b}\tau^+\tau^-$ analysis imposes the fixed cut loose isolation working point, which requires no nearby tracks or energy deposits within a cone of $p_{\mathrm{T}}$-dependent size around the electron candidate [135, 136]. The energy estimate of an electron candidate is obtained from the associated calorimeter clusters. The estimate is calibrated in dedicated $Z \to e^+e^-$ measurements. The $\eta$-$\phi$ direction is estimated from the ID track that best matches the calorimeter cluster.

## 6.4 Photons

Photons are reconstructed in parallel to electrons given their similar ECAL signatures[136]. Prompt photons are characterized by the absence of tracks in the ID and are identified by topological clusters without associated tracks. Converted photons are identified from a characteristic conversion vertex formed by two oppositely charged tracks inside the ID that are matched to two electromagnetic clusters. Although photons are not explicitly studied in this thesis, they are considered in the overlap removal procedure between physics objects.

## 6.5 Muons

Muons that are produced in a hard scatter LHC interaction are most of the time minimal ionising. This means that they leave tracks in the ID, but deposit only small amounts of energy in the calorimeter and traverse it completely. They leave hits in the muon chambers of the MS that covers up to $|\eta| < 2.7$. Most muon candidates are reconstructed by combining tracks from both the ID and the MS [137]. MS tracks are formed by identifying straight segments in individual muon chambers, which are then extrapolated to be approximately consistent with originating from the interaction point.

In total, five different methods are used to increase the efficiency of muon reconstruction, particularly for muons in regions of low MS efficiency ($|\eta| < 0.1$) caused by cables and support structures.

- "Combined" muon candidates are reconstructed by matching an MS track to an ID track. The combined tracks are refitted from the MS towards the ID while accounting for energy losses in the calorimeter. This method yields the most commonly used and purest muon reconstruction technique.

- The "inside-out" method increases the efficiency for low-$p_T$ muons and in regions with limited MS instrumentation. Instead of requiring a fully reconstructed MS track, this approach extends an ID track by associating it with hits in the MS.

- Muon candidates outside the ID coverage but within the MS coverage are reconstructed as "MS extrapolated". An MS track is extrapolated inwards without requiring a matching ID track.

- Candidates leaving the MS acceptance, for example given their low $p_T$, are reconstructed by the "segment-tagged" method. In addition to one ID track, at least one reconstructed track segment is required in the MS, making this method more stringent than the "inside-out" method. The properties of the muon candidate are estimated solely on the ID track.

- In the very central region, muons are reconstructed as "calorimeter-tagged". ID tracks are extrapolated into the calorimeter, and if the calorimeter response is consistent with that of a minimum ionising particle, the corresponding ID track is identified as a muon.

Similarly to electrons, prompt muons from the hard scatter vertex are distinguished from non-prompt and fake muon candidates using dedicated muon identification algorithms. Non-prompt muons arise from the decays of heavy-flavour hadrons, kaons or charged pions. They can often be identified through a displaced secondary vertex or a noticeable kink in the reconstructed trajectory. The loose identification working point used in this thesis considers muon candidates from all five reconstruction methods. The muon momentum is reconstructed from a combined fit of both track types. Muon candidates must have $p_{\mathrm{T}} > 7\,\mathrm{GeV}$, be within $|\eta| < 2.7$. Additional isolation criteria with respect to ID and MS tracks and calorimeter activity in the vicinity of the candidate improve the background rejection. Overall, the imposed criteria lead to about about 97 % of true muons to be identified [137, 138]. The muon reconstruction and identification are calibrated with $Z \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ measurements. In general, the uncertainties related to the muon reconstruction are small. Their energy is known at the per-cent level [138].

## 6.6 Jets

Quarks and gluons that are produced during the collisions hadronise into colour-neutral hadrons due to colour confinement. The decay of the hadrons causes collimated sprays of particles, referred to as "jets", which leave tracks in the ID and energy deposits in the calorimeter system. Reconstruction algorithms estimate the kinematic properties of the quark or gluon that initiated the jet by clustering the constituents into a jet candidate. To ensure theoretical robustness, the reconstruction must be infrared and co-linear safe, i.e. the jet definition is not altered when adding either low energy emissions or co-linear radiation to the initial parton.

In this thesis, jet reconstruction relies on objects defined by the particle flow algorithm [139]. The algorithm combines ID tracks and the topological calorimeter clusters to obtain the four-vectors that are grouped by the clustering algorithm jets. Charged objects are reconstructed from a mixture of ID tracks and calorimeter clusters. The associated calorimeter clusters are removed, while the remaining clusters are assumed to originate from neutral objects. This procedure improves the reconstruction with respect to the sole use of calorimeter clusters particular for charged hadrons with low-$p_{\mathrm{T}}$. Furthermore, pileup contributions are reduced by matching the tracks of charged objects to the primary vertex.

The jets used in this thesis are clustered by the anti-$k_{\mathrm{T}}$ algorithm [140], which uses four-vectors from charged and neutral objects as inputs. Starting with the object with the highest energy, the algorithm combines it with the nearest object as defined by the distance metric $d_{ij}$. The metric expresses the relative distance between two objects $i$ and $j$ in the rapidity-azimuthal $y$-$\phi$ plane as defined in Section 4.2, normalised by a variable radius parameter $R$. The distance is weighted by the squared inverse transverse momentum of the harder object:

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T}i}^2}, \frac{1}{p_{\mathrm{T}j}^2}\right) \times \frac{\left(\Delta y_{ij}^2 + \Delta\phi_{ij}^2\right)}{R^2}. \tag{6.1}$$

The clustering stops if the distance to the next object is larger than the inverse of the squared transverse momentum ($d_{ij} > \frac{1}{p_{\mathrm{T},\,i}^2}$). The object $i$ is then declared a jet and the procedure is iterated until all objects are clustered. Sufficiently isolated jets are reconstructed as cone-shaped jets with radius $R$ in the $y$-$\phi$ plane. In this thesis, jets are reconstructed with $R = 0.4$.

Jets are calibrated to mitigate experimental uncertainties arising from pileup contribution and variations in the calorimeter response. In addition, a calibration procedure ensures that the jet energy scale (JES) and resolution (JER) are consistent between simulation and data [139]. The calibration [141] begins by matching the jet direction to the primary vertex, as jets are initially reconstructed assuming the nominal interaction point. Subsequently, corrections are applied to the transverse momentum to account for the average pileup contribution. Then, a correction based on the jet area is applied, followed by additional corrections that account for the number of reconstructed PV candidates and the expected average number of interactions per bunch crossing.

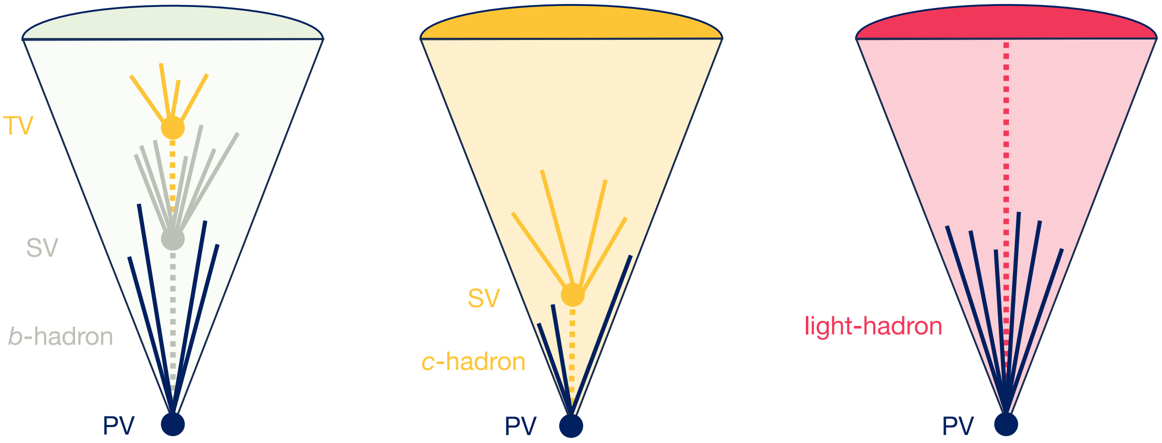
Using simulated events, the reconstructed jet energy is corrected to the particle level energy, and the jet direction is corrected to account for detector biases in the reconstruction. Further corrections address variations in calorimeter response across different jet flavours. To match jets measured in data to simulated jets, the jet energy scale is corrected in data by using scale factors that are obtained in reference measurements, where jets are balanced in the transverse plane against a well measured reference object. The energy resolution of simulated jets is smeared to match the resolution in data. Lastly, jets originating from pileup instead of the primary vertex are rejected using the Jet vertex tagger (JVT) algorithm [142].

Jets in the central region ($|\eta| < 2.5$) must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and pass the tight JVT working point if they have $p_{\mathrm{T}} < 60\,\mathrm{GeV}$. VBF $HH$ events are reconstructed with forward jets. These jets must be in the forward region ($2.5 \leq |\eta| < 4.5$), and have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. If they have $p_{\mathrm{T}} < 120\,\mathrm{GeV}$, they also must pass the loose JVT working point.

### 6.6.1 Flavour tagging

"Flavour tagging" denotes the identification of the flavour of the quark or gluon that initiated a jet. Algorithms exploit distinct features of jets originating from $b$-, $c$- and light quarks and gluons. While hadronically decaying $\tau$-leptons also produce jet signatures, their reconstruction and identification of $\tau$-leptons is discussed separately in Section 6.7. Any jet that passes an operating point of a tagging algorithm is labelled as "tagged", for example "$b$-tagged". In contrast, the term $b$-jet refers to the flavour label. In simulated events, jets are labelled by looking subsequently for $b$-hadrons, $c$-hadrons or $\tau$-leptons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ within a cone of $\Delta R = 0.3$ around the jet axis. Searching in this order, if a corresponding particle is found, the jet is labelled accordingly. Otherwise the jet is labelled a light-flavour jet.

Figure 6.1 illustrates the distinct signatures of jets originating from $b$-, $c$-, and light quarks, as well as gluons, produced at the LHC. Jets arising from the hadronisation of a $b$-quark into a $b$-hadron are referred to as $b$-jets. Hadrons containing a $b$-quark have a mass of approximately $5\,\mathrm{GeV}$ and a mean lifetime of about $1.5\,\mathrm{ps}$, corresponding to a proper decay length of roughly



(a) $b$-jet. (b) $c$-jet. (c) Light jet.

Figure 6.1: Illustration of the visible $b$-, $c$- and light jet signatures in the absence of a magnetic field.

450 µm before they typically decay into a $c$-hadron. This decay can sometimes produce a further displaced tertiary vertex. Typically, a $b$-hadron carries approximately three-quarters of the jet energy and its decay produces a large number of tracks originating from the secondary vertex. The hadronisation of a $c$-quark into a $c$-hadron produces a $c$-jet. With masses of about 2 GeV and lifetimes between 0.2 to 1 ps, $c$-hadrons have shorter decay lengths than $b$-hadrons. They typically carry only about half of the jet energy and decay into fewer and softer particles. In general, the constituents of $c$-jets are softer than those of $b$-jets. The term “light jets” refers to jets initiated by $u$-, $d$- or $s$-quarks or gluons. Light jets arising from quarks feature no secondary vertex as light hadrons are either stable or decay within the calorimeter system. Gluon initiated light jets feature a large number of tracks that exhibit a broad radiation pattern. Hadronically decaying $\tau$-leptons ($\tau_{\text{had}}$) produce jets similar to $c$-jets, owing to their mass of approximately 1.8 GeV and a mean lifetime of 0.3 ps. Most of the time, $\tau_{\text{had}}$-jets have either one or three charged tracks.

This thesis uses the DL1r tagger [143] to identify jets containing a $b$-hadron. DL1r is a deep neural network that combines information from low-level algorithms related to jet properties, including the displacement of tracks and vertices from the primary interaction point. The tagger provides several $b$-tagging efficiency working points, which correspond to the probability of correctly identifying a $b$-jet, evaluated on a simulated sample of 1-lepton $t\bar{t}$ events. The jets used by

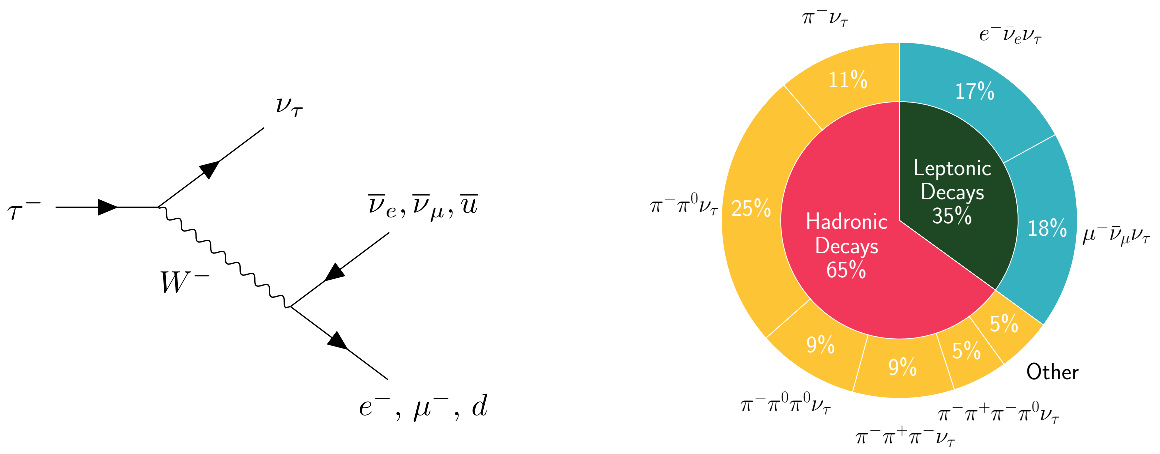
$HH \to b\bar{b}\tau^+\tau^-$ analysis are tagged using the 77 % working point, which corresponds to $c$- and light-jet rejection factors of 6 and 200, respectively. Two additional corrections are applied to $b$-tagged jets. Muon-in-jet correction accounts for muons from $b$- and $c$-hadron decays inside the jet. The muon momentum is added to the jet, and subtracted from the calorimeter energy. Furthermore, the jet $p_{\mathrm{T}}$ is corrected based on simulated jets. The correction depends on the $p_{\mathrm{T}}$ and the presence of nearby leptons [144, 145].

The ATLAS collaboration recently developed a new tagger that is based on a transformer architecture. This "GN2" tagger [146] processes track and jet information and does not rely on the input of low-level tagger information. The transformer architecture additionally allows the training of auxiliary tasks, such as the reconstruction of the internal jet structure. GN2 also enables an approximately 5 % increase in signal efficiency, while improving the $c$-jet rejection by about 30 % and keeping the light-rejection constant.

## 6.7 $\tau$-leptons

The $\tau$-lepton is the heaviest of the charged leptons, with a mass of 1.777 GeV [147]. It has a mean lifetime of approximately 0.3 ps, which allows a $\tau$-lepton with a transverse momentum of about 60 GeV to travel roughly three millimetres in the transverse plane before decaying. A leading order Feynman diagram of the decay is shown in Figure 6.2 next to an illustration of the branching fractions. In about 35 % of cases, $\tau$-leptons decay leptonically to an electron or muon and two neutrinos. The neutrinos cannot be detected due to their very small interaction cross-section with the detector material. Therefore, leptonically decaying $\tau$-leptons are reconstructed similarly to electron and muon candidates with additional contributions to the missing transverse energy.

In the remaining 65 % of cases, $\tau$-leptons decay hadronically, producing jet-like signatures in the detector. Most hadronic $\tau$ decays result in one or three charged pions or kaons, accompanied by up to two neutral pions, as illustrated in Figure 6.2. More rarely, $\tau$-leptons decay into final states containing up to five charged hadrons [147]. The reconstruction is done separately for decays with



(a) (b)

Figure 6.2: Feynman diagram (a) and branching fractions (b) of the $\tau$-lepton decay taken from Reference [147].

one or three charged hadrons, referred to as “1-prong” and “3-prong” decays, respectively [148]. As they are reconstructed from their visible decay products, they are referred to as $\tau_{\text{had-vis}}$ for the remainder of this thesis. The visible decay products are reconstructed from their topological calorimeter clusters that are calibrated to hadronic signatures. A jet is formed using the anti-$k_{\text{T}}$ algorithm with $R = 0.4$. Only jets within $|\eta| < 2.5$ and with $p_{\text{T}} > 20\,\text{GeV}$, excluding the transition region between the calorimeter barrel and end-cap at $1.37 < |\eta| < 1.52$, are considered as $\tau_{\text{had-vis}}$ candidates. The pronged decay structure is exploited by classifying the jet-associated tracks into “core” and “isolation” tracks, as illustrated in Figure 6.3. A Boosted Decision Tree classifies all tracks using information on their transverse momenta and their distance from the $\tau$-decay vertex (TV). The TV must be consistent with a PV candidate and is determined from all ID tracks with $p_{\text{T}} > 1\,\text{GeV}$ that lie inside the jet core. Once the TV is identified, the direction of the $\tau$-candidate is recalculated with respect to it. The number of core tracks is then used to label the $\tau$-candidate as either “1-prong” or “3-prong”. Only candidates with one or three charged-particle tracks and a total charge of $\pm 1$ are considered.

The energy of the $\tau_{\text{had-vis}}$-candidate is estimated from the calorimeter response, which is corrected for pileup contributions. A Boosted Regression Tree determines the energy scale, using additional tracker information which improves the calibration at low transverse momenta. 1- and 3-prong

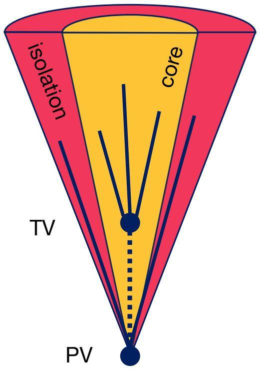


Figure 6.3: Illustration of the reconstruction of the visible decay products of hadronically decaying $\tau$-lepton in the absence of a magnetic field. The sketch shows the division of the jet candidate into a core and an isolation region. The jet starts from the primary vertex (PV). Additionally, the $\tau$-lepton decay vertex (TV) as well as three emerging tracks are indicated.

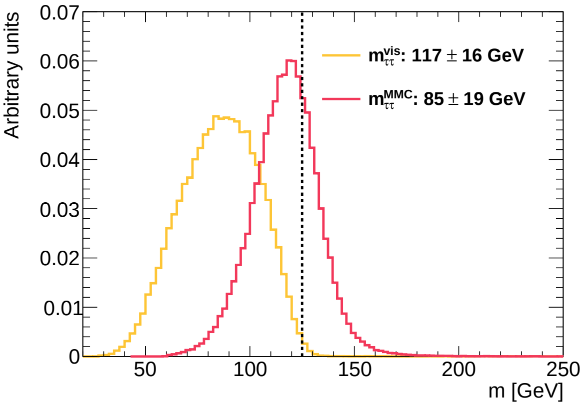


Figure 6.4: Reconstructed mass of the di-$\tau$-pair, using the Missing Mass Calculator ($m_{\tau\tau}^{\mathrm{MMC}}$), and using the visible decay products ($m_{\tau\tau}^{\mathrm{vis}}$) for the ggF $HH$ sample in the high-$m_{HH}$ signal region in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The black dashed line indicates the Higgs boson mass.

$\tau$-candidates are identified separately. Fake-$\tau_{\text{had-vis}}$ candidates from quark or gluon jets are suppressed separately for 1- and 3-prong candidates using algorithms that are based on Recurrent Neural Networks (RNNs). These networks utilise both low- and high-level variables from the Inner Detector and the calorimeter, as well as kinematic properties of the associated tracks and cluster and are trained with simulated $Z/\gamma^* \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ events [149]. Both RNNs feature four working points, candidates must pass the loose identification working points in the $HH \to b\bar{b}\tau^+\tau^-$ analysis. The 1-prong loose working point corresponds to a 85 % signal efficiency at a background rejection rate of 21. For 3-prong candidates, the signal efficiency is 75 % at a background rejection rate of 90. For 1-prong candidates, an additional BDT is used reject true electrons faking $\tau$-leptons.

The invariant mass $m_{\tau\tau}$ of a particle decaying into a pair of $\tau$-leptons ($X \to \tau^+\tau^-$) is reconstructed using the Missing Mass Calculator (MMC) [150]. The algorithm accounts for the presence of the one or two non-detectable neutrinos per $\tau$-lepton decay by constraining their allowed kinematic phase space separately for leptonic and hadronic 1- or 3-prong decays. The MMC significantly improves the mass reconstruction compared to simpler methods that either only use the visible decay products or assume that the neutrinos are emitted collinear to the visible decay products. Figure 6.4 compares the reconstructed di-$\tau$ mass using the MMC ($m_{\tau\tau}^{\text{MMC}}$) with the visible mass ($m_{\tau\tau}^{\text{vis}}$) in the gluon-gluon fusion (ggF) HH sample in the high-$m_{HH}$ signal region for the $\tau_{\text{had}}\tau_{\text{had}}$ channel of the analysis. The MMC improves the mass scale by 27 % and the resolution by 13 %.

### 6.7.1 anti-$\tau_{\text{had-vis}}$

Figure 6.5: Partial requirements for $\tau_{\text{had-vis}}$ and anti-$\tau_{\text{had-vis}}$ candidates.

The background estimation of fake-$\tau_{\text{had-vis}}$ candidates (see Section 7.9) defines anti-$\tau_{\text{had-vis}}$ candidates by inverting the offline $\tau_{\text{had-vis}}$ identification criteria in dedicated control regions. Anti-$\tau_{\text{had-vis}}$ candidates are only selected for events with fewer $\tau_{\text{had-vis}}$ candidates passing the offline $\tau_{\text{had-vis}}$ identification than required (one in $\tau_{\text{lep}}\tau_{\text{had}}$ channel, two in the $\tau_{\text{had}}\tau_{\text{had}}$ channel). In channels for which the $\tau_{\text{had-vis}}$ is identified at trigger level, anti-$\tau_{\text{had-vis}}$ objects must be matched to the $\tau_{\text{had-vis}}$ object at trigger level. Anti-$\tau_{\text{had-vis}}$ candidates must pass the same criteria as $\tau_{\text{had-vis}}$, including the trigger matching described in Section 7.5, except that they fail the loose identification working point, but have a RNN score $> 0.01$, as illustrated in Figure 6.5. The minimum RNN score requirement ensures that primarily quark jets are similar to $\tau_{\text{had-vis}}$ candidates and that the composition of quark- and gluon-initiated jets is similar to that of the signal region.

## 6.8 Missing transverse energy

Neutrinos are the only SM particles that cannot be detected by the ATLAS detector due to their extremely small interaction cross section. They can, however, be inferred from the momentum imbalance in the transverse plane. The colliding partons carry to the best approximation no initial transverse momentum. Therefore, the sum of all transverse momenta of particles must be zero. The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is thus estimated from the vectorial sum of all transverse momenta of all reconstructed and calibrated objects, namely photons, electrons, muons, jets and the visible decay products of hadronically decaying $\tau$-leptons, and an additional term accounting for soft radiation. Anti-$\tau_{\text{had-vis}}$ candidates are treated as $\tau_{\text{had-vis}}$ candidates.

$$E_{\mathrm{T}}^{\mathrm{miss}} = |\,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\,| = |\,-\vec{p}_{\mathrm{T}\,\mathrm{soft}} - \sum_{i \in \mathrm{objects}} \vec{p}_{\mathrm{T}\,i}\,| \tag{6.2}$$

The soft radiation is calculated from the transverse momentum of all well reconstructed ID tracks that are associated to the primary vertex, but not to any object. The calculation of the missing transverse energy is difficult, since it depends strongly on the correct reconstruction and identification of objects and the correct calculation of their transverse momenta or energy.

## 6.9 Overlap removal

Different physics objects are reconstructed independently from one another. To prevent double counting in cases where objects partially or fully overlap, an overlap removal procedure is applied. This defines a prioritisation scheme based on object type and reconstruction quality. For most physics objects, the analysis follows the standard ATLAS procedure and the standard working point [151, 152], using the rapidity-based angular distance $\Delta R_y = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$. Objects are rejected if they can be geometrically matched to another, higher priority object, or if they share charged particle tracks. The $HH \to b\bar{b}\tau^+\tau^-$ analysis employs three additional steps to ensure the priority of $\tau_{\text{had-vis}}$ over $b$-tagged jets over anti-$\tau_{\text{had-vis}}$ candidates over untagged jets.

1. reject any jet within $\Delta R_y < 0.2$ of a $\tau_{\text{had-vis}}$
2. reject any anti-$\tau_{\text{had-vis}}$ within $\Delta R_y < 0.2$ of $b$-tagged jet
3. reject any jet within $\Delta R_y < 0.2$ of a anti-$\tau_{\text{had-vis}}$

This priority has been optimised to give the largest signal yield as it has been shown to maximise the signal acceptance.

# 7 Search for non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state

This chapter presents the latest Run 2 search for non-resonant Higgs boson pair production via gluon fusion and vector boson fusion in the $b\bar{b}\tau^+\tau^-$ final state, which is published in in Reference [12]. Following an introduction and a general overview of the analysis, this chapter provides a detailed description of its individual components, including the event selection, the treatment of systematic uncertainties, and the statistical analysis, as well as the results and their cross-checks that support their robustness. The author's contributions, as one of the main analysts of the search, are emphasised, in particular the derivation of systematic uncertainties, the execution of the statistical analysis, and studies towards a verification measurement of the analysis strategy.

## 7.1 Introduction

The ATLAS and CMS collaborations analysed the LHC Run 2 dataset to search for evidence for Higgs boson pair production and the Higgs self-interaction, considering several final states as described in Section 3.4. Neither experiment observed any statistically significant deviation with respect to the background-only hypothesis.

This thesis describes the legacy ATLAS Run 2 search for $HH \to b\bar{b}\tau^+\tau^-$. It supersedes the first full Run 2 search [153], which comprised a search for non-resonant and resonant $HH$ production.

The legacy search re-analyses the dataset with the sole focus on enhancing the sensitivity for non-resonant $HH$ production. The search also optimised the categorisation of selected events and increased the sensitivity for ggF and VBF $HH$ production. Other improvements include the refinement of multivariate analysis techniques that separate signal from background events, as well as revised simulated predictions for the two main backgrounds, top-quark pair production and $Z$-boson production in association with heavy-flavour quarks ($Z$+HF). The increased size of the simulated samples reduces the impact of limited MC statistics and allows for the use of less conservative multivariate analysis techniques. The analysis sets the currently most stringent limits on the expected signal strength in the absence of Higgs boson pair production. The total uncertainty on the best fit value for the combined signal strength $\mu_{HH}$ of $\Delta_{\text{total}} \approx \pm 1.7$ arises mainly from the limited by the size of the available dataset ($\Delta_{\text{Data stat.}} \approx \pm 1.4$), and to a lesser extent from systematic uncertainties ($\Delta_{\text{Systs.}} \approx \pm 1.1$). The search improves the expected constraint for the coupling modifiers $\kappa_\lambda$ and $\kappa_{2V}$ by 11 % and 19 %, respectively, and provides the first isolated probe of VBF $HH$ production.

## 7.2 Analysis Overview

### 7.2.1 Event selection and categorisation

A variety of triggers selects events with exactly two $\tau$-leptons and two $b$-jets, which are used to reconstruct the decays of each of the Higgs boson into $H \to \tau^+\tau^-$ and $H \to b\bar{b}$, respectively. In total, the analysis defines three channels that are split into three signal regions (SRs) depending on the $\tau$-lepton decays, and one shared control region (CR). The CR improves the background estimate and targets $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$.

An overview of the event selection and categorisation is sketched in Figure 7.1. The $\tau_{\text{had}}\tau_{\text{had}}$ channel considers events with two oppositely charged, hadronically decaying $\tau$-leptons, and accounts for approximately 42 % of $\tau$-lepton pairs. Events were recorded by a combination of different single-$\tau_{\text{had-vis}}$ triggers (STTs) and di-$\tau_{\text{had-vis}}$ triggers (DTTs) that ATLAS used during

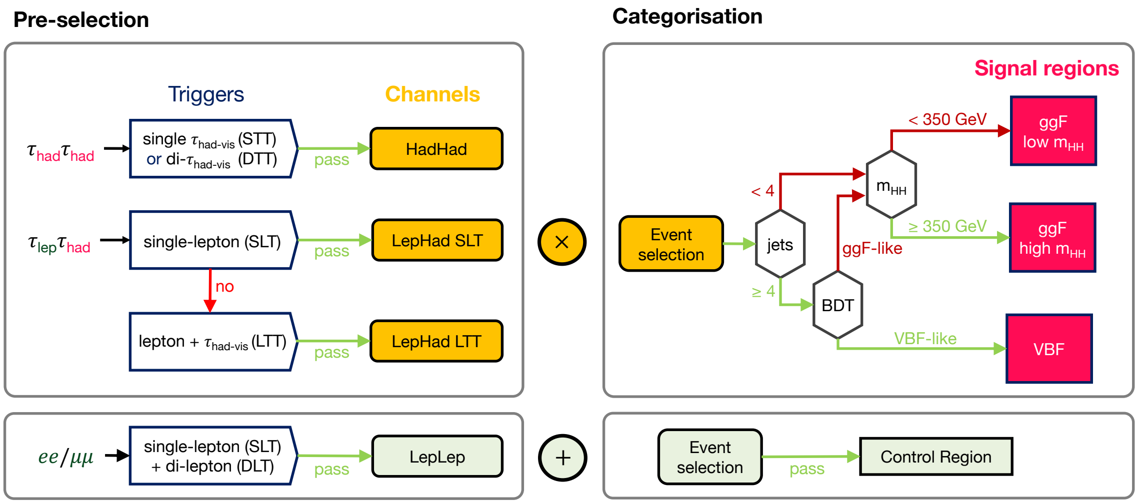


Figure 7.1: Sketch of the event selection and categorisation. Events are pre-selected into three channels depending on the decay of the two $\tau$-leptons and on which trigger the events pass. Following channel specific events selections, events are categorised whether they stem from ggF or VBF $HH$ production into three signal regions. The ggF signal regions are further divided at $m_{HH} = 350\,\text{GeV}$.

different data-taking periods of Run 2 [97, 154]. Events containing an electron or muon are vetoed in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. This ensures that events selected in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are not included in the two $\tau_{\text{lep}}\tau_{\text{had}}$ channels that account for the 46 % of $\tau$-lepton pairs that decay semi-leptonically. The $\tau_{\text{lep}}\tau_{\text{had}}$ channels select events with a pair of oppositely charged $\tau$-leptons, one of which decays leptonically, and the other hadronically. Events passing single-lepton triggers (SLT) are categorised into the SLT $\tau_{\text{lep}}\tau_{\text{had}}$ channel [155, 156]. Events that fail the SLTs, but pass lepton-plus-$\tau_{\text{had-vis}}$ triggers (LTT) are categorised into the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel to increase the signal efficiency [154]. Events for which both $\tau$-leptons decay leptonically are considered by another analysis [60] that uses a combination of single and di-lepton triggers (SLT, DTT).

Events must pass channel specific event selection criteria. They are then categorised based on their production mode into ggF and VBF SRs to provide a stand-alone estimate of VBF $HH$ production and to enhance the sensitivity for $\kappa_{2V}$. Events with fewer than four jets, meaning less than two additional jets accompanying the $H \to b\bar{b}$ decay, are categorised as ggF-like. All others are categorised as ggF or VBF-like by channel specific Boosted Decision Trees (BDT) that use mainly variables describing the jets in the events. The ggF-like events are further split into a low-and a high-$m_{HH}$ regions according to the reconstructed invariant mass of the Higgs boson pair at $m_{HH} = 350\,\text{GeV}$. $m_{HH}$ is defined by the four-vectors $p_\mu$ of the di-$\tau$ pair that is reconstructed with the MMC algorithm as described in Section 6.7, and of the two $b$-jets

$$m_{HH} = \sqrt{(p_\mu^{\tau\tau,\,\text{MMC}} + p_\mu^{b_1} + p_\mu^{b_2})^2}\ . \tag{7.1}$$

The split improves the sensitivity for $\kappa_\lambda$ as it allows differential information on the $m_{HH}$ spectrum to be accessed as described in Subsection 3.3.1. The low-$m_{HH}$ regions have an increased acceptance for large $\kappa_\lambda$ values that lead to a softer $m_{HH}$ spectrum. The high-$m_{HH}$ region targets SM-like $HH$ production, i.e. $\kappa_\lambda \approx 1$. The trigger and event selection is discussed in more detail in Section 7.5.

Figure 7.2 shows a sketch of the transverse view of a $\tau_{\text{lep}}\tau_{\text{had}}$ event in the ATLAS detector. One Higgs boson candidate is reconstructed from two close-by $b$-jets, the second from the visible decay products and $E_{\text{T}}^{\text{miss}}$ of a leptonically and one hadronically decaying $\tau$-lepton. The presence of additional jets is used to determine if the event is associated to ggF or VBF $HH$ production.

### 7.2.2 Main backgrounds

Several non-$HH$ processes can lead to the same final state of two $b$-jets and two $\tau$-leptons. The main backgrounds include the production of top-quark pairs ($t\bar{t}$), $Z$+HF and QCD multi-jets, as well as single top-quark and single Higgs boson production. With the exception of QCD multi-jets, all backgrounds are estimated from Monte Carlo (MC) predictions. To reduce the dependency on MC simulation, the $t\bar{t}$ and $Z$+HF cross-sections and the expected yields in the signal regions are determined in an auxiliary measurement in the control region. Furthermore, quark- or gluon-initiated jets can be misidentified as hadronically decaying $\tau$-leptons. Those so-called "fake-$\tau_{\text{had-vis}}$" can occur in jets from $t\bar{t}$ and QCD multi-jet production. Their simulation is challenging since it depends on the correct simulation of the detector response. Therefore, contributions from fake-$\tau_{\text{had-vis}}$ are estimated from data in dedicated control and validation

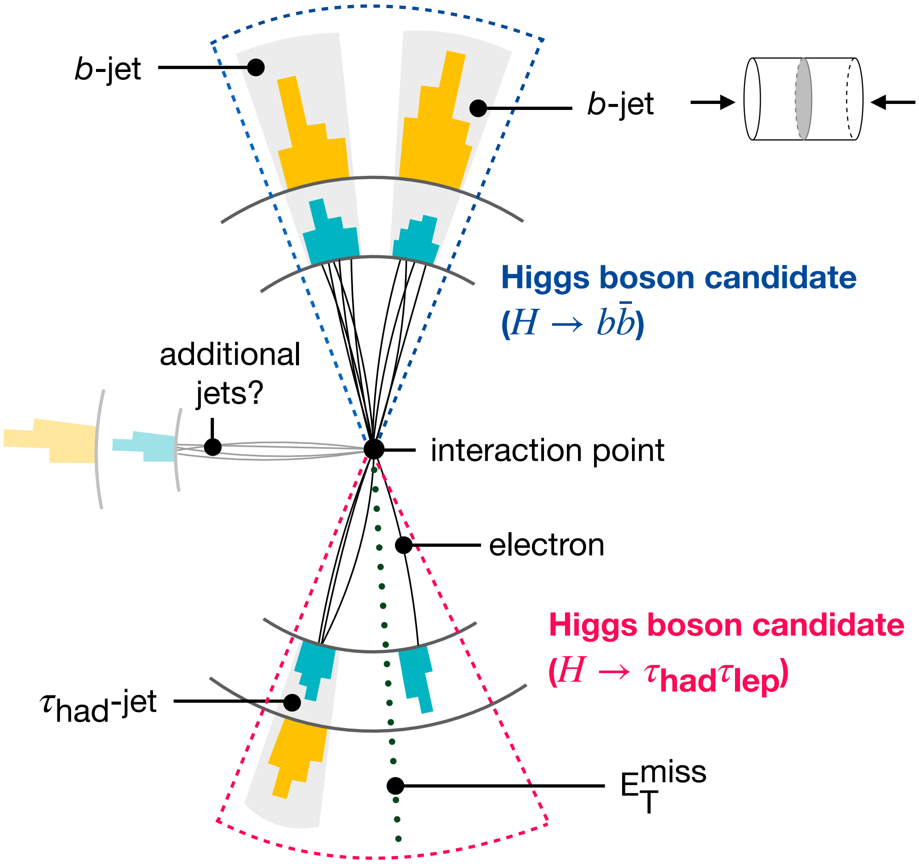


Figure 7.2: Sketched transverse view of an $\tau_{\text{lep}}\tau_{\text{had}}$ event. One Higgs boson candidate is formed from the $H \rightarrow b\bar{b}$ decay that is reconstructed from two $b$-jets. The other Higgs boson is reconstructed from the $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ decay. The hadronically decaying $\tau$-lepton is reconstructed as a jet from the visible decay products, particularly form the three charged tracks, and the calorimeter deposits. The leptonically decaying $\tau$-lepton is reconstructed from an electron and missing transverse energy $E_{\text{T}}^{\text{miss}}$. $E_{\text{T}}^{\text{miss}}$ receives contributions from both the leptonically and the hadronically decaying $\tau$-lepton. Black lines represent ID tracks, blue and yellow boxes depict energy deposits in the electromagnetic and hadronic calorimeters, respectively. The green dotted line shows the $E_{\text{T}}^{\text{miss}}$.

regions. The background estimation is discussed in detail in Section 7.9.

The background composition and relative contributions differ across the analysis channels and directly impact their experimental sensitivities. The $t\bar{t}$ background is approximately an order of magnitude smaller in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel than in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels, resulting in an increased sensitivity in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. This difference originates from the approximately seven times smaller branching fraction of $t\bar{t} \to b\bar{b}\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ compared to $t\bar{t} \to b\bar{b}\tau_{\mathrm{lep}}\tau_{\mathrm{had}} + E_{\mathrm{T}}^{\mathrm{miss}}$. The smaller branching fraction arises because, in many $t\bar{t}$ events, one of the two $W$ bosons from the $t \to Wb$ decay chain decays directly into an electron or a muon (and a neutrino), causing such events to populate the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ category.

### 7.2.3 Signal extraction

The trigger selection and event categorisation aim primarily at increasing the signal acceptance, while only secondarily providing a good signal to background ratio. Therefore, BDT-based classifiers separate the signal from the sum of backgrounds in each signal region, that are discussed in detail in Section 7.10.

A binned profile-likelihood fit of the BDT score in the nine signal regions and the invariant di-lepton mass $m_{\ell\ell}$ in the control region is used to extract the parameter of interest (POI) (see Section 7.12). The analysis is designed to search for evidence of Higgs boson pair production by measuring the signal strength $\mu_{HH}$. It is defined as the ratio of the measured and predicted cross-section $\sigma$ times branching fraction $\mathcal{B}$

$$\mu_{HH}(pp \to HH \to b\bar{b}\tau^{+}\tau^{-}) = \frac{[\sigma(pp \to HH) \times \mathcal{B}(HH \to b\bar{b}\tau^{+}\tau^{-})]_{\mathrm{measured}}}{[\sigma(pp \to HH) \times \mathcal{B}(HH \to b\bar{b}\tau^{+}\tau^{-})]_{\mathrm{predicted}}} . \tag{7.2}$$

In addition, the categorisation of events according to their production allows for separately constraining the signal strength of ggF and VBF $HH$ production, denoted as $\mu_{\mathrm{ggF}\ HH}$ and $\mu_{\mathrm{VBF}\ HH}$, respectively. Furthermore, the signal can also be parametrised in terms of $\kappa_{\lambda}$ or $\kappa_{2V}$ or a combination of thereof as discussed in Subsection 7.3.2.

## 7.3 Data and simulated event samples

### 7.3.1 Data

The search presented in this thesis analyses the full Run 2 proton-proton collision dataset of the LHC that was recorded by the ATLAS detector between 2015 and 2018 at a centre-of-mass energy of $\sqrt{s}$ =13 TeV. Recorded data events are required to have a fully operational ATLAS detector, stable LHC beams and must pass data quality criteria [157]. Selected events correspond to an integrated luminosity of $\mathcal{L} = 140.1 \pm 1.2\,\mathrm{fb}^{-1}$ [158, 159]. The algorithms used to trigger the events are discussed in more detail in Section 7.5.

### 7.3.2 Simulated Samples

Signal and background samples are simulated as described in Section 2.2. Simulated events are used both for the design of the search and for the signal extraction, a summary of all samples is given in Table 7.1. For most samples, the matrix element (ME) is calculated in terms of the strong coupling constant $\alpha_s$ up to next-to-leading order (NLO) in perturbation theory. The shower evolution and hadronisation are simulated by dedicated parton shower (PS) algorithms. The decay of bottom and charm hadrons is modelled with EvtGen [160] or Sherpa [161]. Simulated event yields are normalised to cross-section calculations, which are typically calculated to higher orders than the ME calculation. The mass of the Higgs boson is fixed to $m_H = 125$ GeV for the event simulation and the calculation of the Higgs branching fractions. The impact of the choice of the Higgs boson mass with respect to the best-fit values from References [162, 163] is negligible.

Generated events are passed through a simulation of the ATLAS detector using the Geant4 [164], and can therefore be directly compared to recorded data. The same reconstruction algorithms are used on simulated and recorded events.

| **Process** | **ME generator** | **ME QCD order** | **PDF set** | **PS and hadronisation** | **UE model tune** | **Cross-section order** |
|---|---|---|---|---|---|---|
| **Signal** | | | | | | |
| $gg \to HH$ (ggF) | Powheg Box v2 | NLO | PDF4LHC15nlo | Pythia 8.244 | A14 | NNLO FTApprox |
| $qq \to qqHH$ (VBF) | MadGraph5_aMC@NLO 2.7.3 | LO | NNPDF3.0nlo | Pythia 8.244 | A14 | N$^3$LO(QCD) |
| **Top-quark** | | | | | | |
| $t\bar{t}$ | Powheg Box v2 | NLO | NNPDF3.0nlo | Pythia 8.230 | A14 | NNLO+NNLL |
| $t$-channel | Powheg Box v2 | NLO | NNPDF3.0nlo | Pythia 8.230 | A14 | NLO |
| $s$-channel | Powheg Box v2 | NLO | NNPDF3.0nlo | Pythia 8.230 | A14 | NLO |
| $Wt$ | Powheg Box v2 | NLO | NNPDF3.0nlo | Pythia 8.230 | A14 | NLO |
| $t\bar{t}Z$ | Sherpa 2.2.1 | NLO | NNPDF3.0nnlo | Sherpa 2.2.1 | - | NLO |
| $t\bar{t}W$ | Sherpa 2.2.8 | NLO | NNPDF3.0nnlo | Sherpa 2.2.8 | - | NLO |
| **Vector boson + jets** | | | | | | |
| $W/Z$+jets | Sherpa 2.2.11 | NLO($\leq 2$ jets)<br>LO($3, 4$ jets) | NNPDF3.0nnlo | Sherpa 2.2.11 | - | NNLO |
| **Diboson** | | | | | | |
| $WW, WZ, ZZ$ | Sherpa 2.2.1 | NLO($\leq 1$ jets)<br>LO($2, 3$ jets) | NNPDF3.0nnlo | Sherpa 2.2.1 | - | NLO |
| **Single Higgs boson** | | | | | | |
| ggF | Powheg Box v2 | NNLO | PDF4LHC15nnlo | Pythia 8.212 | AZNLO | N$^3$LO(QCD)+NLO(EW) |
| VBF | Powheg Box v2 | NLO | PDF4LHC15nlo | Pythia 8.212 | AZNLO | NNLO(QCD)+NLO(EW) |
| $qq \to WH$ | Powheg Box v2 | NLO | PDF4LHC15nlo | Pythia 8.212 | AZNLO | NNLO(QCD)+NLO(EW) |
| $qq \to ZH$ | Powheg Box v2 | NLO | PDF4LHC15nlo | Pythia 8.212 | AZNLO | NNLO(QCD)+NLO(EW)$^\dagger$ |
| $gg \to ZH$ | Powheg Box v2 | NLO | PDF4LHC15nlo | Pythia 8.212 | AZNLO | NLO+NLL |
| $t\bar{t}H$ | Powheg Box v2 | NLO | NNPDF3.0nlo | Pythia 8.230 | A14 | NLO(QCD)+NLO(EW) |

Table 7.1: Generators used to simulate the signal and background processes. If not otherwise specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant ($\alpha_S$). The acronyms ME, PS and UE are used for matrix element, parton shower and underlying event, respectively. Details of the simulation of the signal and background samples are described in the text. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the $pp \to ZH$ process already includes the $gg \to ZH$ contribution. The $qq \to ZH$ process is normalised to the NNLO(QCD)+NLO(EW) cross-section for the $pp \to ZH$ process, after subtracting the $gg \to ZH$ contribution. Published in Reference [12].

Contributions from pileup are modelled by adding inelastic scatter events to the current, preceding and succeeding bunch crossings. Pileup events are simulated with the soft QCD processes of PYTHIA 8.186 [165] with the A3 set of tuned parameters [166] and the NNPDF2.3LO set of parton distribution functions (PDFs) [167]. The number of simulated pileup events is matched to the recorded pileup. Systematic uncertainties on the event simulation are estimated with additional samples from alternative generators, discussed in Section 7.11.2.

### Signal samples

Signal samples comprise of SM ggF and VBF $HH$ production simulated events. Other $HH$ production modes are omitted due to their significantly lower cross-sections. One Higgs boson is simulated to decay to a pair of $b$-quarks, the other to a pair of $\tau$-leptons.

SM ggF signal samples are simulated with the POWHEG BOX v2 generator [168] at NLO, assuming a finite top-quark mass using the PDF4LHC15_NLO_30_PDFAS PDF set [169]. Parton showers and hadronisation are simulated using PYTHIA 8.244 with the A14 tune and the NNPDF2.3LO PDF set. The simulated event yield is normalised to the SM cross-section calculated at NNLO times the $b\bar{b}\tau^+\tau^-$ branching fraction $\sigma_{\text{ggF }HH} \times \mathcal{B}(HH \to b\bar{b}\tau^+\tau^-) \approx 2.27\,\text{fb}$ [40, 42].

SM VBF signal samples are generated at leading order with the MADGRAPH5_AMC@NLO 2.7.3 [170] generator and the NNPDF3.0NLO [171] PDF set. PYTHIA 8.244 with the A14 tune and the NNPDF2.3LO PDF set are used to simulate the PS and hadronisation. The samples are normalised to the SM cross-section calculated at (N$^3$LO)times the branching fraction $\sigma_{\text{VBF }HH} \times \mathcal{B}(HH \to b\bar{b}\tau^+\tau^-) \approx 0.13\,\text{fb}$ [40, 43, 44].

### Background samples

**Top quarks** The simulation of top-quark samples assumes a top-quark mass of 172.5 GeV. Example diagrams for $t\bar{t}$ and single top-quark production are shown in Figure 7.3 and 7.4, respectively. In the signal regions, about ten times more events are simulated than expected in data. $t\bar{t}$ and single top-quarks production in the $Wt$ , $s$ and $t$-channels are simulated using

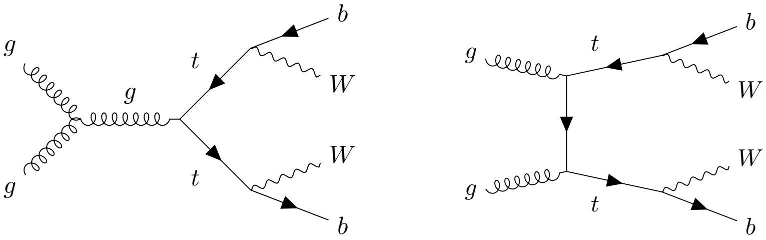


Figure 7.3: Example diagrams for $t\bar{t}$ production. Any of the two $W$ boson can decay to a $\tau$-lepton and a neutrino, yielding two true $\tau$-leptons. However, the decay $W \to qq'$ can be wrongly reconstructed as a hadronically decaying $\tau$-lepton, corresponding to a fake-$\tau_{\text{had}}$.

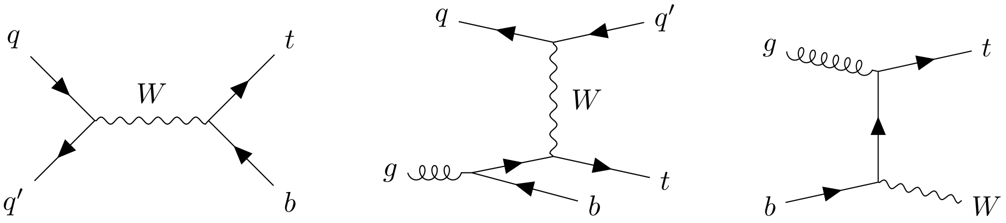


Figure 7.4: Example diagrams for single top-quark production in the $s$-channel (left), $t$-channel (middle) and $Wt$ channel (right).

the Powheg Box v2 generator with the NNPDF3.0nlo PDF set. The matrix elements are calculated to NLO in QCD. Parton showers and hadronisation are simulated with Pythia 8.230 [172] using the A14 set of tuned parameters [173, 174] and the NNPDF2.3lo PDF set. The top-quark decay is simulated using MadSpin [175] conserving spin-correlations. The $t\bar{t}$ cross-section is corrected to NNLO +NNLL, the single top-quark cross sections to NLO theory calculations. The interference between $t\bar{t}$ and single-top events in the $Wt$ channel is handled with the diagram removal scheme applied to single-top events [176].

$t\bar{t}$ production in association with vector-bosons is simulated using Sherpa version 2.2.1 [177] with multi-leg NLO merging for the $ttZ$ production, and Sherpa version 2.2.8 at NLO for the $ttW$ production. The NNPDF3.0nnlo set of PDFs is used with a dedicated parton shower tuning.

**Vector boson + jets** Figure 7.5 shows two example Feynman diagrams of vector boson production in association with jets. Samples are simulated using the Sherpa 2.2.11 generator. Matrix elements are calculated to NLO for up to two jets, and to LO for up to five jets with the

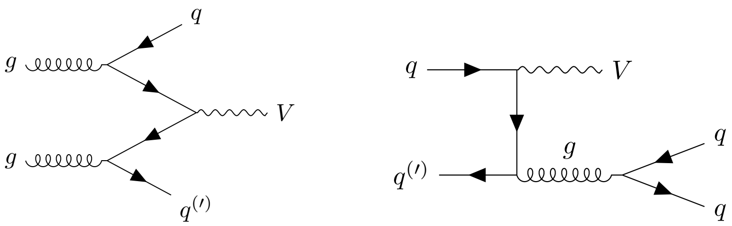


Figure 7.5: Feynman diagrams for V+jets production.

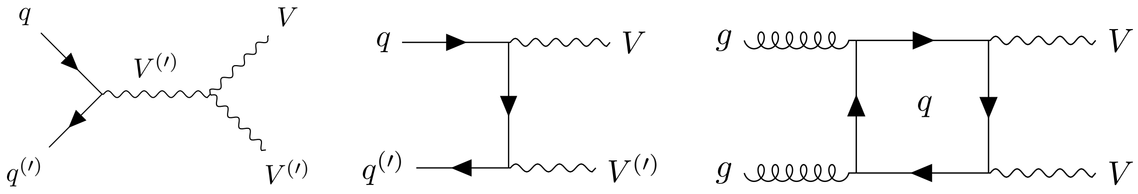


Figure 7.6: Feynman diagrams for diboson production ($VV^{(\prime)}$).

Comix [178] and OpenLoops [179–181] generator tools. The NNPDF3.0nnlo set of PDFs is used with dedicated Sherpa parton showers that are matched to the ME using the the MEPS@NLO prescription [182–185]. A later identified bug affecting the spin correlation of the $Z \rightarrow \tau^+\tau^-$ decay in Sherpa 2.2.11 was studied to have no impact on the results. However, the use of an smaller alternative Sherpa 2.2.1 sample would have doubled the statistical uncertainty. The Sherpa 2.2.11 sample features about 25 times more events in the signal regions than expected in data.

**Diboson** Feynman diagrams for diboson production are shown in Figure 7.6. Diboson events ($WW$, $WZ$, $ZZ$) are generated with Sherpa version 2.2.1 [161]. Both vector bosons are required to be on-shell. One boson is required to decay hadronically, the other leptonically. The NNPDF3.0nnlo PDF set is used with a dedicated Sherpa parton shower tuning. The cross-sections are normalised to NLO.

**Single Higgs boson** Single Higgs boson production is simulated separately for the different Higgs production mechanisms. Individual samples are generated for Higgs boson production in association with a top-quark pair ($t\bar{t}H$), a $Z$ boson ($ZH$), a $W$ boson ($WH$), and single

Higgs boson production from gluon fusion (ggF $H$) and from vector boson (VBF $H$). The corresponding Feynman diagrams are shown in Figure 3.1. Other SM Higgs boson processes have negligible contributions and are thus not considered. The matrix elements are calculated with the POWHEG BOX generator using the NNPDF3.0NLO PDF set.

The ggF $H$ and the VBF $H$ elements are calculated at NNLO and NLO accuracy, respectively. The parton shower and hadronisation are simulated with PYTHIA 8.212 using the AZNLO tune [186] and the CTEQ6L1 PDF set [187]. The ggF $H$ cross-section is corrected to N$^3$LO(QCD)+NLO(EW), the VBF $H$ to NNLO(QCD)+NLO(EW) calculations [40]. The $VH$ matrix elements are calculated to NLO. The same PS and hadronisation modelling as for ggF and VBF single Higgs production is used. The $WH$ and $qq \rightarrow ZH$ cross-sections are corrected to NNLO and NLO for QCD and EW terms, respectively [40]. The $gg \rightarrow ZH$ cross-section is correct to NLO+NLL in QCD [40]. $t\bar{t}H$ matrix elements are calculated to NLO accuracy with the NNPDF3.0NLO PDF set. PS and hadronisation are modelled using PYTHIA version 8.230 with the A14 set of tuned parameters and the NNPDF2.3LO PDF set. The cross-section is set to $ttH$ production NLO calculations [40].

### Alternative signal samples

Values of $\kappa_\lambda$ and $\kappa_{2V}$ that deviate from the SM prediction of 1 can significantly alter the cross-section and signal kinematics. Ideally, the modelling of signal samples would be available over a wide range of the the effective coupling modifiers. However, the generation of a fine grid of samples in the ($\kappa_\lambda$, $\kappa_{2V}$) parameter space is computationally expensive and demands significant resources. Therefore, ggF and VBF $HH$ signal samples for any value of $\kappa_\lambda$ and/or $\kappa_{2V}$ are instead obtained through interpolation using a set of basis samples.

**ggF $HH$** The ggF $HH$ cross-section from Equation 3.6 (Chapter 2) can be re-written as a quadratic expression of $\kappa_\lambda$ and a set of three weights $a_i$:

$$\sigma(\kappa_\lambda) = \kappa_\lambda^2 \times a_1 + \kappa_\lambda \times a_2 + a_3. \tag{7.3}$$

The weights are determined solving a set of three linear equations given by three known cross-sections $\sigma_i(\kappa_{\lambda_i})$ for different values of $\kappa_{\lambda_i}$. Using the set of $\kappa_\lambda \in \{0, 1, 20\}$ and the corresponding cross-sections, the cross-section (Equation 7.3) is expressed for any value of $\kappa_\lambda$ as the linear combination of three basis values

$$\begin{aligned} \sigma(\kappa_\lambda) = &\left(1 + \frac{\kappa_\lambda^2}{20} - \frac{399}{380}\kappa_\lambda\right) \times \sigma(\kappa_\lambda = 0) \;\; + \\ &\left(\frac{40}{38}\kappa_\lambda - \frac{2}{38}\kappa_\lambda^2\right) \times \sigma(\kappa_\lambda = 1) \;\; + \\ &\left(\frac{\kappa_\lambda^2 - \kappa_\lambda}{380}\right) \times \sigma(\kappa_\lambda = 20)\ . \end{aligned} \tag{7.4}$$

The same combination also allows interpolating differential cross-sections and thus signal kinematics. However, only the SM sample is available as a fully simulated and reconstructed sample, and two additional basis samples are generated by reweighing the SM sample. The variation of the signal kinematics in terms of $\kappa_\lambda$ is mostly captured by the distribution of $m_{HH}$. Thus, the two additional samples are obtained by reweighing the SM samples in terms of generator-level $m_{HH}$ [188]. In analogy to Equation 7.3, the differential cross-section per bin of $m_{HH}$ is parametrised as

$$\frac{\mathrm{d}\sigma(\kappa_\lambda)}{\mathrm{d}m_{HH}} = \kappa_\lambda^2 \times a_1(m_{HH}) + \kappa_\lambda \times a_2(m_{HH}) + a_3(m_{HH})\ , \tag{7.5}$$

where the weights $a_i(m_{HH})$ are determined in bins of $m_{HH}$ in the range from 200 to 1000 GeV with a granularity of 10 GeV using two alternative samples ($\kappa_\lambda = 0$ and 20). The two samples are generated with Powheg Box v2 and feature no parton showers and are not passed through the detector simulation. The fully simulated and reconstructed SM sample is reweighed in each $m_{HH}$ bin to $\kappa_\lambda = 0$ and 20, respectively, yielding additional two fully simulated and reconstructed basis samples. They are used alongside the SM sample by the fit model to parametrises the signal shape and kinematics for any value of $\kappa_\lambda$.

The relative contribution of each sample for different values of $\kappa_\lambda$ is shown Figure 7.7. The fully

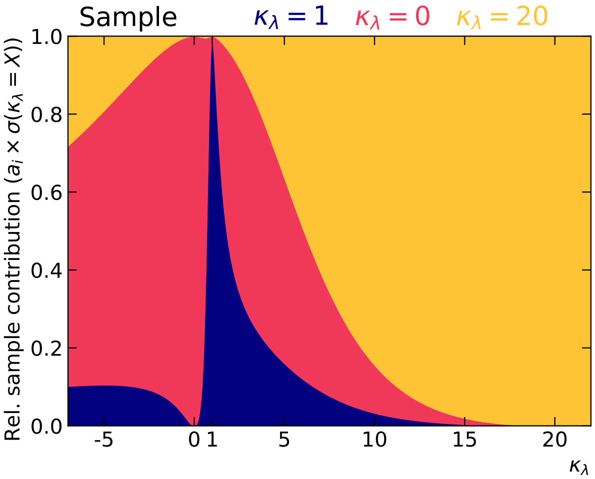


Figure 7.7: Relative contribution of the three samples as a function of $\kappa_\lambda$ in the range $-5$ to 25.

simulated SM sample dominates for values of the coupling modifier in the range $0.5 < \kappa_\lambda < 2$. For values outside this range, the reweighed basis samples dominate. The validity of this re-weighing procedure was assessed by comparing the interpolation to a fully simulated sample generated at $\kappa_\lambda = 10$ with POWHEG BOX v2 at NLO and HERWIG 7 [189, 190]. Overall, a good closure is observed. Residual differences between the interpolated and the $\kappa_\lambda = 10$ samples due the reweighed basis samples are used to derived systematic uncertainties on the reweighing procedure.

**VBF *HH*** Similarly, the VBF $HH$ cross-section can be expressed as a linear combination in terms of the coupling modifiers $\kappa_\lambda$ and $\kappa_{2V}$ that contribute at leading order

$$\sigma_{VBF}(\kappa_\lambda, \kappa_{2V}) = \kappa_\lambda^2 \times a_1 + \kappa_{2V}^2 \times a_2 + a_3 + \kappa_\lambda\kappa_{2V} \times a_4 + \kappa_\lambda \times a_5 + \kappa_{2V} \times a_6 \ . \tag{7.6}$$

The weights $a_i$ are determined from six linear equations in terms of six cross-sections values. For VBF, ten fully simulated and reconstructed VBF $HH$ samples are available for the combinations of ($\kappa_\lambda$, $\kappa_{2V}$) listed in Table 7.2. The signal parametrisation uses the first six samples, the remaining four are used to check the closure of the method. Good closure is observed and hence no residual uncertainties are applied to the interpolation procedure. The derivation and parametrisation is provided in Section A.1.

| | **Parametrisation** | | | | | | **Validation** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{\kappa_\lambda}$ | 1 | 1 | 1 | 2 | 10 | 0 | 1 | 1 | 1 | 0 |
| $\boldsymbol{\kappa_{2V}}$ | 1 | $\frac{1}{2}$ | 3 | 1 | 1 | 0 | 0 | $\frac{3}{2}$ | 2 | 1 |

Table 7.2: Combination of $\kappa_\lambda$ and $\kappa_{2V}$ values used for the alternative VBF $HH$ sample production.

**Single Higgs boson** Single Higgs boson production is susceptible to $\kappa_\lambda$, too. Since diagrams involving the $\kappa_\lambda$ only appear at NLO electroweak order, the effect is significantly smaller. Therefore, only the impact on the cross-sections and branching fractions are considered following Reference [191].

## 7.4 Object reconstruction and selection

$HH \to b\bar{b}\tau^+\tau^-$ events are reconstructed from physics objects that have in turn been reconstructed from the detector signatures recorded by the ATLAS sub-detectors as described in Chapter 6.

Out of all central jets, exactly two are required to pass the $b$-tagging working point. The $H \to b\bar{b}$ candidate is reconstructed from the sum of their four momenta. The $H \to \tau^+\tau^-$ candidate is reconstructed with di-$\tau$ system reconstructed by the MMC [150].

## 7.5 Trigger and event selection

This section describes the trigger and event selection in the three analysis channels. A variety of triggers are used to increase the acceptance for events with two $b$-tagged jets and two $\tau$-leptons. The $HH \to b\bar{b}\tau^+\tau^-$ signal is separated from backgrounds through event selection cuts and, ultimately, BDTs in each signal region.

### 7.5.1 Trigger selection

Events are required to pass the two staged ATLAS trigger system comprising of the hardware-based Level-1 (L1) and the software-based High Level Trigger (HLT) [96–98]. The online trigger algorithms are primarily designed to maximise the total identification efficiency. The trigger requirements, for example a cut on the $p_{\mathrm{T}}$ of an object, balance the total available bandwidth with the largest possible acceptance. Recorded events are fully reconstructed offline with more complex algorithms. The analysis requires that offline reconstructed physics objects are geometrically matched to the objects that passed the HLT triggers. The offline selection criteria are generally chosen to be more stringent, targeting regions where the triggers are well-calibrated and operating on the high-efficiency plateau. This approach ensures that the trigger efficiency is accurately modelled and minimizes systematic uncertainties.

The trigger algorithms were adapted throughout Run 2 to profit from improvements for example in object identification algorithms. In addition, the trigger thresholds were slightly raised to cope with the increasing instantaneous luminosity.

As described before, the analysis categorises events into three signal channels and one control region with orthogonal trigger selections. Two types of triggers with different HLT requirements are used. Triggers requiring the presence of one only object generally have higher thresholds on the $p_{\mathrm{T}}$ of the object, whereas triggers that require two objects can operate with lower thresholds without exceeding the maximal rate. Events are selected by a combination of $\tau_{\text{had-vis}}$, electrons and muon triggers, whose requirements are described in the following. Detailed information is provided in Reference [153].

**$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel** Events in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel are required to pass a combination of single-$\tau_{\text{had-vis}}$ triggers (STTs) and di-$\tau_{\text{had-vis}}$ triggers (DTTs) [97, 154]. STTs are fully efficient for $\tau_{\text{had-vis}}$ candidates with $p_{\mathrm{T}} > 100$ to 180 GeV, depending on the data-taking period. Candidates below these $p_{\mathrm{T}}$ cuts are covered by DTTs, allowing accessing the lower part of the $m_{HH}$ spectrum. The presence of electrons or muons is vetoed to ensure orthogonality with the two $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels.

STTs require at least one $\tau_{\text{had-vis}}$ at the HLT with a minimum $p_{\mathrm{T}}$ between 80 and 160 GeV, depending on the data-taking period. DTTs require a $\tau_{\text{had-vis}}$ pair at the HLT. The $\tau_{\text{had-vis}}$ with the higher $p_{\mathrm{T}}$ is referred to as the "leading" $\tau_{\text{had-vis}}$ and must have $p_{\mathrm{T}} > 35$ GeV. The other "sub-leading" $\tau_{\text{had-vis}}$ must have $p_{\mathrm{T}} > 25$ GeV. Requiring only two $\tau_{\text{had-vis}}$ candidates at trigger level with such $p_\mathrm{T}$ thresholds would exceed the trigger bandwidth. Thus, in order to reduce the trigger rates, additional requirements of one or two jets at the L1 trigger stage were imposed from 2016 onwards. No requirement is imposed on the jet flavour, as the requirement is purely chosen with respect to the trigger rate. During 2016 data-taking, an additional jet with $E_\mathrm{T} > 25\,\mathrm{GeV}$ was required, while in 2017 and 2018 the threshold depended on the offline jet multiplicity, using either a L1 trigger with two additional central jets with $E_\mathrm{T} > 12\,\mathrm{GeV}$ or one with a single jet of $E_\mathrm{T} > 25\,\mathrm{GeV}$ and $\Delta R(\tau_\text{had-vis}, \tau_\text{had-vis}) < 2.8$. This configuration ensures mutually exclusive triggers and removes the need for a combined efficiency evaluation.

To ensure that the events are selected near the trigger efficiency plateaus, offline $\tau_\text{had-vis}$ candidates are required to be within $\Delta R = 0.2$ of the corresponding online object. Additionally, the offline STT $p_\mathrm{T}$ threshold of 100 to 180 GeV is larger than online requirements. STT events must also feature a second offline $\tau_\text{had-vis}$ with $p_\mathrm{T} > 25\,\mathrm{GeV}$. For the DTTs, the leading and sub-leading offline $\tau_\text{had-vis}$ must have $p_\mathrm{T} > 40\,\mathrm{GeV}$, and 30 GeV, respectively. Furthermore, DTT events are required to have offline two jets with $p_\mathrm{T} > 45\,\mathrm{GeV}$, or one jet with $p_\mathrm{T} > 80\,\mathrm{GeV}$ and a maximal distance between the two offline $\tau_\text{had-vis}$ of $\Delta R = 2.5$. If an event passes both the STT and the DTT offline requirements, the STT requirements are used.

**$\tau_\text{lep}\tau_\text{had}$ SLT channel** Events that contain exactly one electron or muon and one $\tau_\text{had-vis}$ are selected either by single-lepton triggers (SLTs) [155, 156] or lepton+$\tau_\text{had-vis}$ triggers (LTTs) [154]. LTTs are only applied to events failing SLT requirements. Since they require an hadronic $\tau$-lepton in addition to the charged lepton, LTTs allow selecting a lower $p_\mathrm{T}$ lepton compared to SLTs. SLTs and LTTs require two additional jets at the L1 and HLT. Their $p_\mathrm{T}$ requirement is lower for SLTs than for LTTs.

SLTs require one electron or one muon at HLT. Depending on the data-taking period, a $p_\mathrm{T}$ threshold of 24 to 26 GeV is required for the electron, and of 20 to 26 GeV for the muon. The offline reconstructed objects must be geometrically matched to the HLT objects. Additionally, offline electrons and muons must pass $p_\mathrm{T}$ thresholds that are 1 GeV higher than the online thresholds such that only events on the trigger plateau are selected. One offline $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.3$ is required.

**$\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel** LTTs require either one electron with $p_{\text{T}} > 18\,\text{GeV}$ or one muon with $p_{\text{T}} > 15\,\text{GeV}$, and one $\tau_{\text{had-vis}}$ with $p_{\text{T}} > 25\,\text{GeV}$. For LTT events with low $p_{\text{T}}$ $\tau_{\text{had-vis}}$, additional jet requirements lower the LTT rate. The offline electrons, muons and $\tau$-leptons must be geometrically matched to the HLT objects. Additionally, offline electrons and muons must pass a 1 GeV higher offline $p_{\text{T}}$ threshold than online. Offline $\tau_{\text{had-vis}}$ must have $p_{\text{T}} > 30\,\text{GeV}$ and $|\eta| < 2.3$. This ensures that only events on the trigger plateau are selected.

**Control region** Events are required to pass a combination of single and di-lepton (DLT) triggers for electrons and muons as described in Reference [60]. The DLTs trigger on pairs of electrons, muons or a pair of one electron and one muon. Depending on lepton flavour and the data-taking periods, the SLT $p_{\text{T}}$ thresholds vary between 21 GeV, and 28 GeV. If no SLT criteria are met, DLTs with $p_{\text{T}}$ thresholds of 13 GeV for the higher $p_{\text{T}}$ lepton, and 9 GeV for the sub-leading lepton are used. The offline $p_{\text{T}}$ threshold are 1 GeV higher for electrons, and 5 % for muons.

### 7.5.2 Event selection

In all three channels, the charge of the $\tau_{\text{had-vis}}$, electron or muon must be opposite to the second $\tau_{\text{had-vis}}$. Events are required to have a $m_{\tau\tau}^{\text{MMC}} > 60\,\text{GeV}$ to reject $\tau$-lepton pairs from fakes, gluon and photon backgrounds. Furthermore, events must feature exactly two $b$-tagged jets passing the DL1r 77 % efficiency working point. To target $H \to b\bar{b}$ decays, the two $b$-tagged jets must be within $|\eta| < 2.5$, and have at least $p_{\text{T}} > 45\,\text{GeV}$ and $p_{\text{T}} > 20\,\text{GeV}$, respectively. In the two $\tau_{\text{lep}}\tau_{\text{had}}$ channels, the invariant mass or the $b$-jet pair must be $m_{bb} < 150\,\text{GeV}$ to reject $t\bar{t}$ events and to define a CR. In all three channels, the majority of events passing the trigger and event selection criteria are from top-quark ($t\bar{t}$, single top-quark), $Z$ + jets backgrounds, and backgrounds in which quark- or gluon initiated jets are falsely reconstructed as a $\tau_{\text{had-vis}}$. A summary of the the

| **$\tau_{\text{had}}\tau_{\text{had}}$ channel** | | **$\tau_{\text{lep}}\tau_{\text{had}}$ channels** | |
|---|---|---|---|
| **STT** | **DTT** | **SLT** | **LTT** |
| ***$e/\mu$* selection** | | | |
| No loose $e$ or $\mu$ | | Exactly one loose $e$ or $\mu$<br>$e$ passes tight identification criteria<br>$\mu$ passes medium identification criteria and $\lvert\eta\rvert < 2.5$ | |
| | | $e$: $p_{\text{T}} > 25$ - $27$ GeV | $18$ GeV $< p_{\text{T}} <$ SLT cut |
| | | $\mu$: $p_{\text{T}} > 21$ - $27$ GeV | $15$ GeV $< p_{\text{T}} <$ SLT cut |
| **$\tau_{\text{had-vis}}$ selection** | | | |
| Two loose $\tau_{\text{had-vis}}$ | | One loose $\tau_{\text{had-vis}}$<br>$\lvert\eta\rvert < 2.3$ | |
| $p_{\text{T}} > 100$ - $180$ $(25)$ GeV | $p_{\text{T}} > 40$ $(30)$ GeV | | $p_{\text{T}} > 30$ GeV |
| **Jet selection** | | | |
| $\geq 2$ jets with $\lvert\eta\rvert < 2.5$ | | | |
| Leading jet $p_{\text{T}} > 45$ GeV | Trigger dependent | Leading jet $p_{\text{T}} > 45$ GeV | Trigger dependent |
| **Event-level selection** | | | |
| Trigger requirements passed<br>Exactly two $b$-tagged jets (DL1r 77 % working point)<br>Opposite-sign electric charges between $\tau_{\text{had-vis}}$ and $e/\mu/\tau_{\text{had-vis}}$<br>Collision vertex reconstructed<br>$m_{\tau\tau}^{\text{MMC}} > 60$ GeV | | | |
| | | $m_{bb} < 150$ GeV | |

Table 7.3: Summary of the event selections, shown separately for events that are selected by different triggers. In cases where pairs of reconstructed objects of the same type are required, thresholds for the (sub-)leading $p_{\text{T}}$ object are given outside (within) parentheses where different event selection thresholds are applied. When the selection depends on the year of data-taking, the possible values of the requirements are separated by commas, except for the jet selection in the LTT and DTT triggers, which use multiple selection criteria. The trigger $p_{\text{T}}$ thresholds shown are applied to the offline physics objects that are matched to the corresponding trigger objects. Published in Reference [153].

event selection criteria in the signal channels is given in Table 7.3.

**$\tau_{\text{had}}\tau_{\text{had}}$ channel** In addition to the trigger and common event selection, both $\tau$-leptons must be oppositely charged and pass the corresponding loose identification criteria.

**$\tau_{\text{lep}}\tau_{\text{had}}$ SLT and LTT channels** Beyond the trigger and common event selection criteria, events must either have one electron passing the tight identification criteria or one muon passing the medium identification criteria. The $\tau$-lepton mass pass the loose identification criteria.

**Control region** In the control region, events must contain exactly two oppositely charged electrons or muons and two $b$-tagged jets. The invariant di-lepton mass $m_{\ell\ell}$ is selected to be in the range around the $Z$-boson mass peak, $75\,\text{GeV} < m_{\ell\ell} < 110\,\text{GeV}$. The invariant mass of the two $b$-tagged jets must be either $m_{bb} < 40\,\text{GeV}$ or $m_{bb} > 210\,\text{GeV}$. This vetoes the Higgs boson mass peak and ensures orthogonality with other analyses targeting $H \to b\bar{b}$ and to the signal regions of the $HH \to b\bar{b}\ell\ell$ analysis. Similar kinematics in the control and signal regions are achieved by requiring $p_{\mathrm{T}} > 45\,\text{GeV}$ for the leading $b$-tagged jet, and $p_{\mathrm{T}} > 40\,\text{GeV}$ for both leptons.

## 7.6 Boosted decision tree strategy

In the $HH \to b\bar{b}\tau^+\tau^-$ analysis, BDTs categorise events into the different SRs, others separate signal from background events. This section gives a general overview of the general training and evaluation used in this thesis.

### 7.6.1 Decision trees and Boosted decision trees

BDTs are based on an ensemble of decision trees, which are supervised learning algorithms. When used as a classifier, a tree learns to subsequently split a dataset into distinct classes as illustrated in Figure 7.8. Assuming two classes, each split of the dataset produces one signal and one background "node". The split on a variable $X$ is chosen such that purity in both nodes is maximised. The purity is calculated with the Gini index $G$, that measures the probability of mis-classification for any randomly classified event. For two classes $S$ and $B$ with $n_s$ and $n_b$ elements, the probabilities of correctly classifying an event as $S$ is then given by $p_S = n_s/(n_s + n_b)$, and mis-classifying as $B$ by $1 - p_S$. The Gini index for a node is calculated as

$$G = (p_S \cdot (1 - p_S)) + (p_B \cdot (1 - p_B)) = 1 - p_S^2 - p_B^2. \tag{7.7}$$

The index ranges between $G = 0$ for a perfectly separated node, and $G = 0.5$ for a maximally impure node. At each node, the cut on $X$ is chosen such, that the weighted sum of the Gini indices of the signal and the background nodes is minimised. The iterative splitting procedure is repeated

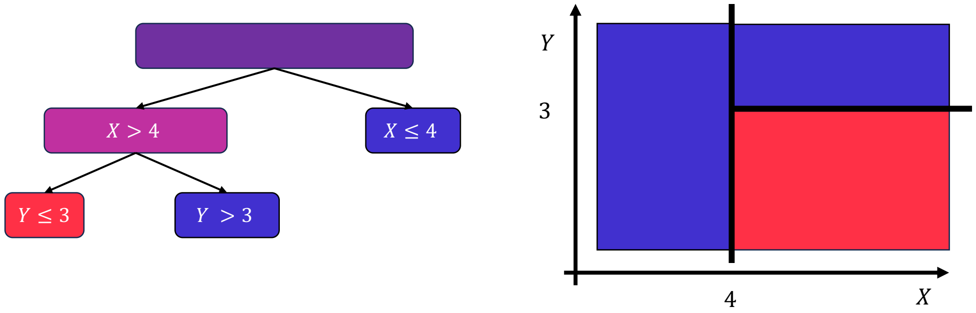


Figure 7.8: Illustration of the working principle of a decision tree.

until a given stopping criterion is reached, for example a maximum number of consecutive splits or a minimum number of events per node.

While this approach gives the locally most optimal split for each individual node, it does not guarantee the overall best separation. Individual decision trees can additionally be susceptible to statistical noise. Thus, the performance is improved and stabilised by a "boosting" procedure that combines many decision trees which are iteratively trained and evaluated into a Boosted Decision Tree (BDT) [192, 193]. Each iterative training step focusses on the mis-classified events from the previous training step, therefore improving the performance of the ensemble. The final BDT output is given by the weighted sum of the individual decision trees.

### 7.6.2 Training and hyper-parameter optimisation

During the design, training and optimisation phase of the $HH \to b\bar{b}\tau^+\tau^-$ analyses and the BDTs, a 3-fold training strategy ensures a reliable and unbiased estimate of the expected sensitivity. The training dataset is divided by event number into three equally sized folds for training, validation, and testing, as illustrated in Table 7.4. Three BDTs are trained employing their respective training folds, using identical input variables and hyper-parameters. The training set is used to optimise the classifier, while the validation set is used to verify the training procedure and mitigate potential overfitting. The final model performance is evaluated on the independent test set. By construction, fit templates derived from the test folds therefore provide unbiased estimates.

| | **Fold 0** | **Fold 1** | **Fold 2** |
|---|---|---|---|
| **BDT 0** | Training | Validation | Testing |
| **BDT 1** | Testing | Training | Validation |
| **BDT 2** | Validation | Testing | Training |

Table 7.4: Partitioning of the dataset for 3-fold training, testing and validation.

The two most significant training hyper-parameters during the BDT training are the total number of individual decision trees per BDT and the maximum number of subsequent cuts, referred to as depth, for each tree. The choice of hyper-parameter can influence the performance significantly. Therefore, they are optimised with respect to the expected binned signal significance $Z$

$$Z = \sqrt{\sum_{i \in \text{bins}} 2\left((s_i + b_i)\log(1 + \frac{s_i}{b_i}) - s_i\right)}, \tag{7.8}$$

with the number of signal $s_i$ and background $b_i$ events in bin $i$. The hyper-parameter optimisation is performed using Bayesian optimisation [194], which efficiently explores the predefined parameter ranges to identify the optimal settings. The same binning algorithm is used as for the multivariate signal extraction described in Section 7.12.2.

### 7.6.3 Input variable selection

A large set of input variables is available to train the BDTs. However, removing redundant or highly correlated observables while retaining a near optimal performance simplifies the analysis. Starting from a baseline selection, a minimal set of variables is iteratively constructed. At each step, one additional variable is considered and retained only if it provides the largest improvement in performance compared to any other candidate variable. If no variable improves the significance, the variable with the least reduction is added. The procedure stops if no further improvement is observed after a predefined number of steps. During the variable selection process, a slightly coarser binning of the BDT output distribution (as described in Subsection 7.12.2) is used to ensure that the observed improvements from additional variables exceed the statistical fluctuations of the sensitivity.

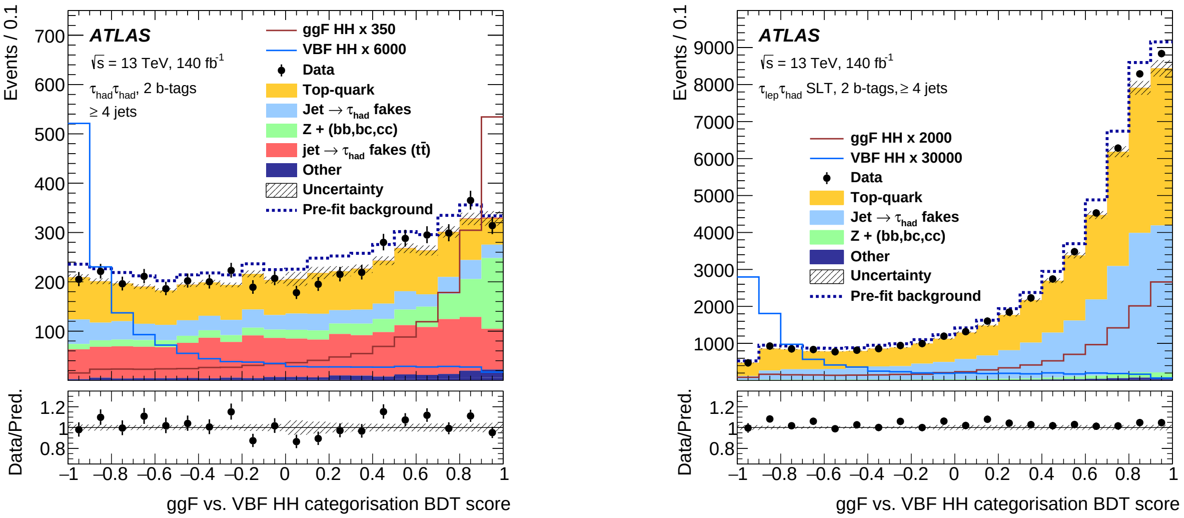
## 7.7 Event categorisation

The event categorisation enhances the sensitivity to VBF $HH$ production, $\kappa_\lambda$ and $\kappa_{2V}$. As sketched in Figure 7.1, per channel, events are categorised into a low-$m_{HH}$, a high-$m_{HH}$ ggF and a VBF signal region.

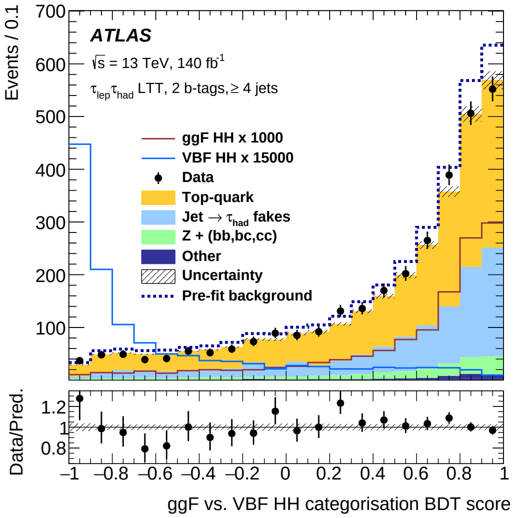
Characteristic event features of VBF $HH$ events are exploited in two steps. First, possible VBF candidate events are identified by VBF jet candidates. They are defined as the two highest $p_\mathrm{T}$ jets that are not associated with the $H \to b\bar{b}$ decay. Events with equal or more than two additional jets are classified by a BDT into either ggF or VBF-like categories. In each channel, BDTs are trained separately to distinguish between ggF $HH$ and VBF $HH$ signal events, using typical VBF-sensitive variables, particularly those related to the two additional jets. Each BDT includes the invariant mass $m_{jj}^{\mathrm{VBF}}$ and the pseudo-rapidity separation $\Delta\eta_{jj}^{\mathrm{VBF}}$ of the two VBF jet. Additional variables such as the separation in $\Delta\phi$ or $\Delta R$ are iteratively selected if they increase the separation power. Each BDT ultimately uses seven variables.

The BDT training follows Section 7.6. The hyper-parameters and the categorisation score thresholds are optimised to maximise the VBF $HH$ signal significance, while preserving ggF $HH$ sensitivity and improving the constraints on $\kappa_\lambda$ and $\kappa_{2V}$. The BDT approach is found to slightly improve the sensitivity with respect to a simple cut based approach on $m_{jj}^{\mathrm{VBF}}$ and $\Delta\eta_{jj}^{\mathrm{VBF}}$. In the $\tau_\mathrm{had}\tau_\mathrm{had}$ channel, about 80 % of VBF $HH$ and 50 % of ggF $HH$ events have VBF jet candidates.

Figure 7.9 shows the predicted and observed BDT classification score distributions for events with two VBF jet candidates for the signal and backgrounds in the three channels. In all three channels, there are significantly more background than signal events. In the $\tau_\mathrm{had}\tau_\mathrm{had}$ channel, the background is composed of top-quark production, events with fake-$\tau_\text{had-vis}$ from both QCD jets and $t\bar{t}$ production, as well as $Z$+HF production. Except for $Z$+HF events, the backgrounds are approximately uniformly distributed across the BDT score range. In the two $\tau_\mathrm{lep}\tau_\mathrm{had}$ channels, the background arises almost entirely from top-quark production and QCD fake-$\tau_\text{had-vis}$ processes. These events are predominantly categorised as ggF-like.



(a) $\tau_{\rm had}\tau_{\rm had}$, cut 0.1. (b) $\tau_{\rm lep}\tau_{\rm had}$ SLT, cut -0.13.



(c) $\tau_{\rm lep}\tau_{\rm had}$ LTT, cut -0.1.

Figure 7.9: Predicted and observed distributions of the categorisation BDT scores in (a) the $\tau_{\rm had}\tau_{\rm had}$ (b) the $\tau_{\rm lep}\tau_{\rm had}$ SLT and (c) the $\tau_{\rm lep}\tau_{\rm had}$ LTT channels, for all VBF candidate events. The background distributions are shown as obtained from the combined likelihood fit to data in the signal vs. background distributions. "Other" refers to contributions from $Z$-boson production in association with less than two jets initiated by $b$ or $c$-quarks, $W$ boson production, diboson production, $t\bar{t}V$ production, and single Higgs boson production processes. The ggF and VBF $HH$ signal distributions are overlaid and scaled to the factor indicated in the legend times the SM expectation. The dashed histogram shows the total pre-fit background. The lower panels show the ratio of data to the total post-fit sum of signal and background. The hatched bands in the upper and lower panels show the combined statistical and systematic uncertainties in the total prediction. The applied cut value is also indicated. Published in Reference [12].

In each channel, events with less than four jets, and events that are categorised by the BDTs as ggF-like are split into a low-$m_{HH}$ and a high-$m_{HH}$ ggF $HH$ signal regions at $m_{HH} = 350\,\text{GeV}$.

In each channel, events with fewer than four jets and those categorised by the BDTs as ggF-like are further divided into low- $m_{HH}$ and high-$m_{HH}$ ggF $HH$ signal regions, using a boundary at $m_{HH} = 350\,\text{GeV}$. The split gives approximately the optimal sensitivity for $\kappa_\lambda$ larger than the SM expectation in the low-$m_{HH}$ region while retaining the sensitivity for SM $HH$ production in the high-$m_{HH}$ region, and corresponds to the maximal interference between the box and the triangle diagram on the differential ggF $HH$ cross-section.

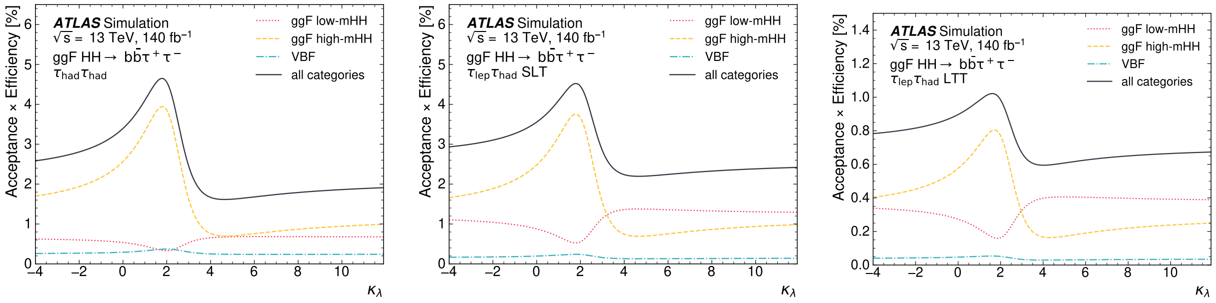
## 7.8 Signal acceptance times efficiency

Table 7.5 lists the total acceptance times efficiency ($\mathcal{A} \times \epsilon$) for SM ggF and VBF $HH$ production. Acceptance refers to the fraction of signal events that satisfy the geometric and kinematic requirements of the analysis, such as detector coverage and object $p_{\text{T}}$ and $\eta$ thresholds defined by the event selection, before accounting for reconstruction and identification efficiencies. Efficiency refers to the probability that events within the acceptance are successfully reconstructed, identified, and selected by the analysis, accounting for the performance of the detector and reconstruction algorithms.

The total $\mathcal{A} \times \epsilon$ for selecting the $HH \to b\bar{b}\tau^+\tau^-$ signal is about 4 % for both the $\tau_{\text{had}}\tau_{\text{had}}$ and the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channels, and about 1 % for the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel. It is dominated by ggF $HH$ production, as its cross-section is an order of magnitude larger than that of VBF $HH$. In all three channels, the $\mathcal{A} \times \epsilon$ is highest in the high-$m_{HH}$ region for SM ggF $HH$ production. Fewer than 0.5 % of ggF $HH$ events end up in the VBF signal region, while most VBF $HH$ events are

| $\boldsymbol{\mathcal{A} \times \epsilon}$ | **ggF $\boldsymbol{HH}$** | | | **VBF $\boldsymbol{HH}$** | | |
|---|---|---|---|---|---|---|
| | $\boldsymbol{\tau_{\text{had}}\tau_{\text{had}}}$ | $\boldsymbol{\tau_{\text{lep}}\tau_{\text{had}}}$ | | $\boldsymbol{\tau_{\text{had}}\tau_{\text{had}}}$ | $\boldsymbol{\tau_{\text{lep}}\tau_{\text{had}}}$ | |
| **Signal region** | | SLT | LTT | | SLT | LTT |
| low-$m_{HH}$ ggF | 0.45 % | 0.68 % | 0.21 % | 0.11 % | 0.25 % | 0.08 % |
| high-$m_{HH}$ ggF | 3.33 % | 3.19 % | 0.72 % | 0.64 % | 0.74 % | 0.18 % |
| VBF | 0.33 % | 0.22 % | 0.05 % | 1.46 % | 1.57 % | 0.44 % |
| Total | 4.11 % | 4.09 % | 0.98 % | 2.31 % | 2.56 % | 0.70 % |

Table 7.5: Acceptance times efficiency for SM ggF and VFB $HH$ production in the nine SRs.

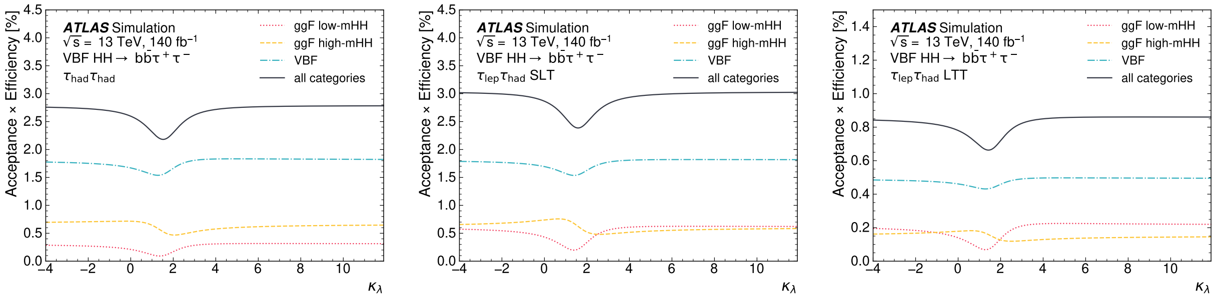


(a) $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ SRs (b) $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT SRs (c) $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT SRs

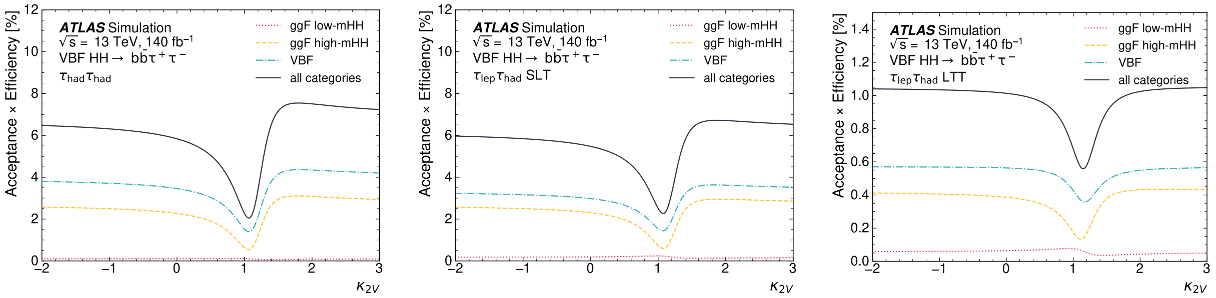
Figure 7.10: Acceptance times efficiency for ggF $HH$ production in the three channels as a function of $\kappa_\lambda$. Published in Reference [12].

categorised in the VBF signal regions. The $\mathcal{A} \times \epsilon$ for VBF $HH$ production is about 2 % in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT channel, and less than 1 % in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT channel.

Most events are lost due to the trigger selection, the requirement of two $b$-tagged jets, and the $\tau_{\text{had-vis}}$ candidate selection. The efficiency of different selection requirements, referred to as "cutflow", is shown for the ggF $HH$ signal in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel in Table 7.6, starting with the theoretical yields for ggF $HH$ production in the $b\bar{b}\tau^+_{\mathrm{had}}\tau^-_{\mathrm{had}}$ final state. The generator filter



(a) $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ SRs (b) $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT SRs (c) $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT SRs.



(d) $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ SRs (e) $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT SRs (f) $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT SRs.

Figure 7.11: Acceptance times efficiency for VBF $HH$ production in the three channels a function of (a-c) $\kappa_\lambda$, and (d-f) $\kappa_{2V}$. Published in Reference [12].

| **Description** | **Number of events** | **Efficiency** [%] | **Relative efficiency** [%] |
|---|---|---|---|
| $HH$ | 4314.9 | | |
| $HH \to b\bar{b}\tau^+\tau^-$ | 315.2 | | |
| $HH \to b\bar{b}\tau^+_{\text{had}}\tau^-_{\text{had}}$ | 132.3 | 100.0 | |
| Generator filter | 102.8 | 77.7 | 77.7 |
| Derivation skimming | 88.3 | 66.7 | 85.9 |
| Object preselection | 40.4 | 30.6 | 45.8 |
| Trigger selection (STT + DTT) | 17.3 | 13.1 | 43.3 |
| Random $\tau$ selection | 17.2 | 13.0 | 99.7 |
| Object selection$^\dagger$ | 16.0 | 12.1 | 92.7 |
| Pileup Correction | 16.3 | 12.3 | |
| 2 Loose $\tau$'s$^\star$ | 13.0 | 9.8 | 79.8 |
| Sub-leading $\tau$ $p_{\text{T}} > 25\,\text{GeV}$ | 12.9 | 9.7 | 98.8 |
| $m_{\tau\tau}^{\text{MMC}} > 60\,\text{GeV}$ | 12.6 | 9.5 | 97.7 |
| $\geq 1$ jet with $p_{\text{T}} > 45$ GeV | 12.3 | 9.3 | 98.2 |
| DTT offline jet cuts | 11.1 | 8.4 | 90.1 |
| Scale Factors | 11.6 | 8.8 | |
| Opposite charge sign | 11.5 | 8.7 | 98.6 |
| Both jets are $b$-tagged | 5.4 | 4.1 | 47.0 |
| ggF low-$m_{HH}$ SR | 0.6 | 0.4 | 10.9 |
| ggF high-$m_{HH}$ SR | 4.4 | 3.3 | 81.1 |
| VBF SR | 0.4 | 0.3 | 7.9 |

$^\dagger$: 2 $\tau$'s, further requirements on jets and $\tau$ - jet overlap removal
$^\star$: one "loose $\tau$" already required in derivation

Table 7.6: Cutflow for the ggF $HH$ signal in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

is applied as part of the MC event generation, and entails requirements such as $p_{\text{T}}$ and $\eta$ for the $\tau_{\text{had-vis}}$ candidates. The derivation skimming is applied when processing AOD to DAOD. Object preselection refers to loose cuts on the physics objects that are applied when deriving the CxAOD format. The trigger selection has a relative efficiency of approximately 43 % in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, and is followed by several requirements on the physical objects. Event selection criteria, such as for example to contain exactly two $\tau$-leptons with opposite-sign charges, that pass the "loose" $\tau$ identification criteria, or to contain exactly two $b$-tagged jets (at the DL1r 77 % working point), further reduce the number of events.

Figure 7.10 and 7.11 show the $\mathcal{A} \times \epsilon$ in the different signal regions and channels as a function $\kappa_\lambda$ for ggF and VBF $HH$ production, and for the latter also as a function of $\kappa_{2V}$. For $\mathcal{A} \times \epsilon$ for ggF

$HH$ production increases in all three channels for $\kappa_\lambda \approx 2$, driven by an increase of the $\mathcal{A} \times \epsilon$ in the high-$m_{HH}$ signal regions. The $\mathcal{A} \times \epsilon$ changes little for different values of $\kappa_\lambda$ for VBF $HH$ production, and increases significantly for $\kappa_{2V} \neq 1$.

## 7.9 Background composition and estimation

This section describes the background composition and estimation methods across the different channels and signal regions. Depending on the process, backgrounds are estimated using simulation, data-driven techniques, or a combination of both.

Figure 7.12 shows the background composition in the analysis regions, as determined by the combined fit to data. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the background is dominated by three main processes: top-quark production, events with fake-$\tau_{\text{had-vis}}$, and $Z$+HF production. The fake-$\tau_{\text{had-vis}}$ contribution is further divided according to its origin, distinguishing between $t\bar{t}$ and multi-jet production. In both $\tau_{\text{lep}}\tau_{\text{had}}$ channels, more than 60 % of background events stem from top-quark processes. In the SLT and LTT channels, the remaining background arises almost entirely from fake-$\tau_{\text{had-vis}}$ events that are almost exclusively from $t\bar{t}$ production. In the LTT channel, some fake candidates stem from multi-jet events. The contribution from $Z$+HF production accounts for approximately 10 % of the background events.

### 7.9.1 Top quark

**Top-quark pair production** $t\bar{t}$ production is a significant background in all regions. The cross-section is approximately 30, 000 times larger than that of $HH$ production, and about 4 % of $t\bar{t}$ events decay to the $b\bar{b}\tau^+\tau^-$ final state. The $t\bar{t}$ contribution is about an order of magnitude larger in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel than in $\tau_{\text{had}}\tau_{\text{had}}$, since, in addition to events where one of the two $W$ bosons decays to a $\tau$-lepton (and a neutrino) that subsequently decays leptonically, events in which one of the $W$ bosons decays directly into an electron or a muon (and a neutrino) also populate the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. Moreover, $t\bar{t}$ production contributes significantly to the background from fake-$\tau_{\text{had-vis}}$ candidates, since quark-initiated jets can be misidentified as fake-$\tau_{\text{had-vis}}$ candidates

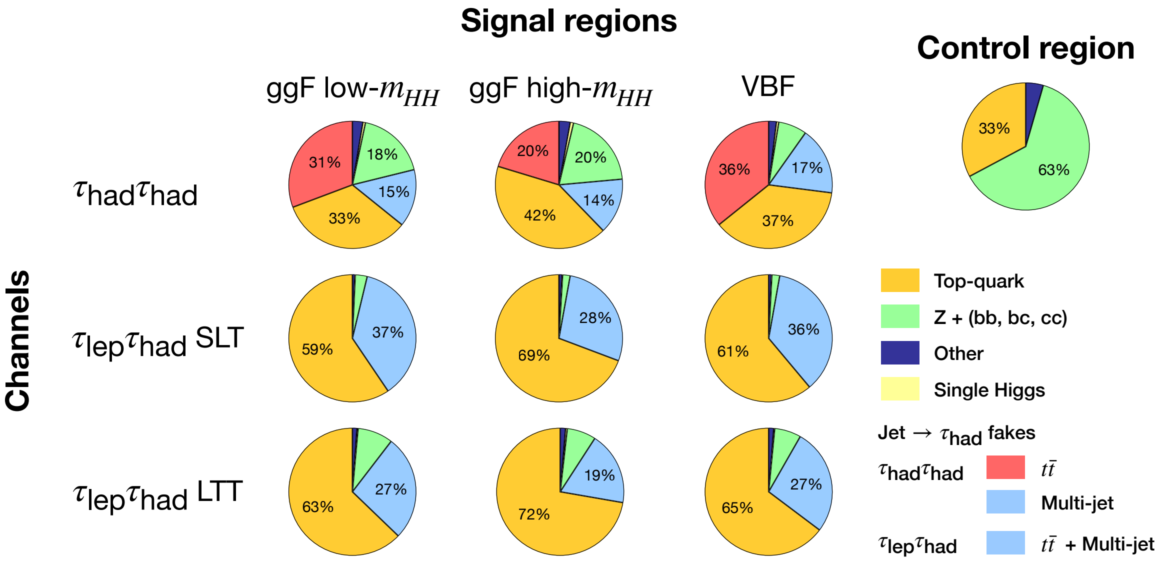


Figure 7.12: Fractional background composition in the analysis regions as determined by the fit to data. For the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the fake-$\tau_{\text{had-vis}}$ background is split into originating from $t\bar{t}$ (red) and multi-jet production (blue). For the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, the combined fake-$\tau_{\text{had-vis}}$ background (blue) mostly consists of fake candidates from $t\bar{t}$ production. "Other" covers contributions from $Z$ boson production in association with less than two jets initiated by $b$ or $c$-quarks, $W$ boson and $t\bar{t}V$ production. Modified from Reference [12].

and $\mathcal{B}(W \to qq') \approx 70\,\%$.

$t\bar{t}$ events with true-$\tau_{\text{had-vis}}$ candidates are estimated from simulation. The MC statistical uncertainty on this estimate, one of the largest sources of systematic uncertainty, is reduced by a factor of two compared to the previous analysis, achieved by increasing the number of generated events by a factor of four, corresponding to about ten times more simulated events than data events. The $t\bar{t}$ normalisation is estimated by the fit in the CR and SRs. The $m_{\ell\ell}$ distribution in the CR is shown in Figure 7.13. The top-quark contribution exhibits a slope towards higher $m_{\ell\ell}$ values due to the phase space selection in the CR. The correction determined by the combined fit approximately agrees with unity within the uncertainty.

**Single-top quark production** The background is estimated from simulation. Contributions from the $Wt$ channel are about a factor five larger than that of the $t$-channel, and about 75 times larger than from the $s$-channel. Single-top quark production makes up between 5 to 7.5 % of the total top quark background in the three channels.

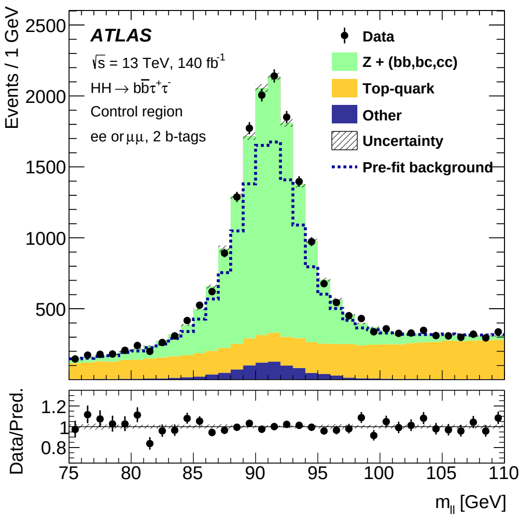


Figure 7.13: Predicted and observed $m_{\ell\ell}$ distribution in the control region. The normalisation, shape and total uncertainty of the $t\bar{t}$ and $Z$+HF backgrounds are drawn as determined from the fit to data. “Other” covers contributions from $Z$-boson production in association with less than two jets initiated by $b$ or $c$-quarks, $W$ boson and $t\bar{t}V$ production and single Higgs boson production processes. The dashed histogram shows the total pre-fit background. The lower panel shows the ratio of data to the total post-fit background. The hatched bands in the upper and lower panels show the combined statistical and systematic uncertainties in the total background. Published in Reference [12].

### 7.9.2 $Z$ + HF

$Z \to \tau\tau$ production in association with jets ($Z$+HF) is a main background in all three analysis channels. In 93 % of times, the selected $H \to b\bar{b}$ candidates in this background are true heavy-flavour jets, predominately $bb$ (85 %), with smaller contributions from $bc$ (8 %) and $cc$. The $Z$+HF background is estimated from simulation. Similar to $t\bar{t}$, the normalisation is also determined by the fit, since MC generators do not predict the cross-section with sufficient accuracy. Experimentally, the inclusive $Z$ boson production cross-section has been measured to agree with SM predictions to high precision. Since the production cross-section is independent of the decay, the normalisation of $Z \to \tau\tau$ + HF can be determined from $Z \to \ell\ell$ + HF in the CR, where a pair of electrons or muons ($\ell \in \{e, \mu\}$) is required. In the CR, the $Z$+HF contribution peaks near the $Z$ boson mass of approximately 90 GeV, as shown in Figure 7.13. The fit determines the $Z$+HF contribution to be underestimated by about 30 %. This is approximately in line with the observed mis-modelling observed in other analyses. More details are discussed alongside the fit results in Section 7.13.

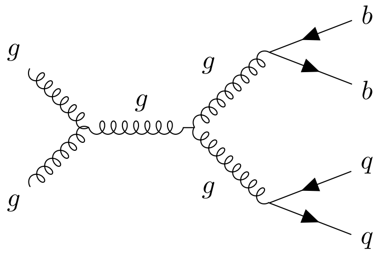


Figure 7.14: Example diagram for QCD multi-jet production with two $b$-jets and two jets which can be mis-reconstructed as a $\tau_{\text{had}}$-candidate.

### 7.9.3 Fake-$\tau_{\text{had-vis}}$

Quark- and gluon-initiated jets can be mis-identified as $\tau_{\text{had-vis}}$-jets, and are commonly referred as fakes. Rarely, also electrons or muons can be mis-identified as fake-1-prong $\tau_{\text{had-vis}}$ candidates. However, this section focuses exclusively on quark and gluon initiated fakes. Most fakes arise from $t\bar{t}$, primarily due to misidentified quark jets, and from QCD multi-jet processes, where both quark- and gluon-initiated jets contribute. An example Feynman diagram for QCD multi-jet production is given in Figure 7.14.

Fake-$\tau_{\text{had-vis}}$ contributions are difficult to simulate accurately. While the underlying processes have very high production cross-sections, the misidentification efficiency is low and highly sensitive to the simulated detector performance. Consequently, even small uncertainties in the simulation can lead to large discrepancies in the predicted yields. Therefore, the contribution in the signal region is obtained by transferring a template from a control region into the signal region via multiplicative “fake-factors” (FF), that are derived with the data-driven “ABCD” method [153].

The ABCD method estimates the contribution of fakes in region A from three orthogonal regions. The method is illustrated in Figure 7.15. In signal region A, contributions from fake-$\tau_{\text{had-vis}}$ candidates are estimated by scaling the contribution in region B by a fake factor. The FF is derived by the ratio of the contributions in two control regions C and D. The fake factor $\text{FF}_X$ can be applied in bins of a variable $X$, for example the $p_{\text{T}}$ of the hadronic $\tau$-lepton decay.

The three regions B, C and D are created by inverting the selections of A on two observables,

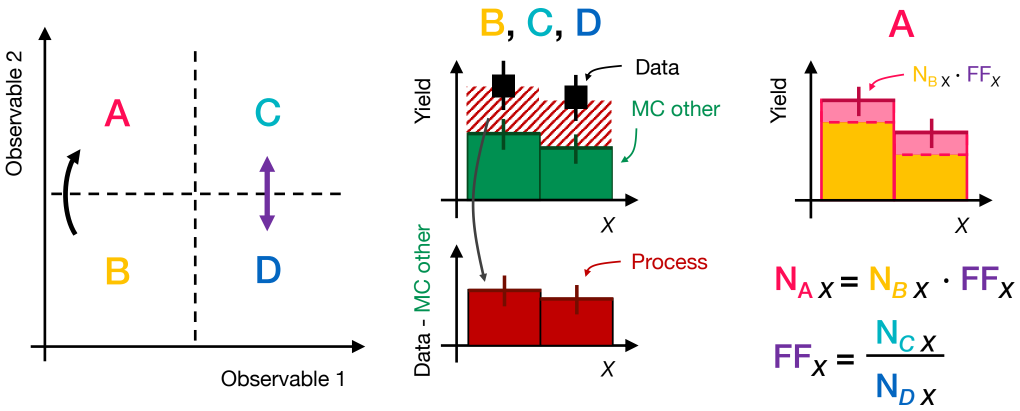


Figure 7.15: Illustration of the ABCD method to obtain the fake factor (FF) in region A using three orthogonal regions B, C and D.

ensuring that the relation

$$\frac{N_{\mathrm{A}}}{N_{\mathrm{B}}} = \frac{N_{\mathrm{C}}}{N_{\mathrm{D}}} \tag{7.9}$$

holds between the number of events $N$ in each region. The inversion of observable 1 must be chosen so as not to introduce significant phase-space differences between A and C, and between B and D, for example in the distribution of some variable $X$. Meanwhile, the inversion of observable 2 should enrich regions B and D in fake-$\tau_{\text{had-vis}}$ contributions. The fake factor is calculated from the ratio of number fake-$\tau_{\text{had-vis}}$ events in regions C and D

$$\mathrm{FF}_X = \frac{N_{\mathrm{C},\,X}}{N_{\mathrm{D},\,X}} \ . \tag{7.10}$$

The number of fake-$\tau_{\text{had-vis}}$ events can, for example, be estimated by subtracting the predicted contribution of true $\tau_{\text{had-vis}}$ events from the observed data. The desired fake-$\tau_{\text{had-vis}}$ contribution in the signal region A is obtained by scaling the contribution in B by the FF

$$N_{\mathrm{A},\,X} = N_{\mathrm{B},\,X} \cdot \mathrm{FF}_X \ . \tag{7.11}$$

The ABCD method is used to estimate contributions from fakes in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels. The method is validated in dedicated control regions. The procedure is slightly varied to account for example for the different background composition or limited available statistics. A short overview is given in the following, detailed information is provided in Section A.2 and References [153, 195].

### Fake-$\tau_{\text{had-vis}}$ estimate in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, fake-$\tau_{\text{had-vis}}$ candidates originate from both the $t\bar{t}$ and the multi-jet backgrounds. Their contributions are simultaneously estimated with separate fake factors, that are combined, weighted by the relative contribution, into a single correction in the region of interest. This combined scaling approach reduces the need to distinguish the origin of fakes in the signal region, which would require accurate modelling of both multi-jet and $t\bar{t}$ fakes. The combined FFs are independently derived and applied in the SLT and LLT channels, and separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates and in bins of the $\tau_{\text{had-vis}}$ $p_{\text{T}}$.

### Fake-$\tau_{\text{had-vis}}$ estimate in the $\tau_{\text{had}}\tau_{\text{had}}$ channel

The different fake composition in the $\tau_{\text{had}}\tau_{\text{had}}$ channel prevents the usage of the combined FF method in this channel. The majority of fakes arise from mis-identified $W \rightarrow q\bar{q}'$ decays from $t\bar{t}$ production. Approximately 85 % of those events are from semi-leptonic $t\bar{t}$ decays, corresponding to mostly one true and one fake $\tau_{\text{had-vis}}$ candidate. Only 15 % of $t\bar{t}$ events feature two fake-$\tau_{\text{had-vis}}$ candidates. Another reason the combined method cannot be applied is the lack of a control region that is both pure in $t\bar{t}$ fakes and sufficiently populated to allow a reliable determination of the fake factor FF. Therefore, contributions from multi-jet and $t\bar{t}$ are estimated individually instead, despite increasing the overall systematic uncertainty on the fake estimate in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Importantly, the sensitivity of the analysis is primarily limited by the available data statistics, and the impact of this choice on the sensitivity is small, as the background can still be effectively distinguished from the $HH$ signal using kinematic differences. Multi-jet contributions are estimated with the ABCD method. Contributions from $t\bar{t}$ are estimated from simulation that are corrected with multiplicative scale factors that are derived from data.

### 7.9.4 Single Higgs and other backgrounds

Backgrounds from other processes, including single Higgs production, are estimated from simulation as described in Section 7.3. Modelling uncertainties related to normalisation, shape and acceptance are described in detail in Subsection 7.11.2.

## 7.10 Multivariate signal extraction

Assuming SM production, only about 4600 $HH$ events were created during Run 2. Therefore, the foremost priority of the event selection and categorisation is to maximise the acceptance times efficiency for signal events, while simultaneously defining control regions for background estimation. Overall, the $HH \to b\bar{b}\tau^+\tau^-$ analysis has an acceptance times efficiency of approximately 4 %, corresponding to an expected signal yield of 13 signal events. The pre-selection ensures that events in the signal regions are generally kinematically compatible with the $HH \to b\bar{b}\tau^+\tau^-$ signal, and fulfil basic identification criteria. Nevertheless, the significantly larger cross-section of background processes leads to a very small signal to background ratio, particular due to $t\bar{t}$ production and processes with $\tau_{\text{had-vis}}$ fakes. The expected background yield from simulation exceeds the expected signal yield by about five orders of magnitude in the in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel, and by about four orders larger in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channels.

The multivariate signal extraction strategy has been significantly revised with respect to the previous analysis. The separation of the signal from the sum of backgrounds exploits distinct kinematic features of signal events. In each signal region, one dedicated classifier BDT is trained.

The BDTs are trained mostly on simulated events following Section 7.6. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, this includes the simulated background with fake-$\tau_{\text{had-vis}}$ candidates as most fakes arise from $t\bar{t}$. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the training uses the data-driven multi-jet fake-$\tau_{\text{had-vis}}$ estimate, and the simulated $t\bar{t}$ events with fake-$\tau_{\text{had-vis}}$ and the data-driven correction factors applied. The sensitivity to $\kappa_\lambda$, VBF $HH$ production and $\kappa_{2V}$ in the different signal regions is increased with different signal samples. The BDTs in the high-$m_{HH}$ SRs are trained with SM ggF signal samples.

The low-$m_{HH}$ SRs have a larger acceptance for $HH$ production with large values of $\kappa_\lambda$. Therefore, the sensitivity for non-SM $HH$ production is increased with ggF signal samples that have been generated at $\kappa_\lambda = 10$. VBF $HH$ signal samples are used in the VBF region. The background yields are scaled to be equal to the signal yield which enhances the classification performance. The relative background composition is adjusted with MC event weights to represent the estimated composition in each signal region.

The training procedure iteratively selects the set of input variables from a larger set of observables describing the event, and individual objects. This maximises the expected improvement in signal sensitivity while keeping the BDTs as simple as possible. In particular, redundant or highly correlated variables are omitted without reducing the separation power. Variables related to the decays of the two Higgs bosons provide good S/B separation. Therefore, the procedure begins with an initial set of variables that includes the invariant mass of the two selected $b$-jets ($m_{bb}$), the invariant mass of the $\tau$-lepton pair as calculated by the MMC algorithm ($m_{\tau\tau}^{\mathrm{MMC}}$), and the invariant mass of the $HH$ system $m_{HH}$. If both $b$-jets originate from the decay of a Higgs boson, the $\Delta R$ between the two outgoing quarks, and thus approximately the jet axes, depends on the transverse momentum $p_{\mathrm{T}}^{H}$ and the mass $m_H$ of the Higgs boson as

$$\Delta R_{bb} \sim \frac{2m_H}{p_{\mathrm{T}}^{H}} \ . \tag{7.12}$$

This dependence can be used to separate backgrounds such as $t\bar{t}$. Since the two $b$-jets arise from two different top-quark decays, and the two $\tau$-leptons stem from two different W-boson decays, their opening angle is in general larger and the correlation between $m_{HH}$ and $\Delta R$ is different than for signal events. Thus, the angular separation of the $b$-jet $\Delta R_{bb}$ and the $\tau$-lepton pair $\Delta R_{\tau\tau}$ is used in most regions[1], as these variables help for example discriminating against $t\bar{t}$ production backgrounds.

Multiple additional variables are related to the kinematics of the leading and sub-leading $b$-jets and

---

[1] $\Delta R_{bb}$ is not used in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT high-$m_{HH}$ and VBF SRs. $\Delta R_{\tau\tau}$ is not used in VBF SR.

$\tau$-leptons and the $H$ candidate. This includes the transverse momentum ($p_{\mathrm{T}}$), the pseudo-rapidity ($\eta$), or the transverse mass ($m_{\mathrm{T}}$). For pairs of $b$-jets and $\tau$-leptons, variables such as $\Delta R$ or the angular separation $\Delta\phi$ are considered. The separation power is further increased with variables describing the hadronic activity, such as the number of jets, or general event properties, such as the $p_{\mathrm{T}}$ of the reconstructed $HH$ system ($p_{\mathrm{T}\ HH}$) which serves as a proxy for the amount of radiation in a $HH$ event. Moreover, characteristic event shapes described by Fox-Wolfram moments, and pseudo-continuos $b$-tagging information are used. In the VBF signal regions, additional variables describing the kinematics of VBF jets and VBF events are included. The full set of variables is found in Reference [12].

The resulting BDTs achieve good separation between the signal and the sum of background processes. Nevertheless, it is not possible to create bins with no background contributions. Re-binning the BDT output score ensures the validity of the asymptotic approximation and sufficient background contributions in the most signal-like bins. The details of the re-binning algorithm and its role in the statistical interpretation are presented in Section 7.12.2.

After re-binning, the sensitivity of each channel is driven by the last two bins in the high-$m_{HH}$ region as they feature the highest signal to background ratio (S/B). Using simulated events, the highest S/B $\approx 1.4/3.1$ is expected in the last bin of the BDT score in the high-$m_{HH}$ signal region of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. Consequently, this SR contributes the most to the overall sensitivity. In this bin, about 40 % of the background arises from single Higgs production, predominantly from ggF $H$ and $ZH$ production, with the Higgs boson decaying into a pair of $\tau$-leptons. For approximately 94 % of the ggF $H$ events in the most sensitive bin, the two $b$-tagged jets are both true $b$-jets. In further 4 % of cases, one of the two $b$-tagged jets is a true $b$-jet. In less than 2 % of events, both $b$-tagged jets are $c$-jets. Backgrounds with a much larger cross-section such as top-quark production and events with fake-$\tau_{\text{had-vis}}$ candidates are more effectively suppressed. The second highest expected S/B occurs in the most signal-like bin of the BDT score distribution in the high-$m_{HH}$ region of the SLT $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel, which is about two times lower ($\approx 0.7/3.1$). In this bin, $Z$+HF accounts for about one third of the background, followed by single-top-quark production contributing about one fifth. Backgrounds with fake-$\tau_{\text{had-vis}}$ candidates are rejected well.

## 7.11 Systematic uncertainties

The analysis is subject to several sources of systematic uncertainty. While the sensitivity is primarily limited by the statistical uncertainty of the available data in the signal regions, both experimental and theoretical systematic uncertainties must also be taken into account, as they can influence the yield and shape of key observables, including the discriminant used in the final fit. Experimental uncertainties arise from the performance of the detector, impacting the reconstruction, identification, and calibration of physics objects. Theoretical uncertainties stem from choices in the underlying theoretical assumptions, affecting, for example, the predicted cross-sections used to normalise simulated samples, as well as the modelling of signal and background processes. This section describes the systematic uncertainties affecting the $HH \to b\bar{b}\tau^+\tau^-$ analysis. Their impact on the likelihood fit leading to the result of this search is described in Section 7.12.

### 7.11.1 Experimental uncertainties

Experimental uncertainties are derived through dedicated performance and calibration measurements led by the ATLAS combined performance groups. The impact on each observable is estimated by varying the simulation settings by the $\pm 1\sigma$ bounds obtained from the calibration measurements. The complexity is decreased by implementing uncertainties as a reduced set for which an analyses is not too sensitive. The reduced set is constructed from a principal component analysis that preserves correlations between the original uncertainties. For example, $HH$ analyses are not particular sensitive to jet energy resolution (JER), scale (JES) or $b$-tagging efficiency uncertainties. Thus, the $HH \to b\bar{b}\tau^+\tau^-$ and other $HH$ analyses employ a reduced set of uncertainties affecting those quantities. An overview of the experimental uncertainties grouped by different physics objects is provided in Table 7.7.

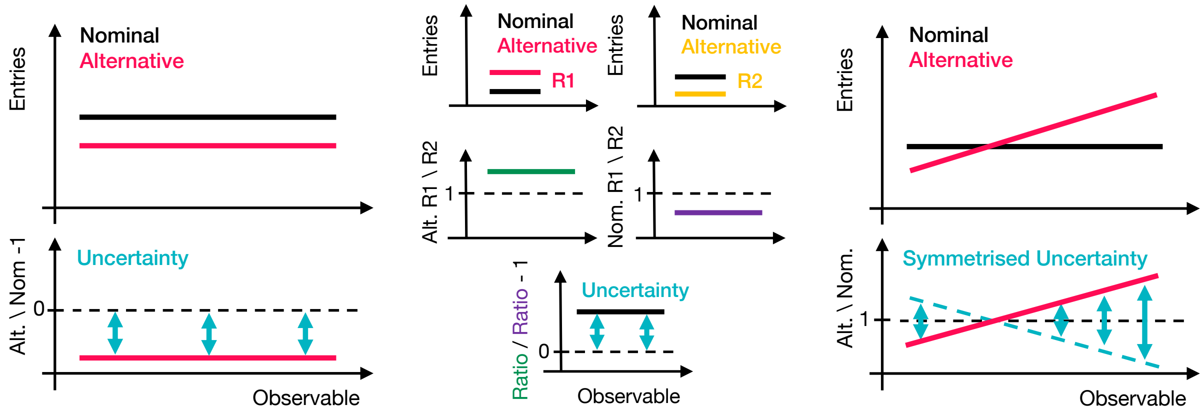
| **Category** | **Affected quantities** | N |
|---|---|---|
| Electrons | Trigger, reconstruction, identification, isolation | 4 |
| Photons | Momentum scale and resolution | 3 |
| Muons | Trigger, reconstruction, isolation, track-to-vertex matching, momentum calibration | 15 |
| $\tau_{\text{had-vis}}$ | Trigger, reconstruction, identification, $\tau$-electron veto, energy scale | 38 |
| Jets | Energy scale, energy resolution, jet vertex tagging | 34 |
| $b$-tagging | $b$, $c$ and light-tagging efficiencies | 11 |
| $E_{\mathrm{T}}^{\mathrm{miss}}$/ $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ | Momentum scale and resolution | 3 |

Table 7.7: Experimental uncertainties per category of physics objects.

With the exception of an improved luminosity estimate, all experimental uncertainties in this analysis remain unchanged with respect to the first Run 2 $HH \to b\bar{b}\tau^+\tau^-$ analysis [153]. The estimate on the full Run 2 integrated luminosity has been improved to 0.83 % [196]. The uncertainty is correlated across all processes for which the normalisation is taken from simulation. The pileup profile of simulated events is reweighed to the measured pileup profile and all related uncertainties are propagated to the final result. The simulated trigger efficiencies for electrons [155], muons [156] and $\tau_{\text{had-vis}}$ [197] and jets [139] are adjusted to the efficiencies measured in data. For muons and hadronic $\tau$-leptons, the related uncertainties are split into statistical and systematic components. The trigger uncertainties are implemented according to the trigger requirements of each channel, and range between $\mathcal{O}(1\,\%)$ to $\mathcal{O}(10\,\%)$. Uncertainties related to the reconstruction, identification and calibration are assigned to electron, [136], photon [136], muon [138], $\tau_{\text{had-vis}}$ [197] and jet [198] candidates. The uncertainties range between $\mathcal{O}(0.1\,\%)$ and $\mathcal{O}(1\,\%)$ and are propagated to the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ [199]. Flavour tagging efficiency uncertainties of $\mathcal{O}(1\,\%)$ to $\mathcal{O}(10\,\%)$ are assigned separately for $b$-, $c$- and light jets [200–202]. The experimental uncertainties have a negligible impact on the final result.

### 7.11.2 Modelling Uncertainties

The simulation of events and the calculation of cross-sections depend on various theoretical assumptions and input parameters of Monte Carlo generators. Uncertainties related to specific choices are quantified as a systematic uncertainties, and are either derived by comparing the



(a) Normalisation (b) Relative acceptance (c) Shape

Figure 7.16: Illustration of three types of modelling uncertainties

predictions of a nominal generator to various alternative generators, or by varying parameters of the underlying event generation in the nominal generator. In general, modelling uncertainties are evaluated individually according to their source and are treated as correlated across all signal regions within a given channel. Their impact on the observables is accounted for either as a normalisation uncertainty, affecting the overall cross section, or as an acceptance uncertainty, which is further categorised into three types: normalisation, shape, and relative acceptance uncertainties. These three types of impacts are illustrated in Figure 7.16.

Cross-section uncertainties are implemented for all processes but $t\bar{t}$ and $Z$+HF production. They are derived per process, and are fully correlated across signal and control regions. For $t\bar{t}$ and $Z$+HF production, no cross-section uncertainties are applied as the normalisations are determined by the fit via unconstrained normalization parameters.

**Normalisation uncertainties** affect the expected yield in all analysis regions. They are estimated from the relative difference in the number of expected events $N$ in a given analysis region, computed using the nominal and an alternative modelling, with are both normalised to the same cross section.

$$\sigma_A = \frac{N_{\text{alternative}}}{N_{\text{nominal}}} - 1 \ . \tag{7.13}$$

**Relative acceptance uncertainties** account for the different acceptance across two analysis regions for a given process. They are derived from the double ratio of the acceptance, or number of events, across two regions R1 and R2

$$\sigma_{A_{\mathrm{R1}}/A_{\mathrm{R2}}} = \left(\frac{N_{\text{alternative R1}}}{N_{\text{alternative R2}}}\right) / \left(\frac{N_{\text{nominal R1}}}{N_{\text{nominal R2}}}\right) - 1 \ . \tag{7.14}$$

Relative acceptance uncertainties are assigned to processes that are better constrained in one region than another. Therefore, they are assigned across signal and control regions for $t\bar{t}$ and $Z$+HF production as their normalisation is better constrained in the control region than in the signal regions.

**Shape uncertainties** affect the observable distributions such as the BDT classifier output in the signal regions. They are derived bin by bin from the ratio of the nominal and alternative process yield, and are implemented symmetrically. The nominal and alternative samples are normalised to the same cross-section to disentangle the impact of acceptance uncertainties. Shape uncertainties are correlated with acceptance uncertainties.

To ensure that only statistically significant shape uncertainties are implemented, the binning of the BDT distribution is modified when deriving the uncertainties. Starting from the most signal-like bin, bins are merged until a specific condition on the relative MC statistical uncertainty is fulfilled and/or a sufficient fraction of events is contained in the bin. The requirements are manually adapted per variation, process and signal region to achieve a smooth and reliable parametrisation of the uncertainty. For signal-like bins, the statistical uncertainty must be less than 5 to 15 %. For background enriched bins, a bin's fractional contribution to the total number of events must exceed 5 %.

Theoretical uncertainties are derived following the recommendations by the different modelling groups for the different processes. They affect various quantities, such as the choice of the renormalisation and factorisation scale, the choice of PDFs and the matching procedure between the hard scatter and the parton shower algorithm. An overview of these uncertainties is presented in the following.

**Generator, matching and merging** While the hard scatter calculation is exact and well defined within perturbation theory, MC generators differ in their treatments of the matching and merging of non-perturbative parton shower emissions with NLO QCD hard scatter calculations. Different techniques exist to match and merge emissions to avoid double counting, which also depend on the choice of the cut-off, factorisation and renormalisation scales. Uncertainties related to the matching and merging are evaluated by comparing the nominal generator, for most processes POWHEG BOX v2 [168], with an alternative generator.

**Parton shower and hadronisation** The simulation of the parton shower and hadronisation needs to consider non-perturbative effects. Different PS algorithms differ in the treatment of soft and co-linear emissions. In PYTHIA [165], the nominal PS algorithm orders emissions by their $p_{\mathrm{T}}$. HERWIG [189, 190] simulates angular ordered parton showers. PYTHIA and HERWIG also employ different hadronisation models. PYTHIA implements the Lund string model, whereas HERWIG employs a cluster model of colour confinement.

**Factorisation and renormalisation scales** The matrix element calculation treats infinities arising from missing higher order QCD contribution using the renormalisation scale ($\mu_{\mathrm{R}}$). Furthermore, the evaluation of the PDF requires the choice of the factorisation scale ($\mu_{\mathrm{F}}$), which also sets the cut-off scale between the perturbative and non-perturbative physics. Typically, both scales are set to the physical scale of the process $Q$, yielding $\mu_{\mathrm{F}} = \mu_{\mathrm{R}} = Q$. The impact of a specific choice of scale is evaluated by varying both scales individually and combined in the range

$$\frac{1}{2} < \frac{\mu}{Q} < 2 \,. \tag{7.15}$$

**Initial and final state radiation** The simulation of the initial and final state radiation (ISR, FSR) highly depends on the choice of $\mu_{\mathrm{R}}$, $\mu_{\mathrm{F}}$, PDFs and the strong coupling $\alpha_{\mathrm{S}}$, as well as the parameter tune of the PS and hadronisation models.

**Choice of the PDF + $\alpha_S$** Uncertainties are assigned to evaluate the impact of the choice of PDF set as different sets employ different extraction methodologies. For example, the PDF4LHC set [169] provides multiple PDFs , which are compared in terms of their central values and shapes. In addition, these uncertainties account for the uncertainty on the strong coupling constant $\alpha_\mathrm{S}$ at a given choice of renormalisation scale $\mu_\mathrm{R}$.

**Underlying event** Soft radiation from minimum bias *pp* collisions is simulated with phenomenological models, which treat the interaction with the hard scatter and the formation of jets from soft processes. Uncertainties are derived by varying the set of model input parameters, referred to as "tunes".

### 7.11.3 Signal modelling

Several systematic uncertainties are considered for the signal processes. Uncertainties due to finite order calculations affecting the acceptance and shape of distributions are derived with samples which are generated with varied event parameters, such as the factorisation or renormalisation scale.

For the ggF $HH$ signal, the parton shower model uncertainty is evaluated by comparing the nominal Pythia sample with an alternative Herwig 7 sample [190]. PDF uncertainties are assessed following the recommendations by the PDF4LHC collaboration, considering 30 eigenvectors for all PDFs and one uncertainty related to the strong coupling [169]. The eigenvectors are available by weight variations of the samples and are combined in quadrature as

$$(\Delta\mathrm{PDF} + \alpha_S) = \sqrt{\sum_i (\delta\mathrm{PDF}_i)^2 + \left(\frac{\alpha_s^{\mathrm{up}} + \alpha_s^{\mathrm{down}}}{2}\right)^2}\,. \tag{7.16}$$

The largest uncertainty affecting the ggF $HH$ signal is the uncertainty on the cross-section prediction, listed in Table 7.8, which is predominantly driven by the combined uncertainty related to the choice of the renormalisation scale and the top-quark mass scheme. Similar uncertainties apply to the signal cross-sections for alternative $\kappa_\lambda$ and $\kappa_{2V}$ values. These follow the

| **Process** | **Scale + $m_{\text{top}}$** | | **PDF + $\alpha_S$** |
|---|---|---|---|
| ggF $HH$ | $-23\,\%$ | $+6\,\%$ | $3\,\%$ |
| VBF $HH$ | $-0.4\,\%$ | $+0.3\,\%$ | $0.21\,\%$ |

Table 7.8: Cross-section uncertainties for the SM $HH$ production.

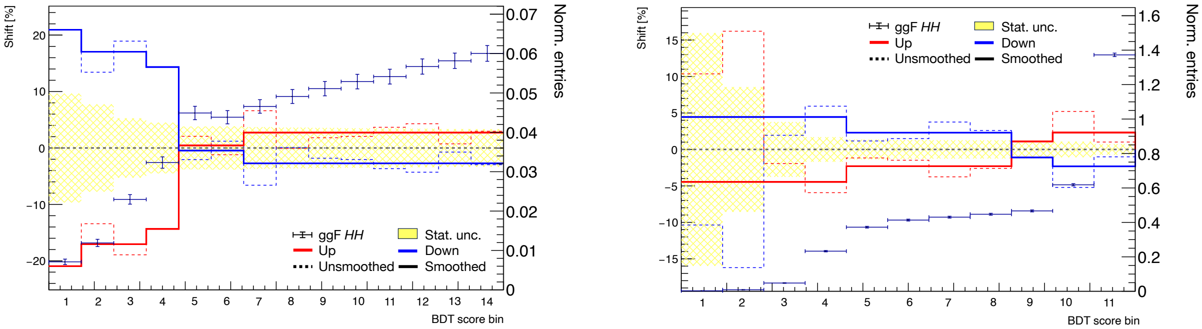
| | **Signal region** | | | | | |
|---|---|---|---|---|---|---|
| **Uncertainty** | low-$m_{HH}$ ggF | | high-$m_{HH}$ ggF | | VBF | |
| $\mu_F/Q \in \{\frac{1}{2}, 2\}$ | $0.5\,\%$ | $-0.6\,\%$ | $-0.1\,\%$ | $0.1\,\%$ | $-1.1\,\%$ | $1.2\,\%$ |
| $\mu_R/Q \in \{\frac{1}{2}, 2\}$ | $-1.2\,\%$ | $1.5\,\%$ | $1.0\,\%$ | $-1.3\,\%$ | $-0.8\,\%$ | $1.2\,\%$ |
| $(\mu_F/Q, \mu_R/Q) \in \{(\frac{1}{2}, \frac{1}{2}), (2, 2)\}$ | $-0.5\,\%$ | $1.1\,\%$ | $0.7\,\%$ | $-1.3\,\%$ | $-1.8\,\%$ | $2.6\,\%$ |
| PS: (HERWIG / PYTHIA) | $\mp 5.8\,\%$ | | $\mp 2.3\,\%$ | | $\mp 0.7\,\%$ | |
| PDF + $\alpha_S$ | $\pm 1.1\,\%$ | | $\pm 0.5\,\%$ | | $\pm 0.3\,\%$ | |
| $\alpha_S$ | $0.1\,\%$ | $-0.2\,\%$ | $0.1\,\%$ | $-0.1\,\%$ | $-0.1\,\%$ | $0.1\,\%$ |
| $\Delta_{\text{stat.}}$ sample: Nom., Alt. | $\pm 1.0\,\%$ | $\pm 1.0\,\%$ | $\pm 0.4\,\%$ | $\pm 0.4\,\%$ | $\pm 1.2\,\%$ | $\pm 1.2\,\%$ |

Table 7.9: Acceptance uncertainties for the SM ggF $HH$ sample in the signal regions in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The PS uncertainties have been symmetrised.

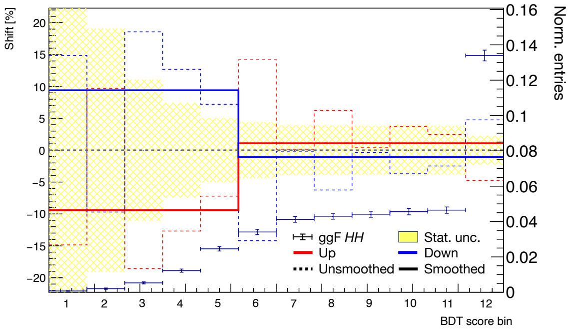
recommendations of the LHC $HH$ working group [40] and remain unchanged from the previous analysis. The same applies to the uncertainties related to the branching fraction for $H \to b\bar{b}$ ($\Delta\mathcal{B} \approx \pm 1.3\,\%$) and $H \to \tau^+\tau^-$ ($\Delta\mathcal{B} \approx \pm 2.0\,\%$).

Table 7.9 lists the acceptance uncertainties for the ggF $HH$ signal in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. With the exception of the parton shower, the uncertainties are similar or smaller than the statistical uncertainty of the samples. No additional parametrisation is implemented for the samples with varied $\kappa_\lambda$ as similar uncertainties are observed when the uncertainties are evaluated for the fully simulated ggF $HH$ $\kappa_\lambda = 10$ sample, or when they are extrapolated to the samples that are generated with $\kappa_\lambda = 0, 20$.

Figure 7.17 shows the shape uncertainty due to the choice of the parton shower model for the three signal regions in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The uncertainty is derived from the ratio of the nominal and alternative samples. The shape uncertainty is smoothed and parametrised to stabilise the fit procedure, and to ensure only statistically significant variations are implemented. The



(a) low-$m_{HH}$ ggF SR (b) high-$m_{HH}$ ggF SR



(c) VBF SR

Figure 7.17: PS shape uncertainty for the ggF $HH$ sample in the signal regions of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The re-binned distribution is shown with error bars. The dashed lines indicate the uncertainty derived from the ratio of the re-binned nominal and alternative distributions. The statistical uncertainty of the ratio is shown in yellow. The solid lines display the symmetrised shape uncertainty, after it has additionally been smoothed as described in Section 7.12.2.

PS uncertainty is implemented symmetrically, by mirroring the “down” variation of (Herwig /Pythia) to −(Herwig/Pythia) as the “up” variation.

The systematic uncertainties for the VBF $HH$ signal are derived in a similar way and are listed in Table 7.10. The uncertainties related to the choice of PDF and $\alpha_S$ are derived with the NNPDF set [171] instead of the PDF4LHC set. Notably, the uncertainty of $\mu_R$ is less than 0.1 % in all three signal regions. The impact of the PS uncertainty is shown in Figure 7.18.

An additional uncertainty is applied to treat the non-closure of the truth $m_{HH}$-based reweighing ggF $HH$ sample to alternative $\kappa_\lambda$ values. The full difference in the normalisation and the shape between the fully-simulated $\kappa_\lambda = 10$ sample and an extrapolated sampled using the three basis samples is taken as the uncertainty. The normalisation uncertainties are listed in Table 7.11 for all

| | **Signal region** | | | | | |
|---|---|---|---|---|---|---|
| **Uncertainty** | low-$m_{HH}$ ggF | | high-$m_{HH}$ ggF | | VBF | |
| $\mu_{\mathrm{F}}/Q \in \{\frac{1}{2}, 2\}$ | 3.3 % | −3.7 % | 0.4 % | −0.4 % | −0.4 % | 0.4 % |
| $\mu_{\mathrm{R}}/Q \in \{\frac{1}{2}, 2\}$ | 0.0 % | 0.0 % | 0.0 % | 0.0 % | 0.0 % | 0.0 % |
| $(\mu_{\mathrm{F}}/Q, \mu_{\mathrm{R}}/Q) < \{(\frac{1}{2}, \frac{1}{2}), (2, 2)\}$ | 3.3 % | −3.7 % | 0.4 % | −0.4 % | −0.4 % | 0.4 % |
| PS: Herwig / Pythia | ±1.3 % | | ±0.6 % | | ±2.6 % | |
| PDF | ±0.8 % | | ±0.5 % | | ±0.3 % | |
| $\alpha_{\mathrm{S}}$ | 0.4 % | −0.1 % | −0.0 % | −0.0 % | −0.1 % | 0.0 % |
| $\Delta_{\mathrm{stat.}}$ sample: Nom., Alt. | ±2.6 % | ±2.6 % | ±1.1 % | ±1.1 % | ±0.7 % | ±0.7 % |

Table 7.10: Acceptance uncertainties for the VBF $HH$ sample in the signal regions in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The PS uncertainties have been symmetrised.

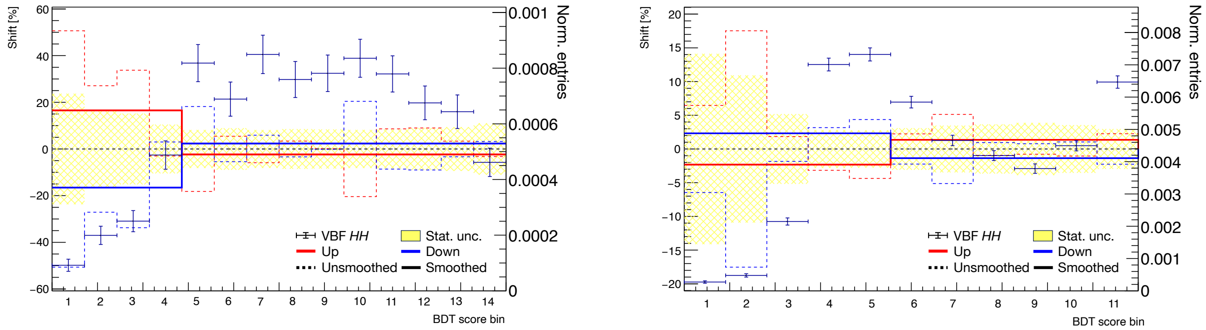
| | **Signal region** | | |
|---|---|---|---|
| **Channel** | low-$m_{HH}$ ggF | high-$m_{HH}$ ggF | VBF |
| $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ | −2 % | 1 % | 25 % |
| $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT | 1 % | 1 % | 20 % |
| $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT | 2 % | 4 % | 40 % |

Table 7.11: ggF $HH$ non-closure uncertainty due to the truth $m_{HH}$-based reweighing for alternative $\kappa_\lambda$ values in all signal regions.

three channels. They are small in the ggF signal regions, but more than 20 % in the VBF signal regions as the truth $m_{HH}$ reweighing fails to correctly predict the number of jets. Nevertheless, the impact on the final result and the confidence interval on $\kappa_\lambda$ is minimal. The associated shape uncertainties are smaller than the statistical uncertainties and are therefore neglected. The VBF $HH$ parametrisation does not rely on extrapolated basis samples given the availability of six fully simulated samples. A potential residual non-closure is evaluated with the four validation samples listed in Table 7.2. No sign of non-closure is observed.

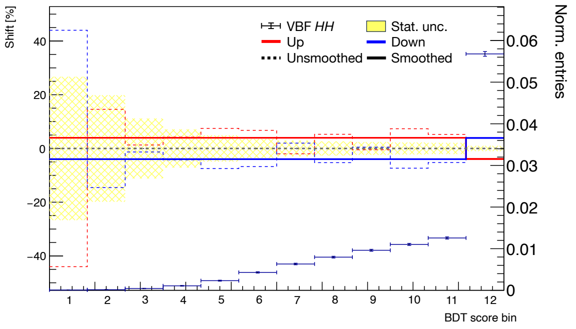
### 7.11.4 Background modelling

**$t\bar{t}$** The $t\bar{t}$ normalisation is determined by a fit to the SRs and the CR, with the CR providing the strongest constraint given its high $t\bar{t}$ purity. Relative acceptance uncertainties across the CR and the SRs are applied to account for differences in the normalisation due to phase space differences.



(a) low-$m_{HH}$ ggF SR

(b) high-$m_{HH}$ ggF SR



(c) VBF SR

Figure 7.18: PS shape uncertainty for the VBF $HH$ sample in the signal regions of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The re-binned distribution is shown with error bars. The dashed lines indicate the uncertainty derived from the ratio of the re-binned nominal and alternative distributions. The statistical uncertainty of the ratio is shown in yellow. The solid lines display the symmetrised shape uncertainty, after it has been additionally been smoothed as described in Section 7.12.2.

They are correlated with the shape uncertainties. Uncertainties related to the matching and merging to the hard scatter with the PS are defined by the difference between the nominal POWHEG BOX v2 + PYTHIA 8 sample [168, 172] and an alternative MADGRAPH5_AMC@NLO+PYTHIA 8 sample [170]. A PS uncertainty is derived from the comparison with an alternative POWHEG+HERWIG 7 sample [190]. In both cases, the nominal and alternative samples are simulated with the ATLAS fast detector simulation (AF2) [33], ensuring that any differences between AF2 and the full detector simulation do not bias the uncertainty estimate. ISR uncertainties are defined with event weights containing three variations of QCD scales and showering parameters. PDF, $\alpha_S$ and FSR uncertainties are also derived by varying event weights. All uncertainties affect both the fractional $t\bar{t}$ contribution and the BDT distribution in the analysis regions. The parton shower (PS) uncertainties have the largest impact, reaching up to 10 % in the VBF signal region of the

$\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel.

**$Z$+HF** Similar to $t\bar{t}$ production, the $Z$+HF normalisation is determined by the fit and is primarily constrained in the CR. Relative acceptance uncertainties account for potential normalisation differences between the distinct phase spaces of the SRs and the CR. These are correlated with shape uncertainties. Historically, uncertainties related to the matching and merging have been one of the largest source and were not fully covered by the internal Sherpa variations. Therefore, in addition to uncertainties from varying matching and merging scales, the nominal sample is compared to an alternative MADGRAPH5_AMC@NLO+PYTHIA 8 [170] sample. Missing higher order QCD contributions are evaluated by varying the renormalisation and factorisation scales. Higher order electroweak corrections have a negligible impact. Uncertainties due to the choice of the PDF and $\alpha_{\text{S}}$ are evaluated by comparing the nominal PDF set to two alternative PDF sets. PDF4LHC uncertainties [169] are found to be negligible. Due to the known $p_{\text{T}}^{V}$ mis-modelling in the nominal MC generator (SHERPA 2.2.11) [203], an additional uncertainty is assigned based on the comparison of the $p_{\text{T}}^{V}$ distribution between data and MC in the CR. In the SRs all uncertainties impact the fractional contribution of $Z$+HF. The largest acceptance uncertainties ($\mathcal{O}(10\,\%)$) arise from the QCD scale variations and the generator comparison. The latter also leads to significant shape uncertainties in the signal regions.

**Single top-quark production** Cross-section uncertainties of 4 to 5 % are applied to the $s$-, $t$- and $Wt$ channels based on NNLO calculations [204, 205]. For the $s$- and $t$-channel, other systematic uncertainties are negligible due to their size, whereas additional uncertainties are applied only to the $Wt$ channel. Uncertainties related to simulation of parton shower, and the matching and merging with the hard scatter are estimated using the difference between a nominal (POWHEG+PYTHIA 8 [168, 172]) and two alternative samples (POWHEG BOX +HERWIG 7 [190] and MADGRAPH5_AMC@NLO+PYTHIA 8 [170]). The nominal and the alternative samples have been processed with the ATLAS fast detector simulation [34]. PDF, ISR and FSR uncertainties are derived by varying QCD scales and shower tunes. The simulation needs to account for the

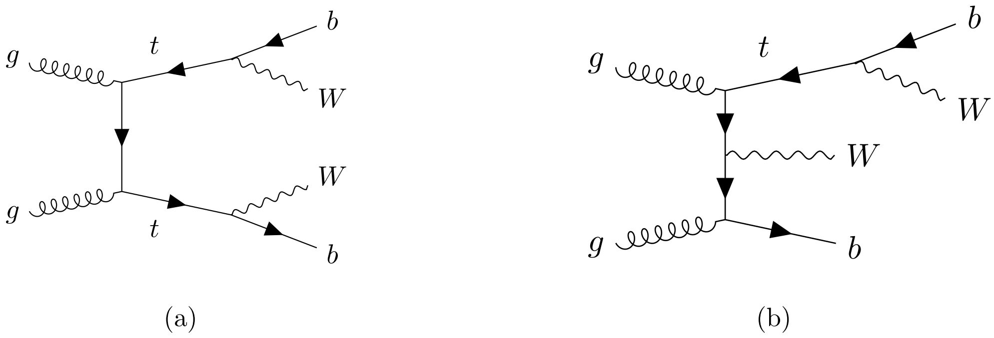


Figure 7.19: Example diagrams for interfering LO $t\bar{t}$ (a) and NLO $Wt$ production (b).

interference between LO $t\bar{t}$ and NLO single-top-quark production illustrated by two example diagrams are shown in Figure 7.19. Different theoretical schemes exist to handle this interference: diagram removal, where $Wt$ diagrams overlapping with $t\bar{t}$ production are omitted, and diagram subtraction, where the overlapping $Wt$ contribution is subtracted from the cross-section [176, 206]. The difference between the two methods is taken as the systematic uncertainty, which is largest in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, reaching up to 34 % in the VBF SR. All uncertainties affect the fractional contribution of the $Wt$ channel, with generator and PS comparisons contributing uncertainties between 15 % and 36 %. QCD variations significantly affect the BDT distribution shapes.

**Single Higgs** Cross-section uncertainties are applied following the recommendations by the LHC Higgs working group [207]. Single Higgs boson production in association with heavy flavour jets, followed by the Higgs boson decay to a pair of $\tau$-leptons ($H \to \tau^+\tau^- + bb, bc, cc$) is difficult to simulate [208, 209]. Thus, an uncertainty of 100 % is assigned for those processes in the ggF, VBF and $WH$ production mode. For $ttH$ and $ZH$, the two heavy-flavour final state quarks arise from LO resonance decays and are therefore much better modelled. Acceptance uncertainties are therefore derived for the processes. For $ttH$, the matching and merging uncertainty is defined by the difference of the nominal sample (Powheg Box v2 + Pythia 8 [168, 172]) with respect to a MadGraph5_aMC@NLO+Pythia 8 sample [170], both processed with the ATLAS fast detector simulation [34]. For $ZH$, they derived by comparing Pythia 8 [172] to Herwig 7 [190]. This comparison is also used to derive PS uncertainties for both, $ttH$ and $ZH$. PDF and $\alpha_S$

uncertainties are derived using the eigenvector variations of the PDF4LHC set [169]. The QCD scales are varied to asses missing higher order contributions. For $ttH$, ISR and FSR variations are assessed by alternative the PS tunes, corresponding to alternative $\alpha_{\mathrm{S}}$ values. For $ZH$, the $VH, H \to b\bar{b}$ analysis [210] found that ISR and FSR uncertainties are negligible. All sources of uncertainties impact the acceptance times efficiency of the processes to fall into the SRs. Only the PS uncertainty has a significant effect on the shape of the BDT distributions.

**Minor backgrounds** The systematic uncertainty estimates are taken from the $VH, H \to b\bar{b}$ analysis [211]. These include $Z$ boson plus light flavour jets, $W$ boson and diboson processes. Only normalisation uncertainties are applied, as shape uncertainties are found to be negligible. The cross-section uncertainties are smaller than the acceptance uncertainties which range between 20 to 50 % depending on the process, channel and region.

**Data-driven backgrounds** The uncertainties for data-driven processes where jets are misidentified as hadronically decaying $\tau$-leptons follow the same procedure of the previous analysis. Uncertainties related to the fractional contribution, and the shape of the BDT are applied in all channels and signal regions.

## 7.12 Statistical Analysis

The analysis uses a binned profile likelihood fit to extract the signal strength and the confidence intervals on the coupling modifiers from data. The fit inputs are the signal versus background BDT classifier distributions in the nine signal regions, and the $m_{\ell\ell}$ distribution in the control region. The fit is implemented using the WorkSpaceMaker framework that is built upon the HistFactory [212], RooFit [213] and RooStats [214] software packages.

### 7.12.1 Statistical model

**Likelihood function** The binned fit corresponds to a counting experiment in each bin $i$. For the signal strength $\mu$ the expected number of events in a bin is given by

$$n_i^{\text{exp}} = \mu\, s_i + b_i \ , \tag{7.17}$$

where $s_i$ and $b_i$ are the number of expected signal and background events. The parameter of interest (POI) is determined by maximising the likelihood function $\mathcal{L}$, which encodes the probability of the observed data given the signal hypothesis. $\mathcal{L}$ is given by the product of Poisson probability terms in each bin. For a single POI, for example the signal strength $\mu$, the $\mathcal{L}$ is given by

$$\mathcal{L}_{\text{phys.}}(\text{data} \mid \mu) = \prod_{i \in \text{bins}} \text{Poisson}\left(n_i^{\text{obs}} \mid \mu s_i + b_i\right) = \prod_{i \in \text{bins}} \frac{(\mu s_i + b_i)^{n_i^{\text{obs}}}}{n_i^{\text{obs}}!} e^{-(\mu s_i + b_i)}, \tag{7.18}$$

where $n_i^{\text{obs}}$ is the observed data events in bin $i$. The likelihood can also be defined in terms of multiple parameters of interest, for example, several signal strengths $\boldsymbol{\mu}$, and multiple POIs can be fitted simultaneously in a joint likelihood framework.

Systematic uncertainties are implemented as a vector of nuisance parameters (NPs) $\boldsymbol{\theta}$. They affect the expected signal and background yields $\{s_i, b_i\} \rightarrow \{s_i(\boldsymbol{\theta}), b_i(\boldsymbol{\theta})\}$. Their central value and uncertainty are typically constrained in auxiliary measurements, as described in Section 7.11, which are implemented as rescaled Gaussian penalty terms with mean zero and unit variance

$$\mathcal{L}_{\text{aux.}}(\boldsymbol{\theta}) = \prod_{\theta_j \in \boldsymbol{\theta}} \text{Gauss}(0 \mid \theta_j, 1) = \prod_{\theta_j \in \boldsymbol{\theta}} \frac{e^{-\theta_j^2/2}}{\sqrt{2\pi}} \ . \tag{7.19}$$

Theoretical modelling uncertainties are implemented in the same way. For example, a prior uncertainty of $\pm 10\,\%$ on a cross-section is implemented such that a change of $\theta \rightarrow \theta + 1$ translates to a $10\,\%$ increase in the associated cross-section. This implementation allows for the joint measurement of the POI and $\theta$, since the fit can both vary the central value of the NP and further constrain its prior, if the data used for the hypothesis testing are sensitive to it. A best fit value $\hat{\theta}$ that deviates from zero is commonly referred to as a “pull”. Some NPs are implemented without auxiliary measurement terms, notably the normalisation of $t\bar{t}$ and $Z$+HF production. They are purely constrained from $\mathcal{L}_{\text{phys.}}$. Their pre-fit values are set to unity, corresponding to the normalisation predicted by simulation.

The expected signal and background yields per bin is estimated from simulation. The statistical uncertainty due to the limited sample size is incorporated into the likelihood function with a simplified version of the Beeston-Barlow [215] procedure. Specifically, one nuisance parameter $\gamma_i$ is assigned per bin to account for the total background uncertainty (rather than assigning one per process) in order to limit the total number of nuisance parameters. The uncertainty on the signal estimate is neglected, as simulating a sufficiently large signal sample is not difficult given the typically high acceptance for signal events of an analysis. The $\gamma_i$-parameters rescale the expected total background yield by $b_i(\boldsymbol{\theta}) \to \gamma_i b_i(\boldsymbol{\theta})$. Due to the use of weighted events the bin yields can take non-integer values. Therefore, the constraint terms are defined by a continuos gamma distribution

$$\mathcal{L}_{\text{MC stat.}}(\boldsymbol{\gamma}) = \prod_{i \in \text{bins}} \frac{(\gamma_i \beta_i)^{\beta_i}}{\Gamma(\beta_i + 1)} e^{-(\gamma_i \beta_i)} \ , \tag{7.20}$$

where $\beta_i = (1/\sigma^{\text{MC}}_{\text{rel. } i})^2$. $\beta_i$ approximates the sum of background events and is calculated from the relative statistical uncertainty on the expected background yield.

Combining the three terms gives the full likelihood as

$$\mathcal{L}(\text{data} \mid \mu, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \mathcal{L}_{\text{phys.}}(\text{data} \mid \mu, \boldsymbol{\theta}, \boldsymbol{\gamma}) \times \mathcal{L}_{\text{aux.}}(\boldsymbol{\theta}) \times \mathcal{L}_{\text{MC stat.}}(\boldsymbol{\gamma}) \ . \tag{7.21}$$

The maximum is numerically determined by minimizing $-\ln \mathcal{L}$ using a gradient descent method. The iterative procedure is implemented in the MINUIT software package [216]. In the following, $\mathcal{L}(\mu, \boldsymbol{\theta})$ refers to $\mathcal{L}(\text{data} \mid \mu, \boldsymbol{\theta}, \boldsymbol{\gamma})$ for simplicity, and $\boldsymbol{\theta}$ includes the full set of systematic uncertainties, including those without an auxiliary measurement term.

The POI is measured from the maximum likelihood ratio $\lambda(\mu)$, defined as

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \ , \tag{7.22}$$

where $\{\hat{\hat{\boldsymbol{\theta}}}\}$ is the set of profiled values that conditionally maximise the likelihood for a fixed value of $\mu$ and $\{\hat{\mu}, \hat{\boldsymbol{\theta}}\}$ are the best fit values that maximise the likelihood overall.

**Discovery of a signal** The claim of a discovery corresponds to the rejection of the background-only hypothesis $H_0 : \mu = 0$. The discovery test statistic $q_0$ is defined as

$$q_0 = \begin{cases} -2\ln\lambda(\mu = 0) & \text{for } \hat{\mu} \geq 0 \\ \\ 0 & \text{for } \hat{\mu} < \mu \ , \end{cases} \tag{7.23}$$

where $\lambda(\mu = 0)$ is the profile likelihood ratio with $\mu = 0$, and $\hat{\mu}$ is the maximum likelihood estimator (MLE) of the signal strength $\mu$. Here, downward fluctuations ($\mu < 0$ ) are treated as maximally compatible with the background-only hypothesis $H_0$ [217]. The probability of observing data equally or less compatible with the $H_0$ hypothesis is expressed as the $p$-value, which is defined as

$$p_0 = \int_{q_{0,\text{obs.}}}^{\infty} f(q_0|\mu = 0)\, dq_0 \ , \tag{7.24}$$

where $q_{0,\text{obs.}}$ is the observed value of the test statistic and $f(q_0|\mu = 0)$ is the probability density function of $q_0$ estimated from MC simulations. In the limit of large sample sizes, the probability density function asymptomatically approaches

$$\frac{1}{2}\delta(q_0) + \frac{1}{2}f_{\chi^2_n}(q_0) \ , \tag{7.25}$$

where $\delta$ is a Dirac $\delta$ distribution and $f_{\chi^2_n}$ is the $\chi^2$ distribution with $n$ degrees of freedom, equal to number of POIs tested. In high energy physics, the discovery test statistic is often expressed in the terms of the discovery significance $Z_0$. The significance is calculated as

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} \; , \tag{7.26}$$

where $\Phi$ is the inverse of the cumulative distribution of the Gaussian normal distribution. By convention, a discovery is claimed for $Z_0 \geq 5$, corresponding to $p_0 = 2.87 \cdot 10^{-7}$ [147].

**Upper limits** In the absence of any observed signal, it is often of interest to determine the largest value of $\mu$ that is still compatible with the observed data. For this purpose, another test statistic is defined as

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \text{for } \hat{\mu} \leq \mu \\ 0 & \text{for } \hat{\mu} > \mu \; , \end{cases} \tag{7.27}$$

where $\lambda(\mu)$ is the profile likelihood ratio and $\hat{\mu}$ is the maximum likelihood estimator of $\mu$. Here, fluctuations above the expected signal are not counted against the signal hypothesis. The $1 - \alpha$ confidence limit $\mathrm{CL}_{s+b}$ on the agreement between data and a value of $\mu$ is given by the $p$-value

$$\mathrm{CL}_{s+b} = p_\mu = \int_{q_\mu,\mathrm{obs.}}^{\infty} f(q_\mu|\mu)\, dq_\mu \;\; > \alpha \; . \tag{7.28}$$

The 95 % upper limit on $\mu$ is determined by finding the largest value of $\mu$ for which $\mathrm{CL}_{\mathrm{s+b}} < 1 - 0.95$.

The method has one shortcoming. In the case of a downward fluctuation below the background-only expectation, the calculation can lead to too stringent exclusion limits, especially for signals that are not well separated from backgrounds. This is, for example, the case for the very small $HH$ signal and the large background in the $HH \to b\bar{b}\tau^+\tau^-$ analysis. Therefore, the $\mathrm{CL}_\mathrm{s}$ prescription [218, 219] is used instead. The observed confidence level for the signal hypothesis is normalised by the confidence level observed for the background-only hypothesis. Limits are obtained from

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{p_{s+b}}{1 - p_b} = \frac{\int_{q_\mu,\mathrm{obs.}}^{\infty} f(q_\mu|\mu)\, dq_\mu}{\int_{q_\mu,\mathrm{obs.}}^{\infty} f(q_\mu|\mu = 0)\, dq_\mu} \;\; > \alpha \; . \tag{7.29}$$

In the large sample limit, the $\mathrm{CL_s}$ limits can be obtained using the asymptotic approximation of $q_\mu$ instead of simulating the probability density function.

**Uncertainty intervals** Confidence intervals on the POIs $\mu$, such as the signal strength or the coupling modifiers are obtained from the likelihood ratio test statistic

$$t_\mu = -2\ln\lambda(\mu)\ . \tag{7.30}$$

In the case of $\kappa_\lambda$ and $\kappa_{2V}$, the POIs are determined by fitting to a linear combination of simulated signal templates, which are described in Subsection 7.3.2. In the limit of large sample sizes, the likelihood ratio test statistic asymptotically follows a $\chi^2$ distribution due to Wilks' theorem, and the likelihood surface is approximately Gaussian near the maximum likelihood estimator $\hat{\mu}$. The $1-\alpha$ confidence limit on the best fit value of the POI(s) $\hat{\boldsymbol{\mu}}$ is given by the values $\boldsymbol{\mu}$ that fulfil the condition

$$-2\Delta\ln\mathcal{L} = -2\,\left(\ln\mathcal{L}(\boldsymbol{\mu}) - \ln\mathcal{L}(\hat{\boldsymbol{\mu}})\right) \leq Y\ , \tag{7.31}$$

where $\ln\mathcal{L}(\hat{\boldsymbol{\mu}})$ is the log-likelihood at its maximum, $\ln\mathcal{L}(\boldsymbol{\mu})$ is the log-likelihood at some test value $\boldsymbol{\mu}$, and $Y$ is the critical value of the $\chi^2$ distribution. The value of $Y$ corresponding to the confidence level $1-\alpha$ and the number of POIs $n$ is given in Table 7.12.

| **CI** | **$-2\Delta\ln\mathcal{L}$** | |
|---|---|---|
| **$1-\alpha$** | $n=1$ | $n=2$ |
| 68.27 % | 1.00 | 2.30 |
| 95.00 % | 3.84 | 5.99 |

Table 7.12: Upper bounds defining confidence intervals. Taken from Reference [147].

**Asimov datasets** During the design and optimisation phase of the analysis, the expected sensitivity is estimated with Asimov datasets [220]. A "pre-fit Asimov" dataset is constructed with nominal SM signal and backgrounds and a fixed signal strength $\mu$, while fixing all NPs to their central input value. A fit of this dataset will, by definition, always extract the input values, but nonetheless allows investigating NP constraints and correlations.

The expected signal significance or expected upper limit is obtain with a “post-fit Asimov dataset”. Unlike the pre-fit Asimov dataset, only the POIs and the NPs related to the signal are fixed to their expected values. The central values of the background NPs are set to the values obtained by an conditional fit to data. Limits obtained using a post-fit Asimov dataset are closer to the observed data results than the expected limits from a pre-fit dataset, particularly when the nuisance parameters deviate significantly from unity, for example, in the case of the background normalization for $Z$+HF production in the $HH \to b\bar{b}\tau^+\tau^-$ analysis.

### 7.12.2 Fit setup

**Fit input** The likelihood function is constructed using the re-binned signal versus background BDT distribution in all signal regions, and the $m_{\ell\ell}$ control region spectrum. Multiple POIs are considered:

- $\mu_{HH}$ - The common signal strength for ggF and VBF $HH$ production, defined as the observed cross-section times branching fraction normalised to the SM prediction (see Equation 7.2).
- $\mu_{\text{ggF}}$ and $\mu_{\text{VBF}}$ - The signal strengths for ggF and VBF production, respectively. These can be fitted either separately or simultaneously.
- $\kappa_\lambda$ and $\kappa_{2V}$ - For the scan of the coupling modifiers, the signal samples are obtained by a linear combination of three ggF and six VBF samples as described in Section 7.3.2. The $\kappa_\lambda$ parametrisation also takes into account changes to the single Higgs boson production cross-section and branching ratios.

During the design and optimisation period, all POIs and signal related NPs are blinded. Starting from the most signal-like bins, bins corresponding to up to 85 % of the expected signal significance are blinded.

**Binning** The BDT output scores are initially binned with 2090 bins in the range from $-1$ (background-like) to $+1$ (signal-like). In the range from $-1$ to 0.99, a uniform bin width of

$10^{-3}$ is used, while the 100 most signal-like bins are assigned a finer width of $10^{-4}$ to enhance sensitivity. The BDT output is subsequently re-binned to ensure a reliable background estimate while retaining the experimental sensitivity in regions with high signal to background ratio. The re-binning procedure is designed to maintain the validity of the asymptotic approximation, to ensure fit stability, and to reduce the overall number of bins. The algorithm iteratively merges bins from most signal- to most background-like if a merged interval satisfies all three of the following criteria:

1. $Z_{\text{merged bin}} = 10\frac{n_s}{N_s} + 5\frac{n_b}{N_b} > 1$.
2. The statistical uncertainty on the MC estimate of the bin is less than 20 %.
3. There are more than three expected background events per merged bin.

Requirement (1) ensures finer binning in regions with a high signal-to-background ratio, preserving sensitivity in the most relevant regions. Requirement (2) addresses the impact of Monte Carlo (MC) statistical uncertainties: studies have shown that uncertainties exceeding 20 % can introduce a bias in the signal strength estimation. Requirement (3) ensures the validity of the asymptotic approximation, which has been verified using MC toy simulations. The expected limits derived from toy samples differ by only a few percent from those obtained in the asymptotic limit, confirming the reliability of the approximation. The $m_{\ell\ell}$ distribution has a bin width of 1 GeV.

**Implementation of nuisance parameters** The impact of a NP on the BDT distributions is encoded as two additional histograms, corresponding to the $\pm 1\,\sigma$ “up” and “down” variations. The fit model employs polynomial interpolation within the $\pm 1\,\sigma$ range and exponential extrapolation beyond it, providing a continuous description of the distribution as a function of the NP. The implementation ensures that the first and second derivatives required for the minimization remain continuous, while also preventing negative yields.

The NPs derived in Section 7.11 are further treated to improve the stability of the fit model. One sided NP are symmetrised. Experimental uncertainties with asymmetric up and down variations are also symmetrized by assigning the average of the two variations, in order to avoid under-constraining the fit. Smoothing is applied to NPs related to systematic uncertainties from reconstruction, identification and selection, which are derived from variations of the underlying four-vectors. Such variations can cause events to migrate between analysis regions or fall outside the acceptance, leading to statistical fluctuations that destabilize the fit. To mitigate this, several smoothing algorithms are employed, with the default algorithm ensuring a monotonic NP effect. The procedure is not applied to theoretical NPs as they are smoothed during their derivation. Lastly, the fit procedure is stabilised and accelerated by removing NPs with a negligible impact on the sensitivity. This so-called “pruning” is applied separately to shape and normalisation uncertainties. NPs are removed if the difference between the nominal and alternative yield or shape variation is small.

Experimental uncertainties are correlated across all fit regions. The same applies to cross-section uncertainties for signal and background processes that are estimated from simulation, and the $t\bar{t}$ and $Z$+HF free floating normalisations. The relative acceptance uncertainty for the two latter processes is correlated with shape uncertainties across the three channels if they originate from the same underlying source. NPs affecting data-driven backgrounds are not correlated.

## 7.13 Results

### 7.13.1 Inclusive signal strength results

| **Best fit** | $\hat{\mu}_{HH}$ | ± | $\Delta\hat{\mu}_{HH}$ |
|---|---|---|---|
| **Combined** | 2.16 | ± | 1.65 |
| $\tau_{\text{had}}\tau_{\text{had}}$ | −0.36 | ± | 1.35 |
| $\tau_{\text{lep}}\tau_{\text{had}}$ SLT | 7.8 | ± | 3.6 |
| $\tau_{\text{lep}}\tau_{\text{had}}$ LTT | 2.6 | ± | 8.1 |

Table 7.13: Best fit results for $\mu_{HH}$ for the combined signal strength, and the signal strength in individual fits in the sub-channels and the CR of the analysis. The uncertainties include statistical and systematic uncertainties.

The best fit result for the inclusive signal strength is $\mu_{HH} = 2.16 \pm 1.65$, where the uncertainty

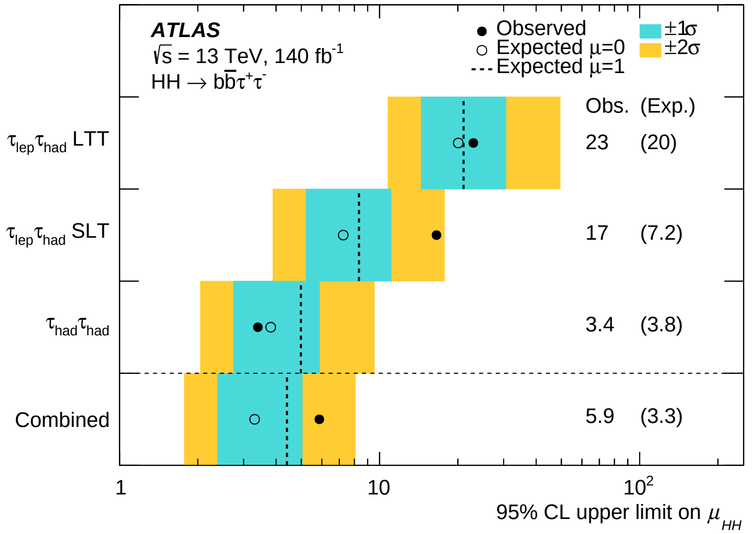


Figure 7.20: 95 % CL upper limits on $\mu_{HH}$ from the individual fits in the sub-channels and the combined fit. Filled circles indicate the observed limits, while open circles represent the expected limits under the background-only hypothesis ($\mu_{HH} = 0$). The expected limits assuming the SM hypothesis ($\mu_{HH} = 1$) are shown as dashed lines. The $\pm 1\,\sigma$ and $\pm 2\,\sigma$ variations on the expected limits, due to statistical and systematic uncertainties relative to the background-only hypothesis, are depicted as turquoise and yellow bands, respectively. Published in Reference [12].

includes both statistical and systematic contributions. The result is primarily limited by the available data statistics, corresponding to an uncertainty of about $(\Delta\mu_{HH})_{\text{Data stat.}} \approx \pm 1.4$. To a lesser extent, the sensitivity is affected by the uncertainty arising from systematic uncertainties of approximately $(\Delta\mu_{HH})_{\text{Systs.}} \approx \pm 1.1$. The finite size of the simulated data corresponds to an uncertainty of approximately $(\Delta\mu_{HH})_{\text{MC stat.}} \approx \pm 0.5$. Table 7.13 summarizes the best-fit values obtained from individual fits to the sub-channels and the $Z$+HF control region of the analysis. The results indicate that the $\tau_{\text{had}}\tau_{\text{had}}$ channel has the best sensitivity and that the large best-fit value of the signal strength $\mu_{HH}$ is driven by an excess in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel. The best fit values of the free floating background normalisations obtained in the combined fit is $t\bar{t} : 0.96 \pm 0.03$ and $Z + HF : 1.34 \pm 0.08$. While the $t\bar{t}$ normalisation is compatible with unity, the $Z$+HF normalisation deviates significantly from unity as has also been observed in other analyses [211].

Figure 7.20 shows the observed and expected limits with respect to the background-only hypothesis ($\mu_{HH} = 0$) on the $HH$ signal strength, both for the combined fit and per channel. Additionally, the expected limits with respect to the SM hypothesis ($\mu_{HH} = 1$) are indicated as dashed lines. The observed limit is 5.9 times the SM expectation, and larger than the expected limit of $3.9^{+1.7}_{-0.9} \times$ SM and corresponds to a $1.4\,\sigma$ excess over the background-only hypothesis. As expected from the comparisons of the uncertainties on the signal strength, the $\tau_{\text{had}}\tau_{\text{had}}$ channel is most sensitive and and provides the strongest limit on the signal strength. In the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel, the observed limit is weaker than the expected. Due to this excess, the observed limit in the combined fit is worse than the observed limit in the $\tau_{\text{had}}\tau_{\text{had}}$ channel alone. Dedicated studies on the excess are presented below. When fitting the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel alone, the excess corresponds to a local significance of $2.7\,\sigma$ relative to the background-only hypothesis, and $2.3\,\sigma$ compared to the SM hypothesis.

The impact of systematic uncertainties is estimated by comparing the expected limit from the full fit to that from a fit where all systematic uncertainties, except for the $t\bar{t}$ and $Z$+HF normalisations and the statistical uncertainty on the MC estimate, are neglected. The expected limit increases by 23 % with respect to the fit with reduced systematic uncertainties.

### Fit cross checks

The complex fit model requires a thorough validation of both the likelihood function and the fit procedure. Initial checks have been performed in a blinded setup. Conditional fits, with the POIs fixed to either the background-only or the SM hypothesis, were conducted to understand the nuisance parameter constraints and pulls.

**BDT post-fit distributions** Figure 7.21 shows the post-fit distributions for signal and backgrounds compared to the data in the nine SRs. The stacked signal and background contributions are scaled to their best fit values. The ratio of data to the summed post-fit signal and backgrounds are shown in the lower panels, along their statistical uncertainties. For comparison, the pre-fit background estimates are indicated with blue dashed lines. Additionally, the SM ggF and VBF $HH$ signal distributions are overlaid. Overall, there is a good agreement between the observed data and the (post-fit) background and signal expectations. The $t\bar{t}$ prediction is slightly reduced due the normalisation and the pulls of NPs lowering the prediction in the background-like regions. The sensitivity is driven by the high-$m_{HH}$ ggF SRs, due to their superior signal to background ratio.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the fit prefers a negative signal strength, as shown by the channel-specific fit. This can be understood from the ratio of the data to the post-fit prediction in the last two bins of the high-$m_{HH}$ region. Nevertheless, the measured data remain statistically compatible with the post-fit predictions. In the VBF SR, there is an under-fluctuation in the second to last bin, which does not affect the best fit value meaningfully due to the low statistical power of this SR.

In the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel, more events than expected are observed in several bins of the high-$m_{HH}$ ggF SR and VBF SR. An excess of events is observed particularly in the most signal-like bin in the high-$m_{HH}$ ggF SR. Consequently, the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel-only fit yields a signal strength of $7.8 \pm 3.6$. In the LTT channel, the data agree very well with the prediction except the most signal-like bin of the VBF SR. Since the sensitivity mostly comes from the high-$m_{HH}$ ggF SR where no excess is observed, the LTT only fit yields a signal strength of about 2.6, albeit with a very large error, that mostly arises from the low signal to background in the channel.

**Post-fit variable distributions** The post-fit distributions for other variables other than the BDT score allow verifying whether the fit is correcting generic mis-modelling or if particular phase space regions are mis-modelled. Figure 7.22 shows the pre- and post-fit distributions of $m_{bb}$ and $m_{\tau\tau}^{\text{MMC}}$ for the most sensitive SR ($\tau_{\text{had}}\tau_{\text{had}}$ ggF high-$m_{HH}$ SR), and of $m_{HH}$ in the combined ggF region. The post-fit prediction is obtained from the sum of the signal prediction, scaled by the best-fit signal strength, and the background predictions. The NPs are applied as obtained by the fit to the prediction with exception of the MC statistical uncertainty described by the $\gamma$-functions, as those are derived specifically for the BDT bins.

Overall, a good agreement between data and the post-fit prediction is observed. The reduction of the $t\bar{t}$ process is larger than the combined free floating normalisation factor of $0.96 \pm 0.03$. This

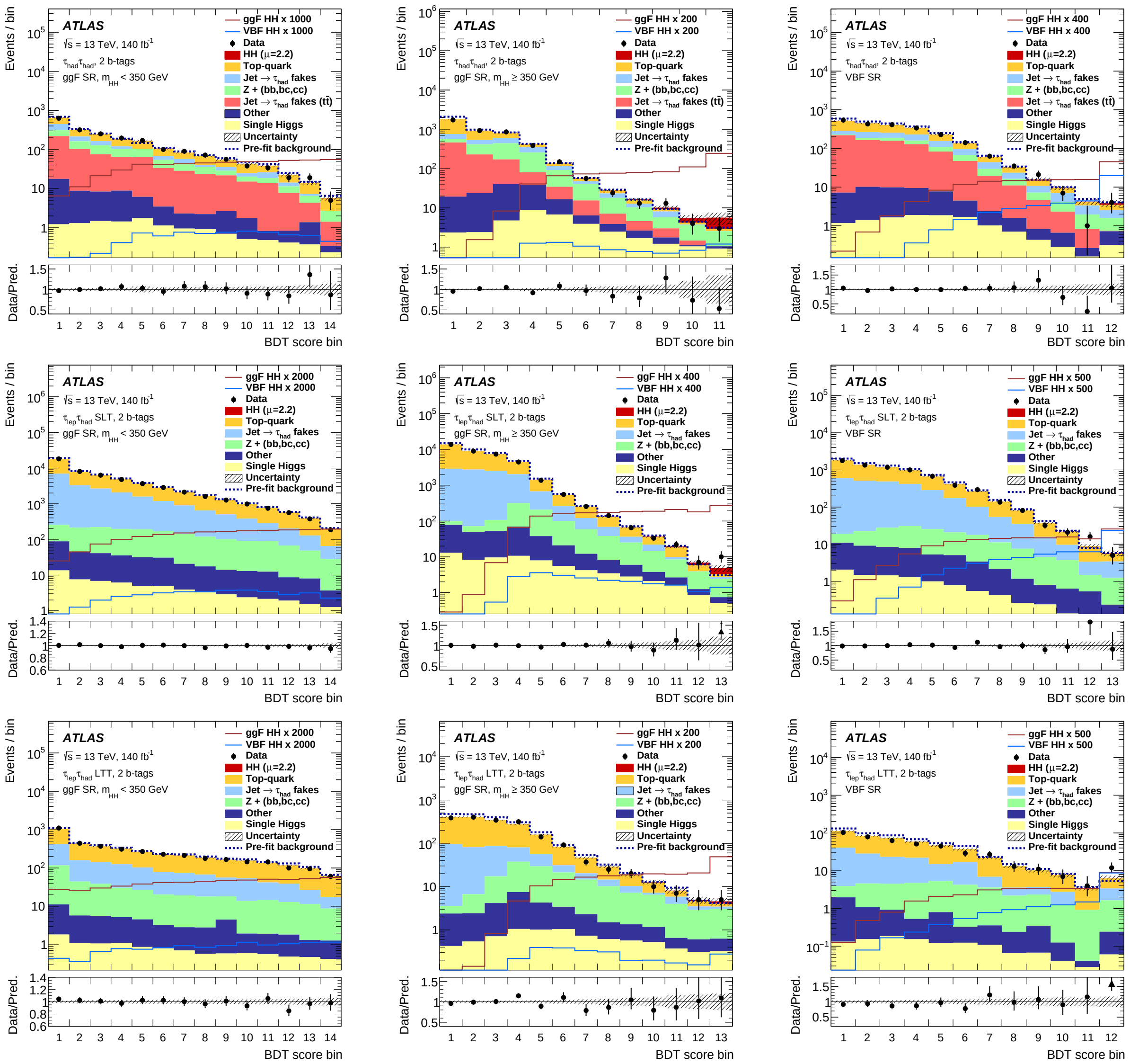
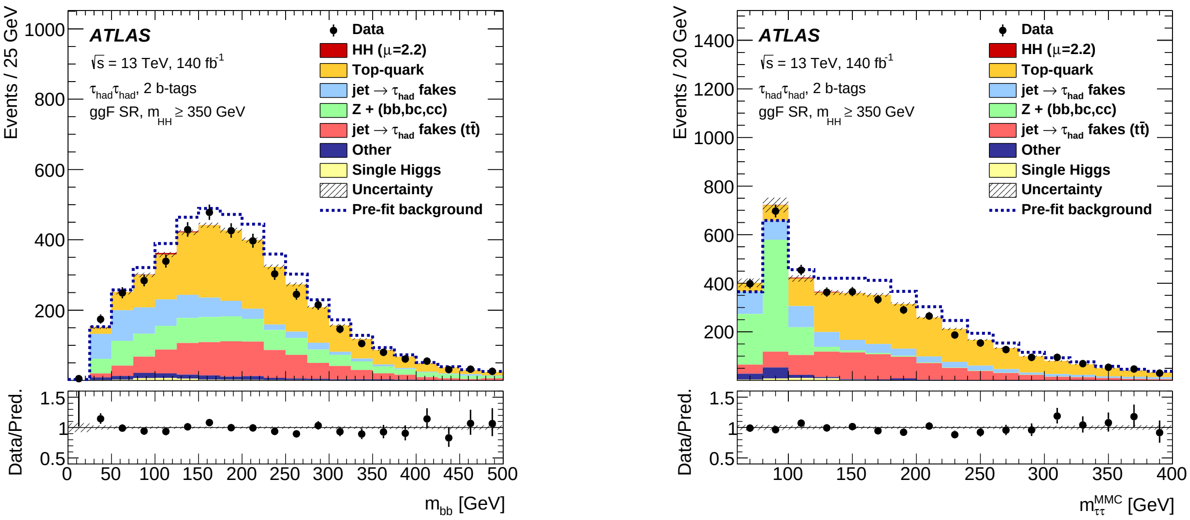


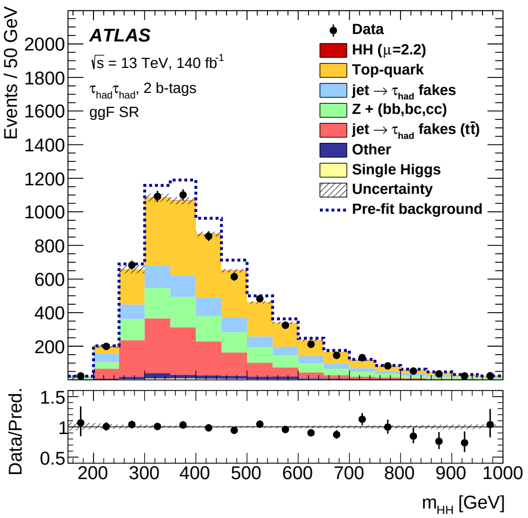
Figure 7.21: Predicted and observed BDT score distributions in the signal regions. The pre-fit estimates are shown as dashed lines. From top to bottom, the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SLT and the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ LTT channels are shown. From left to right, the low-$m_{HH}$ and high-$m_{HH}$ ggF SRs, and VBF SR are displayed. The contributions from various processes are indicated in the legend. "Other" includes contributions from $Z$+light flavour, $W$ boson, $VV$, and $t\bar{t}V$ production processes. For visualisation each bin after the BDT re-binning is displayed with uniform width and labelled with an index The lower panels show the ratio of the data to the post-fit sum of signal and background , along with the statistical uncertainty on the ratio. The CR is shown in Figure 7.13. Published in Reference [12].

is due to NPs that are adjusted with respect to their pre-fit central values.

**NP constraints, pulls and ranking** Given the nature of the profile-likelihood fit, the central values and uncertainties of the NPs can change during the fit if the fitted data are more sensitive to



(a) $m_{bb}$ - $\tau_{\rm had}\tau_{\rm had}$ ggF high-$m_{HH}$ SR (b) $m_{\tau\tau}^{\rm MMC}$- $\tau_{\rm had}\tau_{\rm had}$ ggF high-$m_{HH}$ SR



(c) $m_{HH}$- $\tau_{\rm had}\tau_{\rm had}$ ggF SRs

Figure 7.22: Predicted and observed distributions of (a) $m_{bb}$, and (b) $m_{\tau\tau}^{\rm MMC}$ in the $\tau_{\rm had}\tau_{\rm had}$ ggF high-$m_{HH}$ signal region, and (c) $m_{HH}$ in the combined $\tau_{\rm had}\tau_{\rm had}$ ggF signal region. The contributions from various processes are indicated in the legend. "Other" includes contributions from $Z$+light flavour, $W$ boson, $VV$, and $t\bar{t}V$ production processes. The lower panels show the ratio of the observed data to the post-fit sum of signal and background, along with the statistical uncertainty of the ratio. Published in Reference [12].

them than the constraints provided by the auxiliary measurements included in the likelihood. The post-fit uncertainty $\sigma_{\hat{\alpha}_i}$ of a NP is given by the diagonal of the covariance matrix $\sigma^2_{\hat{\alpha}_i} = \mathrm{cov}(\alpha_i, \alpha_i)$, which is evaluated at the maximum of the likelihood function

$$\mathrm{cov}(\alpha_i, \alpha_j) = -\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_i \alpha_j}\right)^{-1} . \tag{7.32}$$

A nuisance parameter is said to be “constrained” if its post-fit uncertainty is smaller than its pre-fit uncertainty $\sigma_{\alpha_i}$, indicating that the data are particularly sensitive to that NP.

The impact of a NP is measured by its “pull”, which is defined as the difference between the best fit value $\hat{\alpha}_i$ and the pre-fit central value $\alpha_i$, normalised by the pre-fit uncertainty

$$\text{pull}_i = \frac{\hat{\alpha}_i - \alpha_i}{\sigma_{\alpha_i}} \; . \tag{7.33}$$

The correlation between two NPs $\rho_{\alpha_i\alpha_j}$ is calculated from the corresponding covariance matrix entry, divided by the uncertainty of each NP

$$\rho_{\alpha_i\alpha_j} = \frac{\text{cov}(\alpha_i, \alpha_j)}{\sigma_{\alpha_i}\sigma_{\alpha_j}} \; . \tag{7.34}$$

The normalisations of $t\bar{t}$ and $Z$+HF production are implemented as unconstrained NPs and are therefore determined purely from data. Since the best fit value is driven by the CR (shown in Figure 7.13), these normalisations change little when the signal regions in the three analysis channels are included in the fit. The two normalisations are correlated ($\rho \approx 0.57$) as they are sensitive to the same flavour-tagging related uncertainties.

Figure 7.23 shows the NPs which have been determined to have the largest impact on the signal strength. Per NP $\alpha_i$, the post-fit impact is evaluated by varying the best fit result $\hat{\alpha}_i$ within $\pm 1\sigma_{\hat{\alpha}_i}$, and evaluating the corresponding change in $\mu_{HH}$, with respect to the best fit value $\hat{\mu}_{HH}$. The impact on $\mu$ is shown as boxes whose colours indicate the type of uncertainty. The pull and the post-fit uncertainties are also indicated as points and error bars.

The highest ranked systematic uncertainties are related to the modelling of signals and backgrounds, and the statistical uncertainty due to the limited size of the simulated samples (denoted “MC statistics”), particularly in the signal-like bins of the high-$m_{HH}$ SRs across all three channels. In general, the systematic uncertainties are significantly smaller than the dominant uncertainty, which stems from the limited size of the data sample.

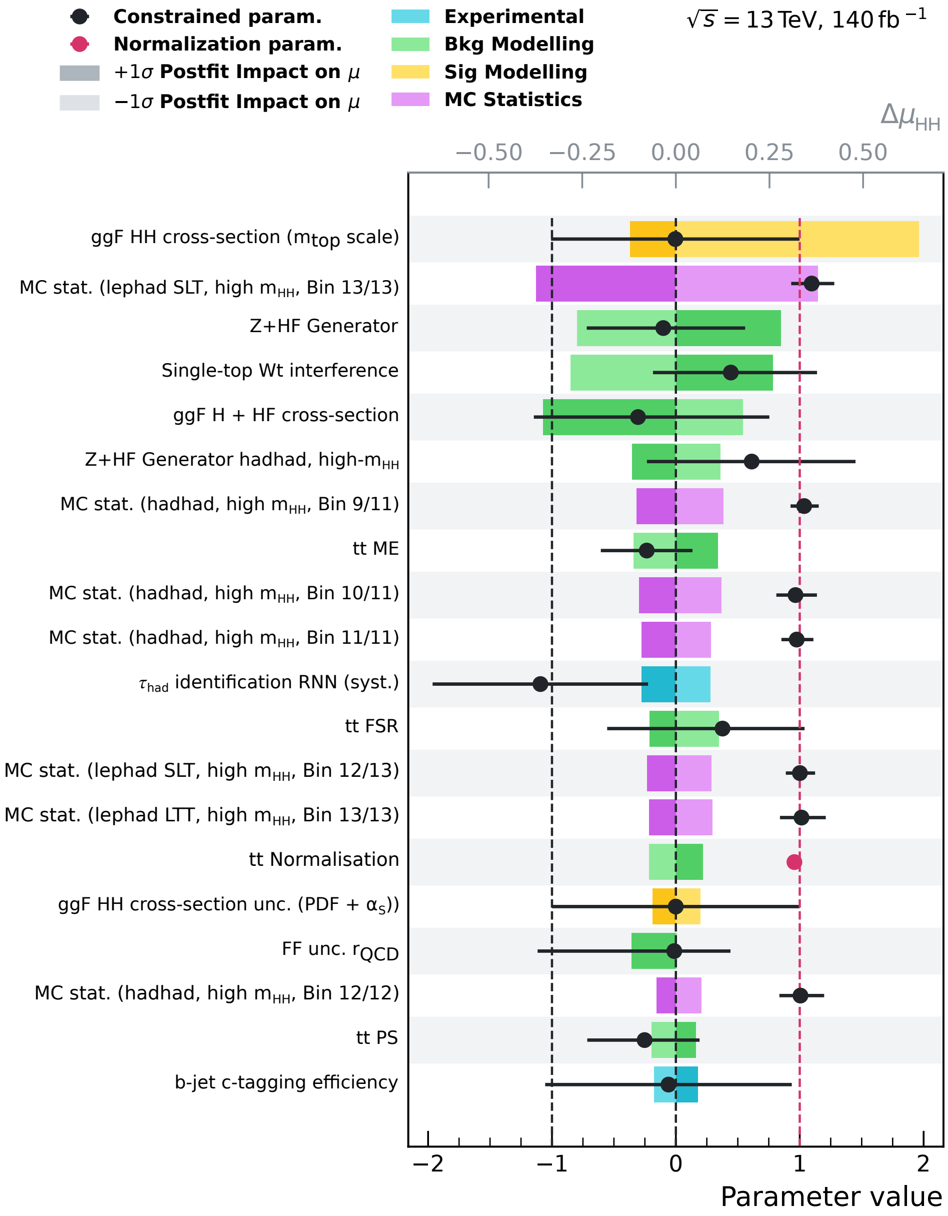
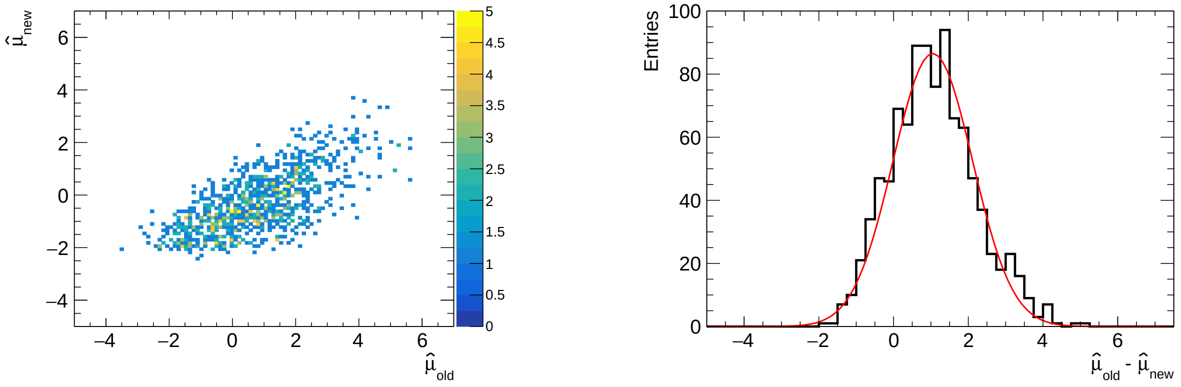


Figure 7.23: The impact of individual NPs on the signal strength $\mu_{HH}$, sorted in decreasing order. The coloured boxes and the top $x$-axis show the post-fit impact on the best-fit value $\hat{\mu}_{HH}$. The black points indicate the pull on NPs expressed in units of standard deviations on the bottom $x$-axis, for NPs that are constrained. MC statistical NPs, which are unconstrained and have a nominal value of 1, are also shown. The error bars show the post-fit uncertainties with respect to the pre-fit constraints. The red points show the post-fit value for the free floating normalisation parameters. The statistical uncertainty is $(\Delta\hat{\mu}_{HH})_{\text{Stat.}} = \pm 1.38$.

The largest systematic uncertainty arises from the cross-section prediction of the ggF $HH$ signal due to the choice of the top-quark mass scale. This is expected, as the large observed signal strength increases the importance of NPs that scale with the signal strength. Notably, this NP is neither pulled nor constrained by the fit. The second highest ranked NP is the statistical uncertainty in the most signal-like bin in the SLT high-$m_{HH}$ SR, that is implemented as a Beeston-Barlow $\gamma$ with a pre-fit value of 1. The NP is slightly pulled towards larger values as more events are observed in that bin than expected and the fit cannot adjust the signal strength to compensate as other channels prefer a lower $\mu_{HH}$. NPs related to the modelling of background processes are pulled and slightly constrained. The pattern of pulls and constraints, as well as their correlations, are similar to those observed in a fit to data with the signal strength fixed to zero. This suggests that the pulls originate from correcting generic modelling issues rather than the presence of signal. This is also true for the most-pulled NP ("$\tau_{\text{had}}$ identification RNN (syst.)"), which is related to the systematic uncertainty on the identification efficiency of $\tau_{\text{had-vis}}$ candidates. This pull arises because the data in the $\tau_{\text{had}}\tau_{\text{had}}$ ggF high-$m_{HH}$ region prefer a lower $t\bar{t}$ yield, which is achieved by adjusting both this NP and the $t\bar{t}$ normalisation.

**Self-consistency** An additional cross-check is the self-consistency of the fit result, both across the different channels, and across all nine SRs. For the former, the combined fit is compared to a fit with three separate POIs, one per channel. For the latter, the comparison is done with a fit with nine POIs, one per SR. The compatibility is evaluated by the difference in the negative log-likelihood values of the best-fit points for the two setups, denoted as $-\Delta\mathcal{L}$. This quantity is approximated by a $\chi^2_n$-distribution with $n$ degrees of freedom, where $n = 3 - 1 = 2$ ($n = 9 - 1 = 8$) for the first across the different channels (the nine SRs). The resulting $p$-values reflect the fact that there is an approximate $2\sigma$ excess in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels, yielding $p \approx 5\,\%$ for the fit with separate POIs per channel and similarly $p \approx 5\,\%$ for the fit with separate POIs per SR.

**Compatibility with the previous Run 2 result** The re-analysis of Run 2 dataset retains the same trigger and event pre-selection as the previous Run 2 result. Most of the simulated events are re-used, with the exception of an expanded di-leptonic $t\bar{t}$ sample and an updated simulation for the $Z$+jets process. Therefore, the compatibility of the two analyses, taking into account statistical and systematic uncertainties, is quantified using a bootstrapping technique [221]. In



(a) (b)

Figure 7.24: (a) Two-dimensional distribution of the best-fit results $\hat{\mu}_{\text{new}}$ and $\hat{\mu}_{\text{old}}$, and (b) their difference ($\hat{\mu}_{\text{old}} - \hat{\mu}_{\text{new}}$) for the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

this procedure, 1000 alternative datasets are generated by replacing the event weights of the data (all initially set to 1) with new weights drawn from a Poisson distribution with a mean of 1. Since the resampling does not include the MC samples, the resulting compatibility estimate represents a conservative lower bound and the actual compatibility may be higher. For each resampled dataset the best-fit signal strength is determined using both the new analysis model ($\hat{\mu}_{\text{new}}$) and the old analysis ($\hat{\mu}_{\text{old}}$).

The statistical compatibility is evaluated separately for the three channels, and for the combined fit. As an example, the two-dimensional distribution of ($\hat{\mu}_{\text{new}}$, $\hat{\mu}_{\text{old}}$) in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, and the distribution of their difference $\hat{\mu}_{\text{old}} - \hat{\mu}_{\text{new}}$, are shown in Figure 7.24. The correlation coefficient is approximately 70 %. A conservative estimate of the compatibility is obtained by taking the mean of the difference distribution and normalising it by its standard deviation, calculated from the corresponding histogram. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the results are statistically compatible within $< 1\,\sigma$, in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel within $< 2.5\,\sigma$, and in the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT channel within $< 0.1\,\sigma$. The combined fit result shows compatibility within $< 0.8\,\sigma$. The reduced compatibility in the SLT channel is driven by the excess observed in the most signal-like bins.

**Investigations regarding the observed excess in the SLT channel** The excess in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel is driven by the most signal-like bin in the high-$m_{HH}$ SR, where ten events are observed. Despite extensive investigations, no clear indication of an error was found, and the excess is therefore considered most likely to be a statistical fluctuation. A summary of the studies performed is given below.

Of the ten events observed in the most signal-like bin, five were also classified in the most signal-like bin of the signal-versus-background classifier used in the previous analysis. The remaining were previously categorised in the second (three events) and third (two events) most signal-like bins.

Since the new BDT has been optimised and uses additional input variables, it is plausible that the new BDT recovers more signal-like events compared to the earlier analysis. Furthermore, it has been verified that the excess is not caused by the updated signal region categorisation or by the use of new simulated samples. The BDT input variables have been checked, and no anomalies or inconsistencies were identified. No evidence of mis-modelling was observed in the signal-like region defined by a BDT score $> 0.5$.

The ten events originate from different LHC runs, and have different properties of the $\tau$-lepton candidates. At most, one or two events could plausibly involve misidentified calorimeter-tagged muons. One such candidate is reconstructed as a 1-prong leptonic $\tau$-candidate at $\eta \approx 0.3$, near a muon spectrometer cable tray that is known to reduce the muon reconstruction efficiency around $\eta \approx 0$. The second candidate is reconstructed as a 3-prong $\tau$-candidate, making this scenario unlike. A tighter overlap removal algorithm between reconstructed muon- and $\tau$-candidates has been tried and found not to affect any of the ten events.

### 7.13.2 Signal strengths for ggF and VBF $HH$ production

The analysis design and fit setup allow extracting the signal strength for ggF and VBF $HH$ production separately. The best fit values for $\mu_{\text{ggF}}$ and $\mu_{\text{VBF}}$ are given in Table 7.14 for two fit types. In the individual case, the other POI is fixed to the SM prediction of 1. In the simultaneous case, both POIs can vary at the same time. Only minor differences are observed between the two fit results, thanks to the inclusion of dedicated VBF $HH$ signal regions in the analysis.

| **Best fit** | | $\hat{\mu}_{\mathbf{ggF}}$ | ± | $\Delta\mu_{\mathbf{ggF}}$ | $\hat{\mu}_{\mathbf{VBF}}$ | ± | $\Delta\mu_{\mathbf{VBF}}$ |
|---|---|---|---|---|---|---|---|
| Individual fit | Combined | 2.1 | ± | 1.7 | 21 | ± | 36 |
| | $\tau_{\text{had}}\tau_{\text{had}}$ | −0.36 | ± | 1.35 | −17 | ± | 35 |
| | $\tau_{\text{lep}}\tau_{\text{had}}$ SLT | 7.9 | ± | 3.7 | 21 | ± | 62 |
| | $\tau_{\text{lep}}\tau_{\text{had}}$ LTT | 1.5 | ± | 7.9 | 373 | ± | 177 |
| Simultaneous fit | Combined | 2.0 | ± | 1.7 | 17 | ± | 36 |
| | $\tau_{\text{had}}\tau_{\text{had}}$ | -0.28 | ± | 1.38 | −11 | ± | 37 |
| | $\tau_{\text{lep}}\tau_{\text{had}}$ SLT | 7.9 | ± | 3.7 | −2.6 | ± | 58 |
| | $\tau_{\text{lep}}\tau_{\text{had}}$ LTT | −3.1 | ± | 7.9 | 394 | ± | 181 |

Table 7.14: Best fit results for the ggF and VBF signal strength POIs in the individual (1D) and simultaneous (2D) fit of $\mu_{\text{ggF}}$ and $\mu_{\text{VBF}}$.

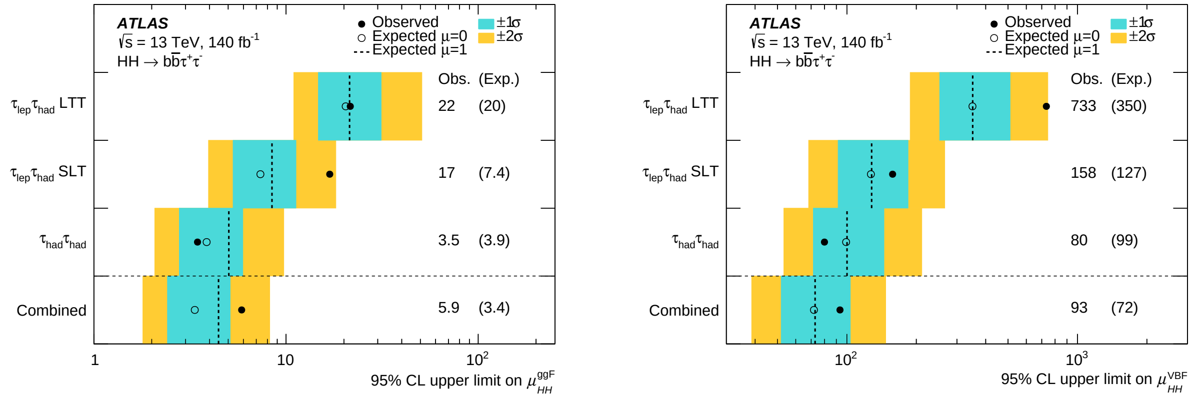
Given the larger cross-section of ggF $HH$ production, the fitted values of $\mu_{\text{ggF}}$ and $\mu_{HH}$ are similar. As for the combined fit, the $\tau_{\text{had}}\tau_{\text{had}}$ channel is the most sensitive for both POIs. The corresponding limits at 95 % CL are shown in Figure 7.25, which are also dominated by the contribution of the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

### 7.13.3 Constraints on the coupling modifiers $\kappa_\lambda$ and $\kappa_{2V}$

The expected and observed values of the negative log-likelihood ratio ($-2\ln\Lambda$) as a function of the coupling modifier $\kappa_\lambda$ are shown in Figure 7.26a. All POIs except the one being scanned are fixed to their respective SM values. At each scan point, the NPs are re-fitted. The combined fit yields a 95 % confidence interval (CI) of $\kappa_\lambda \in [-3.1, 9.0]$, which is slightly larger than the expected interval of $\kappa_\lambda \in [-2.5, 9.3]$ due to the excess in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT channel as described below.

The expected and observed values are shown separately for the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ (SLT + LTT) channels in Figures 7.26b and 7.26c highlighting that the $\tau_{\text{had}}\tau_{\text{had}}$ channel is more sensitive. The second minima observed in the expected $\tau_{\text{had}}\tau_{\text{had}}$ and the combined scan around $\kappa_\lambda \approx 6$ arise due to the quadratic dependency of the cross-section on $\kappa_\lambda$. The $m_{HH}$ split in the signal regions reduces the degeneracy.

The observed scan gives a best-fit value of $\hat{\kappa}_\lambda \approx 0$, and exhibits two local minima, one at $\kappa_\lambda \approx 0$ and the second one at $\kappa_\lambda \approx 6$. The second minimum is not fully degenerate with the first given

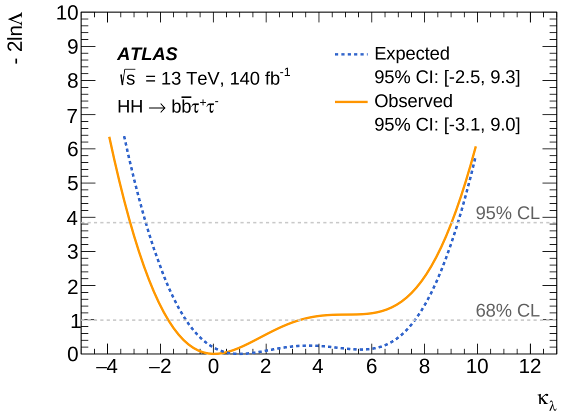


(a) $\mu_{\mathrm{ggF}}$ (b) $\mu_{\mathrm{VBF}}$

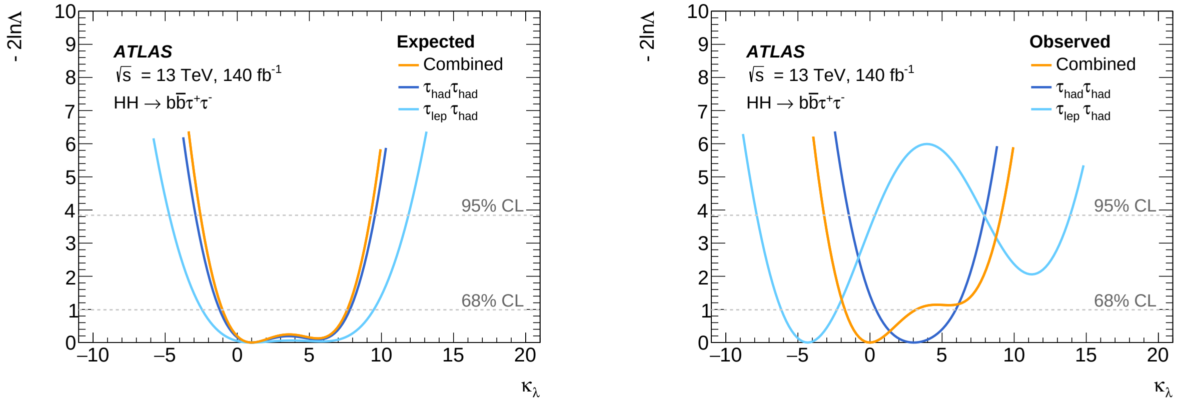
Figure 7.25: 95 % CL upper limits on $\mu_{\mathrm{ggF}}$ and $\mu_{\mathrm{VBF}}$ from the individual fits in the sub-channels and the combined fit. Filled circles indicate the observed limits, while open circles show the expected limits under the background-only ($\mu_{HH} = 0$) hypothesis. The expected limits under the SM hypothesis ($\mu_{HH} = 1$) are shown with dashed lines. The $\pm1\,\sigma$ and $\pm2\,\sigma$ variations of the expected limit with respect to the background-only hypothesis accounting for statistical and systematic uncertainties, are shown as turquoise and yellow bands, respectively. Published in Reference [12].

the difference in the $m_{HH}$ spectra at those values of $\kappa_\lambda$. The shape of the combined scan can be understood from the scans in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels. In the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, the observed $-2\ln\Lambda$ minimum is shifted towards $\kappa_\lambda \approx 2$, corresponding to the point where the cross-section is minimal and consistent with the negative observed $\mu_{HH}$ in this channel. On the contrary, the observed excess (large $\hat{\mu}$) in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels prefers a larger cross-section, producing two minima at approximately $\kappa_\lambda \approx -4$ and $\kappa_\lambda \approx 12$. The excess observed in the high-$m_{HH}$ SR in SLT channel corresponds to a SM-like $m_{HH}$ spectrum. Therefore, the second minium around $\kappa_\lambda \approx 12$ is less pronounced than the $\kappa_\lambda \approx -4$ case, because at $\kappa_\lambda \approx 12$ the $m_{HH}$ spectrum is significantly softer than the SM spectrum.

The impact of systematic uncertainties on the expected $\kappa_\lambda$ confidence interval (CI) is estimated in the same way as for $\mu_{HH}$. A fit without systematic uncertainties, apart from the MC statistical ones, yields a CI that is approximately 9 % narrower. The observed and expected $\kappa_\lambda$ CI are slightly affected by an issue reported in Reference [222] concerning the prediction of ggF $HH$ production in non-SM scenarios. When the ggF HH signal yields in the analysis regions are scaled according to the correction published in Reference [49], the impact on the 95 % CIs is estimated



(a) Observed and expected values for the combined fit.



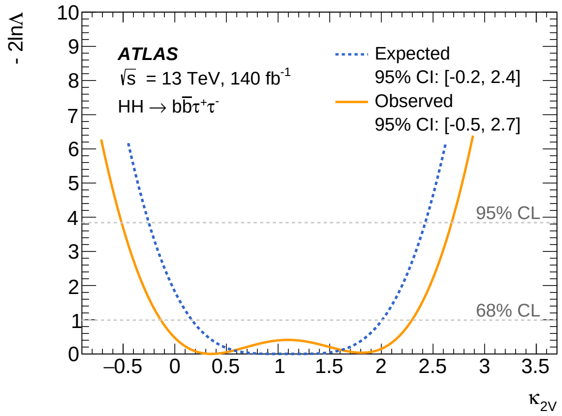
(b) Expected values per channel.

(c) Observed values per channel.

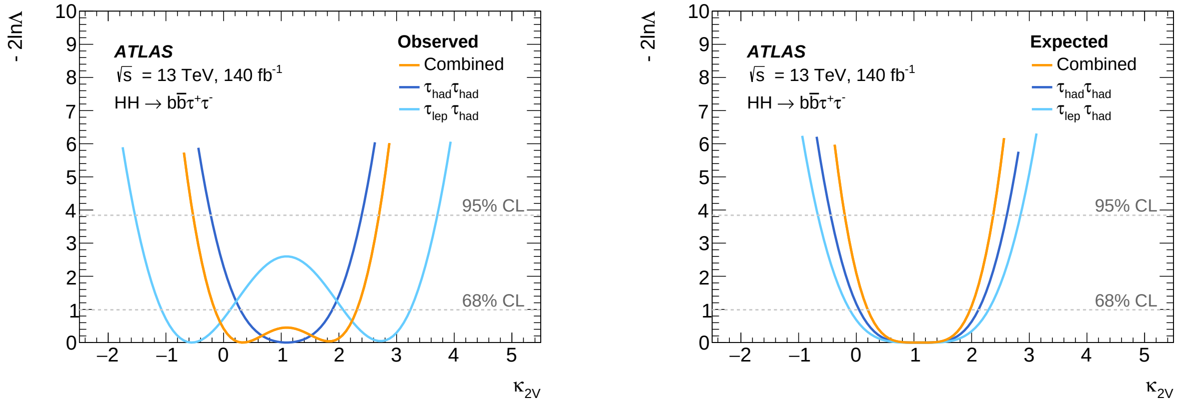
Figure 7.26: Values of the negative log-likelihood ratio as a function of $\kappa_\lambda$. Observed and expected values for the combined fit (a), expected (b) and observed (c) per channel. The expected values are obtained from a fit to an SM Asimov dataset. All POIs except the scanned parameter are fixed to their SM values at each scan point. Published in Reference [12].

to be less than 5 %.

For the coupling modifier $\kappa_{2V}$, the combined fit yields a 95 % CI of $\kappa_{2V} \in [-0.5, 2.7]$. Similar to the $\kappa_\lambda$ scan, the observed CI is slightly wider than the expected interval of $\kappa_{2V} \in [-0.2, 2.4]$ due to the excess in the SLT channel. The impact of systematic uncertainties is estimated to be about 2 %. The expected scan yields a wide minium around the SM prediction. As with $\kappa_\lambda$, the observed and expected $\kappa_{2V}$ scans are dominated by the $\tau_{\text{had}}\tau_{\text{had}}$ channel, where the negative observed signal strength corresponds to a reduced cross-section, consistent with the negative signal strength in that channel. The observed excess in $\tau_{\text{lep}}\tau_{\text{had}}$ in both the SLT and LTT channels broaden the combined scan as the $\tau_{\text{lep}}\tau_{\text{had}}$ scan exhibits two degenerate minima around



(a) Observed and expected values for the combined fit.



(b) Observed values per channel.

(c) Expected values per channel.

Figure 7.27: Values of the negative log-likelihood ratio as a function $\kappa_{2V}$. Observed and expected values for the combined fit (a), observed (b) and expected (c) per channel. The expected values are obtained from a fit to an SM Asimov dataset. All POIs except the scanned parameter are fixed to their SM values at each scan point. Published in Reference [12].

$\kappa_{2V} \approx -0.5$ and $\kappa_{2V} \approx 2.7$. The combined scan therefore features two minima $\kappa_{2V} \approx 0.5$ and $\kappa_{2V} \approx 1.7$. Consequently, the combined scan also features two local minima, at approximately $\kappa_{2V} \approx 0.5$ and $\kappa_{2V} \approx 1.7$.

Figure 7.28 shows the 68 % and 95 % contours from the simultaneous fit of the two coupling modifiers, $\kappa_\lambda$ and $\kappa_{2V}$. The best fit point ($\{\hat{\kappa}_\lambda, \hat{\kappa}_{VV}\} \in \{0.7, 0.4\}$) is compatible with the SM prediction within $1\,\sigma$. The expected and observed two-dimensional CI are similar.

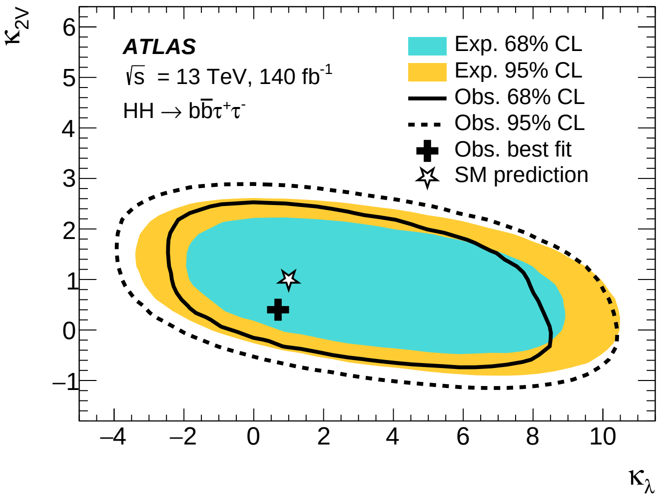


Figure 7.28: Likelihood contours at 68% (solid line) and 95% (dashed line) CL in the $\{\kappa_\lambda, \kappa_{2V}\}$ plane. All other coupling modifiers are fixed to the SM prediction. The expected contours are shown by the yellow and turquoise coloured shaded regions. The SM prediction is indicated by the star and the best-fit value is denoted by the black cross. Published in Reference [12].

## 7.14 VZ verification measurement

This section describes research and development towards a verification measurement of the $HH \to b\bar{b}\tau^+\tau^-$ analysis using $ZH$ and $ZZ$ production which can lead to the same final state with two oppositely charge $\tau$-leptons and $b$-jets. Both processes have been previously observed [223, 224] and the measured branching fractions of the $Z$ and Higgs bosons to $b\bar{b}$ and $\tau\bar{\tau}$ final states are consistent with the Standard Model (SM) [225]. Since the $Z$ boson is only slightly lighter than the Higgs boson, resulting in similar decay kinematics, these processes can be re-measured and used as "standard candles" for the $HH \to b\bar{b}\tau^+\tau^-$ analysis.

| **Process** | $\boldsymbol{\sigma_{pp\to X}}$ |
|---|---|
| $HH$ | 33 fb |
| $ZH$ | 884 fb |
| $ZZ$ | 15600 fb |

(a)

| **Process** | $\boldsymbol{\mathcal{B}}$ |
|---|---|
| $H \to b\bar{b}$ | 58 % |
| $Z \to b\bar{b}$ | 15 % |
| $H \to \tau^+\tau^-$ | 6 % |
| $Z \to \tau^+\tau^-$ | 3 % |

(b)

| **Process** | **Yield ratio** |
|---|---|
| $ZH/HH \to b\bar{b}\tau\tau$ | $\approx 10$ |
| $ZZ/HH \to b\bar{b}\tau\tau$ | $\approx 60$ |

(c)

Table 7.15: (a) Cross-section predictions for 13 TeV [225], (b) branching fractions [225], and (c) the corresponding expected yield ratio in the $b\bar{b}\tau^+\tau^-$ final state.

| | $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ | $b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$ | **Combined** |
|---|---|---|---|
| $HH \to ...$ | 4.00 % | 4.97 % | 4.51 % |
| $ZH \to ...$ | 3.11 % | 5.29 % | 4.25 % |
| $ZZ \to ...$ | 0.34 % | 0.98 % | 0.67 % |

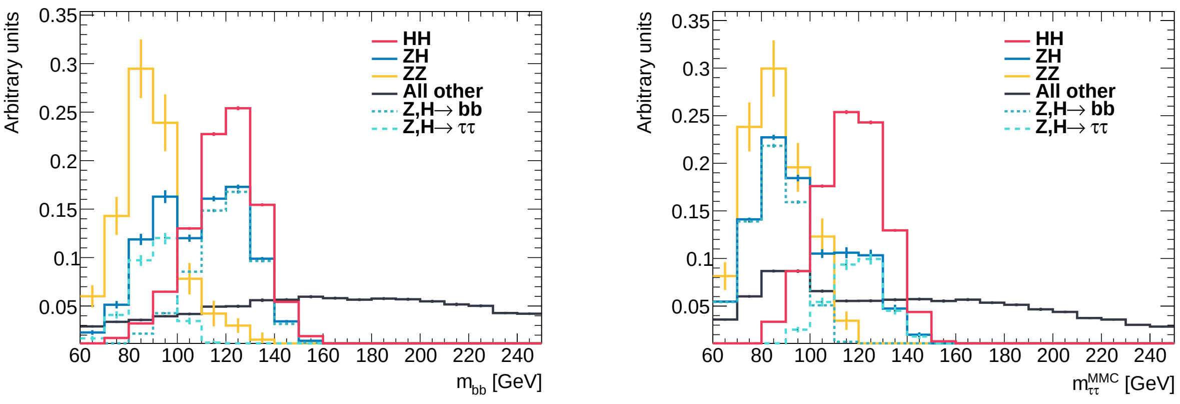
Table 7.16: Acceptance times efficiency ($\mathcal{A} \times \epsilon$) for the $HH$, $ZH$, and $ZZ$ processes to enter the $HH \to b\bar{b}\tau^+\tau^-$ signal regions. The total event yield used in the denominator includes all events produced in the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ final states, and their combination.

As listed in Table 7.15a, the production cross-section of $ZH$ is more than 25 times larger than of $HH$, while the $ZZ$ cross-section is nearly 500 times larger. However, the expected yields of $ZH$ and $ZZ$ production in the $b\bar{b}\tau^+\tau^-$ final state, shown in Table 7.15c, are reduced relative to $HH$ due to the smaller $Z \to \tau^+\tau^-$ and $Z \to b\bar{b}$ branching fractions that are listed in Table 7.15b. While about 7 % of $HH$ events decay into the $b\bar{b}\tau^+\tau^-$ final state, the branching fraction for $ZH$ is about two thirds smaller (2.6 %). Nevertheless, due to the much larger $ZH$ cross-section, roughly ten times more $b\bar{b}\tau^+\tau^-$ events are produced from $ZH$ than from $HH$. In the case of $ZZ$, the branching fraction to $b\bar{b}\tau^+\tau^-$ is even smaller ($< 1$ %), yet the large $ZZ$ cross-section results in approximately 60 times more $b\bar{b}\tau^+\tau^-$ events compared to $HH$.

The actual event yield in the signal regions is significantly affected by the trigger selection. In particular, $Z$ boson decays are slightly softer, and the associated $b$- and $\tau$-candidates more often fail, for example, the hadronic trigger thresholds compared to decays from Higgs bosons. This is especially evident in the significantly lower acceptance times efficiency for $ZZ$ production. The $\mathcal{A} \times \epsilon$ for $ZH$ is also slightly lower than for $HH$, with a small decrease in the $\tau_{\text{had}}\tau_{\text{had}}$ channel but an increase in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. The acceptance times efficiency for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and their combination is summarised in Table 7.16.

The following summarises research and development work towards a measurement of the $ZH$ process in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, performed in a blinded setup. The $ZZ$ measurement is highly susceptible to statistical fluctuations due to the limited size of the available $ZZ$ sample, and is therefore omitted from this study.

The SR BDTs are re-trained to separate $ZH$ as the new signal against the sum of backgrounds,



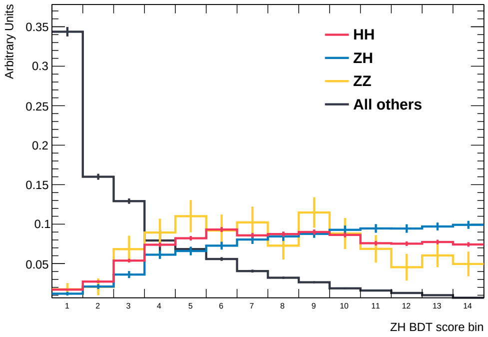
(a) $m_{bb}$ (b) $m_{\tau\tau}^{\mathrm{MMC}}$

Figure 7.29: Unstacked distribution of (a) $m_{bb}$, and (b) $m_{\tau\tau}^{\mathrm{MMC}}$ for $HH$, $ZH$, and the sum of all other processes in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ high-$m_{HH}$ ggF signal region. The $ZH$ contribution is additionally split into $ZH, H \to bb$ and $ZH, H \to \tau\tau$. The uncertainties on the bin predictions reflect the limited size of the simulated samples.

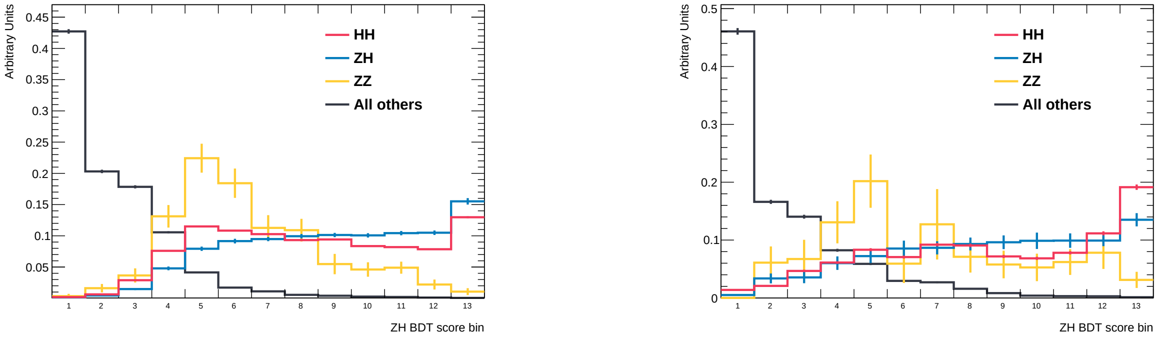
treating the ggF and VBF $HH$ samples as part of the background. Similar to the $HH$ BDTs, the $ZH$ BDTs mostly rely on variables that are sensitive to the mass of the decaying particle, such as $m_{bb}$, $m_{\tau\tau}^{\mathrm{MMC}}$, $\Delta R_{\tau\tau}$ and $\Delta R_{bb}$. Figure 7.29 shows the $m_{bb}$ and $m_{\tau\tau}^{\mathrm{MMC}}$ distributions for $HH$, $ZH$, $ZZ$ and all other backgrounds in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ high-$m_{HH}$ ggF signal region.

The $ZH$ BDTs are trained with the same hyper-parameters as the $HH$ signal BDTs. The small impact of this choice on the performance has been verified by comparing the expected significance on the test and validation dataset while varying the two most significant hyper-parameters, namely the number of trees and the maximal depth of an individual tree.

Figure 7.30 shows the $ZH$ BDT performance for the $ZH$, $HH$ and $ZZ$ processes, as well as for all other backgrounds in the three signal regions in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The same binning algorithm as used for the $HH$ BDT (see Section 7.12.2) is applied here. The BDTs perform well in rejecting the majority of background processes. The $ZZ$ processes are rejected reasonably well, although the lower rejection performance likely stems from its similarity to $ZH$ in terms of $m_{bb}$, $m_{\tau\tau}^{\mathrm{MMC}}$, $\Delta R_{\tau\tau}$ and $\Delta R_{bb}$. The $HH$ process is not rejected effectively either. This is partly likely due to similar kinematic properties to $ZH$, and partly because of its very small $HH$ event yield. Since $HH$ contributes only a tiny fraction of the total background, and since the training



(a) low-$m_{HH}$ ggF SR



(b) high-$m_{HH}$ ggF SR

(c) VBF SR

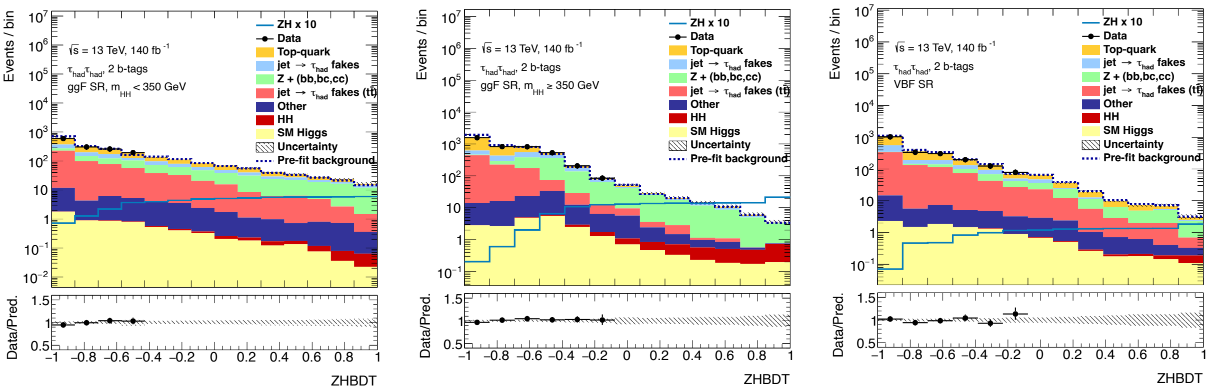
Figure 7.30: $ZH$ BDT distribution in the three $\tau_{\text{had}}\tau_{\text{had}}$ signal regions.

reweighs the backgrounds according to their physical event yields, there is little incentive during the training to improve separation against $HH$. Notably, the inverse relation is not true as the nominal BDTs reject the $ZH$ backgrounds reasonably well as $ZH$ contributes about ten times more to the total background in the signal regions.

The fit of the $ZH$ BDT output score, including the uncertainties from the limited size of the MC dataset, sets an expected limit at the 95 % CL on the $ZH$ signal strength of $\mu_{ZH} < 1.41$. This is about a factor two better than the limit obtained from fitting the $HH$ BDT with $\mu_{HH}$ as the POI (see Table 7.17). The relative small improvement in comparison to the yield differences can be attributed to the lower acceptance times efficiency of the $HH$ trigger and $HH \to b\bar{b}\tau^+\tau^-$ selection for $ZH$ events. The blinded post-fit distribution of the output scores of the $ZH$ BDTs are shown in Figure 7.31. The partly unblinded data points agree well the predictions. $Z$+HF production is the dominant background in the two most sensitive bins in all three SRs.

| **BDT** | **POI** | **Exp. limit at 95 % CL** |
|---|---|---|
| $HH$ vs. rest | $\mu_{HH}$ | $2.93\ ^{+1.15}_{-0.82} \times$ SM |
| $ZH$ vs. rest | $\mu_{ZH}$ | $1.41\ ^{+0.55}_{-0.39} \times$ SM |

Table 7.17: Expected limits on $\mu_{HH}$ and $\mu_{ZH}$ set by fits in the $\tau_{\text{had}}\tau_{\text{had}}$ channel including only the uncertainties related to the limited size of the MC samples.



(a) low-$m_{HH}$ ggF SR (b) high-$m_{HH}$ ggF SR (c) VBF SR

Figure 7.31: Blinded BDT post-fit distributions in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The $ZH$ signal is overlaid, scaled by a factor of 10.

A fit including all systematic uncertainties would require re-deriving all shape uncertainties as a function of the $ZH$ BDT output score. As a first estimate, a fit was performed including all uncertainties except the shape uncertainties of the $ZH$ BDT output. This fits sets a limit of $\mu_{ZH} < 1.60\ ^{+0.63}_{-0.45}$ times the SM expectation.

The presented strategy uses the same data to extract the $HH$ and the $ZH$ signal strengths. This introduces non-trivial correlations between the two POIs, making it impossible to measure them simultaneously in the current setup. The next step towards a simultaneous extraction of the $HH$, $ZH$ (and $ZZ$) signals requires the definition of orthogonal signal regions that suppress most of the correlations between these signals, though background-related nuisance parameters would remain.

A good starting point would be the definition of cut-based signal regions in the $m_{bb}$ - $m_{\tau\tau}^{\text{MMC}}$ mass plane, which exploit the distinctive peaks shown in Figure 7.29. Alternatively, signal regions could also be defined using multivariate analysis techniques such as BDTs or neural networks.

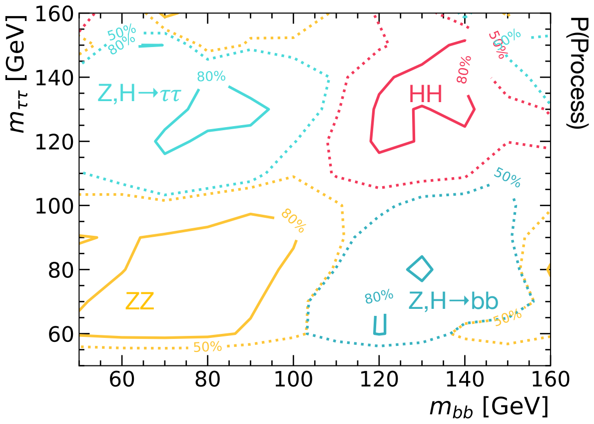


Figure 7.32: Probability contours for $HH$, $Z,H \to \tau\tau$, $Z,H \to bb$ and $ZZ$ in the $m_{bb}$ - $m_{\tau\tau}^{\text{MMC}}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

Figure 7.32 show the 50 % and 80 % probability contours for an event to originate from a given process in the $m_{bb}$ - $m_{\tau\tau}^{\text{MMC}}$ plane as defined as

$$\text{P(process } i) = \frac{N_i}{N_{HH} + N_{ZH} + N_{ZZ}} \,, \tag{7.35}$$

which approximately represent the output of an ideally trained classifier without backgrounds.

## 7.15 Conclusion and outlook

This chapter presents the search for non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state. The search re-analyses the Run 2 dataset which was recorded with the ATLAS detector from 2015 to 2018, and improves the experimental sensitivity with respect to the previous iteration. With an observed limit of $\mu_{HH} < 5.9$ SM at the 95 % CL in the absence of Higgs boson pair production, the analysis is among the three most sensitive searches to constrain $HH$ production and significantly contributes to the ATLAS Run 2 combination (see Section 3.4). The observed limit is weaker than the expected limit, due to a statistical fluctuation in the SLT channel. The expected limit of $\mu_{HH} < 3.3$ SM at the 95 % CL is the most stringent expected $HH$ upper limit with the Run 2 dataset. Compared to the previous search in this final state, the expected limit improves by about 15 %, equivalent to an increase of approximately 60 fb$^{-1}$ of data in this statistically limited analysis.

For the first time, ggF and VBF $HH$ production signal strengths are measured independently thanks to the introduction of a dedicated VBF $HH$ category. Furthermore, the search improves the sensitivity for the coupling modifiers $\kappa_\lambda$ and $\kappa_{2V}$, setting the observed 95 % confidence intervals of $\kappa_\lambda \in [-3.1, 9.0]$ and $\kappa_{2V} \in [-0.5, 2.7]$. The expected intervals $\kappa_\lambda \in [-2.5, 9.3]$ and $\kappa_{2V} \in [-0.2, 2.4]$ represent improvements of 11 % and 19 %, respectively, compared to the previous iteration.

The Run 2 search is predominantly limited by statistical uncertainties. After a three year long shutdown, the LHC resumed operation in 2022, colliding protons at a centre-of-mass energy of 13.6 TeV which increases the SM $HH$ production cross-section by about 11 % compared to the value at 13 TeV [41]. By the end of Run 3 in summer 2026, the recorded dataset is expected to be three times larger than that of Run 2, which will reduce the statistical uncertainty. Future analyses will further profit from improvements of trigger and reconstruction algorithms.

After Run 3, the LHC and the ATLAS detector will undergo another major upgrade in preparation of the High-Luminosity LHC. The HL-LHC is expected to to deliver an additional tenfold increase in the dataset, which will require a substantial upgrade of the ATLAS electronics and sub-detectors, notably the tracker system. An extrapolation of the expected sensitivity of the $HH \to b\bar{b}\tau^+\tau^-$ analysis to these conditions is presented in the next Chapter.

# 8 Projection of the Run 2 result to the High Luminosity LHC period

The main limitation of the Run 2 $HH$ searches is the limited dataset size given the small $HH$ cross-section. Therefore, it is expected that current measurements of the $HH$ signal strength and $\kappa_\lambda$ will become significantly more precise once the HL-LHC dataset [104–106] is accessible. The upgrade plans foresee an increase of the peak instantaneous luminosity by a factor of five, and an increase of the collision energy to 14 TeV resulting in an order of magnitude increase in the delivered proton-proton collision data over the course of a decade [107]. The higher instantaneous luminosity requires an upgrade of the ATLAS detector as described in Section 4.3.

This chapter presents the updated projections for various integrated luminosities ranging from 1000 to 3000 $\text{fb}^{-1}$ collected at a centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$, for which the author served as the principal analyst and studied the impact of advancements in theory, event generation, and detector performance. The results presented in this chapter are expected to be give a valid estimate also in the case that the HL-LHC will operate at 13.6 TeV. The $HH$ cross-section increase of +18 % for 14 TeV would be reduced to +11 % for 13.6 TeV. However, the results show limited sensitivity to the exact value of the cross-section, and the overall impact is estimated to be at the level of a few percent. The study has been published in Reference [13].

For the extrapolation to HL-LHC conditions, the analysis strategy and the underlying algorithm performances remain unchanged. Thus, the presented estimates are conservative. Notably, past pure-luminosity extrapolations of previous $b\bar{b}\tau^+\tau^-$ results have been outperformed, as evident in Figure 8.1.

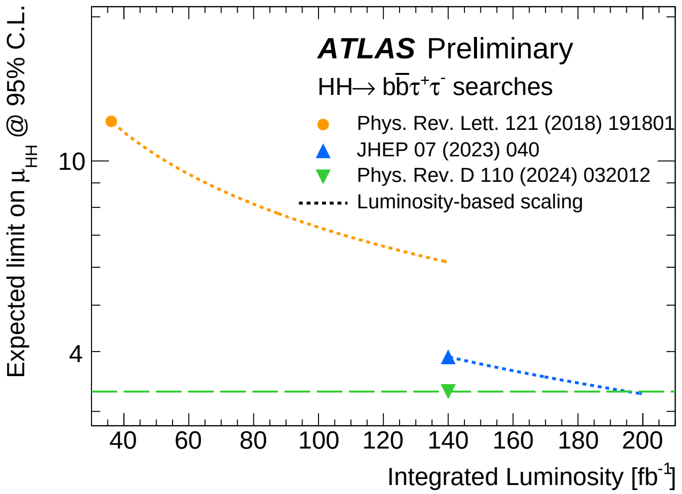


Figure 8.1: The expected limit on $\mu_{HH}$ at the 95 % confidence level for the three Run 2 $HH \to b\bar{b}\tau^+\tau^-$ results [12, 144, 226]. The first two results are extrapolated to higher luminosity values $L$ in dotted lines assuming all uncertainties scale as $\sqrt{L'/L}$, where $L'$ is the measurement luminosity. The green dashed horizontal line is the current best limit on $\mu_{HH}$, measured by [12]. A comparison between the dotted orange line at $L = 140\,\text{fb}^{-1}$ and the blue and green points highlights the gains from improved analysis techniques. The intersection of the horizontal dashed green and the blue dotted line quantifies the improvement from reanalysing the Run 2 dataset in terms of effective increased luminosity. Figure from Reference [13].

## 8.1 Extrapolation procedure

The Run 2 $HH \to b\bar{b}\tau^+\tau^-$ analysis is extrapolated by scaling the yields in each bin of the final signal versus background BDTs of the signal and background processes in each analysis region. This choice leads to conservative estimates, as a larger future dataset will allow for optimized binning strategies with improved signal-to-background discrimination, thereby enhancing sensitivity. In general, the extrapolation procedure assumes that the performance of the upgraded Phase-II detector is at least as good as that of the Run 2 detector. It does not consider the impact of new sub-detectors that for example increase the acceptance of the tracker in $\eta$. Recent detector performance studies [227, 228] support this hypothesis, despite the significantly larger pileup of the HL-LHC.

The effect of an increased integrated luminosity $L'$ with respect to the Run 2 luminosity of $L = 140\,\text{fb}^{-1}$ is studied with a common scale factor of $L'/L$ that is applied to all signal and

| | **Process** | **Scale factor** |
|---|---|---|
| **Signals** | ggF $HH$ | 1.18 |
| | VBF $HH$ | 1.19 |
| **Backgrounds** | ggF $H$ | 1.13 |
| | VBF $H$ | 1.13 |
| | $WH$ | 1.10 |
| | $ZH$ | 1.12 |
| | $ttH$ | 1.21 |
| | Others | 1.18 |

Table 8.1: Scale factors applied to the $HH$ signal and background processes to account for the change in cross-sections due to the increase of centre-of-mass energy from $\sqrt{s} = 13\,\text{TeV}$ to $\sqrt{s} = 14\,\text{TeV}$ at the HL-LHC. Published in Reference [13].

background processes. Different scenarios between 1000 and 3000 $\text{fb}^{-1}$ are studied. Increasing the centre-of-mass energy to $\sqrt{s} = 14\,\text{TeV}$ enhances the inclusive production cross-section of most processes and is accounted for with process-dependent scale factors that are summarised in Table 8.1. The $H$ and $HH$ cross-sections follow the recommendations from the LHC Higgs Working Group [229]. For all other processes, a generic scale factor of 1.18 is applied. This approach, consistent with previous HL-LHC Higgs projections [230], reflects the increased gluon cross-section at higher energy. The 1.18 scaling is correct for $t\bar{t}$ production, however it is slightly too large for the other two main backgrounds, QCD multi-jet production and $Z$+HF. While the results depend little on the exact scale factors, it should be noted that this presents a conservative approach. Since the Run 2 result measured a $Z$+HF normalisation factor that deviated significantly from unity ($1.34 \pm 0.08$), the process is scaled additionally by 1.3.

The extrapolation studies five different scenarios, each incorporating potential improvements over the Run 2 analysis. Each scenario modifies a specific set of systematic uncertainties inherited from Run 2:

**Run 2 systematic uncertainties** This is the most conservative scenario and retains all systematic uncertainties at their Run 2 values.

| **Source** | **Scale factor** |
|---|---|
| **Experimental uncertainties** | |
| Luminosity | 1.0 |
| Electrons and muons efficiency | 1.0 |
| $b$-jet $b$-tagging efficiency | 0.5 |
| $c$-jet $b$-tagging efficiency | 0.5 |
| Light-jet $b$-tagging efficiency | 1.0 |
| $\tau_{\text{had}}$ efficiency (statistical) | 0.0 |
| $\tau_{\text{had}}$ efficiency (systematic) | 1.0 |
| $\tau_{\text{had}}$ energy scale | 1.0 |
| Fake-$\tau_{\text{had}}$ estimation (statistical) | 0.0 |
| Fake-$\tau_{\text{had}}$ estimation (systematic) | 0.5 |
| Jet energy scale and resolution, $E_{\text{T}}^{\text{miss}}$ | 1.0 |
| **Theoretical uncertainties** | 0.5 |
| **MC statistical uncertainties** | 0.0 |

Table 8.2: Summary of scale factors for the relevant experimental and theoretical systematic uncertainties defining the baseline scenario. Published in Reference [13].

**Theoretical uncertainties halved** The impact of future improvements in the theoretical modelling is studied by halving all theoretical uncertainties associated with the signal and background processes. All other uncertainties are kept at their Run 2 values. Halving the uncertainty on the normalisation of single Higgs boson production in the ggF, VBF and $WH$ production modes in association with heavy-flavour quark has the largest impact in this scenario. Recent studies consider a factor two reduction on this uncertainty (currently a prior of 100 %) to be realistic [231].

**MC luminosity scaled** For the Run 2 analysis, a significant portion of the systematic uncertainties originates from statistical limitations, specifically the limited size of the available MC simulations for the background processes, and the statistical uncertainties of the data-driven jet → $\tau_{\text{had}}$ fake estimate. As data-driven uncertainties scale with $\sqrt{L/L'}$, and MC simulation capabilities will improve in the future [232], the MC statisitical uncertainties are scaled down accordingly by $\sqrt{L/L'}$. All other uncertainties in this scenario are kept at their Run 2 values.

**Baseline scenario** This scenario follows the current baseline recommendations for HL-LHC projection studies [233]. It accounts for anticipated improvements in analysis techniques enabled by the larger dataset, as well as the expected reduction in statistical components. All theoretical uncertainties are halved. Statistical uncertainties on the background estimates from simulation and data-driven methods are assumed to be negligible. Uncertainties related to jets, electrons, muons and $E_{\mathrm{T}}^{\mathrm{miss}}$ are kept at their Run 2 values. Flavour-tagging uncertainties for $b$- and $c$-jets are halved, light-tagging uncertainties are kept at their Run 2 values. The statistical components of the $\tau_{\mathrm{had}}$ reconstruction and identification efficiency as well as the energy scale are removed from the fit model, their systematic components are kept the same as for Run 2. The same applies to uncertainties related to the estimation of fake-$\tau_{\text{had-vis}}$ candidates. The scale factors applied to the systematic uncertainties are summarised in Table 8.2.

**Baseline with MC luminosity scaled** This scenario combines the aforementioned baseline scenario and the MC luminosity scaled scenario.

**No systematic uncertainties** This scenario removes all systematic uncertainties and is thus the most optimistic scenario.

## 8.2 Results

### 8.2.1 *HH* production signal strength

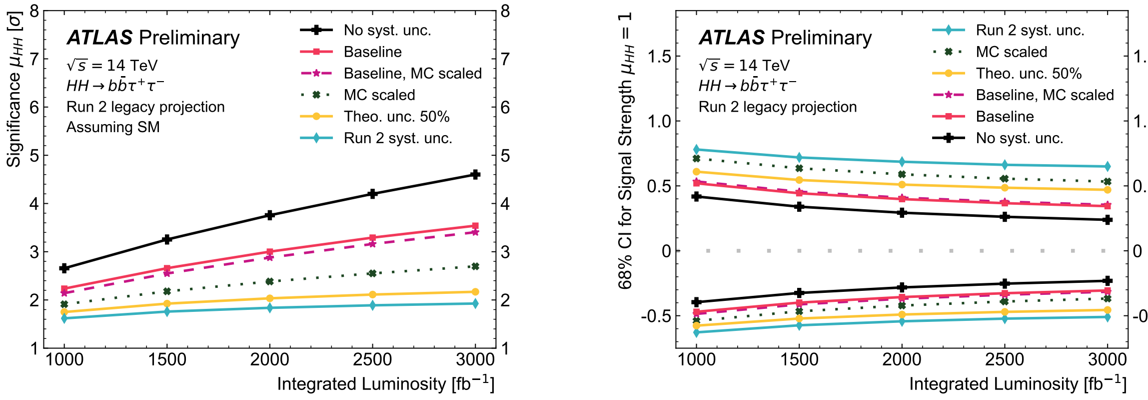
The significance and the uncertainty on the signal strength for the combined results are shown for a luminosity range from 1000 to 3000 fb$^{-1}$ in Figure 8.2. In addition, Table 8.3 shows the expected $HH$ signal significance assuming SM-like Higgs pair production ($\kappa_{\lambda} = 1$) for the different extrapolation scenarios for 3000 fb$^{-1}$ at 14 TeV. The significance for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel as well as the combined result and the relative precision of the combined results is given, too. In all scenarios, the significance is driven by the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel as is expected given the Run 2 results.

| **Uncertainty scenario** | **Significance [$\sigma$]** | | | **$\Delta\mu_{HH}$[%]** |
|---|---|---|---|---|
| | $\tau_{\text{lep}}\tau_{\text{had}}$ | $\tau_{\text{had}}\tau_{\text{had}}$ | Comb. | Comb. |
| No systematic uncertainties | 2.3 | 4.0 | 4.6 | −23 / +24 |
| Baseline | 1.8 | 3.1 | 3.5 | −31 / +35 |
| Baseline with MC luminosity scaled | 1.7 | 3.0 | 3.4 | −32 / +35 |
| MC luminosity scaled | 1.6 | 2.4 | 2.7 | −37 / +53 |
| Theoretical uncertainties halved | 1.0 | 1.9 | 2.2 | −46 / +47 |
| Run 2 systematic uncertainties | 0.9 | 1.8 | 1.9 | −51 / +65 |

Table 8.3: Expected significance and signal strength precision $\Delta\mu_{HH}$ of the SM $HH$ signals at 3000 fb$^{-1}$ and $\sqrt{s}$ =14 TeV for the extrapolation scenarios described in the text. The significance is given separately for the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channel, as well as the combined fit (Comb.). Modified after Reference [13].

In the baseline scenario, evidence (3.0 $\sigma$) for SM $HH$ production is expected to achieved after 2000 fb$^{-1}$. For 3000 fb$^{-1}$, the signal significance is projected to increase to 3.5 $\sigma$ with a signal strength precision of $\Delta\mu_{HH} = {}^{+35}_{-31}\,\%$. In the scenario without systematic uncertainties, a significance of 4.6 $\sigma$ is anticipated after 3000 fb$^{-1}$, with $\Delta\mu_{HH} = {}^{+24}_{-23}\,\%$.

A breakdown of the contributions to the uncertainty on the signal strength $\Delta\mu_{HH}$ is given in Table 8.4 for the baseline and the Run 2 systematic uncertainty scenario for an integrated luminosity of 3000 fb$^{-1}$. Sources are grouped into statistical and systematic uncertainties. The



(a) Significance

(b) Signal strength precision

Figure 8.2: Expected significance (a) and the corresponding expected uncertainty on the signal strength (b) as a function of various integrated luminosities ranging from 1000 fb$^{-1}$ to 3000 fb$^{-1}$ at $\sqrt{s}$ = 14 TeV for the different extrapolation scenarios. Both the significance and uncertainty on $\mu_{HH}$ assume SM-like $HH$ production ($\mu_{HH} = 1$). Published in Reference [13].

| **Source of uncertainty** | **Baseline** | | **Run 2 Syst.** | |
|---|---|---|---|---|
| Total | +0.35 | −0.31 | +0.65 | −0.51 |
| Statistical | +0.24 | −0.23 | +0.24 | −0.23 |
| $\hookrightarrow$ Data stat. only | +0.24 | −0.23 | +0.24 | −0.23 |
| $\hookrightarrow$ Floating normalisations | +0.02 | −0.02 | +0.04 | −0.02 |
| Systematic | +0.25 | −0.20 | +0.61 | −0.46 |
| **Experimental uncertainties** | | | | |
| Electrons and muons | < 0.01 | | < 0.01 | |
| $\tau$-leptons | +0.03 | −0.03 | +0.06 | −0.05 |
| Jets | +0.06 | −0.06 | +0.06 | −0.07 |
| $b$-tagging | +0.02 | −0.02 | +0.04 | −0.03 |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | +0.03 | −0.02 | +0.04 | −0.02 |
| pileup | +0.01 | −0.01 | +0.01 | −0.01 |
| Luminosity | +0.02 | −0.01 | +0.02 | −0.01 |
| **Theoretical and modelling uncertainties** | | | | |
| Signal | +0.12 | −0.05 | +0.39 | −0.07 |
| Backgrounds | +0.19 | −0.17 | +0.37 | −0.30 |
| $\hookrightarrow$ Single Higgs boson | +0.17 | −0.15 | +0.34 | −0.27 |
| $\hookrightarrow$ $Z$ + jets | +0.06 | −0.05 | +0.10 | −0.09 |
| $\hookrightarrow$ $W$ + jets | < 0.01 | | < 0.01 | |
| $\hookrightarrow$ $t\bar{t}$ | +0.02 | −0.02 | +0.03 | −0.02 |
| $\hookrightarrow$ Single top quark | +0.01 | −0.01 | +0.03 | −0.04 |
| $\hookrightarrow$ Diboson | < 0.01 | | < 0.01 | |
| $\hookrightarrow$ Jet $\rightarrow \tau_{\mathrm{had}}$ fakes | +0.05 | −0.05 | +0.09 | −0.08 |
| **MC statistical uncertainties** | < 0.01 | | +0.38 | −0.36 |

Table 8.4: Breakdown of contributions to the uncertainty on the $HH$ signal strength $\Delta\mu_{HH}$ assuming $3000\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 14\,\mathrm{TeV}$, shown for both the baseline extrapolation scenario and the scenario with the current Run 2 systematic uncertainties. The sum in quadrature of different sources may differ from the total due to correlations. Published in Reference [13].

statistical uncertainties are divided into those originating from the limitation of the dataset, those related to the free floating normalisation of the $t\bar{t}$ and $Z$+HF backgrounds. The systematic contributions are divided into subgroups: experimental, theoretical and modelling, as well as those originating from the limited size of the Monte Carlo simulations. Experimental uncertainties are categorized by the various physics objects that they affect, such as electrons, muons, $\tau$-leptons and jets, as well as $b$-tagging and the measurement uncertainty of physical quantities such as missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) or luminosity. Modelling uncertainties are subdivided into contributions from the modelling of the signal and the different backgrounds.

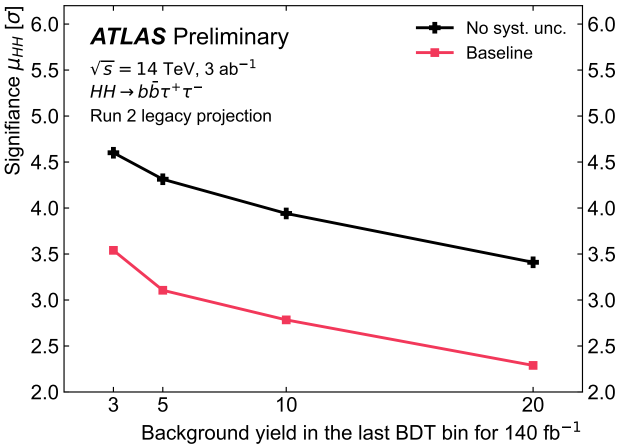
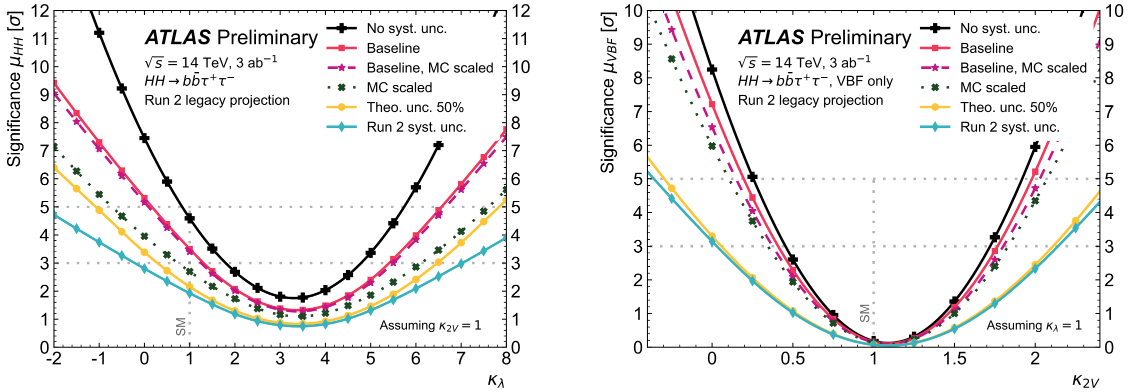


Figure 8.3: Impact of the requirement on the minimum number of expected background events per bin in the BDT score re-binning algorithm on the extrapolated $HH$ signal significance for $3000\,\text{fb}^{-1}$ at $\sqrt{s} = 14\,\text{TeV}$. Published in Reference [13].

The systematic impact of each group of uncertainties is determined by evaluating the quadratic change in the upper and lower uncertainties on the signal strength. All uncertainties of one group is kept constant, all other uncertainties can vary. The quadrature sum of the different contributions is not always equal to the total uncertainty. This is because some uncertainties are correlated, for example the flavour tagging efficiencies and the normalisation of backgrounds. In the baseline scenario, the uncertainty on the signal is evenly split between statistical and systematic uncertainties, while uncertainties from MC statistics are negligible. The main systematic limitation is the modelling of the signal and of the single Higgs production in association with heavy-flavour quarks. Cross-section uncertainties remain the predominant uncertainty affecting the signal modelling. The Run 2 systematic uncertainties scenario is systematically limited, primarily by the size of the MC simulation samples, and the modelling of the signal and backgrounds. For both scenarios, experimental uncertainties related to the reconstruction and identification of physics objects are small.

As mentioned before, the extrapolation procedure uses the same binning as in the Run 2 analysis (see Subsection 7.12.2). Naturally, the increased dataset of the HL-LHC period will allow for a more sensitive binning. While it is not possible to study a more aggressive binning, the impact



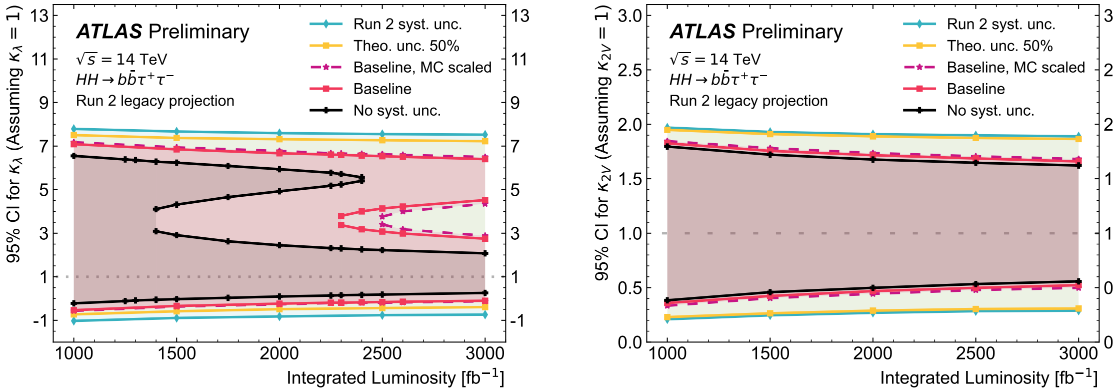
(a) Varying $\kappa_\lambda$ ($\kappa_{2V} = 1$)

(b) Varying $\kappa_{2V}$ ($\kappa_\lambda = 1$)

Figure 8.4: Expected $HH$ (a) and VBF $HH$ (b) significance assuming different values of $\kappa_\lambda$ and $\kappa_{2V}$ for 3000 fb$^{-1}$ and $\sqrt{s}$ =14 TeV for the different extrapolation scenarios. The dashed horizontal lines indicate the thresholds to claim "evidence" ($3\sigma$) and "observation" ($5\sigma$). The dashed vertical line indicates the SM hypothesis with $\kappa = 1$. Published in Reference [13].

of the binning on the signal significance is studied by adopting a more conservative binning strategy. Figure 8.3 shows the impact of the requirement on the minimum number of expected background events per bin in the BDT score re-binning algorithm on the extrapolated $HH$ signal significance. Starting from the currently used threshold of three expected background events per bin at 140 fb$^{-1}$, the requirement is successively increased, which significantly decreases the signal significance.

The expected $HH$ signal significance is highly sensitive to the cross-section, which strongly depends on the value of $\kappa_\lambda$. Therefore, the significance is estimated for the case of non-SM-like Higgs boson couplings ($\kappa_\lambda \neq 1$) for the different extrapolation scenarios at 3000 fb$^{-1}$, as shown in Figure 8.4a. The minimum significance is reached in all extrapolation scenarios around $\kappa_\lambda \approx 3.5$, due to a combination of a reduced cross-section and a soft $m_{HH}$ spectrum, which leads to a challenging signal-to-background ratio. This shift in the $m_{HH}$ spectrum leads to varying sensitivities across the different scenarios. In the baseline scenario, evidence ($3\,\sigma$) for $HH$ production is expected to be reached for $\kappa_\lambda < 1.3$ or $\kappa_\lambda > 5.4$, while observation ($5\,\sigma$) of $HH$ production is expected for $\kappa_\lambda < 0.2$ or $\kappa_\lambda > 6.6$. A similar plot is shown in Figure 8.4b for the expected signal significance for VBF $HH$ production as a function of $\kappa_{2V}$. In the baseline scenario,



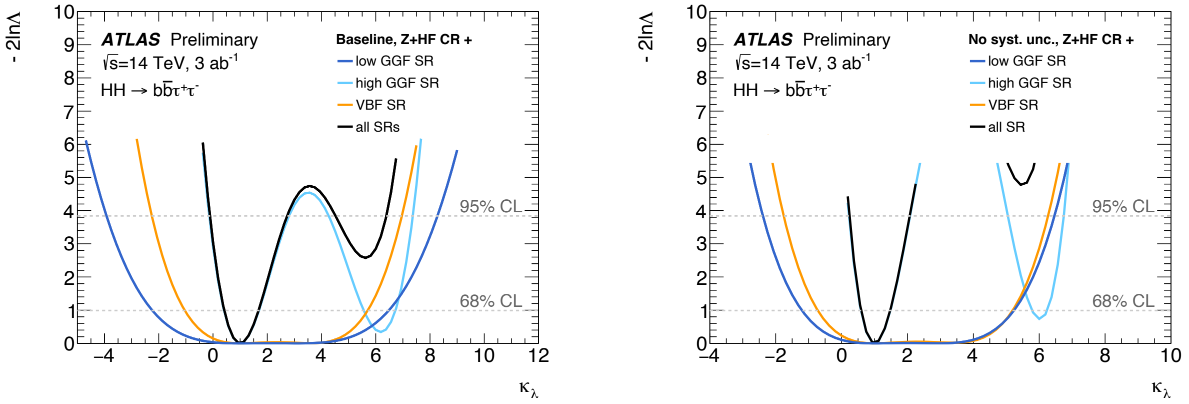
(a) Varying $\kappa_\lambda$ ($\kappa_{2V} = 1$) (b) Varying $\kappa_{2V}$ ($\kappa_\lambda = 1$)

Figure 8.5: Expected range of the 95 % confidence interval on $\kappa_\lambda$ (a) and $\kappa_{2V}$ (b) as a function of various integrated luminosities from 1000 fb$^{-1}$ at $\sqrt{s}$ = 14 TeV for the extrapolation scenarios described in the text. The confidence intervals are obtained from the values of the negative log-likelihood ratio as a function of $\kappa_\lambda$ with the Asimov dataset being created from the SM hypothesis $\kappa = 1$. Published in Reference [13].

evidence for VBF $HH$ production is expected for $\kappa_{2V} < 0.4$ or $\kappa_{2V} > 1.8$, while observation is expected for $\kappa_{2V} < 0.2$ or $\kappa_{2V} > 2.0$. Notably, these ranges are already excluded at the 95 % CL by a search for boosted Higgs boson pair production via vector-boson fusion in the $b\bar{b}b\bar{b}$ final state using LHC Run 2 data [57]. The latest combination of Run 2 $HH$ searches with the ATLAS detector [234] sets the 95 % confidence interval for $\kappa_\lambda$ to be [-1.2, 7.2] and for $\kappa_{2V}$ to [0.6, 1.5].

### 8.2.2 Constraints on Higgs boson coupling modifiers

The 95 % expected confidence intervals on $\kappa_\lambda$ and $\kappa_{2V}$ assuming SM $HH$ production for the various extrapolation scenarios at integrated luminosities ranging from 1000 to 3000 fb$^{-1}$ at 14 TeV are shown in Figure 8.5. In the baseline scenario, for 3000 fb$^{-1}$, $\kappa_\lambda$ is expected to be constrained at the 95 % confidence level to the intervals $[-0.1, 2.7] \cup [4.5, 6.4]$. In the scenario without systematic uncertainties, the confidence interval narrows significantly to $[0.3, 2.1]$. This marks the first time that the double minimum at high $\kappa_\lambda$ is projected to be resolved with a significance greater than $2\,\sigma$ by a $HH \to b\bar{b}\tau^+\tau^-$ extrapolation. This improvement is primarily due to the splitting of the $m_{HH}$ spectrum and the introduction of a new VBF signal region in the legacy Run 2 analysis. For the baseline scenario, the second minimum is excluded with



(a) Baseline

(b) No systematic uncertainty

Figure 8.6: Expected values of the negative log-likelihood ratio as a function of $\kappa_\lambda$ for the baseline extrapolation scenario (a) and the scenario without systematic uncertainties (b) at an integrated luminosity of $3000\,\text{fb}^{-1}$. For each of the scenarios, the contribution of the low-$m_{HH}$ ggF SR, the high-$m_{HH}$ ggF SR and the VBF SR are shown individually as well. They are obtained from a fit to only that region and the $Z+$HF control region. Published in Reference [13].

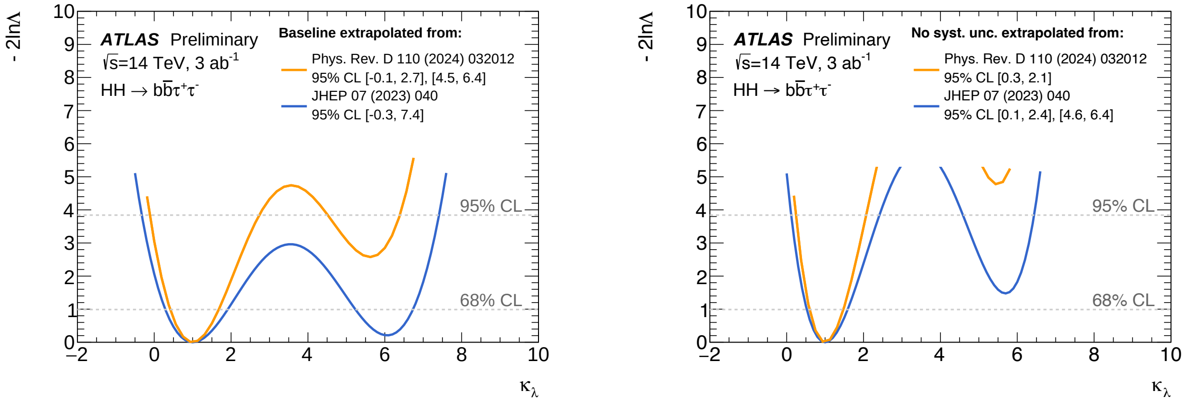
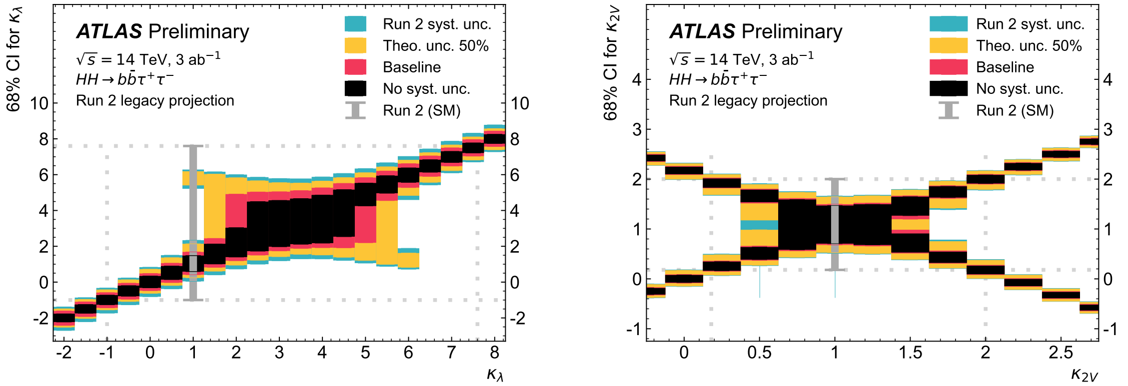


Figure 8.7: Expected values of the negative log-likelihood ratio as a function of $\kappa_\lambda$ for (a) the baseline extrapolation scenario and (b) the scenario without systematic uncertainties. The extrapolation based on search presented in this thesis is compared with an earlier Run 2 based HL-LHC extrapolation. The Asimov dataset for these results has been created using the SM hypothesis ($\kappa_\lambda = 1$). Published in Reference [13].

approximately $1.6\,\sigma$.

The different constraints given by the three signal regions are illustrated for the baseline and the scenario without systematic uncertainties in Figure 8.6. In both scenarios, the high-$m_{HH}$ signal region drives the lower $\kappa_\lambda$ constraint around $\kappa_\lambda \approx 0$, and together with the VBF signal region improves the upper limit on the $\kappa_\lambda$ range. The impact of the new SRs is further emphasized by



(a) Varying $\kappa_\lambda$ ($\kappa_{2V} = 1$) (b) Varying $\kappa_{2V}$ ($\kappa_\lambda = 1$)

Figure 8.8: Expected range of the 68 % confidence interval on (a) $\kappa_\lambda$ and (b) $\kappa_{2V}$ given the used $\kappa$ value of the Asimov dataset for an integrated luminosity of $3000\,\text{fb}^{-1}$ and $\sqrt{s} = 14\,\text{TeV}$. The separate colours correspond to the various extrapolation scenarios. The Run 2 result and CIs [12] are shown in grey as dashed grey lines. Published in Reference [13].

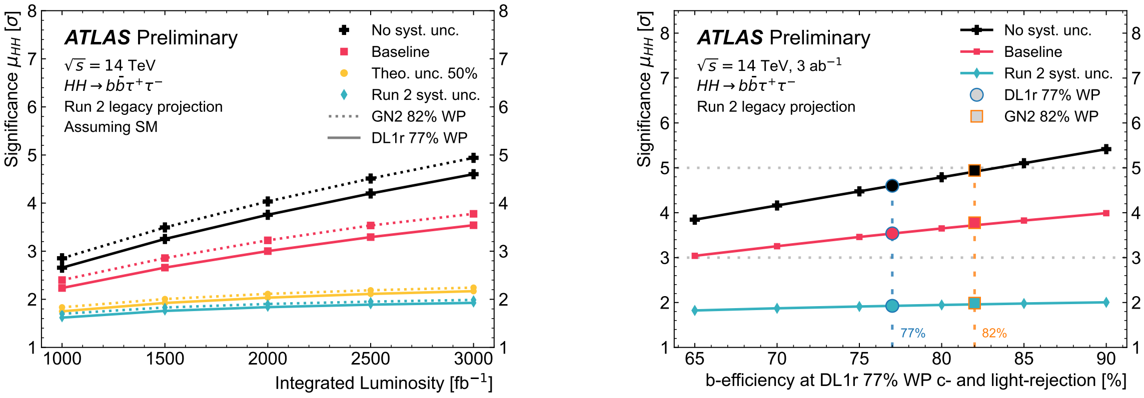
comparing the expected values of the negative log-likelihood ratio as a function of $\kappa_\lambda$ from the extrapolation in this report with those from the previous extrapolation [104]. This comparison, shown in Figure 8.7, is performed for the same two uncertainty scenarios. The 95 % confidence interval on the coupling modifier $\kappa_{2V}$ is $[0.5, 1.7]$ at $3000\,\text{fb}^{-1}$ in the baseline scenario. In the most optimistic scenario without systematic uncertainties, this is reduced to $[0.6, 1.6]$.

The strong impact of $\kappa_\lambda$ on the cross-section and the $m_{HH}$ spectrum also affects how precise a non-SM $\kappa_\lambda$ can be determined. The impact of non-SM $\kappa_\lambda$ values on the expected constraints is obtained by fitting Asimov datasets generated for various $\kappa_\lambda$ values. The expected constraints for $\kappa_\lambda$ as a function of $\kappa_\lambda$ are shown in Figure 8.8a for $3000\,\text{fb}^{-1}$ across the different extrapolation scenarios. For comparison, the current exclusion limits of the Run 2 $b\bar{b}\tau^+\tau^-$ analysis, assuming SM-like coupling ($\kappa_\lambda = 1$), are shown in grey. The figure illustrates a pronounced dependence of the expected constraining power on the true coupling strength of the Higgs boson. Notably, if the true coupling strength is near the point of maximum interference ($\kappa_\lambda \approx 3$), future $HH$ searches may yield larger-than-expected confidence intervals, despite the improved sensitivity - due to the suppression of the signal cross-section in that region. A corresponding plot for $\kappa_{2V}$ is presented in Figure 8.8b.

## 8.3 Improvements to the identification of $b$-jets and $\tau$-leptons

The extrapolations presented in the previous sections are likely conservative. They do not, for example, consider future improvements in the identification of hadronic signatures like $b$-jets or hadronically decaying $\tau$-leptons, which are vital ingredients to the sensitivity of the $HH \to b\bar{b}\tau\tau$ analysis.

The ATLAS collaboration developed a new $b$-jet identification algorithm ("$b$-tagger") based on a transformer architecture [235]. Performance studies of the upgraded ATLAS detector for the HL-LHC show that this new "GN2" tagger is expected to increase the $c$- and light-jet rejection by a factor of four at the same signal efficiency of 70 % [236]. To evaluate the potential impact of the GN2 tagger on extrapolations, its effect is emulated using the 82 % $b$-tagging efficiency working point. This configuration offers a similar light-jet rejection rate but achieves a 30 % improvement in $c$-jet rejection and 5 % higher signal efficiency compared to the 77 % DL1r working point used in the Run 2 analysis. For the simulated backgrounds, events are re-weighted based on the truth



(a) (b)

Figure 8.9: (a) The effect of transitioning from the DL1r $b$-tagger at a 77 % b-tagging efficiency working point (solid lines) to the GN2 tagger at an 82 % working point (dashed lines) on the $HH$ significance presented for different uncertainty extrapolation scenarios as a function of the integrated luminosity for various uncertainty extrapolation scenarios. (b) The impact of varying the $b$-tagging efficiency, while maintaining the current $c$-jet and light-jet rejection rates, at an integrated luminosity of 3000 fb$^{-1}$. The dashed horizontal lines mark the thresholds for claiming "evidence" (3 $\sigma$) and "observation" (5 $\sigma$). Published in Reference [13].

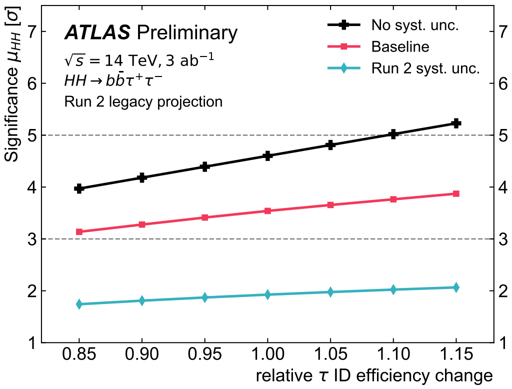


Figure 8.10: Impact of varying the hadronically decaying $\tau$-lepton identification efficiency on the expected $HH$ signal significance at the current background rejection rates for different uncertainty extrapolation scenarios at an integrated luminosity of $3000\,\text{fb}^{-1}$. The dashed horizontal lines indicate the thresholds to claim "evidence" ($3\,\sigma$) and "observation" ($5\,\sigma$). Published in Reference [13].

flavour content, i.e. the flavour label, of the two $H \to b\bar{b}$ jet candidates, taking into account the increased efficiency and modified rejection rates. The data-driven backgrounds (e.g. QCD multi-jet background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel) are conservatively assumed to be $100\,\%$ $b\bar{b}$-labelled events. In general, the fraction of $b\bar{b}$-labelled events exceeds $90\,\%$ for all major backgrounds.

The estimated improvement on the projected $HH$ significances is shown in Figure 8.9a as a function of the integrated luminosity for the different uncertainty extrapolation scenarios. The impact varies across the scenarios. In the baseline scenario, the significance improves by $\mathcal{O}(10\,\%)$. Future $b$-taggers might be able to increase the signal efficiency at the same background rejection that GN2 is expected to achieve today. The impact of an increasing or decreasing $b$-tagging efficiency at fixed $c$- and light-jet rejection on the $HH$ signal significance is shown in Figure 8.9b.

A similar impact is expected from an increased $\tau_{\text{had}}$ identification efficiency at a fixed background rejection is shown in Figure 8.10. For this scaling, a joint scale factor is applied to both 1-prong and 3-prong hadronic $\tau$-lepton decays.

## 8.4 Combination with other final states

As described in Section 3.4, the ATLAS Collaboration searched for $HH$ production in several final states, and published a statistical combination of the Run 2 projections in six final states [237], namely $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ [238], $b\bar{b}b\bar{b}$ [239, 240], multi-lepton [241] and $b\bar{b}\ell^+\ell^- + E_{\mathrm{T}}^{\mathrm{miss}}$ [237]. Following a similar approach, the expected significance of SM $HH$ production is extrapolated to $4.3\,\sigma$ in the baseline scenario, and $6.0\,\sigma$ in the scenario without systematic uncertainties. In both scenarios, the $b\bar{b}\tau^+\tau^-$ final state provides the largest contribution to the overall sensitivity.

The 68 % CI for $\kappa_\lambda$ is extrapolated to $\kappa_\lambda \in [0.58, 1.48]$ in the baseline scenario, and $\kappa_\lambda \in [0.71, 1.33]$ in the most optimistic scenario without systematic uncertainties. Similar to the Run 2 results discussed in Section 3.4, the $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ final states provide the strongest constraints on $\kappa_\lambda$. The degeneracy of the $\kappa_\lambda$ CI around $\kappa_\lambda \approx 5$, as discussed in Figures 8.5 and 8.6, is resolved already for 1000 fb$^{-1}$ when statistically combining all six projected analyses. Similarly, the degeneracy of the 68 % $\kappa_\lambda$ CIs when assuming alternative $\kappa_\lambda$ values is resolved by the combination. $\kappa_{2V}$ is expected to be constrained within $\kappa_{2V} \in [0.85, 1.18]$ in the baseline scenario at the 68 % CL, and $\kappa_{2V} \in [0.87, 1.16]$ in the scenario without systematics. As for Run 2, the sensitivity for $\kappa_{2V}$ is dominated by the $b\bar{b}b\bar{b}$ channel, especially the boosted analysis.

A further statistical combination of the combined ATLAS projection with projections by the CMS collaboration [242] extrapolates the combined expected statistical significance for $HH$ production to $7.2\sigma\,(6.0)\,\sigma$ in the baseline scenario for 3000 fb$^{-1}$ (2000 fb$^{-1}$), with a CMS standalone significance of $4.2\,(3.5)\,\sigma$. The combination with CMS is projected to achieve a sensitivity for $\kappa_\lambda$ of less than 30 % at the 68 % CL. These results demonstrate that, assuming continued improvements in analysis techniques, the observation of $HH$ production by a single experiment is feasible at 3000 fb$^{-1}$, but unlikely at 2000 fb$^{-1}$.

# 9 Conclusion and outlook

This thesis presents the search for $HH$ production in the $b\bar{b}\tau^+\tau^-$ final state with Run 2 dataset recorded by the ATLAS detector, as well as the extrapolation of the results to the HL-LHC period. Furthermore, it describes the Alarm Helper tool that improves the detector safety.

The re-analysis of the ATLAS Run 2 $pp$ collision dataset to search for Higgs boson pair ($HH$) production in the $b\bar{b}\tau^+\tau^-$ final state enhanced the experimental sensitivity to $HH$ production. This analysis ranks among the three most sensitive to constrain $HH$ production and sets an observed limit of $\mu_{HH} < 5.9$ SM at the 95 % CL in the absence of $HH$ production. Due to a statistical fluctuation in the SLT channel, the observed limit is weaker than the expected limit; however, the expected limit still improved by 15 % compared to the previous iteration. Since the analysis is predominantly statistically limited, this corresponds to an increase of approximately $60\,\text{fb}^{-1}$ of data.

The search also introduced a dedicated VBF $HH$ category, enabling the simultaneous measurement of ggF and VBF $HH$ production signal strengths. Moreover, the sensitivity for the coupling modifiers $\kappa_\lambda$ and $\kappa_{2V}$ is improved to $\kappa_\lambda \in [-3.1, 9.0]$ and $\kappa_{2V} \in [-0.5, 2.7]$ at the 95 % confidence level. With respect to the previous analysis in the same channel, the expected intervals of $\kappa_\lambda \in [-2.5, 9.3]$ and $\kappa_{2V} \in [-0.2, 2.4]$ are improved by 11 % and 19 %, respectively. The Run 2 analysis has been improved in expectation of the HL-LHC dataset. Some changes, such as new low-$m_{HH}$ signal regions, have only a small impact at the Run 2 luminosity but will enhance the sensitivity significantly at the HL-LHC.

The LHC resumed operation in 2022 and is expected to run until Summer 2026 at a centre-of-mass energy of 13.6 TeV. The Run 3 dataset is expected to be three times larger than the Run 2 dataset, leading to a significant reduction in statistical uncertainties, supported by improved trigger and reconstruction algorithms. Run 3 will be followed by a major upgrade of the ATLAS detector and the LHC for the HL-LHC programme, which is expected to increase the dataset by an additional factor of ten. The physics potential for measuring Higgs boson self-coupling is estimated by extrapolating the Run 2 $HH \rightarrow b\bar{b}\tau^+\tau^-$ sensitivity in various scenarios to the HL-LHC conditions. The study shows that analysis improvements of Run 2 search improve precision to $\kappa_\lambda$ by nearly a factor of two and make the observation of $HH$ production in the $b\bar{b}\tau^+\tau^-$ channel a realistic goal for the HL-LHC program. Especially when considering the combination of various search channels, Higgs boson self-coupling will likely be measured with a precision of approximately 30 % [242, 243] and will therefore remain a central goal of the field beyond the HL-LHC.

At the time of writing, several future colliders are under discussion. Proposed linear [244, 245] or circular [246, 247] electron-positron colliders ($e^-e^+$) will provide unprecedented experimental precision on Higgs boson properties and its coupling to other particles. However, Higgs boson self-coupling will be only accessible via electroweak corrections at those “Higgs factories”, via associated production ($e^+e^- \rightarrow ZH$) or vector boson fusion [248]. Direct $HH$ production is possible only via $e^+e^- \rightarrow ZHH$ production beyond a centre-of-mass energy of $\sqrt{s} > 500$ GeV, and therefore not possible at circular colliders due to beam energy losses resulting from synchrotron radiation. Linear $e^+e^-$ colliders could access $HH$ production via $e^+e^- \rightarrow HH\nu_e\bar{\nu}_e$ for $\sqrt{s} > 1$ TeV, but will be limited by the delivered luminosity [249]. In combination with the HL-LHC, future $e^+e^-$ colliders will likely be able to achieve a precision between 10 to 20 % on $\kappa_\lambda$ [250, 251]. More precise measurements will require hadron colliders that will combine significantly larger $HH$ cross-sections and higher luminosities. One possibility could be the FCC-hh [252], that would re-use the FCC-ee tunnel, and commence operation in the 2070s. Current studies suggest that a precision of few per-cents will be feasible [253]. Similar precision might be also be reached at muon colliders ($\mu^+\mu^-$) that operate at multi-TeV centre-of-mass energies [254–256]. Until then, the knowledge on Higgs boson self-coupling will be dominated by the HL-LHC times.

Recording the HL-LHC dataset will require a major upgrade of the ATLAS detector, including its Detector Safety System (DSS) soft- and hardware. New electronics and sub-detectors, such as the new all-silicon Inner Tracking Detector [109, 110], will further require an extension of DSS. The alarm analysis demonstrates that most alarms are related to maintenance and repair works. In combination with an ageing detector, the findings suggest that the frequency of alarms is likely to increase, diverting limited resources and potentially impacting the detector or data quality. The new Alarm Helper tool has improved the follow up of alarms and identification of faults, and provides a robust database of alarms and their cause. In the future, this database can be integrated into machine learning and artificial intelligence tools that are currently developed across the HEP community [257–259], offering further improvements in detector safety and operational efficiency.

# A Appendix

## A.1 VBF signal parametrisation

The VBF $HH$ matrix element is calculated from the contributions from the three Feynman diagrams shown in Figure 3.11:

$$M_{\text{VBF }HH}(\kappa_\lambda, \kappa_{2V}, \kappa_V) = \kappa_\lambda \kappa_V M_1 + \kappa_{2V} M_2 + \kappa_V^2 M_3 \ . \tag{A.1}$$

When setting $\kappa_V = 1$, the expression reduces to

$$M_{\text{VBF }HH}(\kappa_\lambda, \kappa_{2V}) = \kappa_\lambda M_1 + \kappa_{2V} M_2 + M_3 \ , \tag{A.2}$$

and the VBF $HH$ cross-section is calculated as

$$\begin{aligned}
\sigma_{\text{VBF }HH}(\kappa_\lambda, \kappa_{2V}) &\propto |M_{\text{VBF }HH}(\kappa_\lambda, \kappa_{2V})|^2 \\
&= |\kappa_\lambda M_1 + \kappa_{2V} M_2 + M_3|^2 \\
&= \kappa_\lambda^2 |M_1|^2 + \kappa_{2V}^2 |M_2|^2 + |M_3|^2 + \kappa_\lambda \kappa_{2V} (M_1 M_2^* + M_1^* M_2) \\
&\quad + \kappa_\lambda (M_1 M_3^* + M_1^* M_3) + \kappa_{2V} (M_2 M_3^* + M_2^* M_3) \\
&= \kappa_\lambda^2 \times a_1 + \kappa_{2V}^2 \times a_2 + a_3 + \kappa_\lambda \kappa_{2V} \times a_4 + \kappa_\lambda \times a_5 + \kappa_{2V} \times a_6 \ .
\end{aligned} \tag{A.3}$$

The weights $a_i$ are determined with the first six basis samples from Table 7.2. Any VBF $HH$ sample with any combination of ($\kappa_\lambda$,$\kappa_{2V}$) is thus interpolated from the linear combination:

$$\begin{aligned}
\text{sample}(\kappa_\lambda, \kappa_{2V}) = &\left( \frac{\kappa_{2V}^2}{5} - \frac{\kappa_{2V}^2}{5} - \frac{\kappa_{2V}\kappa_\lambda}{10} + \frac{\kappa_\lambda}{10} \right) \times \text{sample}(1, 3) \\
&+ \left( \frac{4\kappa_{2V}^2}{5} - \frac{4\kappa_{2V}}{5} - \frac{12\kappa_{2V}\kappa_\lambda}{5} + \frac{12\kappa_\lambda}{5} \right) \times \text{sample}(1, \frac{1}{2}) \\
&+ \left( -\frac{5\kappa_{2V}}{4} + \frac{5\kappa_{2V}\kappa_\lambda}{4} + \frac{\kappa_\lambda}{8} - \frac{\kappa_\lambda^2}{8} \right) \times \text{sample}(2, 1) \\
&+ (-\kappa_{2V} + \kappa_{2V}\kappa_\lambda + 1 - \kappa_\lambda) \times \text{sample}(0, 0) \\
&+ \left( \frac{\kappa_{2V}}{36} - \frac{\kappa_{2V}\kappa_\lambda}{36} - \frac{\kappa_\lambda}{72} + \frac{\kappa_\lambda^2}{72} \right) \times \text{sample}(10, 1) \\
&+ \left( -\kappa_{2V}^2 + \frac{29\kappa_{2V}}{9} + \frac{5\kappa_{2V}\kappa_\lambda}{18} - \frac{29\kappa_\lambda}{18} + \frac{\kappa_\lambda^2}{9} \right) \times \text{sample}(1, 1) \ .
\end{aligned} \tag{A.4}$$

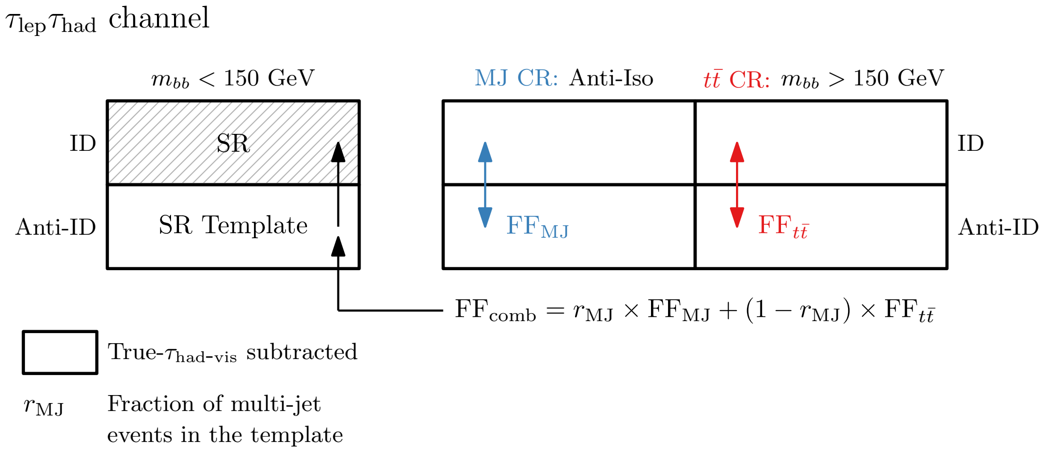


Figure A.1: Illustration of the combined FF factor method in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. Taken from Reference [153].

## A.2 Estimation of the fake-$\tau_{\text{had-vis}}$ background

### Fake-$\tau_{\text{had-vis}}$ estimate in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, fake-$\tau_{\text{had-vis}}$ candidates originate from $t\bar{t}$ and multi-jet backgrounds. The fake contribution is simultaneously estimated for both processes with separate fake factors ($\text{FF}_{t\bar{t}}$, $\text{FF}_{\text{MJ}}$) that are weighted by the relative contribution of the two processes in the region of interest and then combined to produce a single correction. The combined scaling approach reduces the need to distinguish the origin of the fakes in the signal region, which would require accurate modelling of both multi-jet and $t\bar{t}$ fakes. The combined FF approach is independently derived and applied in the SLT and LLT channels, and separately for 1- and 3-prong $\tau_{\text{had-vis}}$ candidates and in bins of the $\tau_{\text{had-vis}}$ $p_{\text{T}}$. The combined method is illustrated in Figure A.1.

The template for the background (region B) is defined by inverting the requirement of exactly one $\tau_{\text{had-vis}}$ in the signal region (region A) by selecting events with exactly one anti-$\tau_{\text{had-vis}}$ candidate, as described in Section 7.4. The process templates in B, C and D are constructed by subtracting all simulated events from data for which the $\tau_{\text{had-vis}}$ is not a fake-$\tau_{\text{had-vis}}$. In other words, only the fake-$\tau_{\text{had-vis}}$ contribution remains in these templates. The template B is scaled by the combined FF

$$\text{FF}_{\text{combined}} = r_{\text{MJ}} \cdot \text{FF}_{\text{MJ}} + (1 - r_{\text{MJ}}) \cdot \text{FF}_{t\bar{t}} \; . \tag{A.5}$$

The weight $r_{\text{MJ}}$ is the expected ratio of multi-jet to $t\bar{t}$ events in B. $r_{\text{MJ}}$ and is determined from simulation. In the SLT channel, $r_{\text{MJ}}$ is approximately zero for any $p_{\text{T}}$. In LTT, $r_{\text{MJ}}$ ranges between 0.2 and 0.3. The multi-jet fake factors ($\text{FF}_{\text{MJ}}$) are calculated in dedicated high multi-jet purity regions ($\text{C}_{\text{MJ}}$, $\text{D}_{\text{MJ}}$). They are defined by inverting isolation criteria for electrons and muons, and dropping the $m_{bb}$ cut. The $t\bar{t}$ fake factors ($\text{FF}_{t\bar{t}}$) are derived in separate regions ($\text{C}_{t\bar{t}}$, $\text{D}_{t\bar{t}}$), that are constructed by inverting the signal regions requirement to $m_{bb} > 150\,\text{GeV}$.

The analysis considers the statistical uncertainties related to the determination of the fake factors and the relative multi-jet contribution. The combined method relies on an accurate modelling of $t\bar{t}$ with true and fake-$\tau_{\text{had-vis}}$ candidates when determining FF and $r_{\text{MJ}}$. The modelling is improved by an additional reweighing of the simulated events to data in a dedicated $t\bar{t}$ control region. The difference between the fake estimate using the corrected and uncorrected $t\bar{t}$ background is applied as a systematic uncertainty in the background estimate. This conservative approach is acceptable

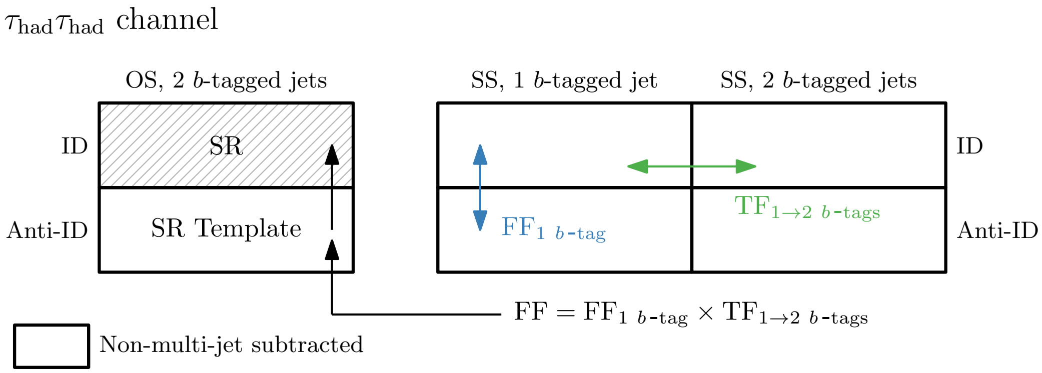


Figure A.2: Illustration of the modified FF method in the $\tau_{\text{had}}\tau_{\text{had}}$ factor. Taken from Reference [153].

as the analysis is not limited by the fake estimate. Conservative modelling uncertainties are also applied for the non-$t\bar{t}$ backgrounds and the ratio $r_{\text{MJ}}$ . Since the values of $\text{FF}_{\text{MJ}}$ and $\text{FF}_{t\bar{t}}$ are similar, the impact is small. The total uncertainty on the combined fake factor varies between channels, being at most 10 % in the SLT channel and at most 25 % in the LTT channel. Closure is checked in the $t\bar{t}$ control region, and validated in the 1-$b$-tagged validation regions.

### Fake-$\tau_{\text{had-vis}}$ estimate in the $\tau_{\text{had}}\tau_{\text{had}}$ channel

**Fake estimate for multi-jets** The modified ABCD method is illustrated in Figure A.2. The template (B), enriched in fakes, is defined by inverting the requirement of two $\tau_{\text{had-vis}}$ candidates per event to one $\tau_{\text{had-vis}}$ and one anti-$\tau_{\text{had-vis}}$. The FF is calculated in two orthogonal regions. They are created by inverting the requirement of opposite charges (OS) between the two $\tau$-candidates to same-signed (SS) as there is no preferred charge correlation for di-$\tau_{\text{had-vis}}$-fakes. To reduce contamination from $t\bar{t}$ events and to increase the available statistics, the control regions $\text{C}_1$ and $\text{D}_1$ require only one b-tagged jet, rather than two, and the corresponding fake factor is denoted as $\text{FF}_{1\ b\text{-tag}}$. A transfer factor $\text{TF}_{1\to 2\ b\text{-tags}}$ is applied to account for the difference between requiring one and two $b$-tagged jets. This transfer factor is determined as the ratio of the $\text{FF}_{1\text{-tag}}$ and the $\text{FF}_{2\text{-tag}}$ obtained in similar control regions requiring 2-$b$-tagged regions, $\text{C}_2$ and $\text{D}_2$.

In all six ABCD regions, the fake-$\tau_{\text{had-vis}}$ template is obtained by subtracting any non-multi-jet simulated processes from data. As the non-multi-jet processes contain a large fraction of $t\bar{t}$ events with fake-$\tau_{\text{had-vis}}$, correction factors for their modelling are derived in a separate step described below, and applied in the subtraction. The fake-$\tau_{\text{had-vis}}$ estimate from multi-jet production is calculated by scaling the template from region B by the FF

$$\text{FF}_{\text{MJ}} = \text{FF}_{1\ b\text{-tag}} \cdot \text{TF}_{1\to 2\ b\text{-tags}} \ . \tag{A.6}$$

FFs are independently derived for the STT and DTT trigger categories, and the different years of data-taking corresponding to different $\tau$-identification procedures at trigger level. Furthermore, they are calculated separately for 1- and 3-prong anti-$\tau_{\text{had-vis}}$ candidates. In the DTT channel, FFs are binned by the $p_{\text{T}}$ and $\eta$ of the anti-$\tau_{\text{had-vis}}$. In the STT channel, the statistical uncertainty is reduced by using the average of the FFs derived separately for events where the anti-$\tau_{\text{had-vis}}$ is leading or sub-leading in $p_{\text{T}}$. Separate TFs are derived for DTT and STT per year of data-taking. These transfer factors are individually determined for 1- and 3-prong and leading and sub-leading anti-$\tau_{\text{had-vis}}$ candidates. In general, the TFs agree with unity within the statistical uncertainty.

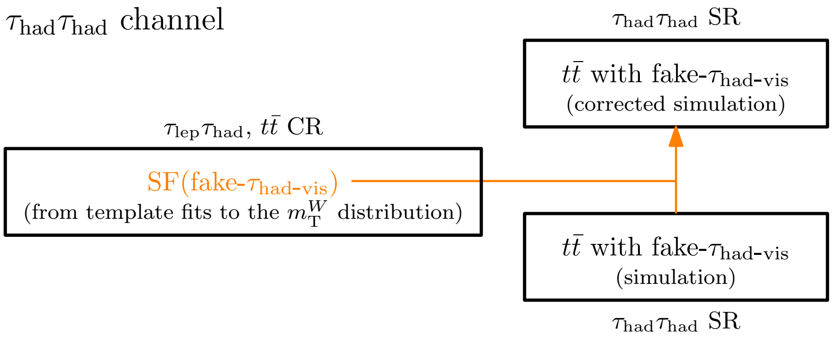


Figure A.3: Illustration of the scale factor method. Taken from Reference [153].

The method considers the statistical uncertainties related to the FFs and TFs, as well as modelling uncertainties of the non-multi-jet background. Furthermore, an uncertainty on potential differences between the fake contribution with OS and SS is derived with an additional 1-$b$-tag OS region that is constructed such to be pure in multi-jet.

**Fake estimate for $t\bar{t}$** Fakes from $t\bar{t}$ are estimated from simulation. The fake-$\tau_{\text{had-vis}}$ misidentification efficiency is corrected with scale factors (SF). These scale factors are determined from a fit to the transverse mass distribution of the $W$ boson, $m_{\text{T}}^{W}$, in dedicated control regions, as illustrated in Figure A.3. The transverse mass is defined using the momentum of the electron or muon and the missing transverse momentum vector, as

$$m_{\text{T}}^{W} = \sqrt{2|\vec{p}_{\text{T}}^{\,\ell}||\vec{p}_{\text{T}}^{\,\text{miss}}|(1-\cos\Delta\phi_{\ell,\,\vec{p}_{\text{T}}^{\,\text{miss}}})}\,, \tag{A.7}$$

where $\Delta\phi$ is the azimuthal angle between the lepton and $\vec{p}_{\text{T}}^{\,\text{miss}}$. The $m_{\text{T}}^{W}$ distribution provides a separation between semi-leptonic and di-leptonic $t\bar{t}$ decays. The former is the source of most $t\bar{t}$ events with fake-$\tau_{\text{had-vis}}$ candidates. Separate SFs are derived for 1- and 3-prong fake-$\tau_{\text{had-vis}}$ candidates and as a function of their $p_{\text{T}}$. The control regions require similar selections as the $t\bar{t}$ control regions defined for the $\tau_{\text{lep}}\tau_{\text{had}}$ fake estimate with the additional requirement of $\tau_{\text{had-vis}}|\eta| < 2.5$ to harmonise with the $\tau_{\text{had}}\tau_{\text{had}}$ channel selection.

The SF determination considers uncertainties related to the detector response and modelling of $t\bar{t}$ and other minor processes. The covariance matrix of the SFs and their statistical and systematic uncertainties is diagonalised. The eigenvectors are applied as independent nuisance parameters to the fit in addition to the general $t\bar{t}$ modelling uncertainties.

# Bibliography

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

[2] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2012.08.021`. URL: `http://dx.doi.org/10.1016/j.physletb.2012.08.021`.

[3] F. Englert et al. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`.

[4] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`.

[5] G. S. Guralnik et al. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`.

[6] R. Adam et al. “Planck2015 results: I. Overview of products and scientific results”. In: *Astronomy & Astrophysics* 594 (Sept. 2016), A1. DOI: `10.1051/0004-6361/201527101`.

[7] G. P. Salam et al. “The Higgs boson turns ten”. In: *Nature Physics* 607 (2022), pp. 41–47. DOI: `doi.org/10.1038/s41586-022-04899-4`.

[8] G. S. Guralnik et al. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`.

[9] F. Arco et al. “Exploring sizable triple Higgs couplings in the 2HDM”. In: *Eur. Phys. J. C* 80.9 (2020), p. 884. arXiv: `2005.10576`.

[10] G. Durieux et al. “Gegenbauer Goldstones”. In: *JHEP* 01 (2022), p. 076. arXiv: `2110.06941`.

[11] F. Haslbeck et al. “The ATLAS Alarm Helper”. In: *EPJ Web Conf.* 295 (2024), p. 02014. DOI: `10.1051/epjconf/202429502014`.

[12] ATLAS Collaboration. “Search for the nonresonant production of Higgs boson pairs via gluon fusion and vector-boson fusion in the $b\bar{b}\tau^+\tau^-$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 110.3 (2024), p. 032012. arXiv: `2404.12660`.

[13] ATLAS Collaboration. *Updated projection of the sensitivity of searches for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state from LHC Run 2 to the High Luminosity LHC with the ATLAS detector.* ATL-PHYS-PUB-2024-016. 2024. URL: `https://cds.cern.ch/record/2910850`.

[14] M. Thomson. *Modern particle physics*. New York: Cambridge University Press, Oct. 2013. DOI: 10.1017/CBO9781139525367.

[15] C. Burgard. *Standard model infographics*. Accessed 20.04.2025. URL: https://texample.net/tikz/examples/model-physics/.

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

[17] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[18] A. Salam et al. "Electromagnetic and weak interactions". In: *Physics Letters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)90711-5.

[19] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[20] J. Riebesell. *Higgs Potential*. Accessed 20.04.2025. URL: https://tikz.net/higgs-potential/.

[21] N. Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[22] M. Kobayashi et al. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[23] D. Denegri. "The discovery of the W and Z". In: *Phys. Rep.* 403-404 (2004), pp. 107–146. DOI: 10.1016/j.physrep.2004.09.006.

[24] CDF Collaboration. "Observation of Top Quark Production in Pbar-P Collisions with the Collider Detector at Fermilab". In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2626–2631. DOI: 10.1103/physrevlett.74.2626.

[25] D0 Collaboration. "Observation of the Top Quark". In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2632–2637. DOI: 10.1103/physrevlett.74.2632.

[26] Super-Kamiokande Collaboration. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[27] A. G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *The Astronomical Journal* 116.3 (Sept. 1998), pp. 1009–1038. arXiv: 9805201.

[28] The Supernova Cosmology Project. "Measurements of $\Sigma$ and $\Lambda$ from 42 High-Redshift Supernovae". In: *The Astrophysical Journal* 517.2 (June 1999), pp. 565–586. arXiv: 9812133.

[29] NNPDF Collaboration. "Parton distributions from high-precision collider data: NNPDF Collaboration". In: *The European Physical Journal C* 77.10 (Oct. 2017). arXiv: 1706.00428.

[30] A. A. Bhatti et al. "Jet Physics at the Tevatron". In: *Annual Review of Nuclear and Particle Science* 60.1 (Nov. 2010), pp. 267–297. arXiv: 1002.1708.

[31] B. Moser. "The Beauty and the Boost: a Higgs Boson Tale - Measurements of Higgs Boson Production at High Energy in Decays to Bottom Quarks and Their Interpretations with the ATLAS Experiment at the LHC". Amsterdam U., 2021. URL: https://cds.cern.ch/record/2803776.

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

[33] ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *The European Physical Journal C* 70.3 (Sept. 2010), pp. 823–874. arXiv: 1005.4568.

[34] ATLAS Collaboration. "AtlFast3: The Next Generation of Fast Simulation in ATLAS". In: *Computing and Software for Big Science* 6.1 (Mar. 2022). ISSN: 2510-2044. arXiv: 2109.02551.

[35] ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". In: *Nature* 607 (2022), pp. 52–59. arXiv: 2207.00092.

[36] S. D. Bass et al. "The Higgs boson implications and prospects for future discoveries. The Higgs boson – its implications and prospects for future discoveries". In: *Nature Rev. Phys.* 3.9 (2021), pp. 608–624. arXiv: 2104.06821.

[37] S. P. Martin. "A supersymmetry primer". In: *Perspectives on Supersymmetry*. World Scientific, July 1998, pp. 1–98. arXiv: 9709356.

[38] ATLAS Collaboration. "Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the ATLAS detector using $\sqrt{s}$ = 7, 8 and 13 TeV $pp$ collision data". In: *Phys. Rev. Lett.* 131 (2023), p. 251802. arXiv: 2308.04775.

[39] D. M. Webber et al. "Measurement of the Positive Muon Lifetime and Determination of the Fermi Constant to Part-per-Million Precision". In: *Phys. Rev. Lett.* 106 (2011), p. 041803. arXiv: 1010.0991.

[40] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports, 2017. DOI: 10.23731/CYRM-2017-002.

[41] A. Karlberg et al. *Ad interim recommendations for the Higgs boson production cross sections at $\sqrt{s} = 13.6$ TeV*. 2024. arXiv: 2402.09955.

[42] M. Grazzini et al. "Higgs boson pair production at NNLO with top quark mass effects". In: *Journal of High Energy Physics* 2018.5 (May 2018). DOI: 10.1007/jhep05(2018)059.

[43] F. A. Dreyer et al. "Precise predictions for double-Higgs production via vector-boson fusion". In: *The European Physical Journal C* 80.11 (Nov. 2020). DOI: 10.1140/epjc/s10052-020-08610-7.

[44] F. A. Dreyer et al. "Vector-boson fusion Higgs pair production at N$^3$LO". In: *Physical Review D* 98.11 (Dec. 2018). DOI: 10.1103/physrevd.98.114016.

[45] D. de Florian et al. "Triple Higgs production at hadron colliders at NNLO in QCD". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). arXiv: 1912.02760.

[46] ATLAS DiHiggs group. *Di-Higgs Branching Ratios*. https://twiki.cern.ch/twiki/pub/AtlasProtected/HHGDD/DiHiggsBRs.pdf. Accessed: 15.02.2025.

[47] B. Di Micco et al. "Higgs boson potential at colliders: Status and perspectives". In: *Reviews in Physics* 5 (Nov. 2020), p. 100045. DOI: 10.1016/j.revip.2020.100045.

[48] S. Heinemeyer et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN, 2013. DOI: 10.5170/CERN-2013-004.

[49] E. Bagnaschi et al. *Higgs boson pair production at NLO in the POWHEG approach and the top quark mass uncertainties*. 2023. arXiv: 2309.10525.

[50] S. Amoroso et al. *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*. 2020. arXiv: 2003.01700.

[51] ATLAS Collaboration. "Combination of searches for Higgs boson pairs in $pp$ collisions at $\sqrt{s}$=13 TeV with the ATLAS detector". In: *Physics Letters B* 800 (Jan. 2020). DOI: 10.1016/j.physletb.2019.135103.

[52] ATLAS Collaboration. *Measurement of off-shell Higgs boson production in the* $H^* \to ZZ \to 4\ell$ *decay channel using a neural simulation-based inference technique in 13 TeV pp collisions with the ATLAS detector*. 2024. arXiv: 2412.01548.

[53] ATLAS Collaboration. “Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector”. In: *Eur. Phys. J. C* 75 (2015), p. 476. DOI: 10.1140/epjc/s10052-015-3685-1. Corrected in “Erratum to: Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016), p. 152. DOI: 10.1140/epjc/s10052-016-3934-y.

[54] ATLAS Collaboration. “Interpretations of the ATLAS measurements of Higgs boson production and decay rates and differential cross-sections in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2024.11 (Nov. 2024). arXiv: 2402.05742.

[55] ATLAS Collaboration. “Combination of searches for Higgs boson pair production in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *CERN-EP-2024-160* (June 2024). arXiv: 2406.09971.

[56] 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”. In: *Phys. Rev. D* 108.5 (2023), p. 052003. arXiv: 2301.03212.

[57] ATLAS Collaboration. “Search for pair production of boosted Higgs bosons via vector-boson fusion in the $b\bar{b}b\bar{b}$ final state using $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *CERN-EP-2024-092* (Apr. 2024). arXiv: 2404.17193.

[58] ATLAS Collaboration. “Studies of new Higgs boson interactions through nonresonant HH production in the $b\bar{b}\gamma\gamma$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 01 (2024), p. 066. arXiv: 2310.12301.

[59] ATLAS Collaboration. “Search for non-resonant Higgs boson pair production in final states with leptons, taus, and photons in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: (May 2024). arXiv: 2405.20040.

[60] ATLAS Collaboration. “Search for non-resonant Higgs boson pair production in the $2b + 2\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 02 (2024), p. 037. arXiv: 2310.11286.

[61] CMS Collaboration. “A portrait of the Higgs boson by the CMS experiment ten years after the discovery”. In: *Nature* 607.7917 (July 2022), pp. 60–68. arXiv: 2207.00043.

[62] 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”. In: *Physics Letters B* 842 (July 2023), p. 137531. arXiv: 2206.09401.

[63] ATLAS Collaboration. “Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at s=13 TeV”. In: *Phys. Lett. B* 843 (2023), p. 137745. arXiv: 2211.01216.

[64] CMS Collaboration. “Constraints on the Higgs boson self-coupling from the combination of single and double Higgs boson production in proton-proton collisions at $\sqrt{s}$=13 TeV”. In: *Physics Letters B* 861 (Feb. 2025), p. 139210. arXiv: 2407.13554.

[65] E. Mobs. *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018*. General Photo. 2018. URL: https://cds.cern.ch/record/2636343.

[66] L. Evans et al. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[67] *LEP design report*. Report. Geneva: CERN, 1984. URL: https://cds.cern.ch/record/102083.

[68] M. Lamont. "The First Years of LHC Operation for Luminosity Production". In: *IPAC 2013*. 2013, MOYAB101. URL: https://cds.cern.ch/record/2010134.

[69] F. Bordry et al. "The First Long Shutdown (LS1) for the LHC". In: *4th International Particle Accelerator Conference*, vol. C130512. 2013, MOZB202. URL: https://cds.cern.ch/record/1575133.

[70] R. Steerenberg et al. "Operation and performance of the CERN Large Hadron Collider during proton Run 2". In: *10th International Particle Accelerator Conference*. 2019, MOPMP031. DOI: 10.18429/JACoW-IPAC2019-MOPMP031.

[71] J. Coupard et al. *LHC Injectors Upgrade, Technical Design Report*. Ed. by J Coupard. 2014. DOI: 10.17181/CERN.7NHR.6HGC.

[72] ATLAS Collaboration. *Public ATLAS Online Luminosity Plots for Run-3 of the LHC*. Accessed 17. October 2025. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3.

[73] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[74] CMS Collaboration. "The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[75] LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[76] ALICE Collaboration. "The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[77] L. Pequenao et al. "How ATLAS detects particles: diagram of particle paths in the detector". 2013. URL: https://cds.cern.ch/record/1505342.

[78] ATLAS Collaboration. *ATLAS inner detector*. Technical design report. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[79] ATLAS Collaboration. "Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment". 2020. URL: https://cds.cern.ch/record/2723878.

[80] ATLAS Collaboration. "Alignment of the ATLAS Inner Detector in Run-2". In: *Eur. Phys. J. C* 80.12 (2020), p. 1194. arXiv: 2007.07624.

[81] ATLAS Collaboration. *IBL Efficiency and Single Point Resolution in Collision Events*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2203893.

[82] A. Ahmad et al. "The Silicon microstrip sensors of the ATLAS semiconductor tracker". In: *Nucl. Instrum. Meth. A* 578 (2007), pp. 98–118. DOI: 10.1016/j.nima.2007.04.157.

[83] ATLAS Collaboration. "Operation and performance of the ATLAS semiconductor tracker in LHC Run 2". In: *JINST* 17 (2022), P01013. arXiv: 2109.02591.

[84] A. Vogel. *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*. Tech. rep. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1537991.

[85] N. Garelli et al. "Performance of the ATLAS Detector in Run-2". In: *EPJ Web Conf.* 164 (2017), p. 01021. DOI: 10.1051/epjconf/201716401021.

[86] ATLAS Collaboration. *ATLAS liquid-argon calorimeter*. Technical design report. Geneva: CERN, 1996. DOI: 10.17181/CERN.FWRW.FOOQ. URL: https://cds.cern.ch/record/331061.

[87] ATLAS Collaboration. *ATLAS tile calorimeter*. Technical design report. Geneva: CERN, 1996. DOI: 10.17181/CERN.JRBJ.7028. URL: https://cds.cern.ch/record/331062.

[88] J. Pequenao. “Computer Generated image of the ATLAS calorimeter”. 2008. URL: https://cds.cern.ch/record/1095927.

[89] N. Nikiforou et al. “Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade”. In: *3rd International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications*. June 2013. arXiv: 1306.6756.

[90] E. Diehl. *ATLAS Muon Detector Commissioning*. Tech. rep. 2009. arXiv: 0910.2767.

[91] ATLAS Collaboration. *ATLAS muon spectrometer*. Technical design report. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331068.

[92] T. Kawamoto et al. *New Small Wheel Technical Design Report*. Tech. rep. CERN, 2013.

[93] G. Avoni et al. “The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS”. In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.

[94] M. Mikuz et al. “Diamond pad detector telescope for beam conditions and luminosity monitoring in ATLAS”. In: *Nucl. Instrum. Methods Phys. Res., A* 579.2 (2007), pp. 788–794. DOI: 10.1016/j.nima.2007.05.297.

[95] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: *The European Physical Journal C* 83.10 (Oct. 2023).

[96] ATLAS Collaboration. “Operation of the ATLAS trigger system in Run 2”. In: *JINST* 15.10 (2020), P10004. arXiv: 2007.12539.

[97] ATLAS Collaboration. “Performance of the ATLAS trigger system in 2015”. In: *The European Physical Journal C* 77.5 (May 2017). arXiv: 1611.09661.

[98] ATLAS Collaboration. “The ATLAS inner detector trigger performance in pp collisions at 13 TeV during LHC Run 2”. In: *The European Physical Journal C* 82.3 (Mar. 2022). arXiv: 2107.02485.

[99] I. Bird et al. “Update of the Computing Models of the WLCG and the LHC Experiments”. In: (Apr. 2014). URL: https://cds.cern.ch/record/1695401.

[100] K. Bos et al. *LHC computing Grid*. Technical design report. LCG. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/840543.

[101] D. Buscher. “Search for Higgs bosons with b-jets in the final state in proton-proton collisions with the ATLAS experiment”. Presented 26 Oct 2016. Freiburg U., 2016. URL: https://cds.cern.ch/record/2232472.

[102] 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”. In: *Journal of High Energy Physics* 2015.1 (Jan. 2015). DOI: 10.1007/jhep01(2015)069.

[103] ATLAS Collaboration. “Luminosity Public Results: Run 2”. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed 21.01.2025.

[104] 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.

[105] 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.

[106] ATLAS Collaboration. *HL-LHC prospects for the measurement of Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state and combination with the $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states at the ATLAS experiment*. ATL-PHYS-PUB-2022-053. 2022. URL: https://cds.cern.ch/record/2841244.

[107] I. Zurbano Fernandez et al. "High-Luminosity Large Hadron Collider (HL-LHC): Technical design report". In: 10/2020 (Dec. 2020). Ed. by I. Béjar Alonso et al. DOI: 10.23731/CYRM-2020-0010.

[108] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.2LBB.4IAL.

[109] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.FOZZ.ZP3Q.

[110] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Strip Detector*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257755.

[111] ATLAS Collaboration. *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2719855.

[112] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285580.

[113] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285583.

[114] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.6QIO.YGHO.

[115] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.2LBB.4IAL.

[116] O. Beltramello et al. "The Detector Safety System of the ATLAS experiment". In: *JINST* 4 (2009), P09012. DOI: 10.1088/1748-0221/4/09/P09012.

[117] S. Grau et al. "CERN Safety Alarm Monitoring". In: *Conf. Proc. C* 110904 (2011), pp. 1674–1676.

[118] A. Ledeul et al. "REMUS: The new CERN Radiation and Environment Monitoring Unified Supervision". In: *ICALEPCS* (2015). DOI: 10.18429/JACoW-ICALEPCS2015-TUD3O03.

[119] S. M. Goodman et al. "Scalable manufacturing of fibrous nanocomposites for multifunctional liquid sensing". In: *Nano Today* 40 (2021), p. 101270. DOI: 10.1016/j.nantod.2021.101270.

[120] A. Barriuso Poy et al. "The detector control system of the ATLAS experiment". In: *JINST* 3 (2008), P05006. DOI: 10.1088/1748-0221/3/05/P05006.

[121] S. Schlenker et al. "The ATLAS Detector Control System". In: *Conf. Proc.* C111010 (2011), MOBAUST02. URL: https://cds.cern.ch/record/1562594.

[122] C. Boyer et al. "The CERN EDMS: Engineering and equipment data management system". In: *8th European Particle Accelerator Conference (EPAC 2002)*. June 2002, pp. 2697–2699.

[123] I. A. Tortajada et al. “ATLAS Technical Coordination Expert System”. In: *EPJ Web Conf.* 214 (2019), p. 05035. DOI: `10.1051/epjconf/201921405035`.

[124] R. Jones et al. “The OKS persistent in-memory object manager”. In: *IEEE Trans. Nucl. Sci.* 45.4 (1998), pp. 1958–1964. DOI: `10.1109/23.710971`.

[125] G. A. Uribe et al. “Graph-based algorithm for the understanding of failures in the ATLAS infrastructure”. In: *J. Phys. Conf. Ser.* 2438.1 (2023), p. 012045. DOI: `10.1088/1742-6596/2438/1/012045`.

[126] BE-OP. “LHC schedule”. `https://edms.cern.ch/project/CERN-0000174580`. Accessed 28.01.2025.

[127] J. Malaquin. “LS2 shutdown twiki”. `https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LS2Shutdown`. Accessed 28.01.2025. 2022.

[128] A. Corso-Radu et al. *The Electronic Logbook for the Information Storage of ATLAS Experiment at LHC (ELisA)*. Tech. rep. Geneva: CERN, 2012. URL: `https://cds.cern.ch/record/1458057`.

[129] C. Garino et al. “Intervention Management from Operation to Shutdown”. In: *”4th International Particle Accelerator Conference”*. May 2013, p. 3705. URL: `https://cds.cern.ch/record/2010151`.

[130] ATLAS Collaboration. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2”. In: *The European Physical Journal C* 77.10 (Oct. 2017). DOI: `10.1140/epjc/s10052-017-5225-7`.

[131] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: `10.1016/0168-9002(87)90887-4`.

[132] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *The European Physical Journal C* 77.5 (2016). DOI: `10.1140/epjc/s10052-017-4887-5`.

[133] ATLAS Collaboration. *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s}$ = 13 TeV*. Tech. rep. Geneva: CERN, 2015. URL: `https://cds.cern.ch/record/2037717`.

[134] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *The European Physical Journal C* 77.7 (July 2017). DOI: `10.1140/epjc/s10052-017-5004-5`.

[135] ATLAS Collaboration. *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*. Tech. rep. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2298955`.

[136] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14.12 (Dec. 2019). DOI: `10.1088/1748-0221/14/12/p12006`.

[137] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 81.7 (July 2021). DOI: `10.1140/epjc/s10052-021-09233-2`.

[138] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV”. In: *The European Physical Journal C* 76.5 (May 2016). DOI: `10.1140/epjc/s10052-016-4120-y`.

[139] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *The European Physical Journal C* 77.7 (July 2017). DOI: `10.1140/epjc/s10052-017-5031-2`.

[140] M. Cacciari et al. “The anti-kt jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063.

[141] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *The European Physical Journal C* 81.8 (Aug. 2021). arXiv: 2007.02645.

[142] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1700870.

[143] ATLAS Collaboration. “ATLAS flavour-tagging algorithms for the LHC Run 2 $pp$ collision dataset”. In: *Eur. Phys. J. C* 83 (2023), p. 681. arXiv: 2211.16345.

[144] 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”. In: *JHEP* 2307 (2023), p. 040. arXiv: 2209.10910.

[145] ATLAS Collaboration. “Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector”. In: *Journal of High Energy Physics* 2017.12 (Dec. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep12(2017)024. arXiv: 1708.03299. URL: http://dx.doi.org/10.1007/JHEP12(2017)024.

[146] A. Duperrin. *Flavour tagging with graph neural networks with the ATLAS detector*. Tech. rep. Geneva: CERN, 2023. arXiv: 2306.04415.

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

[148] ATLAS Collaboration. “Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s}$=8 TeV”. In: *The European Physical Journal C* 75.7 (July 2015). DOI: 10.1140/epjc/s10052-015-3500-z.

[149] ATLAS Collaboration. *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2688062.

[150] A. Elagin et al. “A new mass reconstruction technique for resonances decaying to $\tau\tau$”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (Oct. 2011), pp. 481–489. DOI: 10.1016/j.nima.2011.07.009.

[151] ATLAS Collaboration. *ATLAS Overlap Removal Tool: AssociationUtils*. Accessed 10 April 2025. URL: https://gitlab.cern.ch/atlas/athena/tree/21.2/PhysicsAnalysis/AnalysisCommon/AssociationUtils/.

[152] ATLAS Collaboration. *Recommended Overlap Removal Working Points*. Accessed 10 April 2025. URL: https://indico.cern.ch/event/631313/contributions/2683959/attachments/1518878/2373377/Farrell_ORTools_ftaghbb.pdf.

[153] 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”. In: *Journal of High Energy Physics* 7 (July 2023). arXiv: 2209.10910.

[154] ATLAS Collaboration. *The ATLAS Tau Trigger in Run 2*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2274201.

[155] ATLAS Collaboration. “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *The European Physical Journal C* 80.1 (Jan. 2020). arXiv: 1909.00761.

[156] ATLAS Collaboration. “Performance of the ATLAS muon triggers in Run 2”. In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09015–P09015. ISSN: 1748-0221. arXiv: 2004.13447.

[157] ATLAS Collaboration. "ATLAS data quality operations and performance for 2015–2018 data-taking". In: *Journal of Instrumentation* 15.04 (Apr. 2020). DOI: 10.1088/1748-0221/15/04/p04003.

[158] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: *The European Physical Journal C* 83.10 (Oct. 2023). DOI: 10.1140/epjc/s10052-023-11747-w.

[159] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *JINST* 13.07 (2018). DOI: 10.1088/1748-0221/13/07/P07017.

[160] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). DOI: 10.1016/S0168-9002(01)00089-4.

[161] E. Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Physics* 7.3 (Sept. 2019). DOI: 10.21468/scipostphys.7.3.034.

[162] CMS Collaboration ATLAS Collaboration. "Combined Measurement of the Higgs Boson Mass in $pp$ Collisions at $\sqrt{s}$=7 and 8 TeV with the ATLAS and CMS Experiments". In: *Physical Review Letters* 114.19 (May 2015). DOI: 10.1103/physrevlett.114.191803.

[163] ATLAS Collaboration. "Combined Measurement of the Higgs Boson Mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the ATLAS detector using $\sqrt{s}$=7,8 and 13 TeV $pp$ collision data". In: *Physical Review Letters* 131.25 (Dec. 2023). DOI: 10.1103/physrevlett.131.251802.

[164] S. Agostinelli et al. "Geant 4 simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003). DOI: 10.1016/S0168-9002(03)01368-8.

[165] T. Sjöstrand et al. "A brief introduction to PYTHIA 8.1". In: *Computer Physics Communications* 178.11 (June 2008). DOI: 10.1016/j.cpc.2008.01.036.

[166] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model.* Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2206965.

[167] R. D. Ball et al. "Parton distributions with LHC data". In: *Nuclear Physics B* 867.2 (Feb. 2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003.

[168] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 2010). DOI: 10.1007/jhep06(2010)043.

[169] J. Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001.

[170] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301.

[171] NNPDFCollaboration and R. D. Ball. "Parton distributions for the LHC run II". In: *Journal of High Energy Physics* 2015.4 (Apr. 2015). DOI: 10.1007/jhep04(2015)040.

[172] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. arXiv: 1410.3012.

[173] ATLAS Collaboration. *Summary of ATLAS Pythia 8 tunes.* Tech. rep. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1474107.

[174] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1966419.

[175] P. Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. arXiv: 1212.3460.

[176] S. Frixione et al. "Single-top hadroproduction in association with a W boson". In: *Journal of High Energy Physics* 2008.07 (July 2008), pp. 029–029. arXiv: 0805.3067.

[177] E. Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. arXiv: 1905.09127.

[178] T. Gleisberg et al. "Comix, a new matrix element generator". In: *JHEP* 12 (2008), p. 039. arXiv: 0808.3674.

[179] F. Buccioni et al. "OpenLoops 2". In: *Eur. Phys. J. C* 79.10 (2019), p. 866. arXiv: 1907.13071.

[180] F. Cascioli et al. "Scattering Amplitudes with Open Loops". In: *Phys. Rev. Lett.* 108 (2012), p. 111601. arXiv: 1111.5206.

[181] A. Denner et al. "COLLIER: A fortran-based complex one-loop library in extended regularizations". In: *Comput. Phys. Commun.* 212 (2017), pp. 220–238. arXiv: 1604.06792.

[182] S. Höche et al. "A critical appraisal of NLO+PS matching methods". In: *JHEP* 09 (2012), p. 049. arXiv: 1111.1220.

[183] S. Höche et al. "QCD matrix elements + parton showers. The NLO case". In: *JHEP* 04 (2013), p. 027. arXiv: 1207.5030.

[184] S. Catani et al. "QCD Matrix Elements + Parton Showers". In: *JHEP* 11 (2001), p. 063. arXiv: 0109231.

[185] S. Höche et al. "QCD matrix elements and truncated showers". In: *JHEP* 05 (2009), p. 053. arXiv: 0903.1219.

[186] ATLAS Collaboration. "Measurement of the $Z/\gamma^*$ boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2014.9 (Sept. 2014). arXiv: 1406.3660.

[187] J. Pumplin et al. "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis". In: *Journal of High Energy Physics* 2002.07 (July 2002), pp. 012–012. arXiv: 0201195.

[188] ATLAS Collaboration. *Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2665057.

[189] M. Bähr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. arXiv: 0803.0883.

[190] J Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *The European Physical Journal C* 76.4 (Apr. 2016). DOI: 10.1140/epjc/s10052-016-4018-8.

[191] F. Maltoni et al. "Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC". In: *The European Physical Journal C* 77.12 (Dec. 2017). DOI: 10.1140/epjc/s10052-017-5410-8.

[192] Y. Freund. "Boosting a Weak Learning Algorithm by Majority". In: *Information and Computation* 121.2 (1995), pp. 256–285. ISSN: 0890-5401. DOI: https://doi.org/10.1006/inco.1995.1136.

[193] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2009. arXiv: 0703039.

[194] J. Bergstra et al. *Making a Science of Model Search*. 2012. arXiv: 1209.5111.

[195] C. Deutsch. “Search for Higgs boson pair production in the bbtautau final state with the ATLAS detector”. PhD thesis. Rheinische Friedrich-Wilhelms-Universität Bonn, Dec. 2021. URL: https://cds.cern.ch/record/2788627.

[196] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: *The European Physical Journal C* 83.10 (Oct. 2023). arXiv: 2212.09379.

[197] ATLAS Collaboration. *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2261772.

[198] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *The European Physical Journal C* 81.8 (Aug. 2021). arXiv: 2007.02645.

[199] ATLAS Collaboration. *$E_T^{miss}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2625233.

[200] ATLAS Collaboration. “ATLAS $b$-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 79.11 (Nov. 2019). arXiv: 1907.05120.

[201] 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”. In: *The European Physical Journal C* 82.1 (Jan. 2022). arXiv: 2109.10627.

[202] ATLAS Collaboration. “Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 $\text{fb}^{-1}$ of ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 83.8 (Aug. 2023). arXiv: 2301.06319.

[203] F. A. Di Bello et al. *Modelling and computational improvements to the simulation of single vector-boson plus jet processes*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2765984.

[204] J. Campbell et al. “Single-top-quark production in the $t$-channel at NNLO”. In: *JHEP* 02 (2021), p. 040. arXiv: 2012.01574 [hep-ph].

[205] N. Kidonakis et al. “Higher-order corrections for $tW$ production at high-energy hadron colliders”. In: *JHEP* 05 (2021), p. 278. arXiv: 2102.11300.

[206] M. Villaplana. “Monte Carlo generators for top quark production at the LHC”. In: (2022). URL: https://cds.cern.ch/record/2826417.

[207] LHC Cross-section Working-Group. *LHC Cross Section Working Group Twiki*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGHH.

[208] ATLAS Collaboration. “Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2019.4 (Apr. 2019). arXiv: 1811.12113.

[209] ATLAS Collaboration. “Measurement of the cross-section for W boson production in association with b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2013.6 (June 2013). arXiv: 1302.2929.

[210] ATLAS Collaboration. “Observation of $H \to b\bar{b}$ decays and $VH$ production with the ATLAS detector”. In: *Physics Letters B* 786 (Nov. 2018), pp. 59–86. arXiv: 1808.08238.

[211] ATLAS Collaboration. “Measurements of WH and ZH production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp

collisions with the ATLAS detector”. In: *Journal of High Energy Physics* 2025.4 (Apr. 2025). arXiv: 2410.19611.

[212] K. Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. DOI: 10.17181/CERN-OPEN-2012-016.

[213] W. Verkerke et al. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116.

[214] L. Moneta et al. *The RooStats Project*. 2011. arXiv: 1009.1003.

[215] R. Barlow et al. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[216] F. James et al. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

[217] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. arXiv: 1007.1727. Erratum to: “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 73 (2013), p. 2501. DOI: https://doi.org/10.1140/epjc/s10052-013-2501-z.

[218] A. L. Read. “Modified frequentist analysis of search results (the $CL_s$ method)”. In: (2000). DOI: 10.5170/CERN-2000-005.81. URL: https://cds.cern.ch/record/451614.

[219] A. L. Read. “Presentation of search results: The $CL_s$ technique”. In: *J. Phys. G* 28 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[220] I. Asimov. “Franchise”. In: *If* (1955).

[221] K. G. Hayes et al. “Application of the Bootstrap Statistical Method to the Tau Decay Mode Problem”. In: *Phys. Rev. D* 39 (1989), p. 274. DOI: 10.1103/PhysRevD.39.274.

[222] G. Heinrich et al. “SMEFT predictions for gg → hh at full NLO QCD and truncation uncertainties”. In: *Journal of High Energy Physics* 2022.8 (Aug. 2022). DOI: 10.1007/jhep08(2022)079. arXiv: 2204.13045. Erratum to: “SMEFT predictions for gg → hh at full NLO QCD and truncation uncertainties”. In: *Journal of High Energy Physics* 2023.86 (Aug. 2023). DOI: 10.1007/jhep08(2023)086.

[223] ATLAS Collaboration. “Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector”. In: *Physics Letters B* 786 (Nov. 2018), pp. 59–86. arXiv: 1808.08238.

[224] D0 Collaboration. “Observation of $ZZ$ production in $p\bar{p}$ collisons at $\sqrt{s} = 1.96$ TeV”. In: *Physical Review Letters* 101.17 (Oct. 2008). arXiv: 0808.0703.

[225] B. Mellado Garcia et al. “CERN Report 4: Part I Standard Model Predictions”. In: (2016). URL: https://cds.cern.ch/record/2150771.

[226] 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.” In: *Phys. Rev. Lett.* 121 (2018). Submitted to Phys.Rev.Lett., p. 191801. arXiv: 1808.00336.

[227] 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.

[228] ATLAS Collaboration. *Expected tracking and related performance with the updated ATLAS Inner Tracker layout at the High-Luminosity LHC*. ATL-PHYS-PUB-2021-024. 2021. URL: https://cds.cern.ch/record/2776651.

[229] M. Cepeda et al. “Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC”. In: *CERN Yellow Rep. Monogr.* 7 (2019), pp. 221–584. arXiv: 1902.00134 [hep-ph].

[230] 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”. In: *ATL-PHYS-PUB-2022-005* (2022). URL: https://cds.cern.ch/record/2802127.

[231] S. Manzoni et al. “Taming a leading theoretical uncertainty in HH measurements via accurate simulations for $\mathrm{b\bar{b}H}$ production”. In: *JHEP* 09 (2023), p. 179. arXiv: 2307.09992 [hep-ph].

[232] ATLAS Collaboration. *ATLAS HL-LHC Computing Conceptual Design Report*. CERN-LHCC-2020-015, LHCC-G-178. 2020. URL: https://cds.cern.ch/record/2729668.

[233] ATLAS and CMS Collaborations. “Addendum to the report on the physics at the HL-LHC, and perspectives for the HE-LHC: Collection of notes from ATLAS and CMS”. In: *CERN Yellow Rep. Monogr.* 7 (2019). arXiv: 1902.10229.

[234] ATLAS Collaboration. “Combination of Searches for Higgs Boson Pair Production in $pp$ Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector”. In: *Phys. Rev. Lett.* 133 (10 Sept. 2024), p. 101801. DOI: 10.1103/PhysRevLett.133.101801.

[235] A. Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL].

[236] ATLAS Collaboration. *Neural Network Jet Flavour Tagging with the Upgraded ATLAS Inner Tracker Detector at the High-Luminosity LHC*. ATL-PHYS-PUB-2022-047. 2022. URL: https://cds.cern.ch/record/2839913.

[237] ATLAS Collaboration. *Projected sensitivity of measurements of Higgs boson pair production with the ATLAS experiment at the HL-LHC*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2925853.

[238] ATLAS Collaboration. *Updated projection of the sensitivity of searches for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state from LHC Run 2 to the High Luminosity LHC with the ATLAS detector*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2923694.

[239] ATLAS Collaboration. *HL-LHC prospects for the measurement of Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state and combination with the $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states at the ATLAS experiment*. Tech. rep. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2841244.

[240] ATLAS Collaboration. *HL-LHC prospects for the search of boosted Higgs boson pair production via vector-boson fusion in the $b\bar{b}b\bar{b}$ final state at the ATLAS experiment*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2925813.

[241] ATLAS Collaboration. *Projected sensitivity of searches for Higgs boson pair production in final states with light leptons, taus, and photons with the ATLAS detector at the HL-LHC*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2925675.

[242] ATLAS and CMS Collaborations. *Highlights of the HL-LHC physics projections by ATLAS and CMS*. 2025. arXiv: 2504.00672.

[243] ATLAS Collaboration. *Projected sensitivity of measurements of Higgs boson pair production with the ATLAS experiment at the HL-LHC*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2925853.

[244] P. Lebrun et al. *The CLIC Programme: Towards a Staged e+e- Linear Collider Exploring the Terascale : CLIC Conceptual Design Report*. 2012. arXiv: 1209.2543.

[245] T. Behnke et al. *The International Linear Collider Technical Design Report - Volume 1: Executive Summary*. 2013. arXiv: 1306.6327.

[246] W. Bartmann et al. *Future Circular Collider Feasibility Study Report Volume 1: Physics and Experiments*. 2025. arXiv: 2505.00272.

[247] The CEPC Study Group. *CEPC Conceptual Design Report: Volume 1 - Accelerator*. 2018. arXiv: 1809.00285.

[248] M. McCullough. “Indirect model-dependent probe of the Higgs self-coupling”. In: *Physical Review D* 90.1 (July 2014). arXiv: 1312.3322.

[249] CLIC Collaboration. “Higgs physics at the CLIC electron–positron linear collider”. In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. arXiv: 1608.07538.

[250] M. Berggren et al. *Update of the Higgs Self-coupling Projections from Di-Higgs Production in Detailed Simulation of the ILD Concept*. 2025. arXiv: 2509.14148.

[251] A. Blondel et al. *The FCC integrated programme: a physics manifesto*. 2025. arXiv: 2504.02634.

[252] M. Benedikt et al. *FCC-hh: The Hadron Collider*. Tech. rep. Geneva: CERN, 2019. DOI: 10.1140/epjst/e2019-900087-0.

[253] M. Mangano et al. *Prospects for physics at FCC-hh*. Mar. 2025. DOI: 10.17181/mf0cz-2c516.

[254] C. Accettura et al. *Towards a Muon Collider*. 2023. arXiv: 2303.08533.

[255] D. Buttazzo et al. “Two paths towards precision at a very high energy lepton collider”. In: *Journal of High Energy Physics* 2021.5 (May 2021). arXiv: 2012.11555.

[256] T. Han et al. “Electroweak couplings of the Higgs boson at a multi-TeV muon collider”. In: *Physical Review D* 103.1 (Jan. 2021). arXiv: 2012.11555.

[257] A. Sulc et al. *Towards Unlocking Insights from Logbooks Using AI*. 2024. arXiv: 2406.12881.

[258] D. Dal Santo et al. *chATLAS: An AI Assistant for the ATLAS Collaboration*. 2025. URL: https://cds.cern.ch/record/2935252.

[259] F. Rehm et al. *AccGPT - A chatbot for CERN Internal Knowledge*. Accessed 2.7.2025. 2024. URL: https://indico.cern.ch/event/1423858/.

# Acknowledgements

This DPhil would have not been possible if not for the collaboration and discussions with many wonderful colleagues. Thank you to everyone with whom I had the pleasure to work.

Thank you very much Daniela Bortoletto for your guidance, supervision and the opportunity to conduct my research in your group! Your feedback always gave me new food for thought, and helped me navigate my research and my DPhil.

I would like to express my gratitude to John Alison and Guy Wilkinson for agreeing to be my examiners and for taking the time to read my manuscript.

My journey in high energy physics began many years ago when I met Carlos. I am indebted to you for since becoming my inspiring mentor in HEP and beyond, and for always exposing me to new challenges. I have truly learnt a lot from you. ¡Muchas gracias!

I also cannot overstate how grateful I am for you, Brian. You are a gifted teacher, and your expertise and skills are always an inspiration to me. Many thanks for being an amazing supervisor and for always finding time - even if late at night. Vielen herzlichen Dank!

Mille grazie Valerio for your guidance and supervision - not only through the “analysis-side” of ATLAS - and for your suggestion to research Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ channel!

A big thank you to all my colleagues in the $HH \to b\bar{b}\tau^+\tau^-$ analysis, ATLAS technical coordination, the Oxford ATLAS group and the EP-ADE-TK group. Thank you Kim, Julieta and Sue for your help with all the organisational aspects of this DPhil.

Danke, Thank you, Bedankt, Grazie, Gracias, Merci, Tack and Xièxiè ... to my friends and my family for your support and advice during this journey!

And finally, Margaret, thank you for everything and beyond.