DISSERTATION

$H \to \tau\tau$ cross section measurements using machine learning and its interpretation as global couplings with the ATLAS detector

Frank Sauerburger

Albert-Ludwigs-Universität Freiburg

$H \to \tau\tau$ cross section measurements using machine learning and its interpretation as global couplings with the ATLAS detector

vorgelegt von
Frank Sauerburger

Dissertation
zur Erlangung des Doktorgrades der
Fakultät für Mathematik und Physik der
Albert-Ludwigs-Universität Freiburg

17. Mai 2022

Albert-Ludwigs-Universität Freiburg

Dekan:	Prof. Dr. Michael Thoss
Betreuer:	Prof. Dr. Karl Jakobs
Erstgutachter:	Prof. Dr. Karl Jakobs
Zweitgutachter:	Prof. Dr. Gregor Herten
Datum der mündlichen Prüfung:	28. Juli 2022
1. Prüfer:	Prof. Dr. Karl Jakobs
2. Prüfer:	apl. Prof. Dr. Horst Fischer
3. Prüfer:	Prof. Dr. Stefan Dittmaier
Version:	2.0.0
Datum:	2022-10-04

Contents

Abstract

The Brout-Englert-Higgs mechanism provides a way for particles in the Standard Model to acquire mass. Specifically for fermions, Yukawa interactions with the Higgs boson are introduced to create the mass terms after spontaneous symmetry breaking. This thesis describes the measurement of the Higgs boson production cross section with the $H \to \tau\tau$ decay channel with the ATLAS detector. The analysis uses events from proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The analysis is performed on the full Run 2 dataset corresponding to an integrated luminosity of $139\,\text{fb}^{-1}$. Cross section measurements are presented for the total cross section, for the gluon fusion and vector boson fusion production modes, for Higgs boson production in association with a vector boson or a pair of top quarks, and for nine fiducial regions of phase space in the Simplified Template Cross Section framework. The precision of the analysis is enhanced using Boosted Decision Trees. As an alternative, a deep multiclass neural network is trained, optimized, and evaluated. The multiclass approach achieves an approximately 8 % reduction in the uncertainty of the vector boson fusion cross section measurement. The measured product of total cross section and branching ratio is $0.93^{+0.13}_{-0.12}$ times the Standard Model expectation. A partially unblinded fit is used to validate the fit model and the assumptions on correlations between parameters in the fit.

The results of the $H \to \tau\tau$ analysis are combined with the results from other Higgs boson channels. The combination is interpreted with global coupling parameters in the κ-framework. Several models have been evaluated with resolved or effective loop-induced couplings, and assumptions on contributions from beyond the Standard Model to the total Higgs boson decay widths. The improvements in the $H \to \tau\tau$ analysis compared to previous iterations are directly reflected in a reduction of the uncertainties on the coupling modifiers. All observed couplings are in agreement with the Standard Model.

Zusammenfassung

Der Brout-Englert-Higgs-Mechanismus bietet den Teilchen im Standardmodell eine Möglichkeit Masse zu erhalten. Im Speziellen für Fermionen werden Yukawa-Kopplungen eingeführt, so dass nach der spontanen Symmetriebrechung Massenterme entstehen. Diese Dissertation beschreibt die Messung des Higgs-Boson-Produktionswirkungsquerschnitts im $H \to \tau\tau$ Zerfallskanal mit dem ATLAS-Detektor. Die Analyse verwendet Ereignisse von Proton–Proton-Kollisionen mit einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$. Die Analyse ist auf dem vollen Datensatz der zweiten Datennahmenperiode des Large Hadron Colliders (Run 2) durchgeführt. Der Datensatz entspricht einer integrierten Luminosität von $139\,\text{fb}^{-1}$. Ergebnisse für den totalen Wirkungsquerschnitt, für die Higgs-Boson-Produktionskanäle Gluon-Fusion and Vektorboson-Fusion, für Higgs-Boson-Produktion in Assoziation mit einem Vektorboson oder einem Paar Top-Quarks, und für neun fiduzielle Phasenraumregionen im Simplifed-Template-Cross-Section-Framework werden präsentiert. Die Präzision der Analyse wird mit Hilfe von Boosted Decision Trees erhöht. Als Alternative wird ein tiefes, mehrklassen neuronales Netzwerk trainiert, optimiert und ausgewertet. Der Mehrklassenansatz erreicht eine Reduzierung der Unsicherheiten des Vektorboson-Fusion-Wirkungsquerschnitts um etwa 8 %. Das gemessene Produkt aus Wirkungsquerschnitt und Verzweigungsverhältnis ist $0.93^{+0.13}_{-0.12}$ der Vorhersage des Standardmodells. Ein nur teilweise blinder Fit wird verwendet, um das Fit-Modell und Annahmen zu Korrelationen zwischen Fit-Parametern zu überprüfen.

Die Ergebnisse der $H \to \tau\tau$ Analyse werden mit den Ergebnissen anderer Higgs-Boson-Kanälen kombiniert. Die Kombination wird mit globalen Kopplungsparametern im κ-Framework interpretiert. Verschiedene Modelle mit aufgelösten oder effektiven schleifeninduzierten Kopplungen, und mit Annahmen zu Beiträgen jenseits des Standardmodells zur totalen Higgs-Boson-Zerfallsbreite wurden untersucht. Die Steigerung der Sensitivität der $H \to \tau\tau$ Analyse im Vergleich zu vorherigen Iterationen spiegelt sich direkt in einer Reduktion der Unsicherheiten der Kopplungsparameter wieder. Alle beobachteten Kopplungen stimmen mit dem Standardmodell überein.

Contributions

The work presented in this thesis was performed within the ATLAS Collaboration, a collaboration of approximately 3000 scientific authors. The author of this thesis has contributed significantly to the following analyses and publications in the following ways:

- Development of machine learning approaches in the VBF categories, development and validation of the full analysis chain in all channels, and validation of the likelihood fit with blinded and partially unblinded studies:

  ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2022) 175, arXiv:2201.08269

- Optimization of the SCT noise threshold to improve the tracking performance for Run 2:

  ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker in LHC Run 2*, JINST **17** (2022) P01013, arXiv:2109.02591

- Development of the BDT classifier used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel to achieve the required signal purity:

  ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **805** (2020) 135426, arXiv:2002.05315

- Core analyzer of the $\tau_{\text{lep}}\tau_{\text{had}}$ channel of the analysis and development of BDT classifiers:

  ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072001, arXiv:1811.08856

The author has contributed significantly to the following conference notes:

- Implementation, validation, and running of the κ-combination fit model:

  ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the*

ATLAS experiment, ATLAS-CONF-2021-053, superseded by Nature **607**, 52–59 (2022)

- Development of machine learning approaches in the VBF categories, development and validation of the full analysis chain in all channels, and validation of the likelihood fit with blinded and partially unblinded studies:

  ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2021-044, superseded by JHEP **08** (2022) 175

- Development of the BDT classifier used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel to achieve the required signal purity:

  ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2019-050, superseded by Phys. Lett. B **805** (2020) 135426

- Core analyzer of the $\tau_{\text{lep}}\tau_{\text{had}}$ channel of the analysis and development of BDT classifiers:

  ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying to a pair of tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2018-021, superseded by Phys. Rev. D **99** (2019) 072001

The author of the thesis has submitted the following proceedings accompanying conference contributions:

- F. Sauerburger, ATLAS Collaboration, *Measurement of the Higgs boson coupling to τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC*, PoS **EPS-HEP2021** (2022) 573

- F. Sauerburger, ATLAS Collaboration, *Measurement of the Higgs boson coupling to τ-leptons in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector at the LHC*, PoS **ICHEP2020** (2021) 106

During the work on this thesis, the author has contributed to or realized the following software projects, libraries, and online resources:

- CAF: the Common Analysis Framework, an analysis framework used by several analyses within the ATLAS Collaboration

- uhepp: a collaborative storage format for plots and their visualization to preview, modify, and share typical plots in high-energy physics

- FreeForestML: a Python framework to facilitate machine learning in high-energy physics contexts

- pylorentz: a Python library to perform common operations with 4-momenta

- howlargeisthelhc.com and howlargeisthefcc.com: an interactive map overlaid with the outline of current and future particle accelerators, allowing the user to compare its size to any city
- lhc-xsecs.org: an interactive overview plot of cross sections of common processes at hadron colliders as a function of the center-of-mass energy
- uhepp.org: an interactive online hub to preview and share high-energy physics plots in uhepp format

Notation and conventions

Indices with Greek letters refer to the four spacetime dimensions. A sum over repeated upper and lower Greek indices is implied (Einstein summation convention). The spacetime metric $\mathrm{diag}(1, -1, -1, -1)$ is used. The conjugate transpose of ϕ is denoted as $\phi^\dagger = (\psi^*)^T$. The adjoint spinor of ψ is denoted as $\bar{\psi} = \psi^\dagger \gamma^0$, where γ^i are the Gamma matrices. Multi-dimensional vectors except 4-vectors are written in boldface (e.g., $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$). The non-bold version of a vectorial quantity without an index refers to its absolute value (e.g., $E_{\mathrm{T}}^{\mathrm{miss}}$).

The natural logarithm of x is denoted by $\log(x)$. The notation $O(x)$ is used to express that a quantity has the same order of magnitude as x. For uncertainty treatments, $a \oplus b$ denotes the addition in quadrature $\sqrt{a^2 + b^2}$ assuming statistically independent uncertainties.

Frequently, equations and quantities are expressed in natural units such that velocities are measured as fractions of the speed of light ($c = 1$) and angular momenta as multiples of Planck's constant ($\hbar = 1$). However, for example, dimensions of detector components are expressed in SI units. Following the guidelines set in Reference [1], products of two values use the multiplication sign (e.g., 2×3). Axis labels or table headers show the quantity divided by its unit (e.g., m/GeV) unless a figure from an ATLAS publication is used that uses a different convention.

CHAPTER 1

Introduction

L'essentiel est invisible pour les yeux.

— Antoine de Saint-Exupéry, *Le Petit Prince*, 1943

The discovery of a scalar boson in 2012 by the ATLAS and CMS collaborations at the Large Hadron Collider at CERN was a critical milestone for particle physics [2, 3]. The initial data showed that the new boson is consistent with the long-sought for Higgs boson. The underlaying mechanism was introduced by Brout, Englert, Higgs, and others in 1964 [4–6] to provide a gauge-invariant way to include mass terms for the heavy gauge bosons in the Lagrangian of the Standard Model of particle physics. The same field can be used to provide mass terms also for the fermions. The Higgs boson was the last particle required by the Standard Model that scientists had not been able to observe.

The Large Hadron Collider is not the first collider where scientists have looked for the Higgs boson. Although the model is only complete with the Higgs boson and has various inconsistencies without it, the mass of the Higgs boson is a free parameter of the theory that needs to be determined by experiment. Previous experiments have placed limits on the possible mass range where the boson could hide. Experiments have looked for signatures of a boson in the low-mass range up to 114 GeV and placed lower limits on the mass of the Higgs boson. Theoretical considerations showed that the upper limit of the mass should be around 1 TeV, favoring a light Higgs boson. If it turned out that there is no Higgs boson in that mass range accessible at the Large Hadron Collider, deviations had to become apparent in vector boson scattering, potentially leading the way to a different Standard Model. With the measurement of the Higgs boson mass of around 125 GeV, one of the last free parameters of the Standard Model had been determined, opening the window to a new area of particle physics.

In the Standard Model, the Higgs field is used to generate mass terms for the heavy gauge bosons. We can use the same field and introduce an ad hoc coupling between the Higgs field and the fermions. After spontaneous symmetry breaking, we are left

with an interaction between fermions and the Higgs field, and a mass term for the fermions. Coincidentally, the coupling strength is proportional to the particle's mass. This property of the Standard Model, as well as its experimental confirmation by the ATLAS Collaboration, is illustrated in Figure 1.1. The coupling structure is assumed to apply to all massive particles in the Standard Model. Measuring the coupling strengths for different particles with known masses allows testing this construction experimentally. Testing this relationship is central to this thesis, firstly through measuring the Higgs boson production cross section in processes where the Higgs boson decays to tau leptons, and secondly through the combination of measurements in various different production and decay channels to extract global coupling parameters of the Higgs boson to massive particles in the Standard Model.

The decay to tau leptons is an interesting channel since tau leptons are the heaviest known leptons and one can expect a relatively large decay rate of the Higgs boson to a pair of tau leptons. Specifically, this measurement gives important insight into the lepton Yukawa sector.

The thesis is structured as follows. Chapter 2 introduces the theoretical foundation of the Standard Model and the Brout-Englert-Higgs mechanism, physics at hadron colliders, and machine-learning techniques. The Large Hadron Collider, the ATLAS detector, and particle reconstruction methods are described in Chapter 3. Chapter 4 presents an optimization study to increase the tracking performance of the SemiConductor Tracker with respect to the treatment of detector noise. The decay of the Higgs boson to a pair of tau leptons is studied in Chapter 5. The chapter shows how machine learning can be applied in the measurement and presents a measurement of various cross sections in the decay channel. Chapter 6 goes beyond the decay to tau leptons and describes the combination with other Higgs boson production and decay channels and its interpretation in the κ-framework. The combination of multiple decay channels into a single measurement has the benefit of a larger dataset to reduce statistical fluctuations, constrain systematic uncertainties, and therefore measure properties of the Higgs boson more precisely. Additionally, it is possible to measure effects that emerge as differences between the decay channels or as coherent deviations for specific Higgs boson production modes. The combination in the κ-framework also allows placing limits on decays beyond the Standard Model, such as limits on the branching ratio of decays to invisible particles.

The goal of particle physics is to explore and describe the nature of matter and its interactions. This profound task has implications for the evolution and the ultimate fate of the universe. Although we have now observed all the particles of the Standard Model, it cannot be the final theory of matter and its interactions. The study of particle interactions continues. Although the Standard Model is one of the most stringently tested theories, there are phenomena that currently cannot be explained by the Standard Model, like Dark Matter, Dark Energy, or gravity.

The discovery of the Higgs boson was just the beginning. It is now possible to test precise predictions in the Higgs sector. A deviation of experimental data from theoretical predictions might lead to new physics. The physicist David Griffiths wrote in his famous textbook on particle physics, "Only when (and if) the Higgs particle is actually *found* will it be possible to confirm all this [the Brout-Englert-Higgs mechanism] empirically" [7]. This time is now.

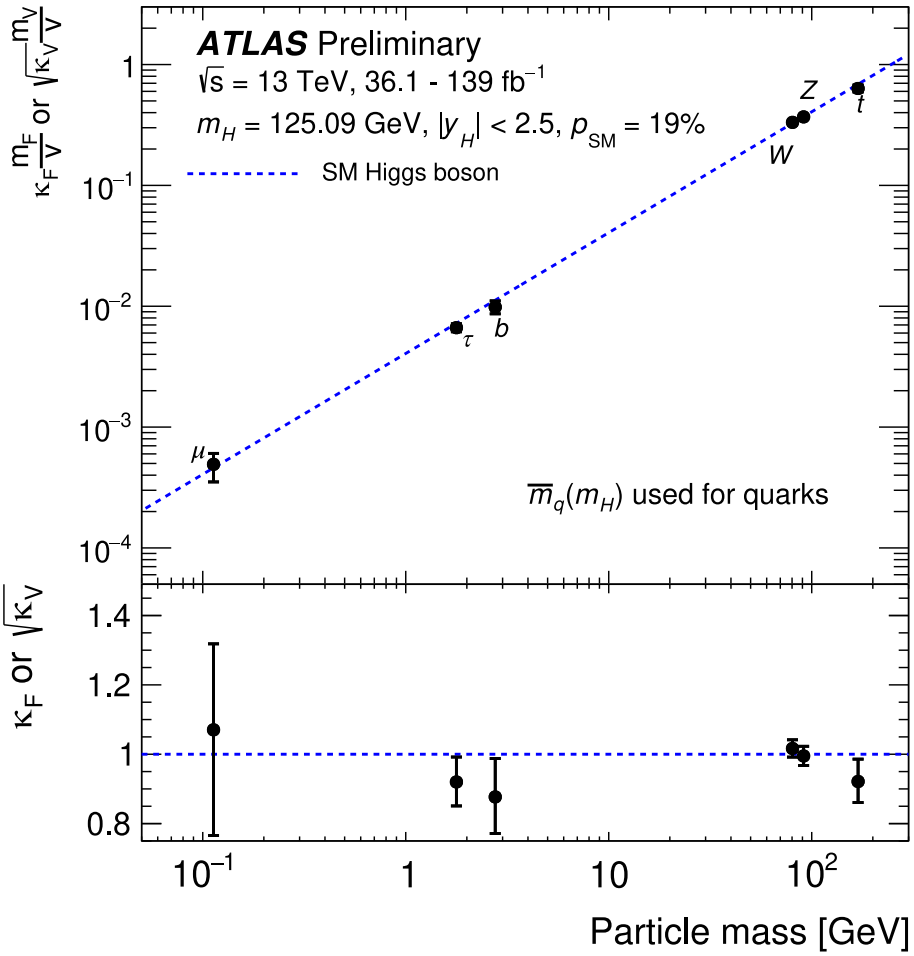


Figure 1.1.: Illustration of the coupling strength between various particles and the Higgs boson as a function of the particles' masses [8]. The blue dashed line indicates the Standard Model prediction. The bottom panel shows the deviation of the measurement from the Standard Model in the κ-framework. Overall the measurement is consistent with the Standard Model at $p_{\text{SM}} = 19\,\%$. This combination includes the results of the $H \to \tau\tau$ measurement presented in Chapter 5. The combination in the κ-framework itself is detailed in Chapter 6.

CHAPTER 2

Theory

There is no difference between Time and any of the three dimensions of Space except that our consciousness moves along it.

— H. G. Wells, *The Time machine*, 1895

The analyses presented in this thesis measure properties of the Standard Model of particle physics (SM) and, at the same time, test the validity of the SM. The SM is formulated as a quantum field theory. This chapter introduces the SM, its relevant theoretical background, and phenomenology of proton–proton collisions. This introduction can only be an overview. The chapter is based on References [7, 9–13]. References to more detailed resources are given in the respective sections.

2.1. Fundamental particle physics

Particle physics aims to describe fundamental building blocks of matter and their interactions at a microscopic level. At first glance, this task is similar to Mendeleev's periodic table that arranges the building blocks of matter at the level of atoms. The structure of the periodic table indicates that the atom is not a fundamental *atomic* particle but is instead built from smaller constituents. The Rutherford experiment delivered the experimental evidence: atoms are built from electrons and a nucleus later found to consist of protons and neutrons. Deep inelastic scattering showed that even protons and neutrons are not elementary particles. Today's SM lists quarks and leptons as the fundamental, point-like building blocks of matter, vector bosons as mediators of fundamental forces, and the Higgs boson as a way through which particles acquire their mass. The following sections introduce the mathematical framework to describe particles and their interactions as fields and introduce the SM with its crucial component: the Higgs boson and the Brout-Englert-Higgs mechanism.

2.1.1. Elementary particles

The matter around us is composed of atoms which in turn are built from electrons (e^-) and up (u) and down (d) quarks. The SM arranges fundamental matter particles in generations and groups them into leptons and quarks. The electron together with the electron neutrino form the first generation of leptons. The aforementioned up and down quarks form the first generation of quarks. For leptons and quarks, the first generation is accompanied by a second and third generation. The particles in these generations have the same properties as their sibling particles in the first generation except that they are heavier[1]. The massive particles in the second and third generations are unstable and eventually decay to particles of the first generation. First-generation particles are stable. The heavier siblings of the up quark are the charm (c) and the top (t) quark. The heavier siblings of the down quark are the strange (s) and bottom (b) quark. The heavier siblings of the electron are the muon (μ) and the tau lepton (τ). In each generation, the massive lepton is accompanied by a neutrino (ν) that we assume to be massless. Experimental evidence shows that neutrinos have a small mass, but to simplify the presentation, we postpone the issue of neutrino masses until Section 2.3.4. The quarks and leptons are aptly summarized as generations:

$$\begin{array}{lccc} \text{Quarks} & \begin{pmatrix} u \\ d \end{pmatrix} & \begin{pmatrix} c \\ s \end{pmatrix} & \begin{pmatrix} t \\ b \end{pmatrix} \\ \text{Leptons} & \begin{pmatrix} \nu_e \\ e^- \end{pmatrix} & \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} & \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \end{array}$$

The fact that each particle is accompanied by an anti-particle with the same mass but opposite charge doubles the number of particles in the SM.[2]

The matter particles have an internal angular momentum, the *spin*, of $\frac{1}{2}\hbar$. All half-integer spin particles are referred to as *fermions*. The SM includes the group of spin-1 particles termed *gauge bosons* or *vector bosons*. The exchange of a vector boson describes an interaction between particles. Of the four fundamental forces, the electromagnetic force, the weak force, and the strong force are described in the SM. Only gravity is not included in the SM. The effects of gravitational attraction are too small, by many orders of magnitude, to have a measurable effect on particle interactions. On the other hand, it is difficult to built a consistent theory including gravity. The kinematic properties of SM interactions depend on the mass of particles. A force acting on two particles is modeled as the exchange of gauge bosons between the two participating particles. The electromagnetic force is characterised by the exchange of photons, the strong force is mediated by gluons, and the weak force is mediated the by $W^\pm$ and Z bosons. Gluons come in eight different kinds. The vector bosons and their mediated forces are summarized in Table 2.1.

For each force, we assign a corresponding charge. The charge of the electromagnetic force is the electric charge Q. Only electromagnetically charged particles, particles with $Q \neq 0$, participate in the electromagnetic interaction. The neutrinos ν, for example, do

[1] Neutrinos take a special role since their masses are not yet experimentally determined.

[2] Again, neutrinos, electrically neutral, represent a special case concerning their anti-particles. Depending on the nature of neutrinos, they could be their own anti-particle.

Table 2.1.: Overview of vector bosons, their mass, the force they mediate and the corresponding charge. Masses for the weak gauge bosons are given according to Reference [14]. Isospin and hypercharge refer to weak isospin and weak hypercharge.

Vector boson	Mass / GeV	Mediated force	Coupled charge
Gluon g	0	Strong force	Color charge
Photon γ	0	Electromagnetic force	Electric charge
$W^{\pm}$	80.379 ± 0.012	Weak force	Isospin
Z	91.1876 ± 0.0021	Weak force	Isospin and hypercharge

not carry an electric charge. Therefore they do not interact electromagnetically and do neither emit nor absorb photons. The fundamental electromagnetic interactions are described by Quantum Electrodynamics (QED).

For the strong force, instead of having positive or negative charges, we assign three different *colors*: red, green, and blue. The *color charge* should not be confused with the actual optical color of particles. The model of color charge builds upon our intuition of additive color mixing. For example, if we combine one unit of red, green, and blue, we end up with a colorless (or white) combination. An essential feature of the color charge model is that we only observe colorless combinations. Quarks carry one unit of color charge (red, green, or blue), while anti-quarks carry one unit of *anti-color* charge (anti-red, anti-green, or anti-blue). Unlike the electrically neutral photon, the gluon, the mediator of the strong force, carries one unit of color and one unit of anti-color. Following group theoretical considerations, the nine possible color combinations of the gluon split into an octet of colored states and a colorless singlet. Experimental data show that only the octet is realized in nature. Since gluons carry color charge, gluons can couple to each other leading to a drastically different structure of the interaction compared to electromagnetism. The interactions with color charge are described by Quantum Chromodynamics (QCD).

Quarks, charged leptons, and neutrinos participate in the weak force. The charges for the weak force originate from the weak hypercharge and the weak isospin. However, these charges *mix* with the electromagnetic charge, and the discussion of electroweak unification is postponed until Section 2.1.2.

The final ingredient in the SM is the Higgs boson, a scalar particle, i.e., a spin-0 particle. The Higgs boson is the most recently discovered particle of the SM, with its experimental observation in 2012 [2, 3]. Its theoretical foundation is attributed to Anderson [15], Englert, Brout [4], Higgs [5, 6, 16], Guralnik, Hagen, and Kibble [17, 18], and was first published in 1963. The role of the Higgs boson in the SM is to provide a mechanism through which particles in the SM can acquire mass. A priori, it is impossible to construct a useful and mathematically consistent (that is renormalizable) theory with massive particles. The Brout-Englert-Higgs mechanism avoids these issues. Initially, we take all particles to be massless. The interaction of particles with the Higgs field allows these particles to acquire mass by modifying the free propagation of the particle through space. The truly massless particles of the SM, the photon, the gluons,

and the neutrinos, do not interact directly with the Higgs field.

2.1.2. Quantum field theory

The SM is formulated as a Quantum Field Theory (QFT). A field maps every point in spacetime to a number or a vector. For example, a temperature field yields the temperature for every point in spacetime. The fields in the SM span the whole universe. The SM contains a collection of different fields; for example, there is a field for up quarks, for electrons, and for the Higgs boson. Excitations in the fields can be seen as particles. The dynamics are fully captured by the Lagrange density $\mathcal{L}$, which compiles all fields into a single model. The fields need to obey a set of conditions. On a fundamental level, the fields need to be Lorentz invariant. Formally, we require the Lagrange density to be invariant under Lorentz boosts. Let us consider the Lagrange density of the fermion field ψ

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi \tag{2.1}$$

where γ^μ refers to the Gamma matrices. The dynamics of the model can be derived from the Lagrange density with the Euler-Lagrange equation. Applying the Euler-Lagrange equation yields the Klein-Gordon equation in the case of a scalar field, the Proca equation in the case of a vector field, and the Dirac equation in the case of a fermion field. The Lagrangian of Equation (2.1) describes massless, free (i.e. non-interacting) fermions.

A specific set of symmetries characterizes the SM. Particles are characterized by their properties under transformation of discrete symmetries, like charge conjugation C or parity conjugation P. At this point, we can introduce transformations $U(\alpha)$ of the fields as elements of a continuous symmetry group and require the equations of motion to be invariant. The invariance is trivially satisfied if the Lagrangian is invariant under these transformations, however, strictly speaking, this is not a mandatory condition. A similar step is known in classical electrodynamics as gauge invariance. In classical electrodynamics, we require invariance under *global* transformations. That means we apply the same transformation at every point in space-time. However, in the SM, we require *local* gauge invariance, i.e., $\alpha = \alpha(x)$. The *local* gauge invariance under

$$\psi \to \psi' = U(\alpha(x))\psi \tag{2.2}$$

is introduced ad hoc and could be taken as a new principle of physics. This principle offers many advantages. What concerns the Brout-Englert-Higgs mechanism, we will see that local gauge invariance is also employed after spontaneous symmetry breaking.

In the SM, we require that the Lagrangian or the equations of motion are invariant under the product symmetry $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ of special unitary groups (SU) and the unitary group (U). The Lagrangian of Equation (2.1) is not invariant under *local* gauge transformations. We can restore the invariance under a local symmetry transformation $U(\boldsymbol{\alpha}) = \exp\left(i \sum_a \alpha_a T_a\right)$ with generators T_a by introducing additional vector fields G^a so that the new covariant derivative $\mathcal{D}\psi$ of the fields transform covariantly,

$$\mathcal{D}\psi \to U(\alpha_a(x))\mathcal{D}\psi \tag{2.3}$$

$$\text{with} \quad \mathcal{D}_\mu = \partial_\mu + \sum_a igT_a G^a_\mu \tag{2.4}$$

which means the derivative transforms like the fields ψ. With this construction, the additional vector fields G^a transform as

$$G^a_\mu \to G^a_\mu - \frac{1}{g}\partial_\mu \alpha_a - \sum_{b,c} f_{abc}\alpha_b G^c_\mu. \tag{2.5}$$

The parameters f_{abc} are the structure constants of the symmetry group and fully specify the local[3] structure of the group. They are related to the generators of the symmetry group by $[T_a, T_b] = i\sum_c f_{abc}T_c$. Remarkably, this procedure is possible for abelian groups, such as $U(1)$, for which the structure constants f_{abc} vanish, and for non-abelian groups, such as $SU(2)$ and $SU(3)$. When we write out all the terms, we see that the introduction of the covariant derivative added interaction terms between the fermion fields and the vector fields. Connecting this to the structure of the SM, we see that the vector fields G^a describe the vector bosons that mediate the forces.

The initial Lagrangian and the derived equations of motion were not invariant under *local* gauge transformations, but by carefully adding the interaction terms we can restore the invariance. In a way, the vector fields are required to "reconcile" [9] the local transformation or phase α of the fermion fields.

So far, we have used the local gauge invariance as a tool to add interaction terms to the Lagrangian. In 1971, 't Hooft showed that imposing local gauge invariance yields renormalizable theories [19]. In renormalizable theories, but also in non-renormalizable theories, we need experimental input to renormalize (i.e., cancel) divergent terms. In the SM, renormalization leads to a redefinition of the electron charge and the dependence of the coupling constants on the energy scale. The charge itself remains an experimental parameter. In non-renormalizable theories, an infinite amount of experimental input would be required, with new divergences appearing at each order in perturbation theory. The description of renormalization is beyond the scope of this introduction and the reader is referred to References [12, 13].

The symmetries of the SM and its structure constants define the structure of the possible interactions. The $SU(3)_c$ group acts on the color states of the quark. We use the irreducible representation of $SU(3)_c$ generated by the Gell-Mann matrices λ_i with $i = 1, 2, \ldots 8$. The requirement of local invariance under $SU(3)_c$ leads to the addition of eight gluon fields. The gluon octet is directly related to the eight generators of the irreducible representation.

With the Glashow-Salam-Weinberg model, the electromagnetic and weak forces are regarded as two manifestations of the same force characterized by the product symmetry group $SU(2)_L \otimes U(1)_Y$. The indices refer to *left* chirality (or handedness) eigenstates of weak isospin and weak hypercharge Y, respectively, and do not coincide directly with the electromagnetic and the weak forces' generators but mix to create these two forces. The weak isospin and weak hypercharge of fermions is listed in Table 2.2. Left chirality states form a doublet χ_L, e.g., $(\nu\ e)^T_L$ for first-generation leptons, that is affected by $SU(2)_L$ transformations. By construction, the right-handed singlet state is not affected by this symmetry. On the other hand, all fermions acquire a phase under $U(1)_Y$ transformations. Imposing invariance under local $SU(2)_L \otimes U(1)_Y$ transformations

[3] Local in this context refers to a small region in the Lie algebra of the group and not the gauge transformation.

Table 2.2.: Summary of fermions and their electroweak charges. For neutrinos ν, the subscript L has been omitted. Isospin and hypercharge refer to weak isospin and weak hypercharge. It should be noted that left and right-handed chirality states are not distinct particles since chirality is not conserved for massive particles.

Particles	Electrial charge Q	Isospin I	Isospin I^3	Hypercharge Y
ν_e, ν_μ, ν_τ	0	$\frac{1}{2}$	$\frac{1}{2}$	-1
e_L^-, μ_L^-, τ_L^-	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1
e_R^-, μ_R^-, τ_R^-	-1	0	0	-2
u_L, c_L, t_L	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$
d_L, s_L, b_L	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$
u_R, c_R, t_R	$\frac{2}{3}$	0	0	$\frac{4}{3}$
d_R, s_R, b_R	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$

forces us to add the triplet vector boson W_i and the singlet B boson. The additional term in the Lagrangian

$$\mathcal{L}_{\text{weak}} = -\frac{i}{2}\left[g\sum_i(\bar{\chi}_L\gamma_\mu\sigma_i\chi_L)W_i^\mu + g'(\bar{\psi}\gamma_\mu Y\psi)B^\mu\right] \tag{2.6}$$

describes the interaction of the additional bosons and the fermions where σ_i are the Pauli matrices and Y is the hypercharge operator. The hypercharge is related to the electrical charge Q and the weak isospin I^3 via the Gell-Mann–Nishijima formula

$$Y = 2Q - 2I^3. \tag{2.7}$$

The fields W_i and B are not the fields of physical particles described in the SM. The first two components form the $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$ bosons. The remaining fields W_μ^3 and B_μ mix and yield the photon field A_μ and the neutral weak bosons field Z_μ:

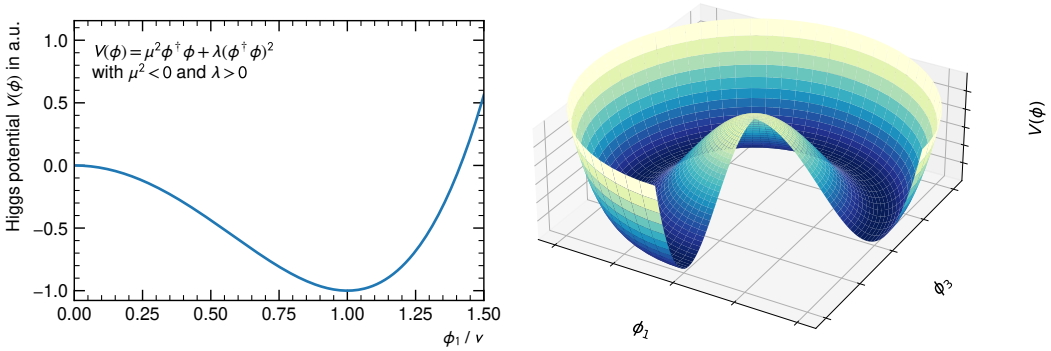
$$A_\mu = +B_\mu\cos\theta_w + W_\mu^3\sin\theta_w \tag{2.8}$$

$$Z_\mu = -B_\mu\sin\theta_w + W_\mu^3\cos\theta_w. \tag{2.9}$$

The weak mixing angle θ_w relates the coupling constants $\tan\theta_w = \frac{g'}{g}$.

2.1.3. The Brout-Englert-Higgs mechanism

So far, we have added interaction terms to the free Lagrangian of Equation (2.1). From experiments, we know that the $W^\pm$ and Z bosons are heavy particles with masses of $m_W \approx 80.4\,\text{GeV}$ and $m_Z \approx 91.2\,\text{GeV}$, respectively, and quarks, electrons, muons, and tau leptons are also not massless. However, gauge invariance prevents us from adding mass terms for the fermions and the vector bosons. For example, a mass term for the



(a) (b)

Figure 2.1.: Visualization of the Higgs potential $V(\phi)$ for a single real component ϕ_1 (a), and for two real components ϕ_1 and ϕ_3 of the Higgs field with a cut-out wedge for illustration purposes (b). All other components are assumed to vanish. The figures show the interesting case of $\mu^2 < 0$ with the minimum of the potential at $\phi_1 = v$. The x-axis is normalized to the vacuum expectation value v.

tau lepton involves the right-handed singlet ψ_R and left-handed doublet ψ_L, which breaks invariance under (global α or local $\alpha = \alpha(x)$) $SU(2)_Y$ transformations $U(\alpha)$:

$$-m_\tau \bar{\psi}\psi = -m_\tau \left(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R\right) \rightarrow -m_\tau \left(\bar{\psi}_R U(\alpha) \psi_L + \bar{\psi}_L U^\dagger(\alpha) \psi_R\right) \tag{2.10}$$

The Brout-Englert-Higgs (BEH) mechanism provides a solution to this by introducing a new $SU(2)$ doublet of complex scalar fields

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.11}$$

The relevant terms in the Lagrangian for this field read

$$\mathcal{L}_H = \frac{1}{2}(\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \underbrace{\left(\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2\right)}_{V(\phi)} \tag{2.12}$$

where the first part is the kinetic term for a scalar field ϕ, and the rest refers to its potential $V(\phi)$ with $\lambda > 0$. In the case $\mu^2 > 0$, the Lagrangian would describe a massive scalar boson with mass $m = \sqrt{2\mu}$ but does not help in adding mass terms for the vector boson fields or fermion fields. The additional terms describe self-interaction.

The interesting case arises with $\mu^2 < 0$. The term does not look like a mass term due to the wrong sign. The potential for this case is shown in Figure 2.1. The Lagrangian, and also the potential $V(\phi)$, are invariant under $SU(2)_L$ symmetry transformations. The minimum of the potential does not coincide with vanishing fields ϕ. However, the rules of the SM are derived as perturbations around the ground state. This approach fails if we expand around the metastable vacuum. To restore perturbation theory, we can substitute the field ϕ and expand the Lagrangian around a point that minimizes

the potential. As illustrated in Figure 2.1(b), the potential is not minimized by a single point. There is a manifold of points with $\phi^\dagger\phi = -\frac{\mu^2}{2\lambda} \equiv v^2$ that minimizes the potential. The value of the field ϕ that minimizes its potential is the vacuum expectation value v. We can choose any of these points to parameterize the new field. It is instructive to choose $\phi = \frac{1}{\sqrt{2}}(0\ v)^\mathrm{T}$. In principle, the choice is arbitrary. With our choice, the four new real fields are $\theta_1(x)$, $\theta_2(x)$, $\theta_3(x)$ and $h(x)$. This procedure is referred to as spontaneous symmetry breaking. The field ϕ is related to the new fields by

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1(x) & +i\theta_2(x) \\ v + h(x) & +i\theta_3(x) \end{pmatrix}. \tag{2.13}$$

Our approach of imposing local gauge invariance of the Lagrangian can finally be exploited in the BEH mechanism. Plugging Equation (2.13) into the Lagrangian (2.12) yields two-point terms involving the fields θ_i that cannot be identified with any of the expected SM interactions. The particles associated with θ_i are referred to as Goldstone bosons. Since we required *local* gauge invariance, we can choose a gauge such that the Goldstone bosons vanish. Therefore, without loss of generality, we can express the field ϕ in terms of a single field $h(x)$ and plug

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.14}$$

into the Lagrangian (2.12). With this choice of ϕ, the only remaining field has weak isospin $I^3 = -\frac{1}{2}$ and hypercharge $Y = 1$, however is electrically neutral $Q = 0$. By expanding the Lagrangian around a point that minimizes the potential, we arrive at a state where we can use a perturbative approach again to derive the dynamics of the model. It looks like the resulting Lagrangian is not invariant under gauge transformations anymore. However, the symmetry is still inherent in the Lagrangian. The symmetry is said to be "hidden" [7] by expressing the Lagrangian in terms of $h(x)$. We could always reintroduce the original parameterization and restore the symmetry.

The point is to have a way to introduce mass terms for gauge bosons, which are otherwise forbidden by gauge invariance. We can achieve this goal by adding an interaction between the vector bosons and the Higgs field. As before, the interaction can be expressed by replacing the derivative ∂_μ with the covariant derivative $\mathcal{D}_\mu$ such that the Lagrangian before breaking the symmetry reads

$$\mathcal{L}_{m_V} = \left| \left(\partial_\mu + i\frac{g}{2} \sum_i \tau_i W_{\mu,i} + i\frac{g'}{2} B_\mu \right) \phi \right|^2 - V(\phi) \tag{2.15}$$

with $|\mathcal{D}_\mu\phi|^2$ as a short-hand notation for $(\mathcal{D}_\mu\phi)^\dagger(\mathcal{D}^\mu\phi)$. Breaking the symmetry spontaneously by using Equation (2.14) and transforming W^3_μ and B_μ into their physical basis Z_μ and A_μ yields the relevant terms in the Lagrangian

$$\mathcal{L}'_{m_V} = 2 \cdot \underbrace{\frac{1}{8} v^2 g^2}_{\frac{1}{2}m_W^2} W^+_\mu W^{-,\mu} + \underbrace{\frac{1}{8} v^2 (g^2 + g'^2)}_{\frac{1}{2}m_Z^2} Z_\mu Z^\mu + \underbrace{m_A^2}_{=0} A_\mu A^\mu + \cdots \tag{2.16}$$

We can identify the terms as the long sought-for mass terms for the vector bosons. Due to the weak isospin and hypercharge of ϕ, this mechanism breaks $SU(2)_L \otimes U(1)_Y$ symmetry and generates mass terms for the W and Z bosons. Our choice of ϕ is electrically neutral, thus the symmetry $U(1)_{\text{em}}$ generated by $Q = I^3 + \frac{Y}{2}$ is unbroken and the photon field A_μ remains massless. With this approach, also gluons are still massless. We did not break the $SU(3)_c$ symmetry. Therefore we did not generate mass terms for the gluons. Additionally, we did break $SU(2)_L \otimes U(1)_Y$ in such a way that only the $W^\pm$ and Z bosons acquire mass, but the photon remains massless. The heavy gauge bosons acquire a third polarization degree of freedom. These degrees of freedom originate from the Goldstone bosons. The heavy gauge bosons gain this degree of freedom by *eating* the Goldstone bosons.

Besides the mass terms listed in Equation (2.16), we get interaction terms between the vector bosons and the Higgs boson of the form VVh. The coupling strength associated with this interaction is proportional to the square of the boson's mass. Additionally, multiplying out all terms in (2.15) produces interaction terms of the form VVh^2 and Higgs boson self-interaction h^3 and h^4. The latter depend on the shape of the Higgs potential.

The proportionality between coupling and vector boson mass is a key feature of the theory. An experimental test of this property is presented in Chapter 6.

The final missing piece is the mass term for the fermions. For this, we can use the same Higgs field ϕ that we have used for the vector bosons and introduce an ad hoc Yukawa coupling between fermions and Higgs fields [20]. For each lepton generation ℓ, the additional terms in the Lagrangian read

$$\mathcal{L}_{m_\ell} = -Y_\ell \left[(\bar{\nu_\ell}\ \bar{\ell})_L \phi \ell_R + \bar{\ell}_R \phi^\dagger \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \right] \tag{2.17}$$

which satisfy the imposed gauge invariance. In contrast to the non-invariant mass term in Equation (2.10), the structure of ϕ has the correct transformation properties such that the combination with ψ_L is gauge-invariant.

Again, applying Equation (2.14) gives two additional types of terms. The terms for the tau lepton field ψ_τ are

$$\mathcal{L}'_{m_\tau} = -\frac{Y_\tau}{\sqrt{2}} v \bar{\psi}_\tau \psi_\tau - \frac{Y_\tau}{\sqrt{2}} \bar{\psi}_\tau \psi_\tau h \equiv -m_\tau \bar{\psi}_\tau \psi_\tau - \frac{m_\tau}{v} \bar{\psi}_\tau \psi_\tau h. \tag{2.18}$$

As we did in the case of the vector bosons, we can identify the first term as a mass term for the tau lepton. The coefficient Y_τ and thus m_τ is a free parameter not predicted by the theory. The second term describes the interaction between tau leptons and a Higgs boson. Again, the coupling strength of this interaction is proportional to the mass of the tau lepton. For each fermion, the coupling is proportional to its mass with the same constant of proportionality. The visualization of this relation was introduced in Figure 1.1. The test of this prediction is subject of Chapters 5 and 6.

2.2. Proton–proton collisions

At the Large Hadron Collider (LHC) [21], protons or heavy ions such as lead (Pb) and xenon (Xe) are accelerated in two circular counter-rotating beams. The particles reach

unprecedented energies. For the studies of the Higgs boson presented in this thesis, the collision of two proton beams are considered. The two proton beams are brought to collision within the ATLAS detector. The energy available in the center-of-mass frame during the interaction converts into new particles. We can detect the spray of particles created in the collision and subsequent decays with the ATLAS detector.

Protons are not point-like, elementary particles. With the energy available at the LHC, we can resolve the substructure of the proton, and observe the scattering of point-like, elementary constituents of the proton, the partons: gluons, valance quarks, and sea quarks. The interaction of two protons at the LHC is dominated by QCD effects.

On a superficial level, the goal of proton–proton collisions is to analyse the outcome and count the number of occurrences for events of interest. The observed number of events can then be compared with theoretical prediction. The computation of cross sections for expected processes is paramount.

The computation of cross sections employs a perturbative approach where possible. The Feynman diagrams and the corresponding terms contributing to the cross section can be computed for different powers of the coupling constants. For practical purposes, cross sections can only be computed to a finite order.

Divergent integrations in the SM require the use of renormalization. Renormalization introduces a dependence of the coupling constants on the energy scale μ_R of the interaction, the so-called running of the coupling constants. The running is governed by the renormalization group equation (RGE). For the strong coupling constant α_S, the right-hand side of the RGE

$$\frac{\mathrm{d}\alpha_S(\mu_R)}{\mathrm{d}\mu_R^2} = \beta(\alpha_s(\mu_R)) \tag{2.19}$$

is expanded in powers of α_S. Predictions of the theory are independent of the renormalization scale μ_R as long as all orders of α_S are taken into account. In practice, however, the series is truncated introducing a dependence on the parameter μ_R. Chapter 5 introduces systematic uncertainties related to the renormalization scale to account for missing higher-order terms.

Due to the non-abelian structure of QCD, the coupling strength increases for small energy scales. For large energies, the coupling strength decreases, such that colored particles are asymptotically free. The computation and generation of proton–proton events is usually split into the matrix element (ME) computation and the parton shower (PS) computation. For the matrix element computation, α_S is small enough to employ perturbation theory. The computation of the PS includes initial state and final state radiation. Since perturbation theory is not applicable for low-energy QCD processes, phenomenological models are employed for PS.

2.2.1. Cross sections

The SM predicts the cross sections of elementary reactions. The cross section of a reaction in proton–proton collisions can be obtained by integrating over the internal distribution of partons within the proton. The Parton Distribution Function (PDF) $f_i(x)$ for parton i specifies the probability to find the parton with a certain momentum fraction x inside the proton. PDFs for quarks and gluons are shown in Figure 2.2. The probability

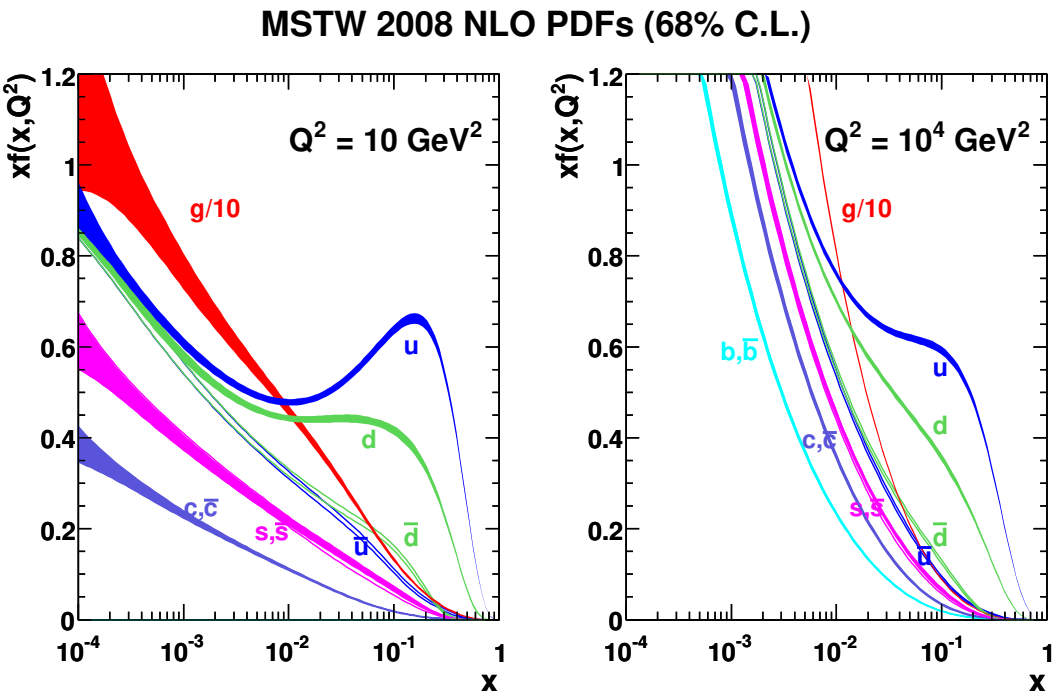


Figure 2.2.: Scaled parton distribution functions for gluons, and valance and sea quarks computed at next-to-leading order. The diagrams show the PDF $f(x, Q^2)$ multiplied by the momentum fraction x of the parton. The PDFs are displayed for two different scales Q^2 [22, 23].

to scatter at a parton with a certain momentum fraction depends on the momentum Q that is transferred in the collision. For collisions with small momentum transfer, e.g. $Q^2 = 10\,\mathrm{GeV}^2$, it is likely to encounter a gluon and a valance quark u or d that carries a large fraction x of the proton momentum. For collisions with large momentum transfer, e.g. $Q^2 = 10^4\,\mathrm{GeV}^2$ the internal structure of the proton and its sea quarks becomes apparent. Intuitively, interactions with large momentum transfers and therefore small de Broglie wavelengths are able to resolve smaller QCD structures inside the proton.

The evolution of the PDF as a function of momentum transfer Q^2 is governed by the DGLAP[4] equation

$$\frac{\mathrm{d}f_i(x, Q^2)}{\mathrm{d}\log Q^2} = \frac{\alpha_S}{2\pi} \int_x^1 \frac{\mathrm{d}z}{z} \sum_j P_{ij}\left(\frac{x}{z}\right) f_j(z, Q^2) \tag{2.20}$$

The right-hand side is a convolution with the splitting functions P_{ij} that describe the probability of parton j with momentum fraction z splitting and yielding parton i with momentum fraction x. The processes responsible for the splitting include gluons radiating a gluon P_{gg}, two quarks pair-produced from a gluon P_{qg}, and quarks radiating a gluon P_{gq} and P_{qq}. The splitting functions are computed up to a specific order in α_S.

[4]Named after Dokshitzer, Gribov, Lipatov, Altarelli, and Parisi

The so-called hadronic cross section σ used for proton–proton collisions as a function of the squared center-of-mass energy s is defined as

$$\begin{aligned}&\sigma(p+p\to c+\mathrm{X};\, s)\\&=\sum_{a,b}\int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b\; f_a(x_a,\mu_F^2)\; f_b(x_b,\mu_F^2)\; \hat{\sigma}(a+b\to c;\,\hat{s},\,\mu_R,\,\mu_F).\end{aligned} \tag{2.21}$$

It is the integration of the partonic cross section $\hat{\sigma}(a+b\to c;\hat{s},\mu_R,\mu_F)$ of the elementary particle reaction over the PDFs of each parton within the protons. The elementary cross section $\hat{\sigma}$ has an explicit dependence on the renormalization scale μ_R, factorization scale μ_F, and the squared center-of-mass energy $\hat{s}=sx_ax_b$ of the partons involved in the hard scattering. Particles participating in the hadronic reaction with momentum smaller than μ_F are considered to be part of the proton and are described by the PDF. Particles with momentum greater than μ_F are taken from the partonic cross section. The factorization scale μ_F is an arbitrary scale and the computed cross section should not depend on its value. However, for all practical purposes, cross sections are computed including only a finite order of terms introducing a dependence of the computed cross section on the factorization scale μ_F. Chapter 5 introduces systematic uncertainties to account for the dependence on μ_F.

2.2.2. Luminosity

The instantaneous luminosity L of an accelerator measures brightness and focus of the incident beams. Higher luminosities result in a higher rate of particle collisions. For circular colliders with n_i particles per colliding proton packet and a collision frequency f, the luminosity is given by

$$L=f\frac{n_1n_2}{4\pi\sigma_x\sigma_y} \tag{2.22}$$

where σ_x and σ_y refer to the vertical and horizontal root-mean-square (RMS) beam size. For the LHC in 2018, with an average collision frequency[5] $f=28.6\,\mathrm{MHz}$ and 1.1×10^{11} protons per bunch, a peak luminosity of $L=1.9\times10^{34}\,\mathrm{cm^{-2}s^{-1}}$ was achieved at the ATLAS interaction point [21, 24]. The integrated luminosity

$$L_{\mathrm{int}}=\int \mathrm{d}t\, L \tag{2.23}$$

is used to refer to the size of a recorded proton–proton event dataset. The integrated luminosity delivered by the LHC during the data-taking period between 2015 and 2018, referred to as Run 2, is $L_{\mathrm{int}}^{\mathrm{Run\,2,\,LHC}}=156\,\mathrm{fb}^{-1}$. Due to data acquisition inefficiencies, the dataset recorded by the ATLAS detector corresponds to an integrated luminosity $L_{\mathrm{int}}^{\mathrm{Run\,2,\,recorded}}=147\,\mathrm{fb}^{-1}$. The ATLAS dataset passing all data quality criteria corresponds to an integrated luminosity $L_{\mathrm{int}}^{\mathrm{Run\,2}}=139\,\mathrm{fb}^{-1}$ [25]. With the integrated luminosity, we can obtain the expected number of events

$$n=\sigma\, L_{\mathrm{int}} \tag{2.24}$$

[5]The average collision frequency takes into account that only 2544 bunches are injected into the LHC out of a total of 35640 radio frequency buckets.

for a process with cross section σ.

2.2.3. Hadronization and Jets

Many processes at hadron colliders include quarks or gluons in the final state of the hard scattering process. Due to the nature of QCD, individual quarks and gluons cannot be observed as independent particles in the detector since they are not colorless. As a quark or gluon moves away from the proton remnants, it is affected by the strong force. With increasing spacial separation, the potential energy increases until it reaches a threshold and new hadrons are created. This process continues and usually a large number of hadrons are created and form colorless sets of quarks and gluons. This fragmentation process is called hadronization. The dynamics are governed by QCD and fall in low-energy regime such that perturbative approaches are not applicable. The details of hadronization are not fully understood and cannot be derived from first principles. To analyse data from experiments, phenomenological models are employed to simulate this process.

The collimated hadrons in the final state are collectively referred to as a *jet*. The exact definition of a jet depends on the reconstruction algorithm, for example, see Section 3.3.3. Jets interact with the detector material and generate secondary particles undergoing subsequent inelastic reactions with the detector. This phenomenon is referred to as a calorimeter shower. The initial energy of the jet is distributed among the secondary particles. The process stops when the energies of the particles fall below the pion production threshold. Hadrons in the jet or the shower can induce nuclear reactions with the detector material. A considerable amount of energy is converted to photons leading to an electromagnetic shower of alternating pair production and bremsstrahlung. The electromagnetic shower stops when the energy per particle falls below the critical energy below which the energy loss due to ionization dominates over energy loss due to bremsstrahlung.

2.3. Higgs boson at the LHC

The Higgs boson was first discovered at the LHC in proton–proton collisions. In this section the production of Higgs bosons at the LHC and the prominent decay channels are briefly described. The two frameworks used in this thesis to study properties of the Higgs bosons are introduced: the κ-framework that parameterizes modified couplings and the Simplified Template Cross Section framework, which defines production cross sections in fiducial regions of phase space.

2.3.1. Higgs boson production and decay

The production of a Higgs boson in proton–proton collisions is a rare phenomenon. Figure 2.3 shows an overview of the cross sections of selected processes as a function of the center-of-mass energy $\sqrt{s}$. The chance of creating a Higgs boson in a single proton–proton collisions at the LHC is less than 10^{-9}.

The four main production modes of the Higgs boson at the LHC with the largest cross sections are gluon fusion (ggF), vector boson fusion (VBF), and the production

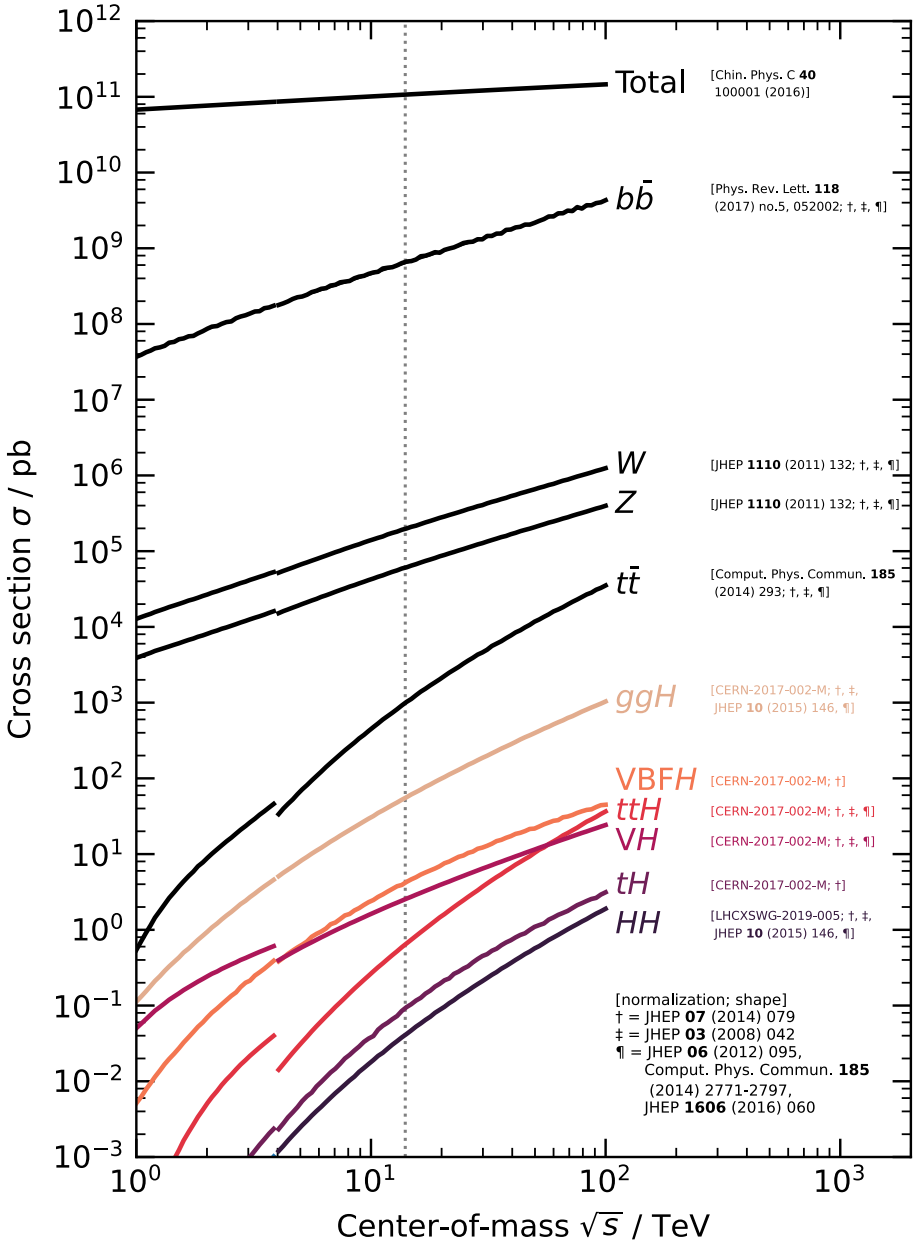


Figure 2.3.: Overview of different cross sections at hadron colliders for pp and $p\bar{p}$ collisions as a function of the collision energy [26]. The discontinuity at $\sqrt{s} = 4\,\text{TeV}$ signifies the change from anti-proton–proton collisions below the threshold to proton–proton collisions above it. The dashed vertical line indicates the center-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ of the LHC. The evolution of the cross section is simulated with MadGraph. Each process is normalized to precision measurements or calculations at the LHC energy.

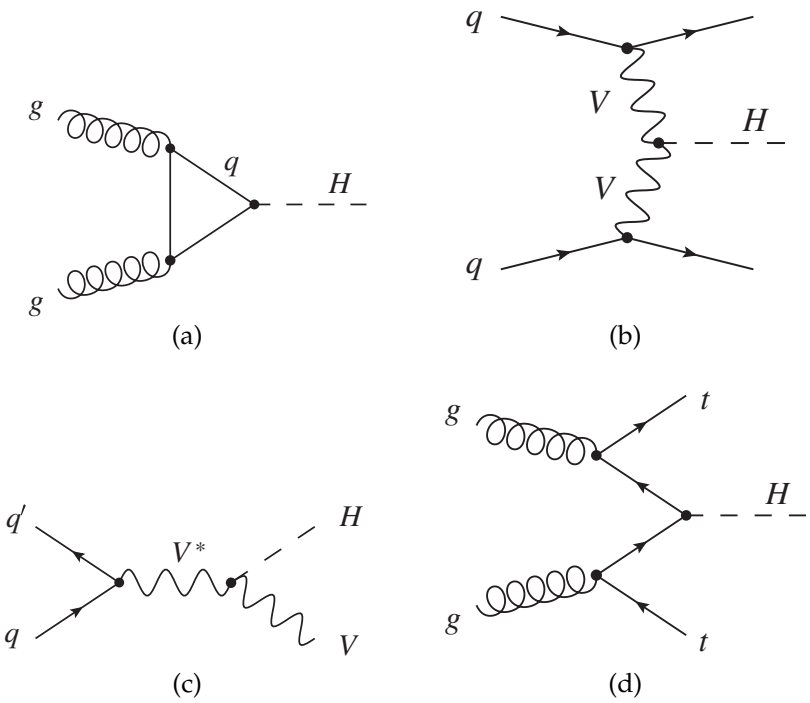


Figure 2.4.: Leading-order Feynman diagrams for the main Higgs boson production modes at the LHC: gluon fusion (a), vector boson fusion (b), in association with a vector boson (c), and in association with a pair of top quarks (d).

in association with a vector boson (VH) or a pair of top quarks ($t\bar{t}H$). In this context, the vector boson V refers only to the weak bosons $W^{\pm}$ and Z. Figure 2.4 shows the leading-order Feynman diagrams of all four production modes.

Due to the high relative abundance of gluons in protons and the observed Higgs mass $m_H = (125.09 \pm 0.24)$ GeV [27], ggF has the largest cross section with a margin of approximately one order of magnitude. The large cross section allows for analyses with larger statistical power and, therefore, more precise measurements. For some analyses, as in the case for the $H \to \tau\tau$ measurement, presented in Chapter 5, we require at least one jet in the final state. This means that for ggF to contribute, we required the radiation of a jet. The Higgs system recoils against this additional jet. This topology is therefore referred to as *boosted*.

On the other hand, VBF processes have a clear and district signature in the detector and can therefore be selected with high efficiency. The two jets in the final state of Figure 2.4(b) appear in the forward and backward directions of the detector with a large separation in rapidity. Additionally, since there are no QCD vertices in the diagram and no color flow between the two jets, we can expect only low QCD activity in the rapidity range between the two jets. The possibility to efficiently select events with this signature enables us to perform precise measurements of this process.

The Higgs boson is a short-lived particle and decays before reaching the innermost detector layers. We observe the detector signature of its decay products and not the direct interaction of a prompt Higgs boson with the detector. The Higgs boson can decay into any massive particle–anti-particle combination if it is kinematically accessible. The decay to a pair of Z bosons, for example, is only possible if one of the Z bosons appears as a virtual particle, i.e., as a particle that is not on its mass shell defined by $E^2 = m^2 + p^2$. As we have seen in Section 2.1.3, the coupling of the Higgs bosons to other particles depends on the particle's mass. Therefore the branching ratio B_c of the Higgs boson decay channel c depend on the known masses of the decay products.

An overview of Higgs boson branching ratios is shown in Figure 2.5. The decay to a pair of top quarks is kinematically suppressed. The decay channel to a pair of bottom quarks has the largest branching ratio. The decay to a pair of tau leptons has the largest branching ratio among all decays to leptons. The Higgs boson does not couple to gluons or photons since both are massless. However, the gluon and the photon appear as decay channels via the indirect coupling to a heavier particle via loop effects. The decay to a pair of gluons is analogous to the ggF production via a fermion loop. The decay channels listed in Figure 2.5 includes final states with unstable particles, which subsequently decay or hadronize.

2.3.2. The κ-framework

With precision measurements of the Higgs boson, we can test SM predictions. When statistically combining independent analyses of different Higgs boson decay channels, it is convenient to quantify possible deviations from the SM in the κ-framework.

In the κ-framework [14], we introduce a set of coupling modifies κ_i that scale the coupling strength between particle i and the Higgs boson. We further assume a single, narrow Higgs resonance, such that we can decompose the cross sections into a production and decay side. Consider the process $i \to H \to f$. The cross section times

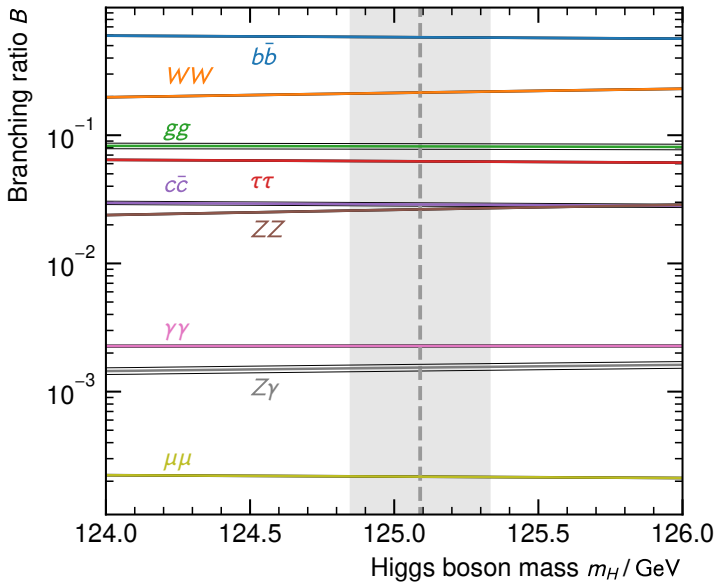


Figure 2.5.: Higgs boson branching ratios B as a function of the Higgs boson mass for various decay channels [28]. The vertical dotted line indicates the measurement of the Higgs bosons mass $m_H = (125.09 \pm 0.24)$ GeV by the ATLAS and CMS collaborations [27]. Theoretical, systematic, and statistical uncertainties are indicated by the transparent bands for the Higgs boson mass and by thin black outlines for the braching ratios.

branching ratio can be parameterized as

$$\sigma_i \cdot B_f = \frac{\sigma_i^{\mathrm{SM}} \kappa_i^2 \cdot \Gamma_f^{\mathrm{SM}} \kappa_f^2}{\Gamma_H^{\mathrm{SM}} \kappa_H^2}, \tag{2.25}$$

where Γ_H^{SM} denotes the total Higgs boson decay width, Γ_f^{SM} the partial decay width to particle f, and σ_i^{SM} the cross section for production via particle i. The SM superscript denotes the SM predictions. The term κ_H adjusts the total Higgs boson width to account for adjustments of partial decay widths. The κ coupling modifiers are therefore defined by

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_i^{\mathrm{SM}}} \quad \text{and} \quad \kappa_f^2 = \frac{\Gamma_f}{\Gamma_f^{\mathrm{SM}}} \tag{2.26}$$

as the ratio of cross sections and decay widths normalized to the SM expectation. Taking $\kappa_i = \kappa_f = 1$ reproduces the SM.

The Higgs boson branching ratio B_{inv} to invisible decays refers to decays beyond the SM (BSM) whose decay products do not interact with the detector and thus are only characterized by the momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ carried away from the momentum balance in the transverse plane, see Section 3.3.6. On the other hand, the branching ratio B_{undet} to undetected decays refers to BSM contributions that non of the included analyses have sensitivity to. These are typically decays to light quark jets that cannot be distinguished from soft QCD jets or decays to undetected BSM particles without a large contribution to $E_{\mathrm{T}}^{\mathrm{miss}}$. Measured coupling modifiers with $\kappa_j \neq 1$ or invisible and undetected branching ratios with $B_{\mathrm{undet}}, B_{\mathrm{inv}} \neq 0$ are signs of BSM physics.

In the κ-framework, the total decay width of the Higgs boson is scaled by κ_H^2. The parametrization for κ_H^2 takes into account the effects from modifications of other coupling parameters and BSM decays to invisible and undetectable particles. The parameter κ_H^2 as a function of the coupling modifiers and the branching ratios B_{undet} and B_{inv} is given by

$$\kappa_H^2 = \frac{\sum_f \kappa_f^2 B_f^{\mathrm{SM}}}{1 - B_{\mathrm{undet}} - B_{\mathrm{inv}}}. \tag{2.27}$$

The current measurements have excellent sensitivity for $\kappa_Z, \kappa_W, \kappa_t, \kappa_b, \kappa_\tau, \kappa_g, \kappa_\gamma$ and B_{inv}, see Chapter 6. However, the analyses included in the combination show only very little sensitivity to the modifiers of the second-generation quarks. Therefore, κ_s is set to be equal to κ_b and κ_c is set to be equal to κ_t. The contributions from first generation quarks and the electron are negligible provided the coupling modifiers κ_u, κ_d, κ_e are of the order of unity.

The gluon, the photon, and also the process $H \to Z\gamma$ do not couple directly to the Higgs boson. The Higgs boson production and decay via these modes is facilitated via a W boson loop or heavy fermion loops. We introduce effective coupling modifiers κ_g, κ_γ and $\kappa_{Z\gamma}$ to scale the cross section or decay widths. Heavy SM particles dominate the contributions to the loop processes. These coupling modifiers are especially sensitive to virtual BSM particles with a strong coupling to the Higgs boson that are otherwise too massive to be created directly.

A limitation of the κ-framework is that it does not account for modifications to the Lorentz structure of the Higgs boson interactions and therefore cannot resolve changes to kinematic distributions beyond a simple scaling of the couplings. However, it has merit in yielding results with a straightforward interpretation and the power to probe for additional BSM particles or test symmetries of the SM.

Chapter 6 presents measurements of parameters of different models in the κ-framework imposing varying assumptions on BSM physics.

2.3.3. Simplified Template Cross Section

Measurements in the κ-framework are expressed as a ratio to the SM prediction. The measured κ modifiers depend on the particular theory used to perform the measurement. The analyses have to be repeated to test alternative or future theories. Additionally, the measured results are subject to uncertainties from the underlying theory. The Simplified Template Cross Section (STXS) framework [29–31] is an alternative approach to measure the properties of the Higgs boson.

Analyses measuring properties of the Higgs boson usually apply a selection on the reconstructed properties of the event. The selected phase space thus depends on the detector effects which makes it difficult to compare the result to theories and to other experiments. Unfolding detector effects leads to the particle or truth level. Typically, particle level is defined by particles with lifetimes longer than 10^{-10} s [32]. At particle level, cross section measurements are performed in phase space volumes defined by particle level quantities. The unfolding procedure introduces only a small theoretical dependence of measured cross sections, firstly, due to the model used to simulate detector effects, and, secondly, due to the extrapolation from reconstructed event selection to particle level phase phase. Phase space regions defined at particle level are referred as *fiducial* phase space regions. Cross section measurements in fiducial phase spaces have a minimal dependency on theoretical predictions.

Machine-learning techniques such as neural networks (NNs), select a region in reconstructed phase space that is optimized to enhance the statistical power of the analysis. However, it is difficult to translate the selected region to particle level in order to minimize dependency on theory. Depending on the concrete analysis or process under consideration, measuring cross sections in fiducial phase spaces can imply reduced sensitivity.

The STXS framework defines regions in particle-level phase space, called *bins*, with the aim to measure exclusive, fiducial cross sections in each bin optimized to take advantage of the statistical sensitivity of the analysis while retaining model independence. STXS bins are a compromise to measure theory-independent, fiducial cross sections while not impeding advanced analysis techniques such as NNs. Besides this compromise, the bin definition should also isolate regions where potential effects from BSM processes are dominant. The process of defining the bins is non-trivial. The bins match as closely as possible the definition of experimental signal regions defined by cuts on reconstructed data to avoid introducing large dependencies due to the extrapolation. The STXS framework defines a global $|y_H| < 2.5$ condition on the rapidity of the Higgs boson matching the typical acceptance of detectors at the LHC to avoid dependencies from the extrapolation to the whole rapidity range.

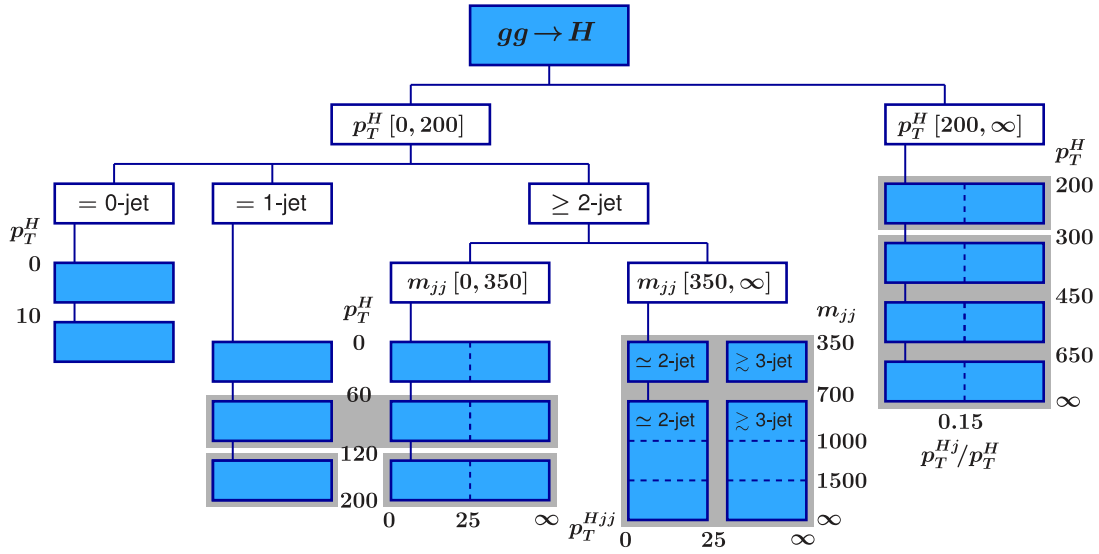


Figure 2.6.: STXS stage 1.2 bin definition for ggF Higgs boson production [30]. The analysis presented in Chapter 5 measures cross sections for the six bins indicated by gray boxes. Some of the STXS bins are merged in the measurement.

The largest difference between STXS and fiducial cross section is that they depend on the Higgs boson production mode. On the other hand, the bins are independent of the Higgs boson decay channel facilitating the combination of different analyses. The definition of the fiducial phase space bins is shared between experiments at the LHC to allow combining measurements in each bin.

The STXS framework defines bins in stages adding granularity at each stage. Each stage further subdivides the existing regions in phase space based on properties of the particle level Higgs boson, associated bosons or properties of reconstruction-level jets. The full stage 1.2 splitting is shown in Figures 2.6–2.9. The first stage, stage 0, only introduces a split based on the production mode where the definition of production mode differs slightly from the definition shown in Figure 2.4. Higgs boson production in association with a vector boson is separated depending on the decay of the vector boson. VH production with hadronic $V \to qq'$ decays and electroweak VBF production are measured in the same stage-0 bin designated for electroweak production with two quarks.

2.3.4. Beyond the Standard Model

Despite the success of the SM, it cannot be the ultimate description of matter and its interactions in the universe. The SM is built in an ad hoc fashion, guided by experimental data [10]. The discovery of the Higgs boson completed the picture of the SM, and to this date, all measurements of properties of the Higgs boson are in spectacular agreement with the SM. However, the large number of free parameters in the SM is unsatisfactory. The SM does not *explain* the fact that fermions come in three generations. Besides these more aesthetic issues, there are measured phenomena that require

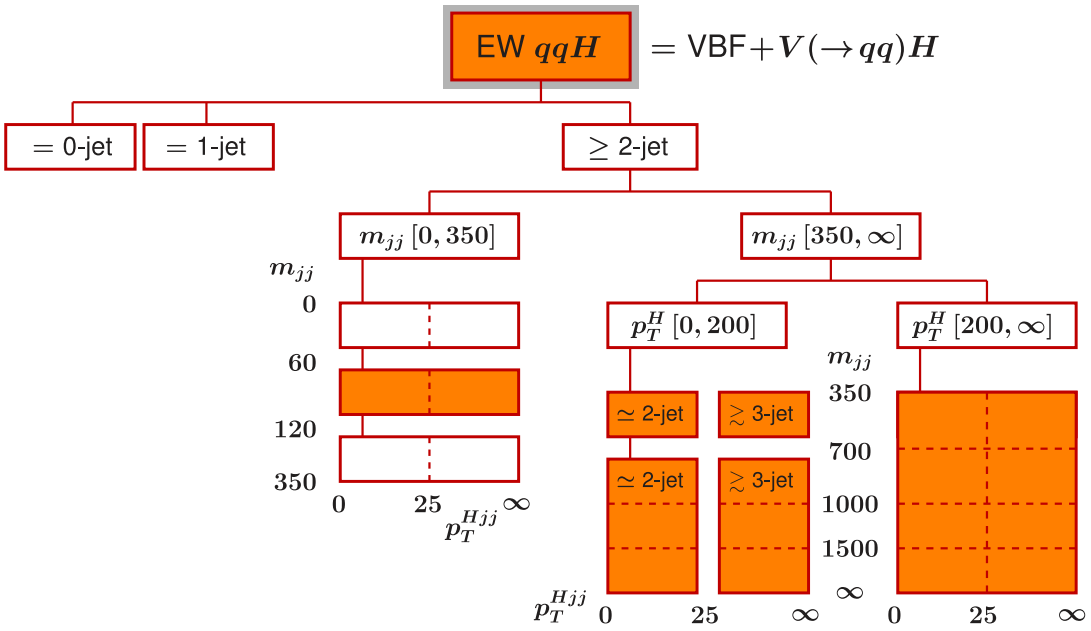


Figure 2.7.: STXS stage 1.2 bin definition for electroweak VBF Higgs boson production and VH production with hadronic V decays [30]. The analysis presented in Chapter 5 measures the inclusive cross section at the stage 0 level.

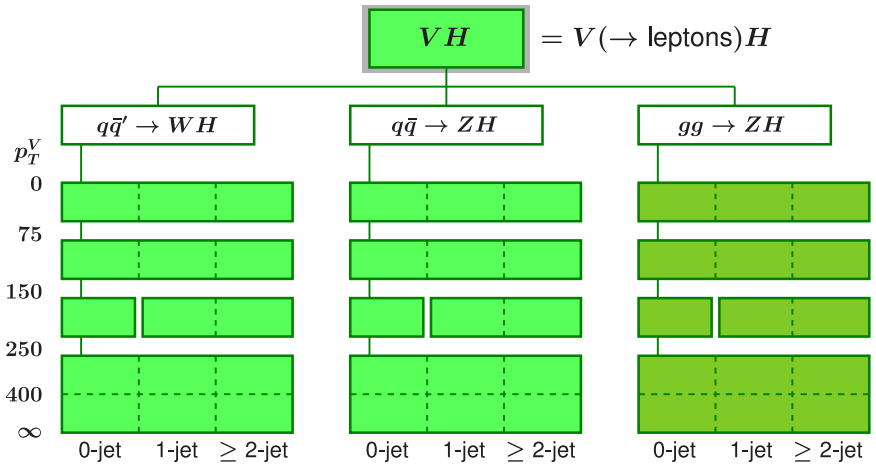


Figure 2.8.: STXS stage 1.2 bin definition for Higgs boson production in association with leptonically decaying vector boson [30]. The analysis presented in Chapter 5 measures the inclusive cross section at the stage 0 level.

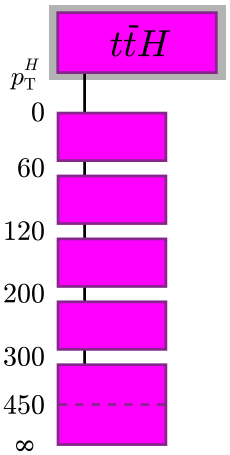


Figure 2.9.: STXS stage 1.2 bin definition for Higgs boson production in association with a pair of top quarks [31]. The analysis presented in Chapter 5 measures the inclusive cross section at the stage 0 level.

an extension of the SM or even a completely new theory.

From the rotation curves of galaxies [33] or the discovery of gravitational lensing [34], we know that the majority of the matter content of galaxies is not due to ordinary, luminous (i.e., electromagnetically interacting) matter. The cosmological ΛCDM model (cosmological constant Λ and Cold Dark Matter) based on many astrophysical measurements predicts that the energy density of dark matter is approximately five times the density of baryonic matter [35]. The nature of dark matter particles remains speculative until it is discovered by an experiment. Searches at colliders, as well as direct and indirect searches have been performed without evidence for dark matter particles. The measurement of potential deviations from the SM described in Chapter 6 places limits on the Higgs boson branching ratios to particles not described in the SM.

The observation of neutrino oscillations [36–38] implies that neutrinos have a mass. In the SM discussed in this chapter, neutrinos were considered massless. As we did for the massive fermions, we can introduce a coupling between the Higgs field and the neutrinos to generate a mass term in the Lagrangian after spontaneous symmetry breaking. The upper limit of the electron neutrino masses in a direct measurement is 0.8 eV [39]. Depending on the implementation, neutrinos could be realized as Majorana particles implying the existence of neutrinoless double β-decay ($0\nu\beta\beta$). In this case, neutrinos would be their own anti-particles however $0\nu\beta\beta$ decays have not yet been observed.

2.4. Machine learning

With technological advances, the number of events and the amount of information collected in each particle collision event grew enormously over the last decades. An example is the proton–proton collision data collected by the ATLAS experiment. On the other hand, the large effort required to build and maintain large-scale experiments

and accelerators like ATLAS at the LHC demands a thorough analysis of the collected data with refined analysis techniques. The explosion of computing power in the last decades and the progress in artificial intelligence delivered a tool that is a perfect match for high-energy physics applications. The large, readily available Monte Carlo (MC) datasets serve as an excellent training input for machine learning applications that are then used to analyze the dataset, test theoretical predictions, and perform measurements. This section introduces the basics of machine learning used for the $H \to \tau\tau$ analysis presented in Chapter 5. The presentation in this section is based on References [40–42].

2.4.1. Supervised learning

One branch of machine learning takes advantage of a dataset with known class labels to train a classifier. The trained classifier is then used to categorize new, previously unseen data into learned classes. During the training phase, the classifier *learns* a function f that maps from the feature space spanned by input variables $\boldsymbol{x}$ to the output space of class labels $\boldsymbol{y}$. The classifier should be able to predict the class label for new, previously unseen data. This approach is called supervised learning.

In high-energy physics, the feature space usually consists of kinematic properties and the measured properties of an event. It is common to use physics processes or groups of processes, such as signal and background, as target classes. The large, high-quality training sets with known labels are readily available from Monte Carlo simulation.

In the $H \to \tau\tau$ analysis presented in this thesis, neural networks (NN) and Boosted Decision Trees (BDT) are used to learn and approximate the function f that predicts the true class labels for each event. Specifically, we can view $f(\boldsymbol{x}; \boldsymbol{\theta})$ as a parametrized function depending on a set of parameters $\boldsymbol{\theta}$. While iterating through the events with known labels, we optimize, i.e., learn the parameters $\boldsymbol{\theta}$. Depending on the internal structure and the parameterization of f, the function reproduces a class of functions. If f was a linear function, the class of functions that we could approximate is limited. We can refer to this property as the *capacity* [40]. Often the capacity scales with the number of parameters.

The capacity of the machine learning model should match the complexity required to approximate the function. If the capacity is too small, we will not be able to optimize the parameters $\boldsymbol{\theta}$ such that we reproduce the labels on the training set with known labels. On the other hand, if the capacity is too large, we achieve almost perfect performance on the training set. However the trained model does not generalize to new, unseen events. For example, if a model's capacity is too high, the model could learn the training data set "by heart", similar to a lookup table, but produces very unreliable predictions for events not included in the training data set. This case is referred to as *overtraining*.

There are several methods to reduce the effective capacity of a model. Specific methods for neural networks are discussed in the next section. To assess the performance on new, unseen data, we can set aside a fraction of the Monte Carlo set as a separate *test set* and use only the remaining events for the training. If the events for the test set are chosen randomly, the events in the training and test set are statistically independent

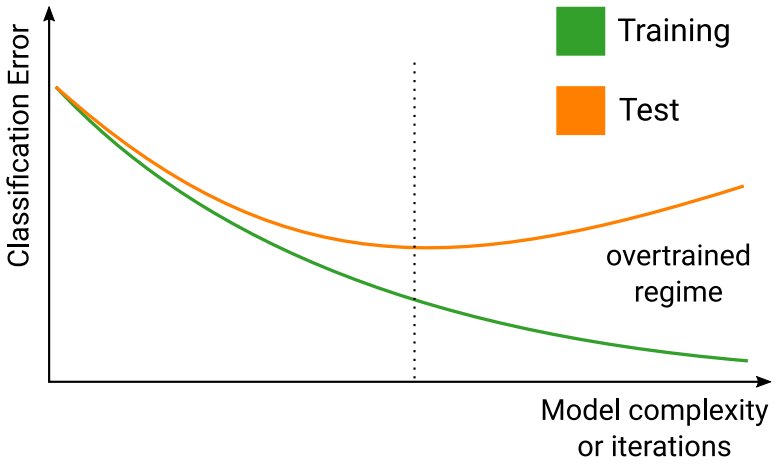


Figure 2.10.: Visualization of classifier performance as a function of the model's complexity or number of training iterations on the training set (green) or the test set (orange). The veritical line indicates the point above which the performance on the validation set degrades with increasing model capacity. The models on the right side of the vertical line are in the overtrained regime.

and identically distributed. This allows us to evaluate the performance on new unseen events. With increasing effective model capacity, the performance on the training set is expected to increase without bounds. For the performance on the test set, we expect an optimal point. If the capacity is increased beyond the optimal point, we enter the overtrained regime as illustrated in Figure 2.10.

Training algorithms and the model parametrization often include a set of hyperparameters. In the case of NNs, this includes the architecture of the network. It is not straightforward to define optimal values beforehand, and the optimal hyperparameter configuration must be found in a trial and error fashion. Using the performance on the test set for this hyperparameter optimization introduces a bias towards the test set. We optimize the model for this specific combination of training and test set and can expect that, in general, the model performs worse on new and unseen events not part of the test or training set. This is because the performance measure is subject to noise, and we select the model with the best performance on the test set. The selected model is probably the best model only by chance.

To prevent this bias, we can split the initial Monte Carlo set into three parts. The first part, the training set, is used to optimize the parameters $\boldsymbol{\theta}$. The learning of parameters $\boldsymbol{\theta}$ on the training set is repeated with different hyperparameter configurations. The performance of each configuration is assessed on the second part of the MC set, the validation set. This allows us to select the hyperparameters that result in the best-performing model. The final part, the test set, can then be used to assess the classifier's performance on unseen events without bias. The classifier's output on the test set is used in the likelihood fit of the physics analysis, where we compare the MC prediction to measured data. The performance on the test set and recorded data events are

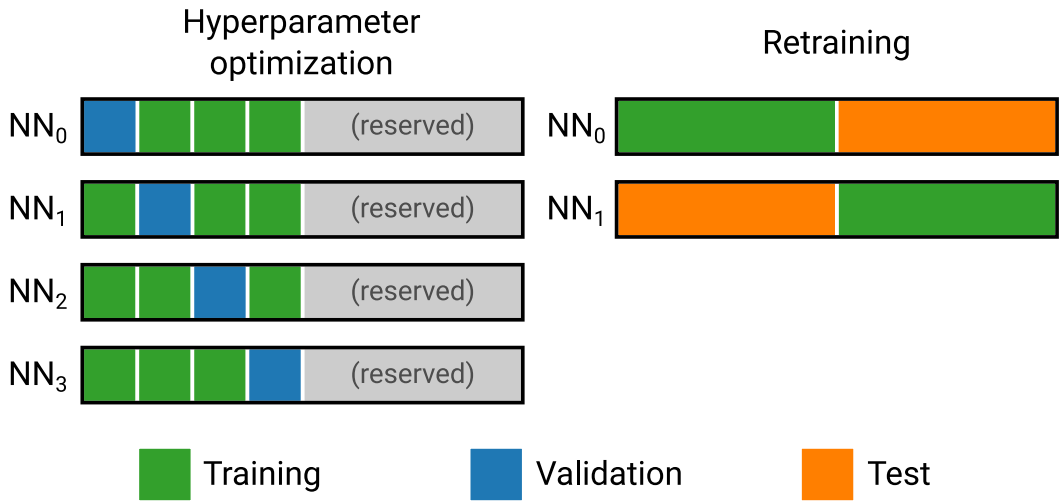


Figure 2.11.: Illustration of the k-fold cross validation scheme (left) and cross evaluation scheme (right) for $k = 4$. Each black rectangle represents the full MC dataset that is split into chunks. The color-coded role of each chunk is different for each NN. The cross validation scheme is used to optimize the hyperparameter configuration that is then used in the cross evaluation scheme to retrain a classifier used for data–MC comparisons in the physics analysis.

expected to be identical assuming perfect MC modeling of the data.

With a three-fold, equal split, we reduce the number of events per set by a factor of three. The reduction in training statistics can impact the best achievable performance of the classifier. Similarly, the reduced statistical power of the validation set introduces noise in the hyperparameter selection leading to potentially suboptimal choices. Smaller test set sizes lead to larger statistical uncertainties on the final physics measurements. To overcome the issue, we can resort to k-fold cross validation combined with cross evaluation, where we first split the Monte Carlo dataset into two halves. The first half is further partitioned into k parts. We train k independent classifiers. For the i-th classifier, the i-th part is used as the validation set, the rest of the first half is used as the training set. The initial second half serves as a reserved test set not used for the hyperparameter optimization. After fixing the hyperparameters, we retrain two classifiers, one on each half, not on the k-fold partition. The training scheme is illustrated in Figure 2.11. The mismatch of dataset sizes between the hyperparameter optimization and the retraining might yield slightly suboptimal networks (the network could have had a higher capacity). However, this is highly preferred over networks biased towards any dataset. With this, we have the validation set for the model selection, and the full Monte Carlo set as a test set. If one fails to use a dedicated set for the training, the model selection, and the final analysis, one should expect that the classifier is biased and leads to discrepancies between data and MC in the likelihood fit.

Some Monte Carlo generators, like Sherpa [43], assign weights to the generated events to quantify over or under-sampled regions in phase space. Event distributions

are expected to match the SM expectation only if the *generator* or *MC weights* are applied. The interpretation of individual events with negative weights is difficult. Negative event weights reduce the expected yield. Events with negative event weights often lead to numerical and mathematical issues when the event weights are propagated to learning algorithms. Therefore, the events with negative weights are ignored during the training. For the $H \to \tau\tau$ analysis presented in Chapter 5, approximately 25 % of the events have a negative weight. This procedure might lead to suboptimal performances, however it cannot lead to discrepancies between Data and MC.

Conceptually, the absolute scale of the weights is insignificant. However, too small or too large weights lead to numerical issues. Good results are usually obtained when we scale the weights such that

- the sum of weights for each output class is an equal fraction, and
- the total sum of weights is equal to the size of the dataset.

Specifically, for m output classes and n total events, the rescaled weight $\tilde{w}_i^{(j)}$ of event i in class j is given by

$$\tilde{w}_i^{(j)} = \frac{n}{m} \cdot \frac{w_i^{(j)}}{\sum\limits_{\text{event } i' \in \text{class } j} w_{i'}^{(j)}} \tag{2.28}$$

where $w_i^{(j)}$ are the original events with $w_i^{(j)} > 0$.

2.4.2. Neural networks

Neural networks were initially developed to mimic the function of the human brain. However, the modern form of neural networks in machine learning has its origin in statistical learning. A simple, connected dense neural network consists of multiple layers where each layer consists of nodes. The first layer functions as an input to the network, the last layers is the network's output. The layers between the input and output are called hidden layers. A sketch of a simple feedforward neural network is shown in Figure 2.12.

Each node, except the input nodes, computes a weighted sum of the outputs of the previous layer with weights $\boldsymbol{\omega}$. The output

$$h_j^{(i)} = a^{(i)}\left(\boldsymbol{\omega}_j^{(i)} \cdot \boldsymbol{h}^{(i-1)}\right) \tag{2.29}$$

of a node j in layer i is a non-linear function $a^{(i)}$ of the weighted sum. Without the non-linear functions $a^{(i)}$, the function computed by the neural network simplifies to a linear function. A modern choice for the activation function is the *rectified linear unit* as shown in Figure 2.13 and defined as

$$\text{ReLU}(x) = \max(0, x). \tag{2.30}$$

Compared to classical choices such as the sigmoid function or tanh, ReLU has the advantage that the gradient does not saturate for large positive input values. For a k-class

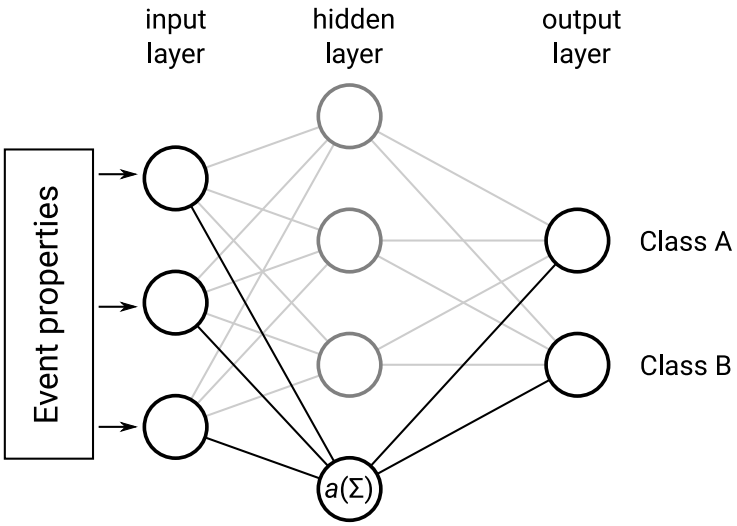


Figure 2.12.: Sketch of a dense feedforward neural network. Information flows strictly from left to right. The network has three input nodes, a single hidden layer with four nodes, and two output nodes. Each node computes the activation function of a weighted sum abbreviated by $a(\Sigma)$. The connections between layers are highlighted for a single node.

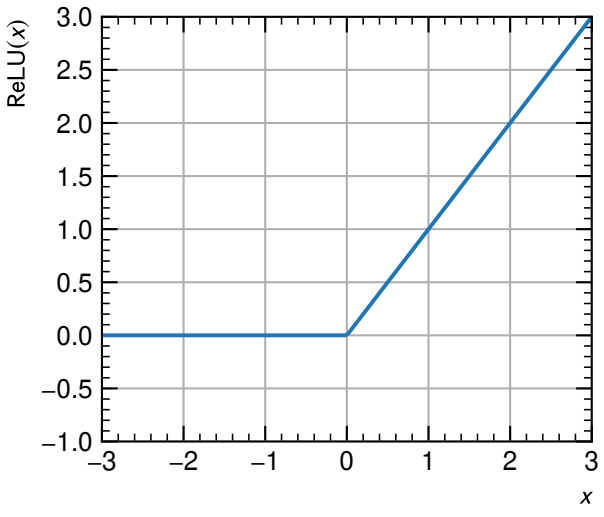


Figure 2.13.: Illustration of the rectified linear unit function $\mathrm{ReLU}(x)$ as a commonly used activation function in neural networks in the range $x \in [-3, 3]$.

classification problem, it is common to use k output nodes. With the softmax activation function

$$\text{softmax}(\boldsymbol{x})_j = \frac{\exp(x_j)}{\sum_i \exp(x_i)}, \tag{2.31}$$

we can interpret the output value of node j as the probability that the event belongs to class j.

The universal approximation theorem states [44, 45] that a network with a single hidden layer can approximate every physical function arbitrarily well. However, with just a single hidden layer, the required size might make this architecture unfeasible. It is therefore beneficial to include more hidden layers of smaller size. The number of layers and the number of nodes per layer are important hyperparameters and control the model's capacity.

Mathematically, learning the parameters $\boldsymbol{\theta}$ is a minimization problem where we minimize a loss function $J(\boldsymbol{x})$ by varying the parameters $\boldsymbol{\theta}$ for samples from the training set. In classification problems, it is common to use the cross entropy

$$J(\boldsymbol{x}) = -\sum_{i \in \text{events}} \sum_{j \in \text{classes}} y_{i,j} \log h_j^{(n)}(x_i) \tag{2.32}$$

as the loss function where $h_j^{(n)}$ refers to the values of the nodes in the output layer. The collection of all weights $\boldsymbol{\omega}_j^{(i)}$ constitutes the set of network parameters $\boldsymbol{\theta}$.

In feedforward networks, as introduced above, we can use gradient descent to optimize the parameters $\boldsymbol{\theta}$. The gradient computation is usually done with back-propagation [46], where we have a forward sweep to compute the loss function and go back through the network to compute the gradient with respect to $\boldsymbol{\theta}$. This can be done efficiently since the function computed by the network $f(\boldsymbol{x}; \boldsymbol{\theta})$ involves function composition at every layer in the network, see Equation (2.29). The gradient factorizes using the chain rule into components that can be computed independently. Modern GPU hardware can be used efficiently to train and evaluate these networks with large datasets.

As stated above, there are different techniques to prevent overtraining with neural networks. The most common methods are regularization [47], dropout [48], and early stopping. Regularization adds a penalty term to the loss function to prevent large weights. With dropout, connections between nodes or entire nodes of the network are randomly removed during training.

The networks trained in this thesis use the early stopping technique. With early stopping, the minimization of the loss requires multiple passes, so-called epochs, over the training set. The performances on the training and validation set are monitored after each iteration. As illustrated in Figure 2.10, each iteration improves the performance on the training set, but the performance on the validation set improves only until a certain point. Once the performance on the validation stops improving, the training is concluded. The early stopping technique therefore prevents entering the overtrained regime.

Another crucial step is the pre-processing of input variables and class weights. In this thesis, we are primarily concerned with continues input variables. We need to

apply a linear transformation to the input variables to prevent a single input variable from dominating the weighted sums in neural network layers. The rescaled variables have a vanishing mean and unit variance. This requirement is specific to neural networks. Other classifiers such as BDT are not affected by the same issue and do not need this pre-processing step.

CHAPTER 3

The LHC and the ATLAS detector

Si tu veux construire un bateau, ne rassemble pas tes hommes et femmes pour leur donner des ordres, pour expliquer chaque détail, pour leur dire où trouver chaque chose. Si tu veux construire un bateau, fais naître dans le cœur de tes hommes et femmes le désir de la mer.

— Antoine de Saint-Exupéry

With highly-energetic particle accelerators, one can study the predictions of the SM and test possible extensions of it. Since the particles under study, like the Higgs boson, are quite heavy, we need high center-of-mass energies and, in turn, high beam energies. At hadron colliders, the colliding particles are extended objects with an internal structure (as opposed to point-like particles). In case of proton–proton collisions, the colliding protons can undergo elastic collisions resulting in a deflection of the two protons. However, with the beam energies reached at the Large Hadron Collider, protons scatter inelastically, breaking the initial protons and creating new particles in the process. The collision of partons within the protons is referred to the hard scattering. The proton remnants constitute the so-called underlying event. A single crossing of two proton packets entails multiple proton–proton interactions referred to as pile-up. The experimental challenge is to build radiation-hard detectors around the interaction point to measure the scatter products, reconstruct the final state, and facilitate statistical tests of the underlying physics. This chapter introduces the Large Hadron Collider and the ATLAS detector that recorded the collision data used in this thesis.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [21] is a circular particle accelerator and storage ring at CERN (European Organization for Nuclear Research) near Geneva. The collider is installed in the 26.7 km long tunnel that previously housed the Large Electron–Positron (LEP) accelerator. The LHC is the final stage in CERN's accelerator chain. Packets of

protons or heavy ions, such as lead (Pb) or xenon (Xe), are injected by the Super Proton Synchrotron (SPS) into the LHC at a beam energy of 450 GeV. The proton packets, the so-called bunches, travel as two counter-rotating beams in separate beam pipes. Radio frequency (RF) cavities located at a single point along the ring provide an oscillating electrical field that accelerates the passing bunches to their target energy. The counter-rotating beams are kept on their circular trajectory by 1232 superconducting dipole magnets. The niobium–titanium (NbTi) magnets are cooled with superfluid helium to temperatures below 2 K to reach magnetic field strengths above 8 T. The dipole magnets provide the magnetic field with opposing orientation for the two closely spaced beam pipes. The LHC is the most powerful particle accelerator with a design beam energy of 7 TeV and center-of-mass collision energy of 14 TeV. During the first data-taking period of the LHC between 2010 and 2012, referred to as Run 1, a center-of-mass energy of 7 TeV, and later 8 TeV was reached. During the second data-taking period, Run 2, between 2015 and 2018, protons were collided at a center-of-mass energy of 13 TeV. This thesis utilizes data from this data-taking period.

Proton and ion bunches are brought to collision at four interaction points around the ring. The locations of the interaction points coincide with the four main LHC experiments: ALICE [49], ATLAS [50], CMS [51], and LHCb [52]. While ATLAS and CMS are two large-scale, general-purpose experiments, ALICE studies primarily the properties of quark-gluon plasma and LHCb properties of the bottom quark and its hadrons.

In Run 2, the proton bunches travel at almost the speed of light with a spacing of 25 ns. This corresponds to a spatial distance of approximately 7.5 m. Consequently, bunch crossings occur at a frequency of 40 MHz. For technical reasons, not every possible bunch position (every tenth RF bucket of the 400.79 MHz RF system) is filled with protons leading to empty bunch crossings in which no proton-proton collisions are expected. The number of bunch crossings depends on the fill pattern. In 2018, a total of 2544 bunches were injected per beam.

Immediately after the injection into the LHC, each filled bunch consists of $O(10^{11})$ protons. Quadrupole and higher-order magnets focus the beams to a small transverse area at the interaction points. According to the luminosity equation for circular colliders, Equation (2.22), this increases the number of expected events leading to multiple inelastic proton–proton interactions during a single bunch crossing. The multiplicity is referred to as the number of pile-up interactions, and the interactions other than the one of interest are said to be pile-up events. The beam focus and separation parameters are updated continuously during an LHC fill to cap (*level*) the delivered instantaneous luminosity at at most $L = 1.9 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ [24, 53] to keep the number of pile-up interactions at a level suitable for the ATLAS detector. Towards the end of a fill, after $O(10\,\mathrm{h})$, the number of protons per bunch is sufficiently reduced such that the maximal delivered luminosity is below $1.9 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ and continuous to drop until it is decided to end the run and refill the machine. The evolution of the average number of pile-up interactions during a typical physics run and the evolution of the total integrated luminosity over the full Run 2 period is shown in Figure 3.1.

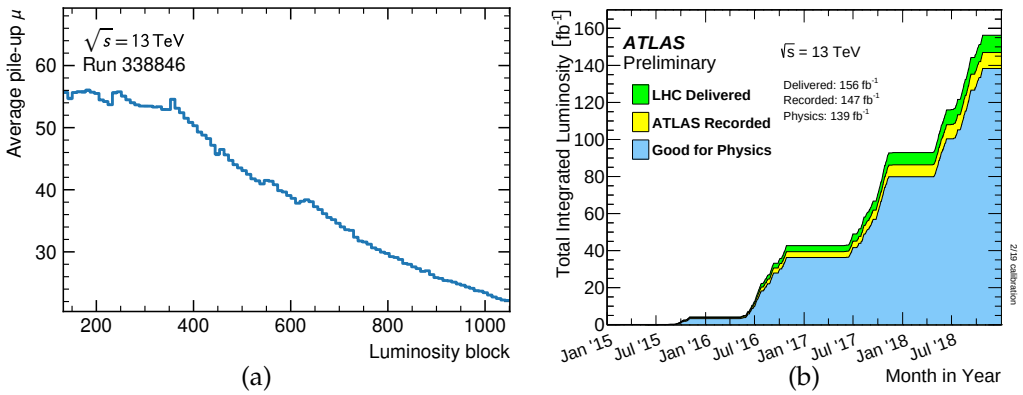


Figure 3.1.: Evolution of average number of pile-up interactions during run 338846 (a) and total integrated luminosity over Run 2 (b) [24, 25]. Each luminosity block corresponds to approximatly one minute in time. At the beginning of the run, until approximately luminosity block 350, the instantaneous luminosity is levelled to limit the number of pile-up interaction to $\mu \approx 55$. The instantaneous luminosity departs from the flat evolution as the number of protons per bunch decreases. The integrated luminosity used for physics analysis is smaller than the intergrated luminosity delivered by the LHC due to detector inefficiencies, and data quality criteria.

3.2. The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector [50] is one of the two large-scale, general-purpose detectors at the LHC. The detector is located at Point 1 of the LHC ring in its cavern approximately 100 m below the surface close to the main CERN site. The ATLAS detector is a cylindrical particle detector built with a forward-backward symmetry and an onion-style design covering a solid angle of almost 4π. With a length of 44 m and a diameter of 25 m, the ATLAS detector is the largest detector located on the LHC ring. The goal of the detector is to provide good tracking performance, particle identification, and energy and momentum resolution for 40 MHz collision rates with $O(1000)$ secondary particles per collision. A sketch of the ATLAS detector with its subcomponents is shown in Figure 3.2.

A dedicated coordinate system is used to describe the detector components and study particle trajectories. The origin of the coordinate system coincides with the nominal interaction point. The z-axis is parallel to the direction of the beam, with its axis pointing towards the city of Geneva. The x-axis points to the center of the LHC ring. The y-axis points upwards, thus resulting in a right-handed coordinate system.

Often it is convenient to switch to spherical coordinates, with the azimuthal angle ϕ measuring the angle in the transverse (x, y)-plane and the polar angle θ measuring the angle towards the beam axis. Instead of θ, often the pseudorapidity

$$\eta = -\log \tan \frac{\theta}{2} \tag{3.1}$$

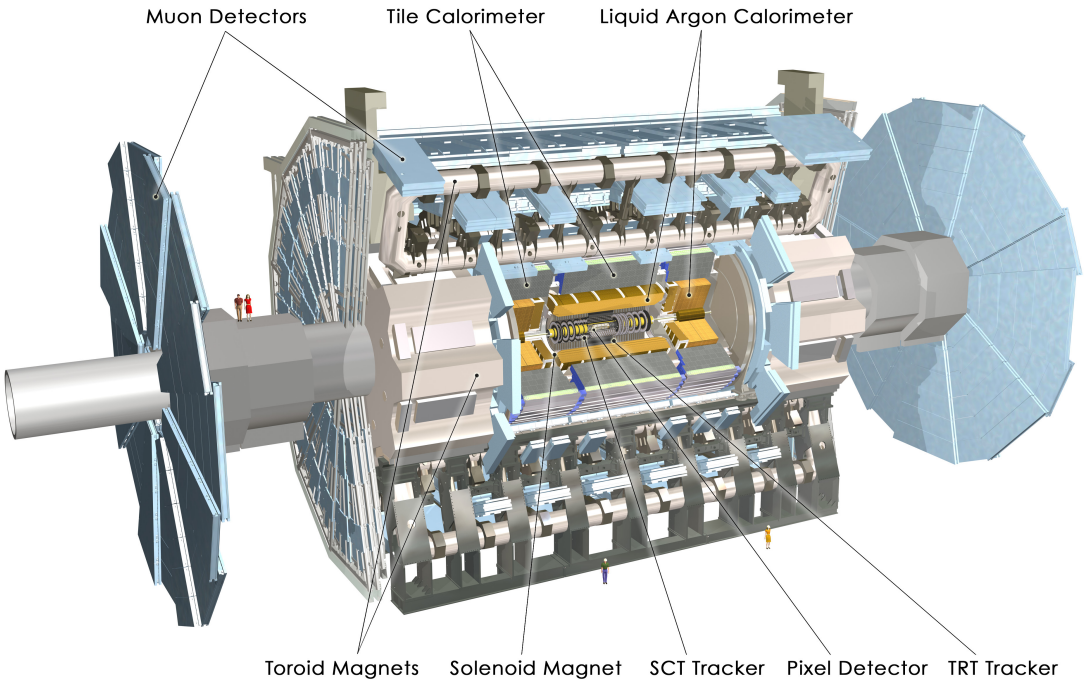


Figure 3.2.: Illustration of the ATLAS detector [54] with its individual detector components and magnets.

is used. For massless particles, the pseudorapidity equals the rapidity

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right). \tag{3.2}$$

Rapidity differences are invariant under Lorentz boosts along the beam axis, which is convenient since the center-of-mass of the interacting partons has an unknown boost along the beam axis. For particle trajectories at the LHC, the particle masses are typically small compared to the total energy and, therefore, the pseudorapidity η is a good approximation. Distances in the cylindrical (ϕ, η)-plane are measured with $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

The ATLAS detector is built in cylindrical layers with end-caps on each side. The inner-most component, the Insertable B-Layer (IBL), is a pixelated silicon detector directly attached to the beam pipe. The Pixel detector is located around the IBL. Further out, the Inner Detector (ID) is comprised of a silicon strip detector, the SemiConductor Tracker (SCT), and the Transition Radiation Tracker (TRT) consisting of straw tubes. The ID is enclosed by the two calorimeters of the ATLAS detector: first the electromagnetic calorimeter (ECal) and further out the hadronic calorimeter (HCal). The outermost component of the ATLAS detector is the muon spectrometer (MS) which drives the overall size of the detector. The ID is immersed in a magnetic field from a superconducting solenoid magnet. The MS is immersed in a toroidal magnetic field created by 24 superconducting coils, eight in the barrel region and eight in each end-cap region. The components are introduced in the next sections. The description is based on

Reference [50] unless otherwise noted and follows the trajectory of scattering products from the interaction point outwards.

3.2.1. IBL and Pixel

The Insertable B-Layer (IBL) [55, 56] and the Pixel Detector are pixelated detectors close to the beam pipe. Due to the proximity of the detector layers to the interaction point, a high granularity is required to resolve individual tracks and provide good position resolution. The IBL was installed during the shutdown of the LHC in 2013 and 2014 and was therefore not available in Run 1.

The centers of the IBL modules have a radial distance of $\rho = 33.25\,\text{mm}$ to the beam axis. The small separation improves the reconstruction of secondary vertices and the impact parameter of a collision which is beneficial for tagging events with long-lived secondary particles, such as b quarks that travel a measurable distance before they decay. The IBL pixel size is $50 \times 250\,\mu\text{m}^2$ in the (ϕ, z)-plane.

In addition to the pixel layer of the IBL, the pixel detector consists of three pixel layers between $\rho = 50.5\,\text{mm}$ and $\rho = 122.5\,\text{mm}$ in the barrel region. The coverage of $|\eta| < 2.5$ is achieved by closing the barrel openings with two end-cap disks per side. The size of the pixels is $50 \times 400\mu\text{m}^2$ in the (ϕ, z)-plane in the barrel region and in the (ϕ, ρ)-plane for the end-cap disks.

In total, the IBL and Pixel components have approximately 80×10^6 readout channels. The IBL and Pixel detectors measure binary hit information, as well as, the time-over-threshold (TOT) related to the energy deposited in the pixel cell.

3.2.2. Semiconductor tracker

The silicon-based strip detector SemiConductor Tracker (SCT) encloses the Pixel detector. The barrel region consists of four double-sided layers. The strips on one side of the layers are parallel to the beam axis. The other side is rotated by an angle of 40 mrad to enable a position measurement in the z-direction. The layers extend from $\rho = 299\,\text{mm}$ to $\rho = 514\,\text{mm}$. Each module consists of two daisy-chained 64 mm long strips with a strip pitch of $80\,\mu\text{m}$. Both end-cap sides consist of nine double-sided disks. The strips on one side extend radially with an average strip pitch of $80\,\mu\text{m}$. The angle between the front and backside is 40 mrad. Similar to the Pixel detector, the SCT extends up to $|\eta| < 2.5$. In total, the SCT provides 6.3×10^6 readout channels. In contrast to the IBL and the Pixel detector, the SCT measures only binary hit information for each strip.

The SCT can be operated with different timing pattern requirements. The status of the strip in the previous and the next bunch crossing might be considered to determine if a hit is present in a strip. The notation for timing patterns is a triplet of either `0` (no hit), `1` (hit), and `X` (any or *don't care*). The first (last) position in the triplet denotes a condition on the strip during the previous (following) bunch crossing. The middle position represents the condition on the strip for the bunch crossing in question. For example, during Run 2, the SCT is operated in `01X` (edge-sensing) mode, which requires the absence of a hit in the previous bunch crossing, but no condition on the following bunch crossing. This terminology is relevant for the SCT noise study in Chapter 4.

3.2.3. Transition radiation tracker

The outer part of the ID, the Transition Radiation Tracker (TRT), consists of straw tubes to measure tracks in the transverse plane. The barrel area consists of 144 cm long tubes parallel to the beam axis with a diameter of 4 mm, arranged in 73 planes. In the end-cap regions, 37 cm long radial tubes are arranged in wheels. Each tube has a spacial accuracy of 130 μm. The relatively low resolution is compensated by a high number of hits. On average, a charged particle causes ionization in 36 tubes [50]. The space between the tubes is filled with materials of varying refractive indices which leads to transition radiation when traversed by charged particles. With a 70 % Xe gas mixture, the signal response for transition radiation is larger than the signal from ionization. The readout provides two discrimination thresholds to facilitate the distinction of transition radiation and ionization signals. The TRT provides coverage of $|\eta| < 2$. In total, the TRT has approximately 350×10^3 readout channels.

The emission of transition radiation depends on the relativistic factor $\gamma = \frac{E}{m}$ of a particle. This dependence is exploited for particle identification [57]. Typically, seven to ten high-threshold hits from transition radiation are detected for electrons with $E > 2\,\mathrm{GeV}$ [50]. At the design luminosity of the LHC $L = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, due to the large number of secondary particles, the occupancy of the TRT tubes can reach 60 %, posing a challenge for track reconstruction [58].

3.2.4. Integrated inner detector

The Inner Detector (IBL, Pixel, SCT, and TRT) is immersed in a magnetic field with $B = 2\,\mathrm{T}$ created by a superconducting solenoid between TRT and electromagnetic calorimeter. The field parallel to the beam pipe results in a curved trajectory of charged particles in the (ϕ, ρ)-plane which enables the measurement of their momentum. By reconstructing the bending radius r of a track, the transverse momentum p_T of singly charged particles is given by [57]

$$p_\mathrm{T} = \frac{0.3\,\mathrm{GeV}}{\mathrm{T}\cdot\mathrm{m}} \cdot B \cdot r. \tag{3.3}$$

The precision of the radius measurement deteriorates with increasing momentum as the track has a lower curvature. Therefore, the uncertainty on the momentum increases with increasing momentum. The relative resolution for momentum measurements in the ID is [50]

$$\frac{\sigma(p_\mathrm{T})}{p_\mathrm{T}} = \frac{0.05\%\, p_\mathrm{T}}{\mathrm{GeV}} \oplus 1\%, \tag{3.4}$$

where $a \oplus b$ denotes the addition in quadrature for statistically independent uncertainty sources.

3.2.5. Electromagnetic calorimeter

The electromagnetic calorimeter (ECal) sits outside the superconducting solenoid magnet. The ECal is a sampling calorimeter with liquid argon (LAr) as active material

and lead as absorber. The absorber medium and kapton electrodes are built in an accordion shape. The barrel part extends up to $|\eta| < 1.476$. The two end-caps cover $1.375 < |\eta| < 3.2$.

In the barrel region, The ECal is split into three cylinders that differ in granularity. The inner-most part (EM1) has a granularity of 0.0031×0.982 and in the (η, ϕ)-plane. This first part amounts to 4.2 radiation lengths (X_0). The very fine granularity in η-direction contributes to the identification of neutral pion decays since it allows the independent detection of both photons in the decay chain $\pi^0 \to \gamma\gamma$. This property is also beneficial for the identification of hadronically decaying τ leptons, see Section 3.3.5.

The middle part (EM2) of the ECal has an eight-fold decrease in η-granularity and a four-fold increase in ϕ-granularity compared to the inner layer. With 16 radiation lengths the middle part comprises the majority of the barrel ECal.

The outer part (EM3) has a reduced granularity of approximately 0.1×0.1 in the (η, ϕ)-plane. The thickness of the outer part corresponds to 2 radiation lengths

Each end-cap ECal consists of a 63 cm thick wheel extending from $\rho = 33$ cm to $\rho = 209.8$ cm. Over most of the η-range, the thickness of the end-caps is between 24 and 38 radiation lengths. The end-caps are split into three layers up to $|\eta| < 2.5$ with granularities matching the ones from the barrel region. Beyond $|\eta| = 2.5$, the wheels are divided into only two layers with a coarser granularity.

The relative resolution $\sigma(E)/E$ of an energy measurement with the ECal increases with increasing energy of the incident particles due to the statistical nature of the electromagnetic showers. The relative resolution of an energy E measurement is given by [50]

$$\frac{\sigma(E)}{E} = 10\%\sqrt{\frac{\text{GeV}}{E}} \oplus 0.7\%. \tag{3.5}$$

3.2.6. Hadronic and forward calorimeters

The ECal is enclosed by the hadronic calorimeter (HCal). The HCal is required to be thick enough to absorb and contain hadronic showers to prevent the punch-through of particles into the muon system.

The barrel section uses steel as an absorber material. The active material is scintillating tiles. The barrel HCal is therefore referred to as the Tile Calorimeter. It is segmented into three layers with 1.5, 4.1, and 1.8 interaction lengths (λ) in the central region. The Tile Calorimeter covers $|\eta| < 1.7$. It extends from $\rho = 2.28$ m to $\rho = 4.5$ m and is split into 64 azimuthal modules. The readout is facilitated with wave-length shifting fibers.

The end-cap HCal is built using copper (Cu) absorbers with LAr as the active medium. Each end-cap side consists of two wheels, each made of 32 identical wedge-shaped modules. The wheels are segmented into two layers, resulting in four HCal layers per side. The HCal wheels sit directly behind the ECal wheels. The end-caps cover $1.5 < |\eta| < 3.2$ and overlap with the barrel and forward part. Radially, the end-caps extend from $\rho = 47.5$ cm to $\rho = 203$ cm.

High particle fluxes and radiation levels are expected in the forward regions for $|\eta| > 3.2$ from elastic collisions and scattering with small momentum transfer. This region is instrumented with the dedicated forward calorimeter. The forward calorimeter

(FCal) extends the coverage up to $|\eta| < 4.9$. To prevent neutron contamination of the ID volume, the FCal is set back by 1.2 m. This limits the overall length and requires compensation by a denser material. The FCal is split into three layers. The inner layer is optimized for electromagnetic showers and uses copper as an absorber. The outer two layers measure hadronic showers and use tungsten (W) as an absorber. All layers have tubes parallel to the beam axis filled with rods to provide a narrow cylindrical volume for LAr as the active medium.

Similar to the electromagnetic calorimeter, the relative resolution of an energy measurement increases with increasing energies of incident particles. The relative energy E resolution in the barrel and end-cap region is [50]

$$\frac{\sigma(E)}{E} = 50\,\%\sqrt{\frac{\mathrm{GeV}}{E}} \oplus 3\,\% \tag{3.6}$$

and

$$\frac{\sigma(E)}{E} = 100\,\%\sqrt{\frac{\mathrm{GeV}}{E}} \oplus 10\,\% \tag{3.7}$$

in the forward region.

3.2.7. Muon system

The muon spectrometer (MS) comprises the outer-most shell of the ATLAS detector and defines its overall size. Over a large momentum range, muons interact only minimally with the detector material and escape the ID and calorimeters without being absorbed. The goal of the MS is to provide high precision track measurements of muons and trigger information. A benchmark measurement for the ATLAS detector design is the decay of a Higgs boson to four muons using only the MS.

The MS is immersed in an eponymous, toroidal magnetic field. In the barrel region for $|\eta| < 1.4$, the field is provided by a large, superconducting air-core toroid magnet. The magnet is built from eight coils giving the detector its characteristic shape. For $1.6 < |\eta| < 2.7$, superconducting, air-core magnets inserted into the end-caps extend the volume immersed in a magnetic field. The design ensures a magnetic field that is mostly orthogonal to the expected muon trajectories allowing a momentum measurement independent from the ID. The bending power is determined by the integral $\int \mathrm{d}s\, B$ along the muon trajectory s. In the barrel region with $|\eta| < 1.5$, the bending power ranges from 1.5 to 5.5 Tm. In the end-cap region with $1.6 < |\eta| < 2.7$, the bending power is between 1 and 7.5 Tm. In the transition region between end-cap and barrel, the integrated magnetic field is usually lower.

The MS in the barrel region consists of three layers located between $\rho = 5$ m and $\rho = 10$ m. The end-caps consist of three layers attached to three wheels per side. The outer-most wheels are located at $|z| = 21.5$ m. The majority of the precision space-point measurements stem from monitored drift tubes (MDT) with a resolution of 35 μm per chamber. For the end-caps with $2 < |\eta| < 2.7$, the precision tracking is facilitated with cathode strip chambers (CSC) that provide a high granularity and accommodate the higher expected muon flux in the end-cap regions. The resolution of the CSC is 40 μm. An optical alignment system is used to ensure a relative positioning accuracy between

adjacent modules of 30 μm. High momentum muons are used to ensure the overall alignment requirement of a few millimeters over the whole detector volume.

The relative momentum resolution for muons with a high transverse momentum is [50]

$$\left.\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right|_{p_{\mathrm{T}}=1\,\mathrm{TeV}} = 10\,\%. \tag{3.8}$$

Analogous to the ID, the relative resolution improves for smaller momenta. In the region $|\eta| < 2.4$, the MS is instrumented with resistive plate chambers (RPC) in the barrel region and thin gap chambers (TGC) in the end-caps. The RPC and TGC chambers are used to deliver trigger information. The MS trigger system also provides a bunch crossing identifier.

3.2.8. Trigger system

With typical instantaneous luminosity during Run 2, tens of proton–proton collisions occur at a frequency of up to 40 MHz. It is infeasible to transmit and record the detector response of every bunch crossing. In Run 2, the ATLAS detector uses a two-stage trigger system to reduce the data and event rate. The trigger system is a crucial component to maximize the use of the available bandwidth and to identify signatures of interesting events in real-time before the information is discarded.

The first stage of the trigger system (L1) is realized in custom hardware using Field Programmable Gate Arrays (FPGAs). The L1 trigger system uses information from the calorimeters and the muon system. Events accepted by the L1 trigger system are read out and transferred to the high-level trigger. The maximal L1 trigger output rate is 100 kHz. The system must reach a decision to accept the event within 2.5 μs [50].

The high-level trigger (HLT) [59] consists of a computer farm having access to information from all detector components. The HLT performs a partial reconstruction of the events accepted by the L1 trigger similar to the offline event reconstruction. Events that pass the HLT are recorded permanently. The average output rate is 1 kHz. The reconstruction time is a function of the instantaneous luminosity. With the conditions in 2015, the average processing duration reaches approximately 235 ms.

The central trigger system splits the data taking process into luminosity blocks as a measure in time [60]. Each luminosity block is approximately one minute long and the detector and data taking conditions including the instantaneous luminosity are assumed to be constant over the whole luminosity block. A new luminosity block is started prematurely whenever the conditions change.

3.3. Particle reconstruction and identification

The detector response is processed by reconstruction and identification algorithms to convert the raw hit and calorimeter cell information into physics objects like electrons or jets with kinematic properties. Depending on the type of the physics object, a different subset of detector components needs to be considered. The reconstruction and identification procedure is presented in the following sections for each object type. Unless otherwise noted, this section is based on Reference [21].

3.3.1. Track reconstruction

On a superficial level, the track reconstruction algorithm consists of three stages. The first step is to cluster the hit information from the silicon layers. In the next stage, patterns are recognized to form seed tracks. The last stage selects between several possible track candidates and thereby solves the inherent ambiguity of hit-to-track assignment. The reconstruction algorithm is described in detail in References [61, 62].

Adjacent pixels or strips that are above their discrimination threshold are grouped into clusters. The clusters are converted in turn to three-dimension space-points. For the SCT, the cluster information from both sides is combined to derive the space-point location. In the high-density track environment of Run 2, it is challenging to distinguish clusters with close proximity. Clusters from two charged particles that overlap and cannot be distinguished are called merged clusters. If two candidate tracks compete for a cluster that originated from a single charged particle, it is termed a shared cluster.

To maximize the track quality, the algorithm starts with identifying seed tracks using only the SCT, then only the pixel detector, and finally the combination of both. A combinatorial Kalman filter is used to build track candidates from the seed tracks, including information from the remaining detector components. The procedure can lead to a set of candidate tracks that compete for the same shared cluster. The next step solves this ambiguity.

A track score is assigned to each track that quantifies the likelihood that the track corresponds to the trajectory of a charged particle. Each cluster assigned to a track increases the track score. The track score is decreased if the track passes active detector material without a cluster. The absence of a cluster is called a *hole*. Holes in inactive regions of the detector, e.g., due to dead modules or at the edge of a sensor, do not affect the track score. This property of the algorithm is exploited in the SCT noise study in Chapter 4 for the mark-as-dead strip treatment. Finally, the χ^2 of the track fit, and the momentum of the track is taken into account. Low momentum tracks are penalized since they typically correspond to incorrect cluster assignments. The track scoring is repeated iteratively when the shared cluster assignment is updated. A neural network is consulted to distinguish shared and merged clusters [63]. Unlike in the case of a shared cluster, two tracks with a merged cluster are not penalized.

The final selection requirements impose conditions on the number of shared clusters, transverse momentum, the pseudorapidity range, the number of pixel and SCT hits, the number of holes in both silicon detectors, the impact parameter, and the longitudinal distance measured with respect to the beamline. Tracks that pass these criteria are subject to a high precision track fit taking all available information into account.

3.3.2. Electron and photon reconstruction

Electrons and photons lose a significant amount of their energy due to bremsstrahlung and pair production, respectively, before they reach the ECal. The radiation of bremsstrahlung photons is problematic since they can split into subsequent electron–positron pairs. Usually, this radiation is limited to a narrow cone and deposits the energy in the same calorimeter cluster. Electrons and photons have a similar signature in the ECal. Therefore, the reconstruction of photons and electrons proceeds in parallel.

The reconstruction of electrons and photons uses a variable-sized cluster termed supercluster [64]. With superclusters energy loss from bremsstrahlung and pair production can be better recovered compared to the previously used fixed-sized sliding window approach [65, 66]. The algorithm selects connected calorimeter cells above a noise threshold and all its neighboring cells. If the supercluster exhibits two local maxima the cluster is split into two. The track reconstruction follows the standard ID track reconstruction using regions of interest from fixed-size ECal clusters. The track fitting is modified to better account for conversion tracks. Electrons are built from superclusters that are matched to a track. Photons are built form superclusters that match a track from a conversion vertex or do not match any track. A dedicated energy calibration is used for electrons and photons.

3.3.3. Jets

During the hard-scattering, colored particles emerge from the interaction region and hadronize. In the detector, this is visible as a large number of particles usually referred to as a jet. The precise definition of a jet depends on the algorithm used to identify the jet. The algorithm is required to be invariant under collinear radiation of the emerging particles and invariant under soft radiation. In a jet, a variety of physics objects can occur, such as charged and neutral hadrons but also electrons, photons, and muons. One class of algorithms suitable to cluster the energy deposits and tracks to form jets are the anti-k_t algorithms [67]. The algorithms iteratively cluster objects i defined by their 4-momenta p_i into cone-like objects. The algorithm defines two distance measures between pairs of objects i, j under consideration

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2p},\ p_{\mathrm{T},j}^{2p}\right) \cdot \frac{\Delta R_{ij}^2}{R^2} \tag{3.9}$$

$$d_{i\mathrm{B}} = p_{\mathrm{T},i}^{2p} \tag{3.10}$$

where the subscript B refers to the beam line and R is a free radius parameter of the algorithm. The momenta are taken to the $2p$-th power, where p is an integer parameter of the algorithm. The algorithm proceeds as follows: Sort the distances d_{ij} and $d_{i\mathrm{B}}$ for all objects under consideration. If the smallest distance is of type $d_{i\mathrm{B}}$, declare object i a jet and remove it from the list. Otherwise, cluster the objects i and j with the smallest distance together into a combined object and remove it from the list. The steps are repeated until every initial object is either clustered or declared a jet. In anti-k_t algorithms, p is negative. We use $p = -1$.

The algorithm can be applied to different input collections. The main part of this thesis uses PFlow jets that were introduced during Run 2 with $R = 0.4$ [68]. The energy deposition in topological clusters [69] of the calorimeter that match a track in the ID is replaced by the momentum measurement from the ID. The anti-k_t algorithm is applied to clusters that do not match any track and all tracks from the primary vertex. PFlow jets are an improvement over previous jet algorithms used within the ATLAS Collaboration as they combine knowledge from both ID and calorimeters. Jets are calibrated with a series of corrections to restore their scale to that of jets formed from stable hadrons with the same algorithm [70].

3.3.4. Muon reconstruction

The reconstruction of muon tracks starts with a search considering hits only in the MS. The found tracks are then subject to a combination with information of the other detector systems. A detailed description of the muon reconstruction in the ATLAS Collaboration can be found in Reference [71].

Initially, track segments are reconstructed from hits within a single module. Track candidates are built by combining track segments from different layers of the MS. For this, a combinatorial search is performed starting from the middle layer of the MS. In the barrel, at least two segments are required for a track. Segments are selected based on a combination of hit multiplicity, fit quality, and geometric considerations. A global χ^2 fit is applied to the hits of candidate tracks. Outlier hits are discarded while also recovering hits that are compatible with the trajectory. Tracks are required to satisfy quality of fit criteria.

For the analysis of Higgs boson decays to a pair of tau leptons, muon track candidates from the MS are extrapolated to the ID and matched with ID tracks. The combination is refit with a global track fit.

3.3.5. Hadronic tau leptons

The tau lepton with a mass of $m_\tau \approx 1.777\,\text{GeV}$ decays to lighter particles after a mean lifetime of approximately $3 \times 10^{-13}\,\text{s}$ [14], which amounts to an average travelled distance of approximately 0.5 mm for a tau with 10 GeV momentum. We can distinguish leptonic tau decays $\tau \to \ell\bar{\nu}_\ell\nu_\tau$ with $\ell = e, \mu$ and hadronic tau decays $\tau \to \text{hadrons}\,\nu_\tau$. Leptonically decaying taus are reconstructed as electrons or muons in the ATLAS detector. The small displacement of the tau vertex from the primary vertex has not been exploited for leptonic decays in the analysis presented in this thesis, so the same reconstruction is used as for prompt electrons and muons. Hadronically decaying taus, denoted τ_{had}, usually produce a shower in the ECal that extends to the HCal. The visible decay products of a hadronically decaying tau are referred to as $\tau_{\text{had-vis}}$, which does not include the tau neutrino. For the remainder of this section, the tau neutrino is neglected. The majority of hadronic tau decays can be categorized in five decay modes as defined in Reference [72] with either one or three charged hadrons ($h^\pm$, primarily $\pi^\pm$ but also $K^\pm$ and other mesons) and up to two neutral pions (π^0). For the analysis presented in Chapter 5, we are primarily interested in the charged hadron multiplicity, the so-called prongness. The charged hadrons shower and deposit their energy inside a narrow cone in the ECal and HCal. The fractional energy deposition between the ECal and HCal layers fluctuates. The neutral pions decay to a pair of photons in 99 % of the cases [14]. The fine granularity of the EM1, see Section 3.2.5, permits resolving each photon separately, depending on the decay plane and the momentum.

A special algorithm is used to reconstruct hadronically decaying taus [73]. A recurrent neural network (RNN) [40] is used to identify reconstructed hadronically decaying taus and reject potential contamination from quark and gluon-initiated jets [74]. These two algorithms are detailed in the following. Alternative reconstruction and identification algorithms, as well as efficiency measurements, can be found in References [72, 73, 75].

The reconstruction is seeded with jets from the anti-k_t algorithm with $R = 0.4$. For

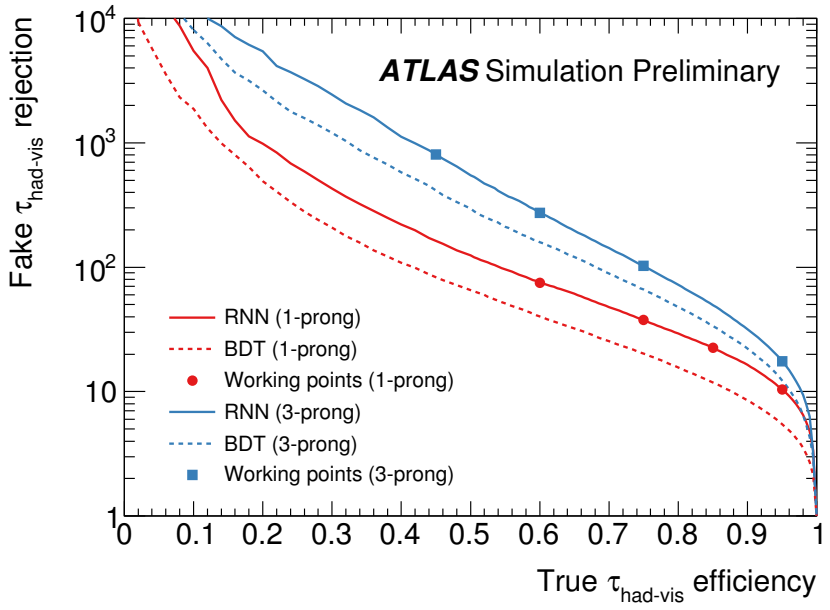


Figure 3.3.: Background rejection as a function of hadronic tau identification efficiency for the RNN used in this thesis and a previously used BDT approach. The dots indicate the selected efficiency working points [74].

a seed jet to serve as a tau candidate, it needs to satisfy $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$. The barycenter of calibrated topological clusters of the calorimeter is used to define the jet direction. Tracks within a cone of $R = 0.2$ and $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ are used to find the secondary vertex with the largest momentum sum. Boosted Decision Trees are used to classify tracks into a *core* and an *isolation* region. The number of charged tracks in the core region determines the prongness of the τ candidate. Finally, an energy calibration employing a Boosted Regression Tree is applied.

An RNN is used to identify hadronically decaying taus and to distinguish them from quark or gluon-initiated jets. The network is split into three input branches. The first two branches use a Long Short-Term Memory (LSTM) architecture [40] to accommodate a variable number of tracks and clusters. The last input branch uses high-level quantities. The network is trained on an MC sample with $\gamma^* \to \tau\tau$ events. Its performance is assessed on a $\gamma^* \to \tau\tau$ and a dijet sample.

The identification algorithm defines multiple efficiency working points. The tight working point is defined with a signal efficiency of 60 % (45 %) and achieves a quark and gluon jet rejection factor[1] of 70 (700) for one-prong (three-prong) taus, see Figure 3.3. An additional Boosted Decision Tree is used to veto electrons misidentified as one-prong hadronically taus.

[1]The rejection factor is defined as the average number of jets out of which only a single jet passes the selection.

3.3.6. Missing transverse momentum

The momenta carried by the partons in the colliding protons follow the PDFs introduced in Section 2.2.1. The center-of-mass of the hard scatter system is at rest with respect to the transverse plane, but due to the unknown momentum fractions of each parton, the center-of-mass system has an unknown Lorentz boost along the beam axis. Imposing momentum conservation on the decay process, we can measure momentum carried away by undetected particles as a momentum imbalance in the transverse plane. The missing transverse energy $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is a vector in the transverse plane quantifying momentum that is missing to restore the transverse momentum balance. Typical examples of SM processes with sizable $E_{\mathrm{T}}^{\mathrm{miss}}$ involve neutrinos that escape the detector without interaction.

The quantity $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is computed from calibrated hard scattering objects (electrons, muons, hadronic taus, photons, and jets), and a soft term from all tracks originating from the primary vertex not associated with any hard scatter object. The missing transverse momentum

$$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} = -\left(\sum_{i \in \text{hard scattering}} \boldsymbol{p}_T^i + \sum_{j \in \text{soft term}} \boldsymbol{p}_T^j \right) \tag{3.11}$$

is a two-dimensional vector in the transverse plane often described by its magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$ and its ϕ-orientation. Although $E_{\mathrm{T}}^{\mathrm{miss}}$ is a projection of a three-dimensional momentum, for historical reasons, it is referred to as an *energy*.

3.3.7. Mass reconstruction

The final states of $Z \to \tau\tau$ and $H \to \tau\tau$ are characterized by $E_{\mathrm{T}}^{\mathrm{miss}}$ attributed to the neutrinos created in the subsequent τ lepton decays. The neutrinos escape the detector without interaction and carry away momentum. The initial proton-proton system has no transverse momentum component, which makes it possible to identify $E_{\mathrm{T}}^{\mathrm{miss}}$ as the vectorial sum of momenta carried by the neutrinos in the final state, see Equation (3.11). Depending on the combination of τ decays, we expect between two (fully hadronic decay) and four (fully leptonic decay) neutrinos in the final state. Without further assumptions or approximations, it is not possible to disentangle the contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$ or reconstruct the individual momenta of the neutrinos.

The invariant mass of the initial Z or H boson is important information for the $H \to \tau\tau$ analysis. Reconstructing the invariant boson mass requires knowledge of the 4-momenta of all decay products, including the invisible neutrinos. Three alternative methods to overcome this limitation are described briefly. The methods compute an estimator of the boson's mass.

The simplest approach is to compute the invariant mass $m_{\tau\tau}^{\mathrm{vis}}$ of the visible decay products: the light leptons ($\ell = e, \mu$) in case of leptonically decaying taus and the visible part of the hadronically decaying taus ($\tau_{\text{had-vis}}$). The visible mass $m_{\tau\tau}^{\mathrm{vis}}$ of the di-τ system is characterized by a large bias towards smaller masses since this approach neglects the effect of the neutrinos in the final state. The distribution of $m_{\tau\tau}^{\mathrm{vis}}$ is relatively wide, directly reflecting the distribution of neutrino 4-momenta.

A more refined approach, the collinear mass approximation [76], uses the assumption that the τ leptons from Z or H boson decays have a large momentum-to-mass ratio

and therefore a large Lorentz boost. The Lorentz boost of the τ leptons restricts their subsequent decay products to a narrow cone. The collinear mass approach further approximates the directions of the neutrinos and the visible decay product to be identical, i.e., collinear, and that the neutrinos are the only source of $E_{\mathrm{T}}^{\mathrm{miss}}$. The assumptions are partially motivated by the large rest mass of the Z ($m_Z \approx 91\,\mathrm{GeV}$) and Higgs bosons ($m_H \approx 125\,\mathrm{GeV}$) compared to the mass of the two τ leptons ($m_\tau \approx 1.8\,\mathrm{GeV}$). However, the approximation works best in case of boosted Z and H bosons which introduces an additional Lorentz boost for the tau leptons. It is common to parametrize the momentum of the visible decay products

$$p_i^{\mathrm{vis}} = x_i p_i^{\tau} \tag{3.12}$$

as a fraction x_i of the tau lepton momentum p_i^{τ} for each tau, $i = 1, 2$. With this parametrization, the sum of expected transverse neutrino momenta can be equated with the measured $E_{\mathrm{T}}^{\mathrm{miss}}$

$$\sum_{i=1,2} \left(\frac{1}{x_i} - 1 \right) \boldsymbol{p}_{\mathrm{T},i}^{\mathrm{vis}} = \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} \tag{3.13}$$

in the transverse plane and solved for the unknown parameters x_1 and x_2. The solutions can be used to determine the tau momentum, including invisible decay products in the collinear approximation, and thus reconstruct the initial mass $m_{\tau\tau}^{\mathrm{col}}$. The collinear mass approximation is characterized by a smaller bias compared to the naive visible mass $m_{\tau\tau}^{\mathrm{vis}}$ and a better resolution.

The most involved and computationally expensive approach is the Missing Mass Calculator (MMC) [77]. The algorithm uses parametrized probability density functions of kinematic properties of the visible and invisible decay products derived from simulated $Z \to \tau\tau$ events. From this, a global probability P can be computed for any event topology with hypothesized neutrino momenta. At the LHC, the limited detector resolution and momentum mismeasurement introduce additional contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$. The algorithm accounts for this by smearing the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the candidate topology with an estimate of its resolution. Finding the most likely configuration is equivalent to maximizing the event probability P with respect to the neutrino momentum hypotheses. The MMC employed in Chapter 5 uses a Metropolis-Hastings algorithm [41, 78] to perform a random walk in the multidimensional parameter space. Each step depends only on the current position, thus forming a Markov chain. The transition probability is larger towards positions that maximize the probability P. The momentum configuration that maximizes P is taken to compute the invariant mass $m_{\tau\tau}^{\mathrm{MMC}}$ of the di-τ system.

The MMC is able to reconstruct the invariant mass of Z or H bosons decaying to a pair of taus with very little bias and sufficient resolution to achieve good separation of the two processes. A comparison of reconstructed masses applying the three methods is shown in Figure 3.4. The comparison uses events from all signal regions defined in Chapter 5. The bosons in a large fraction of the events have a transverse boost of $p_{\mathrm{T}}^{H,Z} > 100\,\mathrm{GeV}$ improving the resolution of $m_{\tau\tau}^{\mathrm{col}}$.

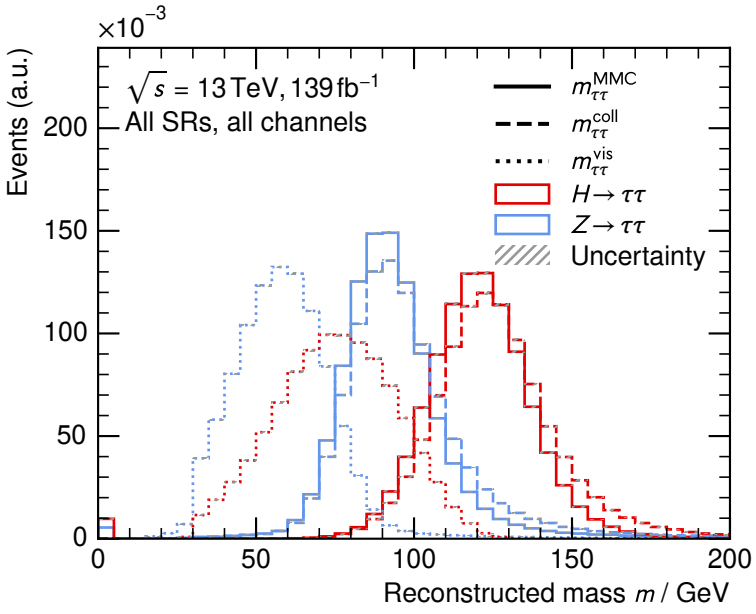


Figure 3.4.: Comparision of different methods to reconstruct the di-τ system mass in Z and Higgs boson decays.

CHAPTER 4

SCT noise threshold optimization

Understand well as I may, my comprehension can only be an infinitesimal fraction of all I want to understand.

— Ada Lovelace

Like every ATLAS detector component, the SCT is affected by noise. Noise can lead to a situation where the signal from a strip exceeds the discrimination threshold, although no charged particle traversed the detector medium. The recorded hits caused by noise are termed fake hits. With almost 6.3×10^6 strips in total, the effects of noise need to be understood and considered during the operation of the detector and the reconstruction of data. Noisy strips are identified and masked such that hits in noisy strips are not considered during reconstruction. This procedure and its parameters predate Run 1. This section details the masking procedure, optimizes its performance, and studies possible alternatives that might be more suited for Run 2 or even Run 3 conditions. The study uses data from typical LHC fills during Run 2. Potential optimizations were considered during Run 2, could be implemented in a reprocessing of Run 2 data, or could be used as a baseline for Run 3 planned to start in 2022.

4.1. Noise and strip masking procedure

For the SCT strip detector, different noise sources need to be considered. The measured signal is generated by electron–hole pairs in the silicon when charged particles interact with the detector medium. The induced mirror charges [57] in the readout electrodes are amplified with a charge-to-voltage circuit and discriminated against an adjustable voltage threshold. The signal discrimination is performed with an analog comparator that compares the amplified voltage of each strip to a per-strip threshold voltage created by a digital-to-analog converter. This design allows adjusting the discrimination threshold on a per-strip basis by reprogramming digital registers, the so-called `TrimDAC` registers.

The total number of created electron–hole pairs is subject to statistical fluctuations. The analog signal amplification and processing chain introduces additional noise (thermal noise, shot noise, and the frequency-dependent $1/f$-noise). Reference [57] offers a more detailed discussion of noise originating from the statistical charge generation process and the electronics. The amount of noise increases over the lifetime of the SCT with increasing instantaneous radiation and accumulated radiation damage to the modules.

The statistical component and noise introduced by the analog circuitry play an indirect role in the discussion of this study. The dominant cause of a noisy strip is a single event upset (SEU) in the digital discrimination circuit. An SEU occurs when charged particles interact with the digital circuit and change the binary state of a register.

We are concerned with SEUs that change the registers that define the discrimination threshold, for this study. If a bit flip occurs in the `TrimDAC` register that defines the threshold for a strip, electrical noise in the strip can cause the signal to frequently (or continuously) exceed the threshold without any generated electron–hole pairs from charged particle interaction. The nature of this kind of noisy strips is different from noise from statistical or electrical sources. A strip starts to be noisy with an SEU and continues to be noisy until the threshold registers are reprogrammed and recalibrated every few days or weeks.

Noisy strips can introduce issues in the data reconstruction. The SCT is monitored to identify and mask noisy strips. The occupancy y, defined as the probability to observe a hit in a strip, is a crucial quantity to identify noisy strips. In the prompt calibration loop of the SCT, the occupancy is estimated individually for each strip with

$$y = \frac{\text{number of hits}}{\text{number of bunch crossings}} \tag{4.1}$$

considering only empty bunch crossings, i.e., time windows when gaps in the bunch trains coincide within the detector such that no proton–proton collisions occur inside the detector. Hits registered during an empty bunch crossing are considered to originate from noise.

The SCT was designed with an occupancy $y^{\text{design}} < 5 \times 10^{-4}$ [50] for fully irradiated modules. Based on y^{design}, the threshold

$$q = 3 \times 10 \times y^{\text{design}} = 1.5\,\% \tag{4.2}$$

for the occupancy was defined. Any strip whose occupancy y exceeds the threshold q is marked as noisy in the calibration loop. The factor 3 in Equation (4.2) is motivated by the fact that hit detection considers three timing bins as introduced in Section 3.2.2 on the operation modes of the SCT. The quantity y mentioned in Equation (4.1) is subject to statistical fluctuations. The factor 10 is introduced as a ten-fold margin to avoid continual classification of strips as noisy.

Hits in strips that were marked as noisy are removed before reconstructing tracks. This means noise hits do not interfere with track reconstruction, and noise hits are effectively treated. At the same time, erroneously removed true hits, i.e., hits from charged particle interactions in strips that were marked as noisy, manifest themselves as a missing hit, a so-called *hole*, during track reconstruction and can lead to a worse

Table 4.1.: Summary of physics runs reprocessed for the strip noise threshold optimization. The duration of the run is quantified as the number of its luminosity blocks. The minimal and maximal number of pile-up interactions μ are rounded to the nearest integer.

Run number	Year	Month	Duration	min μ	max μ
338834	2017	October	90	37	57
338846	2017	October	907	22	58
351698	2018	June	268	23	54

tracking performance. The abundance of these artificial holes is expected to be far smaller than the abundance of fake hits from noise without the noisy strip masking.

This study assesses the impact of different noisy strip thresholds q and alternative treatments of noisy strips to avoid the problem of artificial holes. The parameters are adjusted to optimize the tracking performance. This optimization is expected to be a trade-off between tracking precision, i.e., the geometric accuracy of reconstructed tracks, and tracking efficiency, i.e., the probability of finding the track of a charged particle interaction.

The study is performed by using three exemplary runs as listed in Table 4.1. For each run, the calibration loop is rerun with the default threshold $q = 1.5\,\%$ and the alternative thresholds $q = 0.15\,\%, 0.5\,\%, 5\,\%$ to obtain lists of noisy strips depending on the threshold. The events from each run are then reconstructed with the alternative masked strip lists. The number of noisy strips as a function of the threshold is shown in Figure 4.1. A priori, it is not clear if a higher or lower threshold leads to an improvement of the tracking performance. A higher noise threshold leads to more fake hits that might reduce the track performance, but on the other hand, a higher threshold also reduces the number of artificial holes. The inverse is true for a lower threshold.

4.2. Figure of merit

The goal of the study is to optimize the tracking performance. Two figures of merit are used to assess the performance:

- the number n of tracks passing the tight selection criteria as defined in Table 4.2,
- the residual distance R measured in μm, called the *residual*, between the SCT strip and the intersection of the fitted track and SCT module plane.

As stated earlier, the simultaneous optimization of both quantities might not be possible. The optimization should maximize the number of tracks and minimize the residual. In the case of a correlation between both figures of merits, the optimization transforms into a trade-off decision.

The number of selected, good tracks is readily available by inspecting the reconstructed event. It is assumed that the number of tracks is approximately proportional to the tracking efficiency and, therefore, a good proxy.

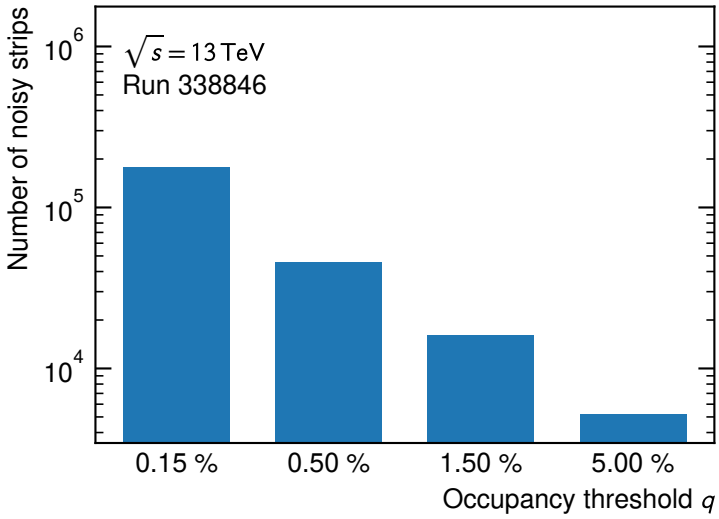


Figure 4.1.: The number of strips identified as noisy in run 338846 using four different noise thresholds q. In total, there are approximately 6.3×10^6 strips in the SCT.

Table 4.2.: Summary of the selection criteria for tight and loose tracks as defined in Reference [79].

	Loose track	Tight track
Transverse momentum p_{T}	> 400 MeV	> 400 MeV
Pseudorapidity $\lvert\eta\rvert$	< 2.5	< 2.5
Number of hits in silicon detectors	≥ 7	≥ 9 (for $\lvert\eta\rvert < 1.65$)
		≥ 11 (for $\lvert\eta\rvert > 1.65$)
Number of shared modules	≤ 1	≤ 1
Number of holes in silicon detector	≤ 2	≤ 2
Number of holes in pixel detector	≤ 1	0
Number of hits in IBL or 2$^{\mathrm{nd}}$ pixel layer		≥ 1

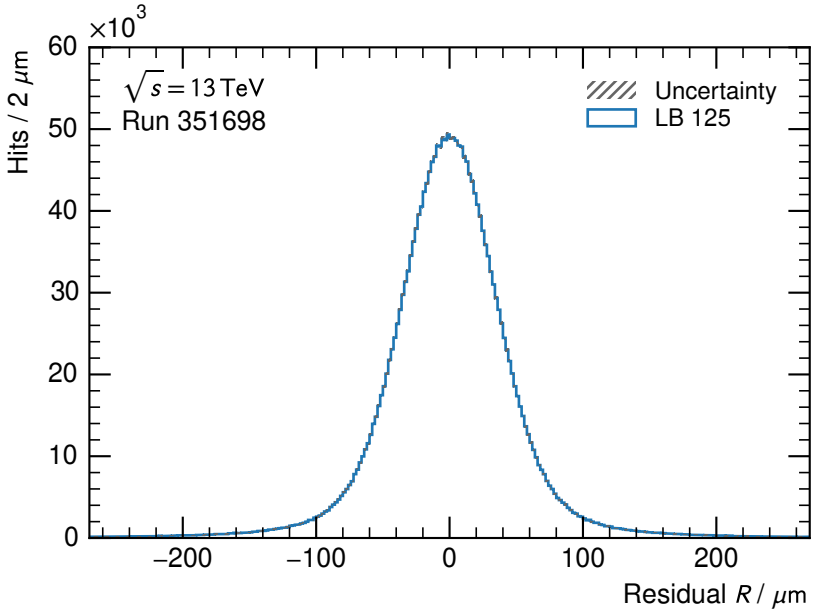


Figure 4.2.: Typical distribution of residuals for a single luminosity block of run 351698. The mean of the distribution is $-0.10\,\mu$m with an RMS width of $56.52\,\mu$m. The statisticaly uncertainties in each bin are too small to be visible.

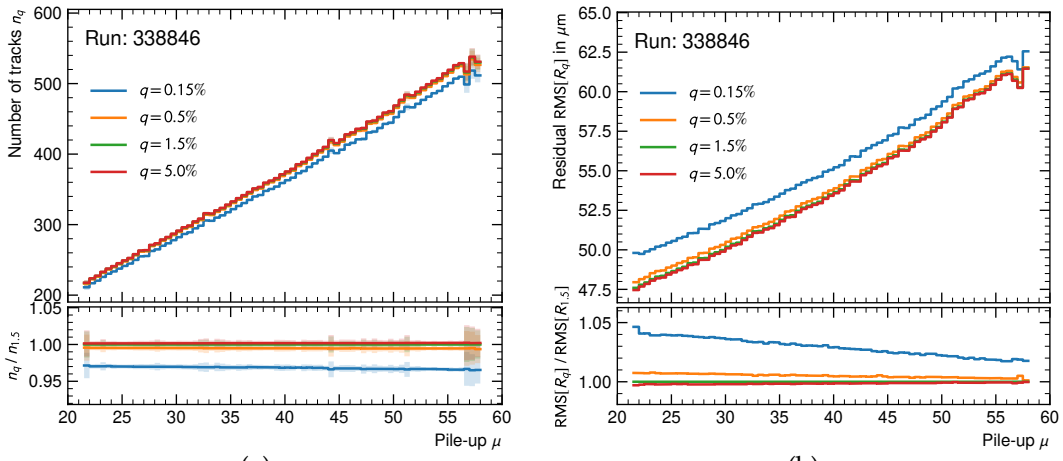
The residual is used as a measure for the quality of the fitted track. With the residual, we can detect cases when a track connects hits from two different charged particles leading to charge misidentification and other issues. As shown in Figure 4.2, the residual is an oriented quantity with its directions defined by the SCT module geometry and orientation. For this study, we are only interested in the width of the distribution taking outliers into account. Therefore we use the root-mean-square (RMS) of the residuals R_i of hit i

$$\mathrm{RMS}[R] = \sqrt{\frac{1}{n_{\mathrm{hits}}} \sum_{\mathrm{hits}} (R_i - \bar{R})^2} \tag{4.3}$$

centered around the average residual $\bar{R}$ as the figure of merit. Often the width of the distribution RMS$[R]$ is also referred to as the residual.

The values for the residual and the number of tracks are obtained as averages for all events in a luminosity block. For each luminosity block, the values are paired with the average number of pile-up interactions μ. This enables the evaluation of the optimization as a function of pile-up interactions μ. Due to beam leveling, most luminosity blocks have a high number of pile-up interactions. Therefore the study has more statistical power in the regime of many pile-up interactions.

The presentation here focuses on the performance in the barrel region of the SCT. It has been confirmed that the conclusions drawn by considering the barrel region are also valid for the SCT end-caps. The tracks used in this study pass the tight track



(a) (b)

Figure 4.3.: The number of reconstructed tracks (a) and the residual in the barrel region (b) for run 338846 for different noise thresholds as a function of the number of pile-up interactions. The shaded bands indicate the statistical uncertainty.

selection criteria. It has been checked that the results obtained with tight tracks also apply for the loose track selection as defined in Table 4.2.

4.3. Threshold dependence

Figure 4.3 shows the figures of merit as a function of the pile-up μ for the different thresholds q for a typical physics run, here run 338846. The equivalent plots for the other runs listed in Table 4.1 are shown in Figure A.1 in the appendix.

It can be observed that lowering the threshold decreases the number of tracks and increases the residual. A lower threshold leads to more strips marked as noisy, which leads to more true hits being removed from noisy strips. The overall smaller number of hits causes a smaller number of tracks to pass the track identification criteria. On the other hand, fewer hits will lead to tracks connecting hits from different charged particle interactions and, therefore, a larger residual.

The default threshold and the threshold $q = 5\,\%$ have similar performance characteristics. Their figures of merit differ only by 1 %. The most significant difference from the performance with the default threshold can be observed with the lowest threshold $q = 0.15\,\%$. The lowest threshold leads to a 3 % decrease in the number of tracks with only a small dependency on the pile-up μ, and a 2 % to 5 % increase in the residual depending on the pile-up μ.

Considering both figures of merit, the tracking performance increases with an increased threshold. Therefore, only the default threshold and the noise threshold $p = 5\,\%$ are considered for the alternative noisy strip treatment studied in the next section.

The number of tracks appears to be approximately proportional to the average number of pile-up interactions μ. The residual shows a similar dependence on the pile-up μ. The dependence is much larger than the differences introduced by varying the noise

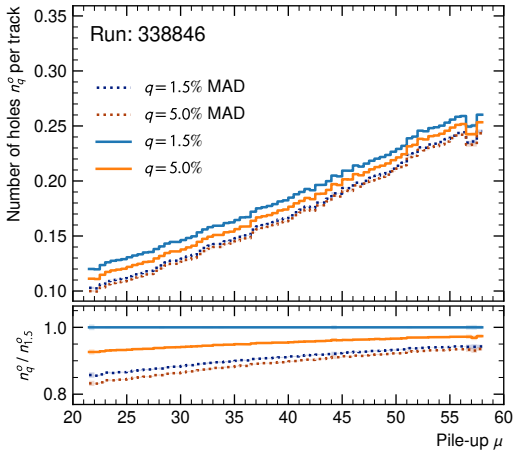


Figure 4.4.: Number of holes per track for the default noisy strip treatment and the alternative mark-as-dead procedure for run 338846 as a function of the number of pile-up interactions. The shaded bands indicate the statistical uncertainty, however, for most bins, the uncertainty is too small to be visible.

threshold q. This effect is not naively understood and will be discussed in Section 4.5.

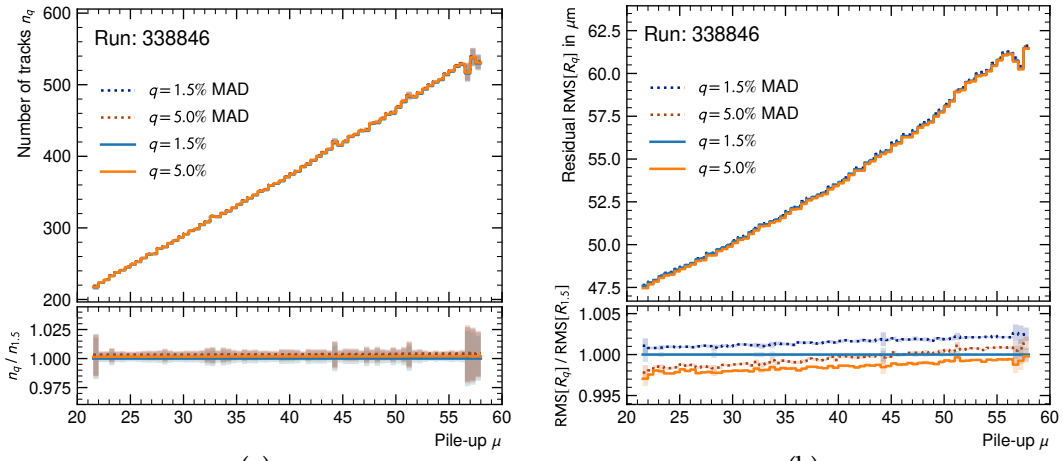
4.4. Alternative noisy strip treatment

As already mentioned, for the track reconstruction, hits in SCT strips marked as noisy are removed. The reconstruction algorithm has no information about the list of noisy strips, such that there is no difference between a non-noisy strip without a hit and a noisy strip with removed true or fake hits. If a charged particle traverses a noisy strip, the strip masking procedure will inadvertently create a hole.

An alternative procedure is called mark-as-dead (MAD). The idea is to mark noisy strips as *dead*, which is a property of each strip propagated to the reconstruction algorithm. Missing hits in dead strips do not penalize a track fitted through it, see Section 3.3.1. The alternative procedure is expected to decrease the number of holes per track and increase the track efficiency. For this study, the runs listed in Table 4.1 are reprocessed with the default threshold and the alternative threshold $q = 5\,\%$, both with the alternative treatment strategy.

Figure 4.4 shows the number of holes n^o per track. As a sanity check, it can be verified that for a given noise threshold q the number of holes decreases by switching to the MAD strategy. The figures of merit for the default and the MAD strategy for run 338846 are shown in Figure 4.5. The number of tracks n shows no significant dependence on the choice of the procedure considering the statistical uncertainties. The residual increases between 0.1 % and 0.3 % depending on pile-up μ when using the alternative MAD procedure. The effect is smaller for a low number of pile-up interactions and for the $q = 5\,\%$ threshold. It has been confirmed that the other runs in Table 4.1 agree with the example shown in Figures 4.4 and 4.5.

In conclusion, the alternative procedure reduces the number of holes per track. How-



(a) (b)

Figure 4.5.: The number of tracks (a) and the residual (b) for the default noisy strip treatment and the alternative mark-as-dead procedure for run 338846 as a function of the number of pile-up interactions. The shaded bands indicate the statistical uncertainty.

ever, this does not lead to the expected performance improvement. Considering the residual R, using the alternative procedure MAD results in a worse performance. The results suggest that the current reconstruction algorithm can handle artificially introduced holes from noisy strips very well.

4.5. Pile-up μ dependence

Figure 4.3(b) shows the residual R for different noise thresholds q as a function of the number of pile-up interactions μ. The residual R shows a much larger dependence on μ than on the noise threshold q.

One can argue that with an increasing number of tracks, from pile-up interactions and the hard scattering, the chance increases that two tracks are close in the detector and a hit is associated with the wrong track. This effect increases the residual R. However, one would also expect an increase in hits per track. Figure 4.6 shows the number of hits n^{h} per track decreasing with an increasing number of pile-up interactions. The findings were confirmed internally by the ATLAS tracking group.

The number of hits per tracks as a function of pile-up can be understood by taking additional effects into account. In Run 2, the SCT is operated in edge-sensing mode, i.e., the SCT timing pattern is `01X`. If a track creates a hit in a strip, it *shadows* a hit in the next bunch crossing in the same strip leading to a hole. With an increasing number of pile-up interactions, the chance of a pile-up track shadowing a hit in the next bunch crossing increases. This effect reduces the number of hits per track with increasing pile-up μ.

The combination of these two effects, pile-up hits increasing the residual R and pile-up tracks shadowing hits, are sufficient to explain the aforementioned observations qualitatively.

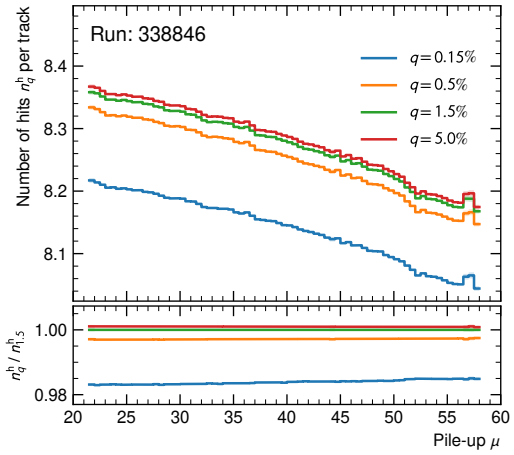


Figure 4.6.: The number of hits per track for different noise thresholds q for run 338846 as a function of the number of pile-up interactions. The shaded bands indicate the statistical uncertainty.

To verify that the unexpected dependence of the residual R on pile-up is modeled correctly in events from Monte Carlo simulation, simulated $t\bar{t}$ events are created for eight fixed values of the number of pile-up interactions $\mu = 20, 25, \ldots 55$. The residual R as a function of pile-up μ is shown in Figure 4.7 for run 338846 and the simulated $t\bar{t}$ sample. The simulated residual is 2 % to 5 % smaller than the one measured in data. Qualitatively, the increasing trend is modeled correctly in Monte Carlo.

4.6. Conclusion

The results of the noisy strip optimization are summarized in Figure 4.8. The summary plot shows the two figures of merit for different noise thresholds q, different physics runs, and the alternative noisy strip treatment procedure. The values are averaged over the pile-up interval $\mu \in [52, 54]$. The points should be in the lower right corner to improve the tracking performance for both figures of merit. We see an improvement in both figures of merit by raising the threshold from $q = 1.5\,\%$ to $q = 5\,\%$. However, the improvement is small compared to other effects, such as run-to-run fluctuations.

The alternative noisy strip masking procedure improves the tracking performance with regard to the number of tracks but deteriorates the performance with respect to the residual R. It is ambiguous whether the alternative procedure leads to a better tracking performance.

Weighting the small relative improvement to the large amount of computing resources required to reprocess the full Run 2 dataset with the updated threshold, it has been decided to keep the threshold at $q = 1.5\,\%$ for Run 2.

In the upcoming data-taking period of the LHC, Run 3, the average amount of pile-up events, the fluence, and therefore also the radiation damage is expected to increase. It might be necessary to increase the noise threshold q to prevent marking a large number of strips as noisy. In that context, the study provides valuable insight. Increasing the

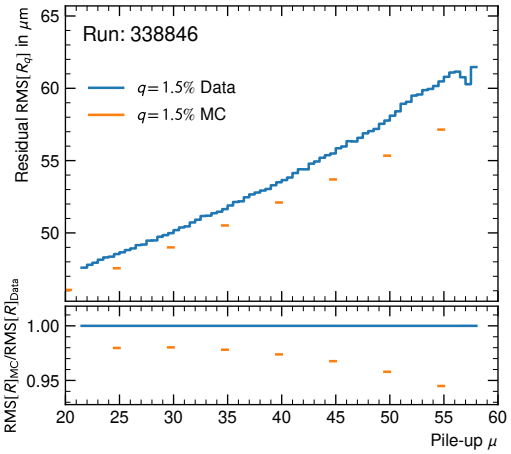


Figure 4.7.: Residual of reconstructed tracks for default noise thresholds of $q = 1.5\,\%$ for run 338846 as a function of pile-up interactions and simulated $t\bar{t}$ events for selected numbers of pile-up interactions. The statistical uncertainty is too small to be visible.

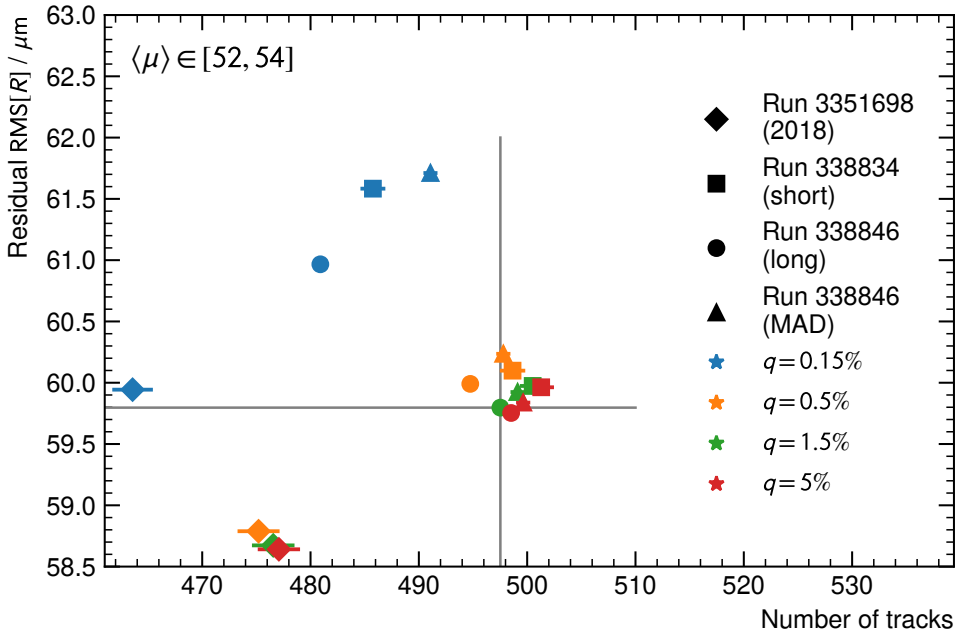


Figure 4.8.: Summary of both figures of merit for all reprocessed runs in the pile-up interval $\mu \in [52, 54]$ with tight track selection and varying noise thresholds. The best performance is achieved in the lower right corner. The default performance for run 338846 is highlighted by the gray cross. The statistical uncertainty on the figures of merit is indicated by error bars. In most cases, the uncertainty is smaller than the marker symbol.

threshold will reduce the number of noisy strips, and at the same time, the study shows the increase does not negatively impact the tracking performance.

CHAPTER 5

Measurement of the Higgs boson production cross section

Broadly speaking, data is structured information with potential for meaning.

— Scott Murry

Measuring the coupling between τ leptons and the Higgs boson is an important test of eletroweak (EW) symmetry breaking in the SM. Since the mass of the τ lepton is known precisely [14], the SM gives testable predictions about the coupling strength. In this analysis, the product of the branching ratio of a Higgs boson decaying to a pair of taus and the production cross section of the Higgs boson in pp collisions is measured. At the time of writing, measuring this decay channel gives the most precise results for the Higgs coupling to leptons since the τ lepton, being a third-generation charged lepton, is by far the heaviest lepton.

Measuring the cross section has also been pursued in Run 1 by the ATLAS and CMS collaborations. The Run 1 dataset recorded by the ATLAS detector yielded evidence for the decay at a significance of $4.5\,\sigma$ ($3.4\,\sigma$ expected) [80]. The combination with the Run 1 CMS result [81] achieved the first experimental observation of the Higgs boson decay to a pair of τ leptons at a significance of $5.5\,\sigma$ ($5.0\,\sigma$ expected) [82]. The ATLAS single-experiment observation was achieved by combining results derived from the data recorded in 2015 and 2016, corresponding to an integrated luminosity of $36\,\text{fb}^{-1}$, and the Run 1 measurement [83]. The decay was observed at a significance of $6.4\,\sigma$ ($5.4\,\sigma$ expected). The larger observed significance is attributed to the larger-than-expected significance in Run 1. The focus of previous publications was on the observation of the decay. The focus has now shifted to performing precision measurements which allows the comparison to SM predictions.

The analysis presented in the following and published in Reference [84] measures the cross section times branching ratio of $pp \to H \to \tau\tau$ using the full Run 2 dataset with an integrated luminosity of $139\,\text{fb}^{-1}$. Unlike the Run 1 and partial Run 2 publications by the ATLAS Collaboration, this analysis targets all four major Higgs boson

production mechanisms at the LHC, namely ggF, VBF, VH, and $t\bar{t}H$. For the VH production mode, the analysis is restricted to hadronic decays of the weak boson. The analysis is furthermore limited to fully hadronic decays of the associated top quarks and tau leptons for the $t\bar{t}H$ production mode. The leptonic modes of these processes are explored in separate ATLAS analyses [85–87]. The analysis of the full Run 2 dataset is an opportunity to study $H \to \tau\tau$ and measure the cross section with unprecedented precision.

The analysis on the full Run 2 dataset is an improvement over the previous Run 2 publication. Besides the 4-fold increase in measured data events, the analysis grew also in terms of its complexity, its techniques, and its computational requirements. The 4-fold increase in terms of integrated luminosity is accompanied by an at least 4-fold increase in the number of simulated MC events adding to the required computing power. The number of measured parameters of interest and the number of measured production modes is increased by a factor of three and two, respectively. The use of several machine-learning techniques adds to the intricacy of the analysis. Furthermore, the complexity of the fit model and its run time increased due to the treatment of $Z \to \tau\tau$ background and the overall gain in statistical power that allows to define more analysis categories and requires more careful assessment of systematic uncertainties. Consequently, the effort is visible in a more precise measurement of the Higgs boson production cross section times branching ratio.

Besides the complexity of the analysis, the measurement is challenging for a number of other reasons. The main reasons are the large number of background events from $Z \to \tau\tau$, neutrinos in the final state leading to a poor mass resolution, and misidentified objects (*fakes*). The τ lepton in the final state is a short-lived particle that usually decays before it reaches the detector. It can decay either leptonically into an electron or muon and two neutrinos ($\tau \to \ell\bar{\nu}_\ell\nu_\tau$), or into hadrons and a neutrino ($\tau \to \text{hadrons}\,\nu_\tau$). Leptonically decaying taus are reconstructed as electrons or muons. They are treated as prompt electrons or muons in the analysis. The possibility to measure a vertex displacement stemming from the τ lifetime is not pursued in this analysis. For the remainder of the thesis, leptonically decaying τ leptons are denoted τ_{lep}, or more specifically, τ_e and τ_μ. Electrons and muons are collectively referred to as light leptons or ℓ. Hadronically decaying τ leptons, on the other hand, have a complex signature in the ID and calorimeter that makes identification and reconstruction difficult, see Section 3.3.5. Hadronically decaying τ leptons are referred to as τ_{had}.

The remainder of this chapter describes the analysis in more detail, discusses the use of machine learning to improve the sensitivity, and presents its final result.

5.1. Analysis strategy

The branching ratio of a tau lepton to light leptons is approximately 35.7 % [14]. The relative contributions to di-τ decays arising from different combinations of leptonic and hadronic decays are shown in Figure 5.1. The composition of background processes depends on the combination of hadronic and leptonic decays of the two τ leptons. The analysis is therefore split into three analysis channels, depending on the multiplicity of leptonically (or hadronically) decaying τ leptons: $\tau_e\tau_\mu$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_{\text{had}}\tau_{\text{had}}$. In contrast to the first Run 2 measurement, the same-flavor $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$ channels have been

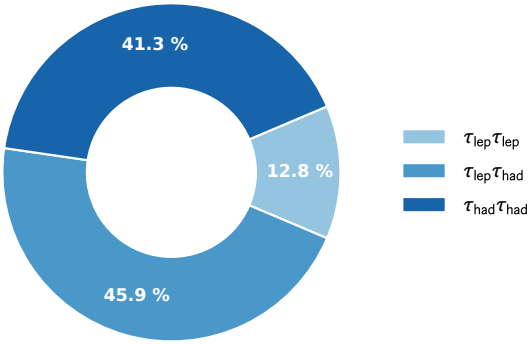


Figure 5.1.: Expected fractions of leptonic and hadronic tau lepton decay multiplicities in $H \to \tau\tau$ decays. The $\tau_{\text{lep}}\tau_{\text{lep}}$ contribution includes same-flavor $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$ decays.

dropped from the analysis. The decision is motivated by the relatively small branching ratio of the fully leptonic channel and the small sensitivity of the same-flavor channel due to the large background from $Z \to \ell\ell$. The event selection for each channel is summarized in Section 5.3.2.

The largest background process in the signal regions of all channels is $Z \to \tau\tau$. The process has the same final state as the Higgs boson signal except for the spin correlation of the two taus not exploited in this analysis and the difference in invariant mass $m^{\tau\tau}$. Getting a precise estimate for Z+jets background events in the signal region is paramount for the analysis. The analysis employs data-driven techniques to avoid reliance purely on Monte Carlo simulated samples. The procedure is described in Section 5.4.1.

Another significant contribution in the signal regions originates from events with a falsely identified hadronically decaying tau, electron or muon. In all three channels, data-driven methods are implemented to estimate the number of events in the signal region. For the $\tau_e\tau_\mu$ channel, the data-driven method is concerned with events where a jet is misidentified as an electron or muon. Muons are identified with high accuracy such that fake muons are rare. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the data-driven method estimates the contribution from events with a jet misidentified as a hadronically decaying τ. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the data-driven method is used to predict the number of events where one or both identified hadronically decaying τ originate from jets. The data-driven methods are described in Section 5.4.2.

Backgrounds with smaller contributions to the signal regions, such as events from $t\bar{t}$ and diboson production are taken from simulation. A detailed description of the MC samples used in this analysis can be found in Section 5.2. Typical tree-level Feynman diagrams for all background processes are shown in Figure 5.2.

The signal region categories follow the selection criteria defined in the STXS framework. For VBF, VH, and $t\bar{t}H$ the categories correspond to the stage 0 phase spaces. For these three Higgs boson production modes, machine learning classifiers are used to

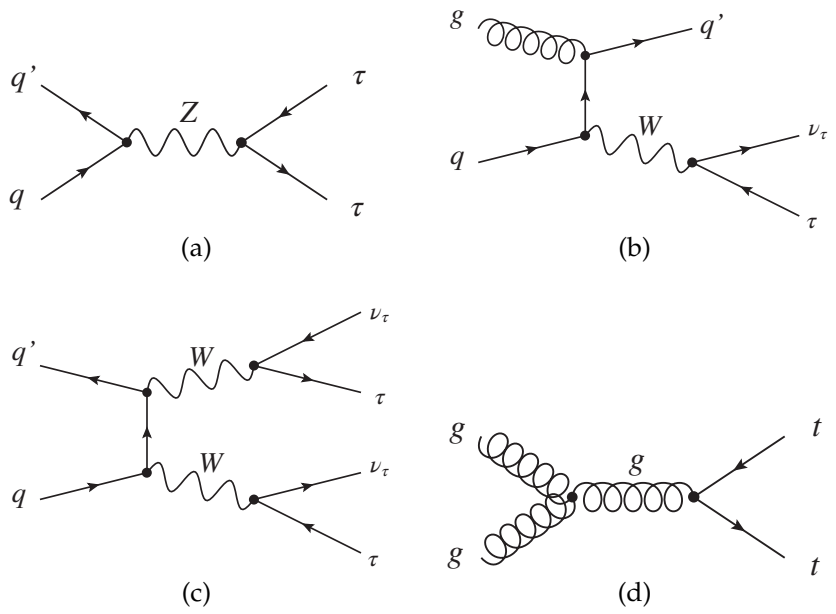


Figure 5.2.: Example leading-order production Feynman diagrams of the main background of the analysis: $Z \rightarrow \tau\tau$ with the same final state as the Higgs boson signal (a), W+jets with a misidentified jet as a tau or an electron (b), events with two bosons (c), and $t\bar{t}$ with subsequent decays to b quarks and $W \rightarrow \tau\nu_\tau$ (d).

increase the signal sensitivity. Six additional analysis categories are defined for each channel that target merged stage 1.2 STXS bins of the ggF production mode, see Figure 2.6. The signal region categories defined in this way are used for a measurement of the total cross section times $H \to \tau\tau$ branching ratio, for the cross section split by production mode, and a nine-bin STXS measurement.

5.2. Samples

The background and signal prediction is taken from Monte Carlo (MC) simulation for most of the processes. The exception to this is the data-driven fake method described in Section 5.4.2. The normalization of Z+jets is measured in situ in control regions as described in Section 5.4.1. However, in both cases, the contamination from other processes is estimated with MC samples. Table 5.1 summarizes the MC used to model the signal background processes of this analysis. For all samples, a detailed detector simulation is performed based on Geant4 [88].

MC events for the main Higgs production mode at the LHC, ggF, with an expected cross section of 48.6 pb, are generated using Powheg Box v2 [89–93] in combination with a reweighting technique [94–97] to achieve next-to-next-to-leading order (NNLO) accuracy in QCD. The parton shower is created with Pythia 8.230 [98] using the AZNLO tune [99]. For this sample, the PDF4LHC15NNLO PDF set [100] is used. The MC sample is normalized to the cross section computed at next-to-next-to-next-to-leading order (N3LO) in QCD plus next-to-leading order (NLO) electroweak corrections [28, 101–110].

The events for the VBF production mode with an expected cross section of 3.78 pb are generated using Powheg Box v2 at NLO in QCD. The computation takes effects due to the finite heavy quark masses and soft-gluon resummation up to next-to-leading logarithms (NNLL) into account. The parton shower is created with Pythia 8.230 using the AZNLO tune. For this sample, the PDF4LHC15NLO PDF set is used. The MC sample is normalized to a cross section computed at NNLO in QCD and NLO electroweak corrections [111–113].

The MC dataset for Higgs boson production in association with a vector boson (VH, 2.25 pb) is simulated using Powheg Box v2 at NLO accuracy for VH with an additional jet. The parton shower is created with Pythia 8.230 using the AZNLO tune. For this sample, the PDF4LHC15NLO PDF set is used. The sample is normalized to the cross section computed at NNLO in QCD with electroweak corrections at NLO. The loop-induced $gg \to ZH$ is normalized separately at NLO and NNLL accuracy in QCD [114–119].

The events for the Higgs boson production mode in association with a pair of top quarks ($t\bar{t}H$, 0.51 pb) or a pair of bottom quarks ($b\bar{b}H$, 0.64 pb) are generated using Powheg Box v2 at NLO in QCD. For $t\bar{t}H$, the decay of charm and bottom quarks is simulated with EvtGen 1.6.0 [120]. The parton shower is created with Pythia 8.230 using the A14 tune [121]. The NNPDF3.0NNLO PDF set [122] is used for this sample. The dataset is normalized to the cross section computed at NLO in QCD. In the case of $t\bar{t}H$, the normalization takes additional electroweak corrections at NLO into account [28, 123–127].

The events from the tH process with a cross section of 0.092 pb are considered in

the analysis but expected to be negligible. The sample is simulated with the MadGraph5_aMC@NLO 2.6.2 generator [128]. The parton shower is created with Pythia 8.230 using the A14 tune. The CT10 PDF set is used for the sample. The MC sample is normalized to the cross section computed at NLO in QCD [129, 130].

Background events with a weak boson and jets from QCD processes (V+jets) are generated using Sherpa 2.2.1 [131] at NLO with up to two parton emissions and at leading-order (LO) for up to four parton emissions. The computation employs the Comix [132] and OpenLoops [133–135] libraries. In contrast to V+jets QCD processes, the events from electroweak$V$+jets processes are generated at leading order with up two parton emissions. The matrix elements are matched with the Sherpa parton shower [136] following the MEPS@LO prescription [137–140]. The sample uses the NNPDF3.0NNLO PDF set. The cross section of the MC sample is normalized to NNLO predictions [141].

The $t\bar{t}$ background is generated using Powheg Box v2 at NLO. The decay of bottom and charm quarks is treated as in the $t\bar{t}H$ case. The parton shower, the hadronization, and the underlying event are created with Pythia 8.230 using the A14 tune. The sample uses the NNPDF3.0NNLO PDF set. The dataset is normalized to NNLO in QCD, including the resummation of soft-gluon terms at NNLL from Top++ [142–148].

The single top background uses the Powheg Box v2 generator at NLO in QCD, employing the five flavor scheme. The sample uses the NNPLDF3.0NLO PDF set. The parton shower is created with Pythia 8.230 using the A14 tune. The cross section of the sample is normalized to predictions at NLO in QCD.

The events of the diboson background process are generated with Sherpa 2.2.1 or 2.2.2, depending on the process, at NLO in QCD with one additional parton emission and LO for up to three emissions. Loop-induced $gg \to VV$ processes are considered at LO with up to one parton emission. The dataset uses the NNPDF3.0NNLO PDF set. The MC sample cross section is normalized to NLO predictions.

The pp data events used in the analysis were recorded in the data-taking period of the ATLAS detector between 2015 and 2018. Only events that satisfy quality conditions are considered in the analysis. The data set corresponds to an integrated luminosity of 139 fb^{-1}.

5.3. Object definition and event selection

The analysis uses the object reconstruction defined in Section 3.3 and applies additional criteria. The definition of the physics objects in the next section is used to count the number of objects. The multiplicities of electrons, muons, and hadronic taus are used to assign events to the analysis channels. The specific channels usually apply additional tighter object selection criteria. A description of the triggers used for the analysis in each channel can be found in Reference [84].

5.3.1. Object definition

The interaction vertex with the largest sum of track momenta associated with the vertex is chosen as the primary vertex of the hard interaction. Other vertices are assumed to be pileup interactions or from long-lived secondary particles.

Table 5.1.: Summary of Monte Carlo samples used in the analysis. The table shows the generators used to compute the matrix element, the parton shower and the properties of the proton PDF. Finally, the accuracy of the cross section normalization is stated.

Process	Generator	PDF ME	PDF PS	Tune	Normalization
ggF	Powheg Box v2	PDF4LHC15NNLO	CTEQ6L1	AZNLO	N3LO QCD + NLO EW
VBF	Powheg Box v2	PDF4LHC15NLO	CTEQ6L1	AZNLO	NNLO QCD + NLO EW
VH	Powheg Box v2	PDF4LHC15NLO	CTEQ6L1	AZNLO	NNLO QCD + NLO EW
$t\bar{t}H$	Powheg Box v2	NNPDF3.0NNLO	NNPDF2.3LO	A14	NLO QCD + NLO EW
tH	MadGraph5_aMC@NLO	CT10	NNPDF2.3LO	A14	NLO
$b\bar{b}H$	Powheg Box v2	NNPDF3.0NNLO	NNPDF2.3LO	A14	NLO
V+jets QCD	Sherpa 2.2.1	NNPDF3.0NNLO	NNPDF3.0NNLO	Sherpa	NNLO
V+jets EW	Sherpa 2.2.1	NNPDF3.0NNLO	NNPDF3.0NNLO	Sherpa	LO
$t\bar{t}$	Powheg Box v2	NNPDF3.0NNLO	NNPDF2.3LO	A14	NNLO + NNLL
Single top	Powheg Box v2	NNPDF3.0NNLO	NNPDF2.3LO	A14	NLO
Diboson	Sherpa 2.2.1	NNPDF3.0NNLO	NNPDF3.0NNLO	Sherpa	NLO

Electrons are required to pass the loose identification working point. Kinematically, they are required to have transverse momentum $p_T > 15\,\text{GeV}$ and be within the central detector region ($|\eta| < 2.47$) covered by the ID. For the signal region, the crack region between barrel and end-cap with $1.37 < |\eta| < 1.52$ does not provide precise electron identification and energy measurements. Electrons in this range are excluded. Electrons in the $\tau_e\tau_\mu$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels are additionally required to pass the medium identification working point and the loose isolation working point [64]. For electrons with $p_T = 15\,\text{GeV}$, the identification efficiency is 80 %. The isolation efficiency is 90 %. The transverse momentum threshold is higher than the trigger threshold to ensure to work in a region with constant efficiency (trigger plateau). It has not been attempted to distinguish taus decaying to electrons from prompt electrons.

Muons are required to pass the loose identification working point [149], have a minimum transverse momentum of $p_T > 10\,\text{GeV}$, and be in the central detector region $|\eta| < 2.5$. The identification criteria have an efficiency of 97 %. For muons in the $\tau_e\tau_\mu$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels, the tight isolation requirement is applied with an efficiency of 85 % for muons at the momentum threshold [149]. Similar to the electrons, the momentum threshold is increased in the leptonic channels. This ensures that the muons are in the efficiency plateau of the muon triggers. No effort has been made to identify muons from taus decays compared to prompt leptons.

The visible decay products $\tau_{\text{had-vis}}$ from hadronic tau decays are reconstructed starting from anti-k_t jets. The $\tau_{\text{had-vis}}$ are required to pass the transverse momentum threshold $p_T > 20\,\text{GeV}$ and be in the central detector region with $|\eta| < 2.47$, excluding the transition region between the barrel and end-cap ECal with $1.37 < |\eta| < 1.52$. The sum of the charges of each track associated with the hadronic $\tau_{\text{had-vis}}$ must be ± 1. In the case of a single track, the $\tau_{\text{had-vis}}$ is referred to as 1-prong tau. In the case of three tracks, the $\tau_{\text{had-vis}}$ is referred to as 3-prong tau. Reconstructed taus with other track multiplicities are rejected. The visible decay products need to pass a very loose RNN score with an efficiency of 95 % [74].

In the fully hadronic channel $\tau_{\text{had}}\tau_{\text{had}}$, the hadronic taus are required to match the hadronic tau candidates identified by the trigger. The p_T threshold is tightened to be in a region with constant tau trigger efficiency. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, only the τ_{had} with the largest transverse momentum is considered. All other tau candidates are treated as jets. To reject events with an electron misidentified as a 1-prong tau, 1-prong τ_{had} are required to pass the medium electron-BDT working point with an efficiency of 85 %. Additionally, to achieve a larger background rejection, the momentum threshold is raised to $p_T > 30\,\text{GeV}$. In both channels, the medium identification working point is required with an efficiency of 75 % and 60 % for 1-prong and 3-prong taus, respectively [74].

Jets are reconstructed using an anti-k_t algorithm described in Section 3.3 with the radius parameter $R = 0.4$. Jets need to exceed the transverse momentum threshold of $p_T > 20\,\text{GeV}$. A Jet Vertex Tagger [150] is used to assign jets to the interaction vertices. Jets with $p_T < 60\,\text{GeV}$ and $|\eta| < 2.5$ that are not associated with the primary vertex are removed. Similarly, the forward Jet Vertex Tagger [151] is used to remove jets with $p_T < 60\,\text{GeV}$ and $|\eta| > 2.5$ that are not associated with the primary vertex.

Jets with secondary vertices from long-lived b-hadrons are tagged with the DL1r algorithm [152, 153]. The b-tagged jets must pass the momentum threshold of $p_T >$

20 GeV and be within the central region of the detector $|\eta| < 2.5$ instrumented by the ID. In the leptonic channels $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$, events with b-tagged jets are rejected using the 85 % efficiency working point. In the fully hadronic channel, the 70 % working point is used to improve the sensitivity.

5.3.2. Event selection

The event selection differs between the three analysis channels due to the different signal final states but also due to different backgrounds relevant to the channels.

In the fully leptonic channel $\tau_e\tau_\mu$, exactly one electron and one muon are required. The same flavor channel with two electrons or two muons is not considered in this analysis due to the relatively small sensitivity of the channel. The light leptons must have opposite charges to reject events from V+jets and $t\bar{t}$ background. The invariant mass $m^{e\mu}$ of the two leptons must be within the interval $[30, 100]$ GeV. The signal region overlaps partially with the $H \to WW^*$ ATLAS analysis [154]. To ensure that the analyses can be used independently in a statistical combination, we enforce orthogonality by requiring the collinear mass m_{coll} to satisfy $m_{\text{coll}} > m_Z - 25\,\text{GeV} \approx 66.2\,\text{GeV}$. The inverse condition is applied in the $H \to WW^*$ analysis. Events from processes with top quarks are suppressed by rejecting events with b-tagged jets (b-veto). At least one jet with transverse momentum $p_{\text{T}} > 40\,\text{GeV}$ is required to further suppress background processes. Additionally, the angular cuts $\Delta R^{e\mu} < 2$ and $\Delta\eta^{e\mu} < 1.5$ are applied to the leptons rejecting events with two light leptons from independent decays and implicitly requiring a minimal momentum for the boosted Higgs boson.

In the semileptonic channel $\tau_{\text{lep}}\tau_{\text{had}}$, exactly one light lepton and one hadronic tau must be present in the event. Analogous to the $\tau_e\tau_\mu$ channel, the light lepton and the tau must have opposite charges. The transverse mass

$$m_{\text{T}} = \sqrt{2p_{\text{T}}^{\ell}E_{\text{T}}^{\text{miss}}(1 - \cos\Delta\phi)} \tag{5.1}$$

of the light lepton and $E_{\text{T}}^{\text{miss}}$ system is used to reject events with leptonic W decays where $\Delta\phi$ denotes the angle between the light lepton and $E_{\text{T}}^{\text{miss}}$ in the transverse plane. Events are required to satisfy $m_{\text{T}} < 70\,\text{GeV}$. At least one jet with $p_{\text{T}} > 40\,\text{GeV}$ must be present. Background from $t\bar{t}$ processes is rejected by vetoing any event with a b-tagged jet. The angular conditions from the fully leptonic channel are also applied in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel using $\Delta R^{\ell\tau} < 2.5$ and $\Delta\eta^{\ell\tau} < 1.5$.

In the fully hadronic channel $\tau_{\text{had}}\tau_{\text{had}}$, exactly two hadronic taus passing the requirements defined in Section 5.3.1 must be in the event. One tau lepton must satisfy $p_{\text{T}} > 40\,\text{GeV}$. Contributions from W+jets and $t\bar{t}$ background processes are suppressed by selecting events with opposite charges of the taus. The low p_{T} threshold of the $\tau_{\text{had-vis}}$ makes it mandatory to add the conditions $\Delta R^{\tau\tau} > 0.6$ on the separation of the taus and to require at least one jet with $p_{\text{T}} > 70\,\text{GeV}$ within $|\eta| < 3.2$. Besides the required minimal angular separation, the same angular conditions as in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel are applied to the two hadronic taus using $\Delta R^{\tau\tau}$ and $|\eta^{\tau\tau}|$. The fully hadronic channel is used to measure the $t\bar{t}H$ production cross section. Except for the signal category designated for $t\bar{t}H$ measurements, events with a b-tagged jet are rejected.

Since the final states in all channels involve at least two neutrinos, a small amount of $E_{\text{T}}^{\text{miss}}$ is expected in signal events. The events used in the analysis must have a min-

imal missing transverse momentum of $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$. Finally, conditions on the momentum fractions $x_{1,2}$ carried by the visible tau decay products computed with the collinear approximation, see Equation (3.12), are applied. For all channels, the momentum fractions must be at least 0.1. The momentum fraction of the leptonic decay must be smaller than 1.0 (1.2) in the $\tau_e\tau_\mu$ ($\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$) channel. The momentum fraction of the $\tau_{\text{had-vis}}$ must be smaller than 1.4. These conditions ensure the good performance of the mass reconstruction.

The $m_{\tau\tau}^{\mathrm{MMC}}$ mass distribution for the inclusive signal regions VBF, VH, and ggF are shown in Figure 5.3 combined for all channels.

5.3.3. Event categorization

The event categorization is driven by the phase space regions defined in STXS stage 1.2 while some of the bins are merged to optimize the expected statistical power in each bin. The signal categories defined in this section are used for the global cross section measurement, for the measurement of the cross section per Higgs boson production mode, and for the nine-bin STXS measurement. Additionally, the nine-bin STXS measurement is used in the Higgs boson combination described in Chapter 6.

Like the STXS bin definition, the categories are first split by production mode. Categories for VBF, VH, and $t\bar{t}H$ are further split into signal-enriched and background-enriched subcategories using machine-learning techniques. Events in the boosted categories targeting boosted ggF topologies are split based on Higgs boson p_{T} and jet multiplicity. The categorization is performed as a chain. If an event fails the section criteria for one category, it is tested against the remaining categories in the chain one by one. Events can only end up in a single exclusive category or may not be considered.

For events from the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel to enter the $t\bar{t}H$ category, conditions on the number of jets and b-tagged jets are imposed. The target topology $t\bar{t}H$ with fully hadronic tau and top quark decays consists of six jets, with two of them tagged as b-quark jets. To account for inefficiency in the reconstruction algorithms, tagging algorithms, and acceptance cuts (e.g., p_{T}^{j}), exactly one of these multiplicities is allowed to be lower by one unit. Therefore, either at least five jets with at least two of them b-tagged are required, or at least six jets with at least one of them b-tagged are required.

Additionally, in the $t\bar{t}H$ category, BDTs are used to reject background events from $Z \to \tau\tau$ and $t\bar{t}$, respectively. Several methods have been assessed to combine the scores from the two BDTs. No improvement over simple rectangular cuts could be found. The BDTs are used to define a signal-enriched category and a background-enriched category. Events are assigned to either category. Events that fail the selection imposed by either or both BDTs are assigned to the background-enriched region. The signal and background-enriched categories are denoted ttH_1 and ttH_0, respectively. The $t\bar{t}H$ signal purity in the two categories is 92 % and 74 %, respectively.

For events from all three analysis channels to enter the VBF categories, jets must not be b-quark tagged. In alignment with the STXS bin definitions, the condition $m^{jj} > 350\,\mathrm{GeV}$ is placed on the invariant mass of the two leading jets, which for VBF events are usually the two jets in the forward and backward directions. Furthermore, a condition on the transverse momentum of the subleading jet of $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is implemented. To further exploit the characteristic topology of VBF events and reject background events

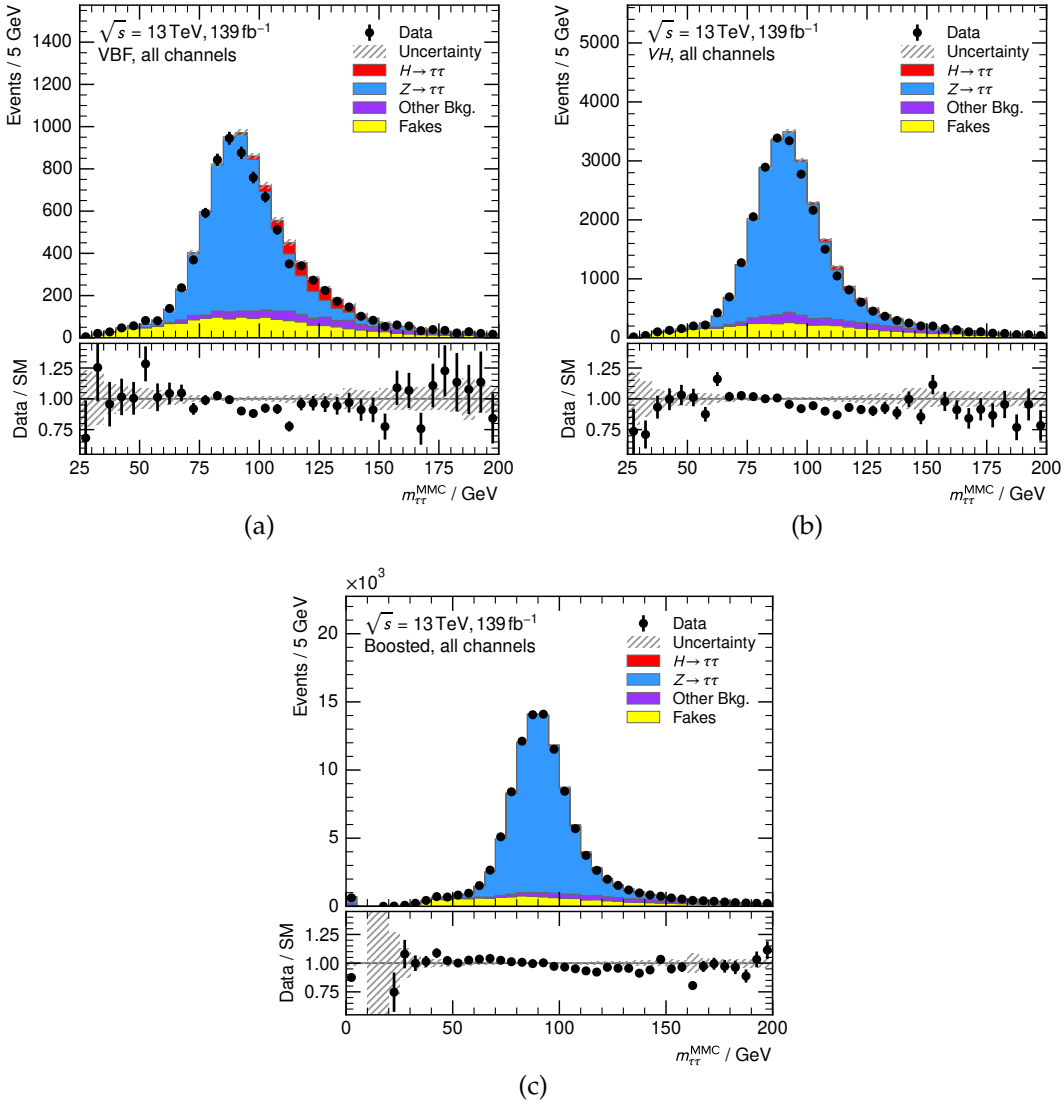


Figure 5.3.: Invariant mass $m_{\tau\tau}^{\text{MMC}}$ distributions in the VBF (a), *VH* (b), and ggF (c) signal region categories. The histograms show the combination of all analysis channels. The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

from $Z \to \tau\tau$, the two forward and backward jets must be in opposite hemispheres (split at $\eta = 0$) and have a large separation in pseudorapidity of $\Delta\eta^{jj} > 3$.

Furthermore, the decay products of the Higgs boson must be inside the pseudorapidity gap of the two forward and backward jets. From previous rounds of the analysis, it is known that systematic uncertainties related to contaminations from ggF and $Z \to \tau\tau$ processes are a limiting factor. To reduce the impact, a BDT, and alternatively, a NN is trained, exploiting the VBF event topology to reject these two background and signal processes, see Section 5.5. In case of the BDT, the signal and background-enriched categories are denoted VBF_1 and VBF_0, respectively.

The *VH* category is comprised of events from all analysis channels that feature two jets from the subsequent decay of the weak boson V. The criterion on the invariant mass of the two jets is $60\,\text{GeV} < m^{jj} < 120\,\text{GeV}$. The transverse momentum of the subleading jet must satisfy $p_\text{T} > 30\,\text{GeV}$. To enhance the purity of *VH* signal compared to other signal processes, a BDT is trained to reject all other Higgs production modes. The BDT is optimized to minimize the uncertainty of the measured *VH* cross section. The BDT is not used to reject background processes. Based on the BDT score, events are either assigned to a signal-enriched or background-enriched category. The signal and background-enriched categories are denoted VH_1 and VH_0, respectively.

Lastly, any event that did not match any of the aforementioned categories is considered for the *boosted* categories used to measure the ggF cross section with a boosted Higgs boson. The transverse momentum of the Higgs boson p_T^H is estimated using the vectorial sum in the transverse plane of the visible Higgs boson decay products and E_T^miss. Events that do not satisfy $p_\text{T}^H > 100\,\text{GeV}$ are discarded. Following the STXS bin definition, categories are defined for $100\,\text{GeV} < p_\text{T}^H < 120\,\text{GeV}$, $120\,\text{GeV} < p_\text{T}^H < 200\,\text{GeV}$, $200\,\text{GeV} < p_\text{T}^H < 300\,\text{GeV}$, and $300\,\text{GeV} < p_\text{T}^H$. Due to the expected statistical power and in alignment with the STXS framework, the two categories with the lowest p_T^H are subdivided by the multiplicity of jets with $p_\text{T}^j > 30\,\text{GeV}$. Events with a single jet and events with at least two jets are placed in separate categories. In total the analysis consists of six boosted categories per channel. The categories are enumerated in order of increasing p_T^H. For the categories with $p_\text{T}^H < 200\,\text{GeV}$, the suffix _eq1J (_ge2J) is added in case of exactly one (two or more) jets.

5.4. Background estimation

As already discussed in the introduction of this chapter, the major background processes are $Z \to \tau\tau$ and events with misidentified τ leptons. Besides these backgrounds, contributions from processes with top quarks, two weak bosons (diboson), or $Z \to \ell\ell$ need to be considered in the signal regions. The distributions in the signal region are estimated using MC samples. Figure 5.4 shows the $m_{\tau\tau}^\text{MMC}$ distribution in selected analysis categories for each channel independently to illustrate the modeling of background processes.

5.4.1. Z+jets estimate

The dominant background process for the $H \to \tau\tau$ analysis is $Z \to \tau\tau$. The contribution in the signal regions is typically around 60 % but can go up to approximately

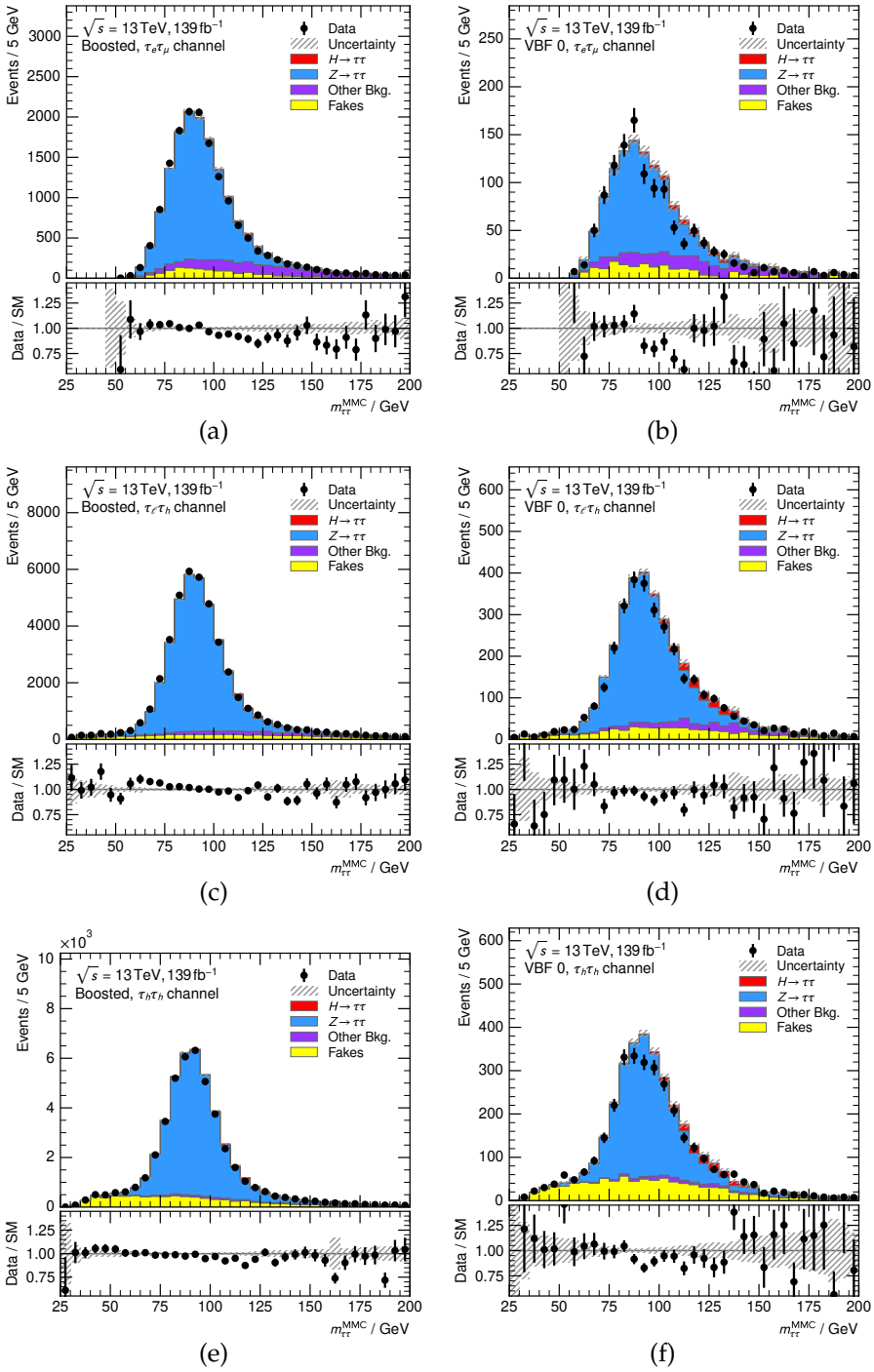


Figure 5.4.: Mass distributions in the sum of all boosted categories (left) and the VBF_0 category (right). The histograms are split into the analysis channels $\tau_e\tau_\mu$ (top), $\tau_{\text{lep}}\tau_{\text{had}}$ (middle), and $\tau_{\text{had}}\tau_{\text{had}}$ (bottom). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively. The difference between data and the background estimation around the Z peak (f) is attributed to the normalization of the Z+ jets process later measured in situ.

93 % in the $\tau_{\text{had}}\tau_{\text{had}}$ boost_2 category. The tree-level process is fully described by the electroweak interaction, has been measured with high precision at LEP, and can be computed with high precision. For the analysis, however, we require at least one jet in the event. The origin of the additional jet in the $Z \to \tau\tau$ sample stems from quark or gluon emission in the initial state (or potentially from pile-up). These emissions are predominantly described by QCD and pose a challenge for MC generators. The jet multiplicity, their momentum spectra, and the overall normalization are usually not modeled well in the Z+jets sample, thus affecting the number of events passing the analysis event selection.

A common analysis technique is to define control regions pure in Z+jets to measure the normalization and therefore mitigate this issue. For $H \to \tau\tau$, applying this method is difficult due to the similarities between the process and the Higgs boson signal. The $Z \to \tau\tau$ background process differs in the invariant mass of the di-τ system, but due to the width of the Z boson and the mass reconstruction resolution, the mass criterion is not sufficient to define a control region pure enough in $Z \to \tau\tau$. The final states also differ in their spin correlation between the two taus since the Z is a spin-1 particle and the Higgs boson is a scalar with spin 0. However, it is difficult to measure the helicity of the taus.

To profit from the recorded dataset and reduce the dependence on the modeling in simulation, a simplified technique called kinematic rescaling [84, 155], described in the following was employed for this analysis.

The kinematic rescaling defines a control region consisting only of $Z \to \ell\ell$ events with $\ell = e, \mu$. The control region has no overlap with the different flavor $\tau_e\tau_\mu$ channel. Besides the difference in lepton flavor, the control regions are designed to mimic the event selection of each signal region category.

Applying the same kinematic requirements on the visible decay products of a decaying tau in the signal region and the light lepton in the control region does not populate the regions with events from the same Z boson phase space since neutrinos carry away undetected momentum in the case of the tau decay. The visible decay products have lower p_{T} in $Z \to \tau\tau$ due to the momentum carried by neutrinos. Extrapolating measurements in control regions to signal regions must be done with care, and kinematic rescaling was employed instead.

With the kinematic rescaling, the momenta p^ℓ of prompt light leptons are rescaled to mimic energy loss from momentum carried away by neutrinos in tau decays. The rescaled momentum is sampled from a momentum distribution obtained from simulated tau decays. The light leptons with the momentum rescaling applied are therefore made to *look* like visible decay production of taus. Dedicated rescaling distributions are derived for the individual lepton flavors. This rescaled momentum is used to recompute derived quantities of the event, like the variant di-τ mass $m_{\tau\tau}^{\text{MMC}}$. The event selection of the control regions imposes the same kinematic conditions on the rescaled momentum as the signal regions on the τ lepton momentum. This ensures that the signal regions and control regions are populated with events from the same Z boson phase space. The procedure takes into account the effects of the trigger, identification, and reconstruction efficiencies. The rescaling is done independently for each channel. Appropriate control region plots and closure tests can be found in Reference [84].

With kinematic rescaling, it is possible to measure the normalization of Z+jets in-

situ for the same region in phase space covered by the signal region. The approach is assumed to reduce the uncertainty of measured cross sections since some of the systematic uncertainties affect the signal and control region in the same way. This cancellation is used to measure and constrain systematic uncertainties in control regions and extrapolate the effects to the signal regions. The measured normalization factors range from 0.92 to 1.18 with the strongest constraint of 1.03 ± 0.07 in the $\tau_{\text{lep}}\tau_{\text{had}}$ boost_0_ge2J category.

5.4.2. Fake estimation

Background events where an object is misidentified are termed *fakes*. In a considerable number of events, jets or electrons are misidentified as hadronically decaying taus. In the $\tau_e\tau_\mu$ channel, jets misidentified as electrons or muons need to be considered. Misidentification is usually not modeled well in MC samples, making more advanced data-driven methods necessary. Three different methods are employed for the three channels to target different scenarios of misidentification. All methods have in common that events estimating the fake contribution in the signal region are recorded data events, selected with a different event selection and event weight to account for kinematic differences. The three methods are summarized in the following. Table B.1 summarized the number of expected fake events for each channel and category. In the low-p_{T} boosted categories of the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the fraction of events with a misidentified $\tau_{\text{had-vis}}$ reaches 35.5 %.

Fake estimate in the $\tau_e\tau_\mu$ channel

The fake estimation in the $\tau_e\tau_\mu$ channel is based on the Matrix Method [156]. For each lepton in the signal region, a looser event selection is defined by removing the requirement on the isolation. QCD jets tend to have more activity in the detector around the central jet cone, therefore, removing the isolation criterion produces an event selection enriched in fake events. Additionally, for the loose selection, the medium identification requirement is dropped. The regular signal region is referred to as the tight event selection. The combination of loose and tight of both leptons introduces four regions with event yields N_{LL}, N_{TL}, N_{LT}, and N_{TT}, where the latter is the regular signal region. The idea of the matrix method is to express these yields as a linear combination of events with two real leptons N_{rr}, two fake leptons N_{ff}, and the two mixed combinations N_{rf} and N_{fr}. The relation is characterized by the efficiencies

- ϵ_{r} for a real lepton, and
- ϵ_{f} for a fake lepton

that passed the loose selection to also pass the tight selection. Their counterparts are denoted $\bar{\epsilon}_i = 1 - \epsilon_i$ referring to the probability that a real (fake) lepton that passed the loose selection fails the tight selection. The data-driven method employed in the $\tau_e\tau_\mu$

channel is based on the eponymous relation summarized in matrix form

$$\begin{pmatrix} N_{\mathrm{TT}} \\ N_{\mathrm{LT}} \\ N_{\mathrm{TL}} \\ N_{\mathrm{LL}} \end{pmatrix} = M \begin{pmatrix} N_{\mathrm{rr}} \\ N_{\mathrm{rf}} \\ N_{\mathrm{fr}} \\ N_{\mathrm{ff}} \end{pmatrix} = \begin{pmatrix} \epsilon_{\mathrm{r}}\epsilon_{\mathrm{r}} & \epsilon_{\mathrm{r}}\epsilon_{\mathrm{f}} & \epsilon_{\mathrm{f}}\epsilon_{\mathrm{r}} & \epsilon_{\mathrm{f}}\epsilon_{\mathrm{f}} \\ \bar{\epsilon}_{\mathrm{r}}\epsilon_{\mathrm{r}} & \bar{\epsilon}_{\mathrm{r}}\epsilon_{\mathrm{f}} & \bar{\epsilon}_{\mathrm{f}}\epsilon_{\mathrm{r}} & \bar{\epsilon}_{\mathrm{f}}\epsilon_{\mathrm{f}} \\ \epsilon_{\mathrm{r}}\bar{\epsilon}_{\mathrm{r}} & \epsilon_{\mathrm{r}}\bar{\epsilon}_{\mathrm{f}} & \epsilon_{\mathrm{f}}\bar{\epsilon}_{\mathrm{r}} & \epsilon_{\mathrm{f}}\bar{\epsilon}_{\mathrm{f}} \\ \bar{\epsilon}_{\mathrm{r}}\bar{\epsilon}_{\mathrm{r}} & \bar{\epsilon}_{\mathrm{r}}\bar{\epsilon}_{\mathrm{f}} & \bar{\epsilon}_{\mathrm{f}}\bar{\epsilon}_{\mathrm{r}} & \bar{\epsilon}_{\mathrm{f}}\bar{\epsilon}_{\mathrm{f}} \end{pmatrix} \begin{pmatrix} N_{\mathrm{rr}} \\ N_{\mathrm{rf}} \\ N_{\mathrm{fr}} \\ N_{\mathrm{ff}} \end{pmatrix}. \tag{5.2}$$

The efficiency ϵ_{r} is measured using MC samples in a region enriched in real leptons from Z boson decays. The efficiency ϵ_{f} is measured in data in a region enriched in fake leptons by requiring two leptons with the same charge. The efficiencies are determined as a function of the lepton p_{T} and η. Inverting the matrix M results in expressions for the total number of single and double fake events N_{rf}, N_{fr}, and N_{ff}, from which one can construct the total contribution of events with electrons in the signal region (SR)

$$N_{\mathrm{fakes}}^{\mathrm{SR},e\mu} = N_{\mathrm{rf}} + N_{\mathrm{fr}} + N_{\mathrm{ff}}. \tag{5.3}$$

The contribution of fake events in the signal region is then given by recorded data events passing the loose event selection weighted such that the above relation holds.

Fake estimate in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel

Jets misidentified as hadronically decaying τ leptons is an important background in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel. Due to the use of the electron-BDT, electrons misidentified as $\tau_{\text{had-vis}}$ are negligible and modeled directly with MC samples. Jets misidentified as τ leptons originate from QCD multijet background and W+jets events. The data-driven fake factor method is employed to estimate their contribution to the signal region (SR). For the fake factor method, an anti-tau region is defined by inverting the identification requirements on the hadronic τ candidate. A minimal RNN identification score of 0.01 is imposed to counteract the prevalence of gluon-initiated over quark-initiated jets if no condition was applied. The fake factor method takes the relative fraction of quark and gluon-initiated jets into account. Recorded data events $N_{\mathrm{Data}}^{\text{anti-}\tau}$ in this anti-tau region are taken as the fake events in the signal region weighted by a fake factor $\mathcal{F}$. The contribution $N_{\mathrm{MC,\,not\,jet\,fake}}^{\text{anti-}\tau}$ in the anti-tau region not originating from jets faking hadronic τ leptons is estimated with MC and receives a negative weight to remove its contribution. Mathematically, the total fake yield is given by

$$N_{\mathrm{fakes}}^{\mathrm{SR},\tau_{\mathrm{lep}}\tau_{\mathrm{had}}} = \mathcal{F} \times \left(N_{\mathrm{Data}}^{\text{anti-}\tau} - N_{\mathrm{MC,\,not\,jet\,fake}}^{\text{anti-}\tau} \right). \tag{5.4}$$

To determine the fake factor $\mathcal{F}$, two control regions are defined that are enriched in QCD multijet events and W+jets events, respectively. The QCD multijet control region is defined by inverting the lepton isolation criteria and characterized by gluon-initiated jets. The W+jets is defined by inverting the requirement on the transverse mass to $m_{\mathrm{T}} > 70$ GeV and is characterized by quark-initiated jets. The control regions are subdivided into the tau and anti-tau regions, as introduced above. For each control region, the fake factor is derived as the ratio of events that pass the tau selection criteria over the number of events in the anti-tau region. Any contribution in the control regions not

originating from a jet faking a tau is estimated with MC and subtracted before taking the ratio. The fake factors derived in each control region i separately are defined as

$$\mathcal{F}_i = \frac{N^{\tau,i}_{\text{Data}} - N^{\tau,i}_{\text{MC, not jet fake}}}{N^{\text{anti-}\tau,i}_{\text{Data}} - N^{\text{anti-}\tau,i}_{\text{MC, not jet fake}}}. \tag{5.5}$$

The fake factor $\mathcal{F}_W$ from the W+jets enriched control region and the fake factor $\mathcal{F}_{\text{QCD}}$ from the QCD multijet enriched region are used in a linear combination to form the final fake factor

$$\mathcal{F} = R_{\text{QCD}}\mathcal{F}_{\text{QCD}} + (1 - R_{\text{QCD}})\mathcal{F}_W \tag{5.6}$$

used in the analysis. The weights of the linear combination are chosen to account for the expected ratio R_{QCD} between gluon and quark-initiated jets in the signal region. Across the categories and for 1-prong and 3-prong taus, R_{QCD} ranges between 0.02 and 0.11. The fake factors $\mathcal{F}_W$ and $\mathcal{F}_{\text{QCD}}$ differ by typically 30 %.

The fake factors are computed separately for each signal region category, for one and three-pronged τ leptons, and as a function of p_{T}.

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

Conceptually, the data-driven fake estimation for the $\tau_{\text{had}}\tau_{\text{had}}$ channel is similar to the fake factor method employed in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. The fake factor method in the $\tau_{\text{had}}\tau_{\text{had}}$ channel is used to estimate the contribution of objects falsely identified as hadronically decaying taus. The main source in this channel is QCD multijet events. The situation, however, is more complex since we need to consider single and double tau fakes.

The contribution in the SR is derived from anti-tau regions, where at least one of the taus fails the medium identification criterion, but at least one tau passes the loose identification criteria. The second condition must be imposed for technical reasons since events without any loose id tau lepton are removed from the dataset. Similar to Equation (5.4), the contribution from prompt hadronically decaying taus is estimated in MC and subtracted from data.

The fake factor method for the $\tau_{\text{had}}\tau_{\text{had}}$ channel uses two sets of fake factors to address the combinatorial complexity of single and double fakes. The fake factors are derived in the W+jets control region of the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. The criteria on the hadronically decaying tau have been modified to match those of the $\tau_{\text{had}}\tau_{\text{had}}$ channel to make the fake factors obtained in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel applicable for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The condition $p_{\text{T}} > 40\,\text{GeV}$ and the η requirement on the leading jet are kept to increase the statistics. The fake factors are parametrized as a function of p_{T} and η of the tau candidates.

5.5. Multivariate analysis

Today, large precision MC datasets are the perfect opportunity for machine learning applications to aid measurements of rare processes like the VBF Higgs production that occurs only in $O(10^{-10})$ cases, see Figure 2.3.

This section first describes the Boosted Decision Tree (BDT) used in Publication [84] to define a very pure VBF category. The second part of this section presents an alternative method to enhance the measurement sensitivity of the VBF cross section using a multiclass neural network.

5.5.1. VBF Tagger

The measurement of the VBF cross section is limited by systematic uncertainties from background processes and events from other Higgs boson production modes in the VBF phase space, primarily from ggF. Effectively suppressing background processes can thus improve the measurement precision. The event selection and categorization described in Sections 5.3.2 and 5.3.3 defines an inclusive VBF signal category that achieves a VBF signal purity of 3.8 %. Machine learning tools can be employed to further increase the purity and statistical power. Machine learning classifiers can learn correlations between input variables that are difficult to exploit *by hand* when defining an event selection using rectangular cuts on kinematic variables.

Using machine learning classifiers, such as a BDT, can introduce unwanted sensitivity biases in the newly defined signal categories. For example, the classifier could distort the reconstructed MMC mass of the selected signal, often referred to as *sculpting*. The mass distribution is used in the likelihood fit described in Section 5.7. In principle, sculpting is not an issue if the distortion affects the MC sample and data in the same way. However, it is still preferred to reproduce the original shape of the Higgs boson peak in the mass spectrum even after applying any machine learning classifier.

One way to counteract sculpting is by selecting a suitable set of input variables. The VBF structure of the processes shown in Figure 5.5 for the Higgs boson signal and electroweak Z production provides the possibility to identify these events by only considering the two jets in the forward and backward directions. Due to the conservation of momentum, knowing the momenta of the two forward and backward jets allows computing the transverse momentum of the Higgs boson system (or Z boson in case of electroweak Z production). However, the invariant mass of the Higgs boson (or Z boson) system is not accessible from the forward and backward jets. Using only information from the forward and backward jets as the input variable of a machine-learning classifier is expected to prevent sculpting of the MMC mass.

The BDT used in the VBF category for the full Run 2 analysis [84] uses only the kinematic properties of the two leading jets and the magnitude of the vectorial sum of momenta of the two leading jets, visible tau decay products τ_i and $E_{\mathrm{T}}^{\mathrm{miss}}$, defined as

$$p_{\mathrm{T}}^{Hjj} = \left| \boldsymbol{p}_{\mathrm{T}}^{\tau_0} + \boldsymbol{p}_{\mathrm{T}}^{\tau_1} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} + \boldsymbol{p}_{\mathrm{T}}^{j_0} + \boldsymbol{p}_{\mathrm{T}}^{j_1} \right|. \tag{5.7}$$

Plugging-in the definition of $E_{\mathrm{T}}^{\mathrm{miss}}$, Equation (3.11), into Equation (5.7) shows that all momenta cancel except for the momentum of any jet with lower momentum than the sub-leading jet. In an idealized VBF event, the variable p_{T}^{Hjj} vanishes. The quantity describes additional jets or QCD activity in the event and can be used by a machine-learning algorithm to reject non-VBF backgrounds.

During the training, all $Z \to \tau\tau$ samples and ggF signal samples are considered as background. This includes the electroweak Z boson production. Events from the VBF

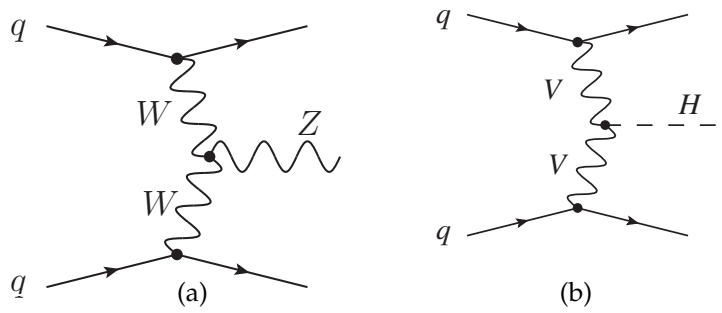


Figure 5.5.: Feynman diagrams of the VBF production of a Z boson (a) and Higgs boson (b) with the two characteristic jets in the forward directions.

Table 5.2.: Training parameters of the BDT as defined in Reference [157]. The number of trees used as a free parameter to be optimized.

Parameter	Value
Minimal node size	2.5 %
Maximal tree depth	3
Number of cut candidates	20

signal samples are considered as signal. Any event with a negative event weight is ignored during the training. The event weights for the background and signal class are rescaled such that both classes have the same sum of weights.

The BDT is trained in TMVA [157]. The training parameters are summarized in Table 5.2. The number of trees n_{trees} is initially kept as a variable hyperparameter that requires further optimization. The training dataset is randomly split into four chunks to facilitate 4-fold cross validation. In a rolling fashion, four BDTs were trained on different input chunks using two training chunks for each BDT. The remaining chunks can be used as validation sets and test sets. This ensures that events used during the training of the BDT are not used again in the optimization or in the final analysis. Similarly, the events in the test set used in the analysis were not previously used during training or optimization. This method effectively prevents biases on data stemming from overtraining. The parameter n_{trees} is optimized on the validation set to maximize the area under the receiver operating characteristic (ROC) curve. The best performance is obtained for $n_{\text{trees}} = 1000$.

In the analysis, the BDT is used to define two VBF subcategories, one enriched in VBF signal, the other enriched in $Z \to \tau\tau$ background and ggF signal. Events passing a BDT score threshold enter the signal-enriched category. Events that fail the condition constitute the background-enriched category. For this, the BDTs are operated on their test set. For samples not used in the training, e.g., the diboson background, events are randomly but reproducibly assigned to one of the four BDTs. The same method is used for data events. The cut on the BDT score to define the signal-enriched category is determined considering systematic uncertainties. The best performance was obtained

with a cut at 0.15. The distribution of the output score is shown in Figure 5.6.

The subcategory enriched in VBF signal events has a very high contribution from VBF signal events out-weighting background events in the signal-sensitive bins. The agreement of the background prediction and data has been verified in control regions and in the sidebands of the MMC distribution in the signal region. Figure 5.7 shows the two VBF subcategories for all channels combined.

5.5.2. Deep neural network

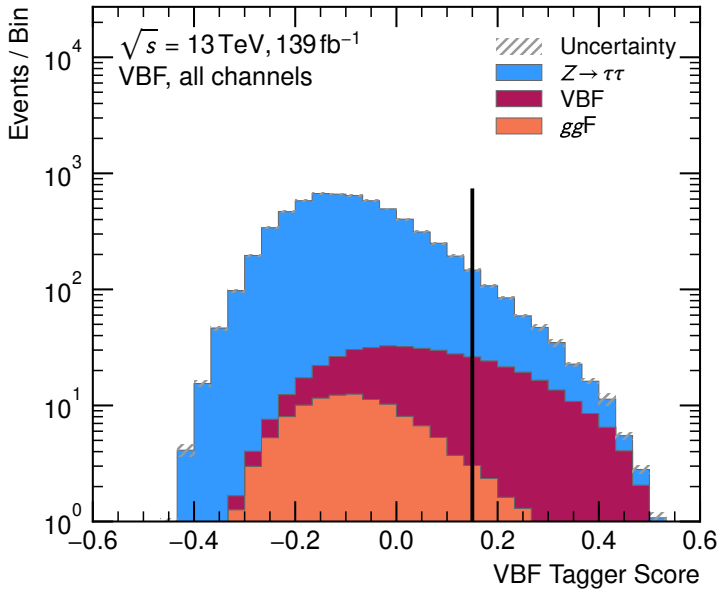
Similar to the use of the BDT in the previous section, the goal of the application of a neural network is to reduce the impact of processes with large systematic uncertainties, namely ggF and $Z \to \tau\tau$. With neural-network-based approaches, it is straightforward to extend the classification to more than two classes. As an alternative to the aforementioned VBF tagger, a NN is trained to classify events into either VBF Higgs boson signal, ggF signal, or $Z \to \tau\tau$.

As demonstrated in Reference [158], deep neural networks can learn from low-level quantities, such as object 4-momenta or even the detector hits. Traditionally, the event selection is based on high-level quantities, such as transverse momenta, invariant masses, or angular differences between objects. BDTs or shallow neural networks are usually not able to compute complex high-level quantities from low-level input variables in order to correctly classify events. With deep neural networks, the network can learn to compute useful quantities in hidden layers. Since high-level variables are derived from their low-level counterparts, a deep neural network learning complex correlations between its low-level inputs can even outperform a classifier trained on high-level quantities. The definition of high or low-level variables depends on the context. The choice of input variables can also be seen as a trade-off. Lower-level input variables give the network the opportunity to learn correlations that are otherwise hidden in high-level variables. On the other hand, high-level inputs benefit from the physics insight into the classification problem and require less complex networks and usually fewer training events to train a performant classifier.

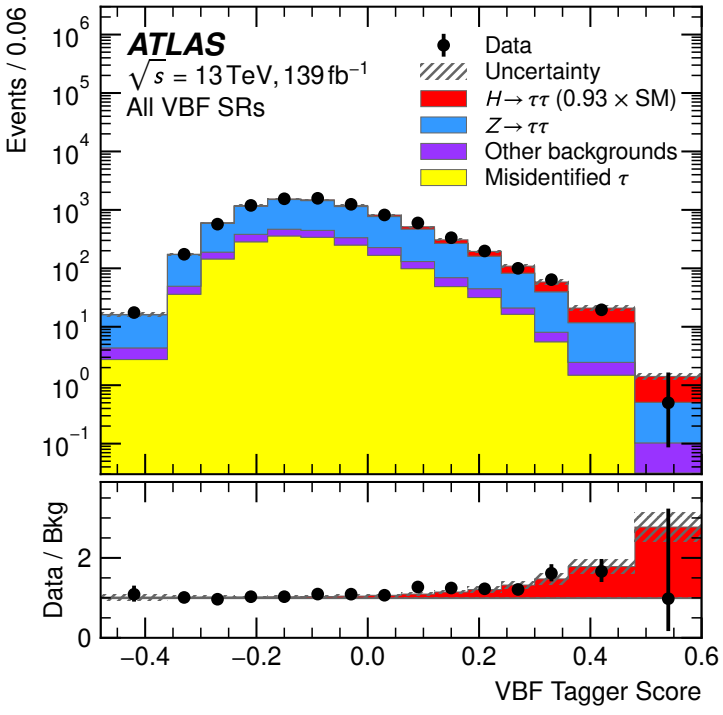
The neural network employed in the VBF category uses the low-level 4-momenta of the two leading jets and p_{T}^{Hjj} as defined in Equation (5.7) as input variables. The 4-momenta are parametrized by ϕ, η, and p_{T}, implicitly assuming that their mass is zero. Therefore, the input layer consists of seven input nodes. This set of variables is chosen to exploit correlations in the jet 4-momenta and therefore learn the VBF event topology while preventing the network from learning the invariant mass of the Higgs boson system. The input variables are pre-processed as outlined in Section 2.4. Figure 5.8 shows the distribution of the input variables and the invariant mass of the two jets m^{jj} in the inclusive VBF signal region for all channels combined. Figures B.1–B.3 in the appendix show the inputs distributions for each channel separately.

The network uses three output nodes with softmax activation, to classify events into the three classes VBF signal, ggF signal, and $Z \to \tau\tau$. The output value of each node can then be interpreted as the probability of an event belonging to the corresponding class. The categorical cross entry as defined in Equation (2.32) is used as the loss function.

Between the input and output layer, a variable number of densely connected hidden layers are placed. ReLU activation is used for the hidden layers. The number of layers

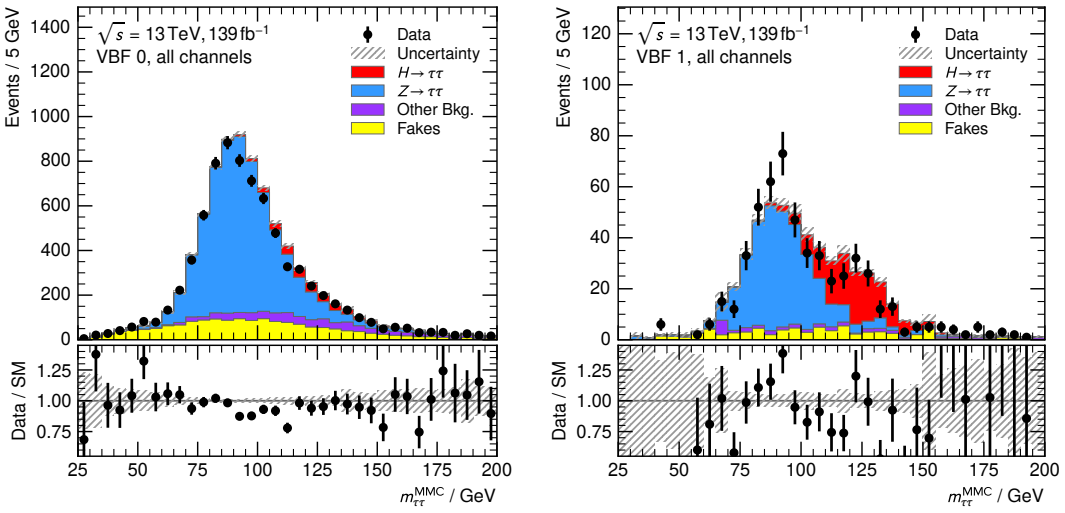


(a)



(b)

Figure 5.6.: BDT output for the signal (VBF) and background (ggF and $Z \to \tau\tau$) classes (a), and with all processes and data [84] (b). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.



(a) (b)

Figure 5.7.: VBF signal event categories defined by the BDT. The histograms show the background-enriched categories (a) and the signal-enriched categories (b) for all analysis channels combined. The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

and the number of nodes per hidden layer are treated as hyperparameters subject to optimization.

The hyperparameter optimization is facilitated using the open-source parameter optimization framework Optuna [159]. The framework uses an iterative search in hyperparameter space to minimize an objective function using a Tree-structured Parzen Estimator [160]. For each proposed hyperparameter configuration, four neural networks are trained in a 4-fold cross validation scheme as described in Section 2.4. The networks are trained for 300 epochs. The objective function, in this case, is the negative area under the ROC curve that shows the inverse background efficiency as a function of the signal efficiency where the signal is the VBF signal class and the background is the union of the ggF and $Z \to \tau\tau$ classes.

The hyperparameter space is spanned by the number of hidden layers, the number of nodes in the first and last hidden layers, and the learning rate. The number of hidden layers is limited to at least two and at most six layers. The number of nodes in the first hidden layer is an integer in the interval $[20, 250]$. The number of nodes in the last hidden layer is an integer in the interval $[5, 100]$. The number of nodes for other hidden layers is interpolated between the number of nodes in the first and last hidden layer leading to a trapezoidal shape of the network. The learning rate is also optimized from values between 10^{-5} and 10^{-2} with the logarithm of the learning rate following an initially uniform distribution.

The number of events per class in the available MC data set is listed in Table 5.3. The data set is split for training, validation, and testing according to the procedure outlined in Section 2.4.

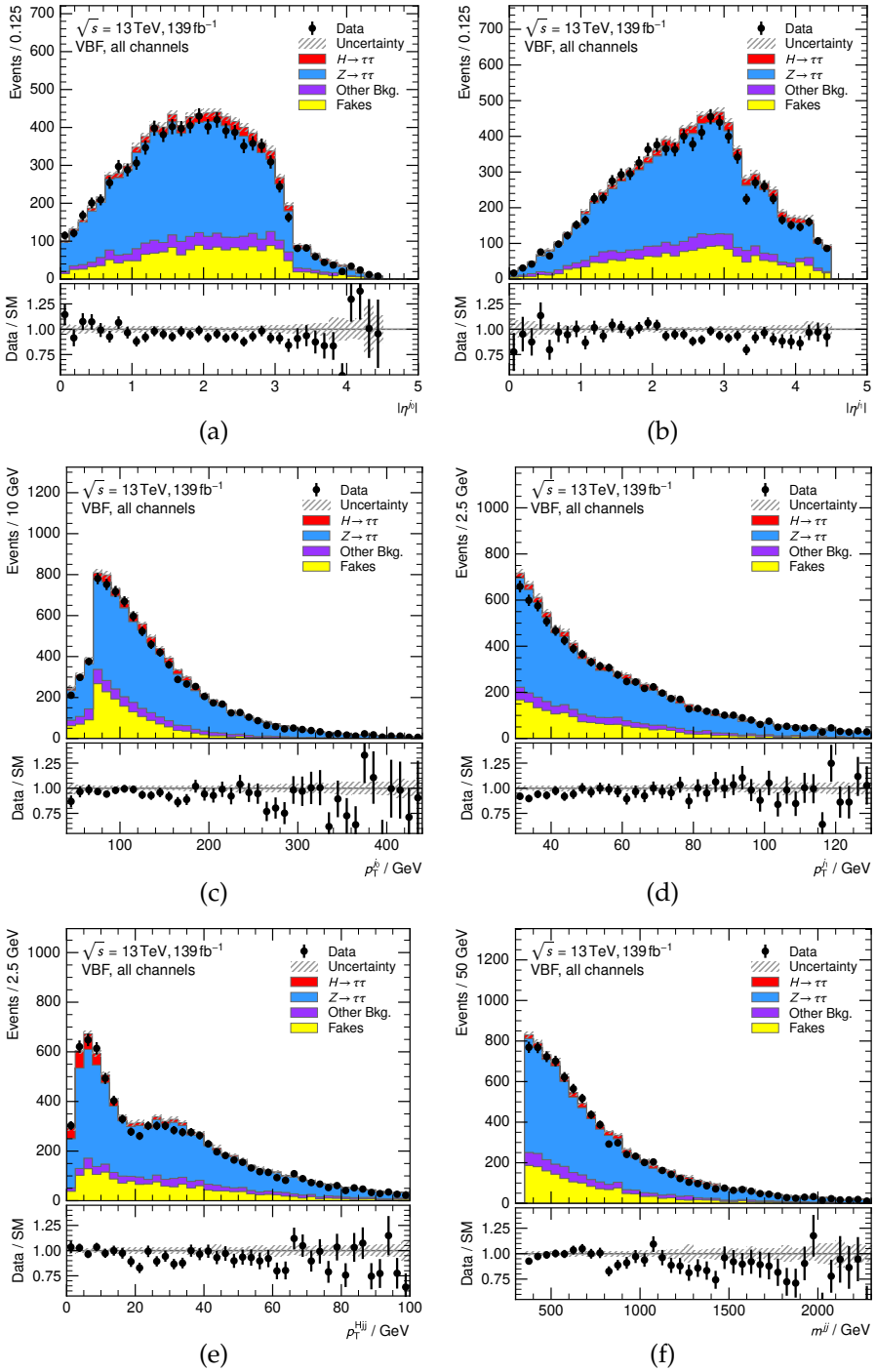


Figure 5.8.: Selected input variables of the neural network in the VBF inclusive signal region for all channels combined (a)–(e), and the invariant mass of the two jet system m^{jj} (f). The step in the $p_{\mathrm{T}}^{j_0}$ distribution (c) is due to the higher jet threshold in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

Table 5.3.: Summary of event counts in the MC training dataset.

Process	Number of events
VBF	234173
ggF	25326
$Z \to \tau\tau$	207400

Table 5.4.: Optimized hyperparameter values and their search ranges.

Parameter	Search range	Optimal value
Number of hidden layers	$[2, 6]$	3
Number of nodes in first hidden layer	$[20, 250]$	244
Number of nodes in last hidden layer	$[5, 100]$	96
Learning rate	$[10^{-5}, 10^{-2}]$ (log)	$\approx 1.77 \times 10^{-3}$

The neural network is implemented in Keras [161] interfaced to Tensorflow [162] as backend. The training is facilitated in the FreeForestML framework developed for neural network training in high-energy physics contexts. Alternatively, the analysis and the training was implemented using Apache Spark. However, this approach suffered from inefficient interfaces between frameworks and formats and was ultimately abandoned in favor of FreeForestML with Keras and Tensorflow. The hyperparameter optimization is performed on GPUs and CPUs in a distributed and dockerized [163] fashion using a central database to store evaluation results. In total, 150 hyperparameter configurations have been assessed. Each hyperparameter configuration is represented as a line in Figure 5.9. The plot shows that the optimization procedure favored three hidden layers and a rather high learning rate. Small learning rates or a large number of hidden layers achieve relatively poor performance. Most of the lines cluster towards the lower end of the objective axes demonstrating the search procedure of the optimization framework and indicating that a large range of parameters gives comparable performances.

The optimal parameters found by the hyperparameter search and the search ranges are listed in Table 5.4. The network favors an architecture where the number of nodes reduces toward the output layers. It is interesting to note that the optimization yielded a relatively small number of hidden layers, resulting in five layers in total, including the input and output layers.

With the optimized configuration, the loss on the validation set is monitored during the training of 300 epochs to employ early stopping. The loss function on the training and validation set as a function of the training epochs is shown in Figure B.4. After 100 epochs, no significant improvement of the validation loss is observed. As explained in Section 2.4, the two networks are retrained with the optimal configuration on the first and second half of the event set with 100 epochs.

As an initial benchmark to compare the performance of the NN approach to the BDT-based VBF tagger, one can compare their ROC curves. Figure 5.10 shows the inverse

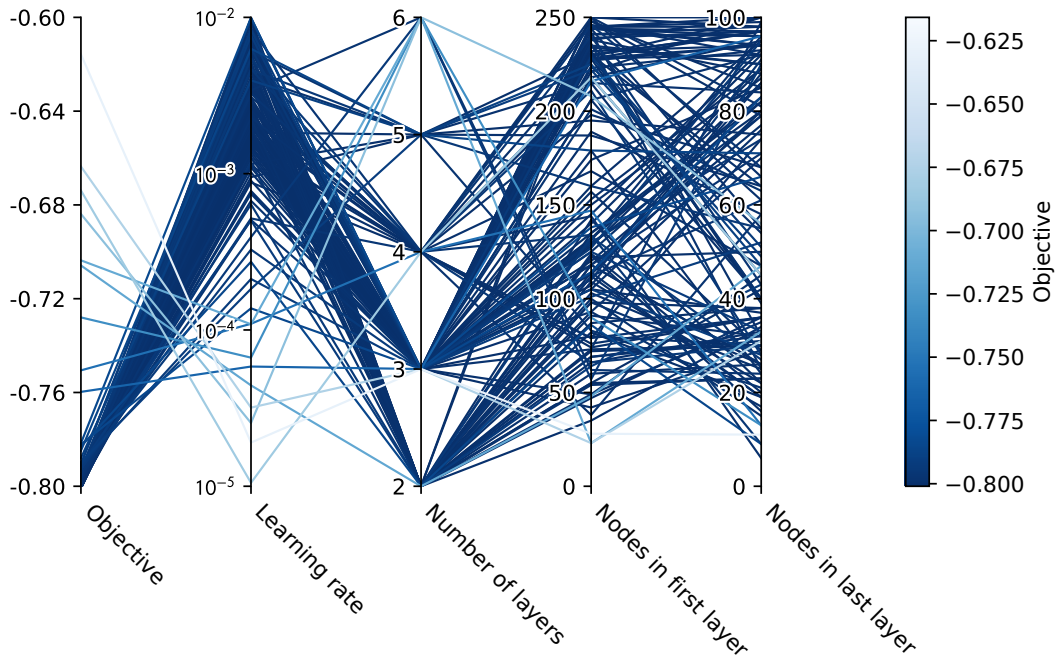


Figure 5.9.: Visualization of all hyperparameter configurations and its performance in a parallel plot. Each configuration is presented as a line connecting all axes. Darker values correspond to configurations with more performant networks.

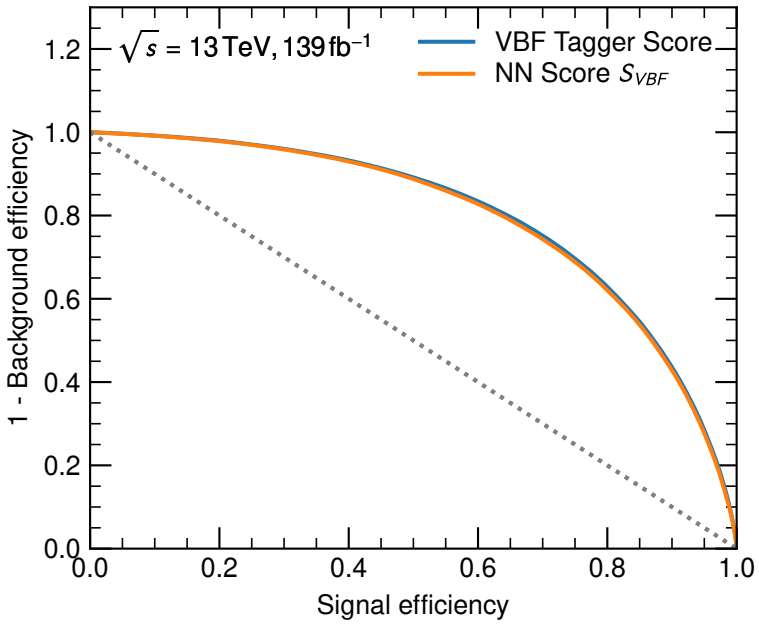


Figure 5.10.: Comparison of the ROC curves for the BDT-based classifier and the multiclass NN used as a binary classifier.

background efficiency as a function of the signal efficiency. Similar to the optimization procedure, the VBF class is treated as a signal and the union of ggF and $Z \to \tau\tau$ as background. The fact that the multiclass network is used as a binary classifier to facilitate the comparison means that this is not a full and fair evaluation. However, Figure 5.10 serves as an initial check. Both classifiers show similar background efficiencies for a given signal efficiency. The ROC curves are almost identical with areas under the curve of $C_{\text{BDT}} = 0.797 \pm 0.007$ for the BDT and $C_{\text{NN}} = 0.793 \pm 0.005$ for the NN. The small difference between the ROC values and the fact that the two classifiers are trained on an input variable set with similar information content is an indication that both classifiers have approximated the probability density function of the input variables sufficiently well such that the discrimination power cannot be improved any further without changing the input variable sets.

A possible limitation of the neural network and the BDT might be the size of the training dataset. The optimized configuration is trained on various random slices of the training dataset covering a broad range of sizes to assess the impact of the artificially reduced training set size. The area under ROC curve as a function of the training set size is shown in Figure B.5. The performance is flat for a wide range of input sizes. The performance deteriorates only with less than 150×10^3 available training events. For the NN training, the classifier is not limited by a lack of training events.

There are several options to use the multiclass network in the physics analysis. Since the classifier output is independent of $m_{\tau\tau}^{\text{MMC}}$, it is beneficial to use the $m_{\tau\tau}^{\text{MMC}}$ distribution in the likelihood fit. The output scores of the NN are shown in Figure 5.11. In analogy to the BDT, we use the VBF classification score S_{VBF} to define the category

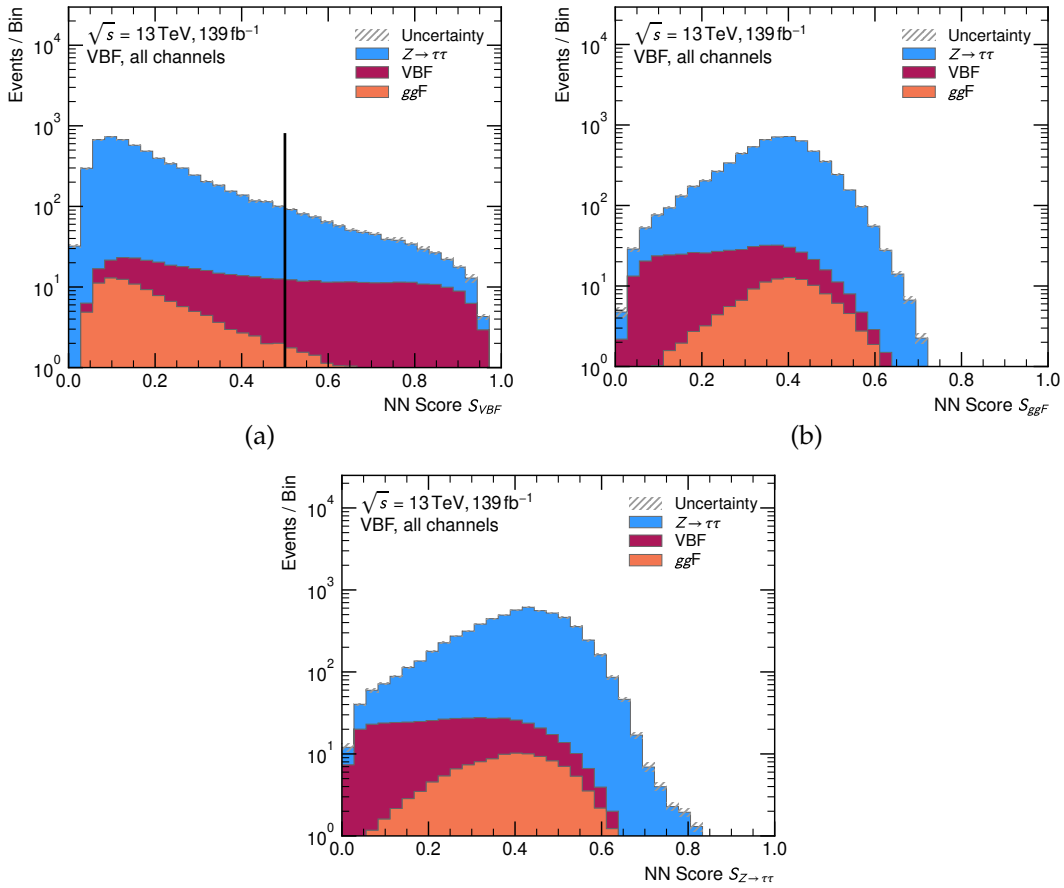


Figure 5.11.: Output scores of the NN for the training classes VBF (a), ggF (b), and $Z \to \tau\tau$ (c) for all analysis channels combined. Note that each event appears in all three histograms. The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

enriched in VBF signal. Events with $S_{\text{VBF}} > 0.5$ form the VBF signal enriched region. For the remaining events, the greater of the two remaining scores $S_{Z\to\tau\tau}$ or S_{ggF} determines the event's category. The categories defined with this prescription are shown in Figure 5.12. Figures B.6–B.8 show the output categories defined by the NN for each channel separately.

5.6. Systematic uncertainties

The cross section measurement depends on the number of observed events and is thus subject to statistical uncertainties. However, the MC generation, the operation of the LHC and the ATLAS detector, and the detector reconstruction and simulation are all subject to systematic uncertainties. Systematic uncertainties can affect the background and signal model in two ways depending on the nature of the uncertainty.

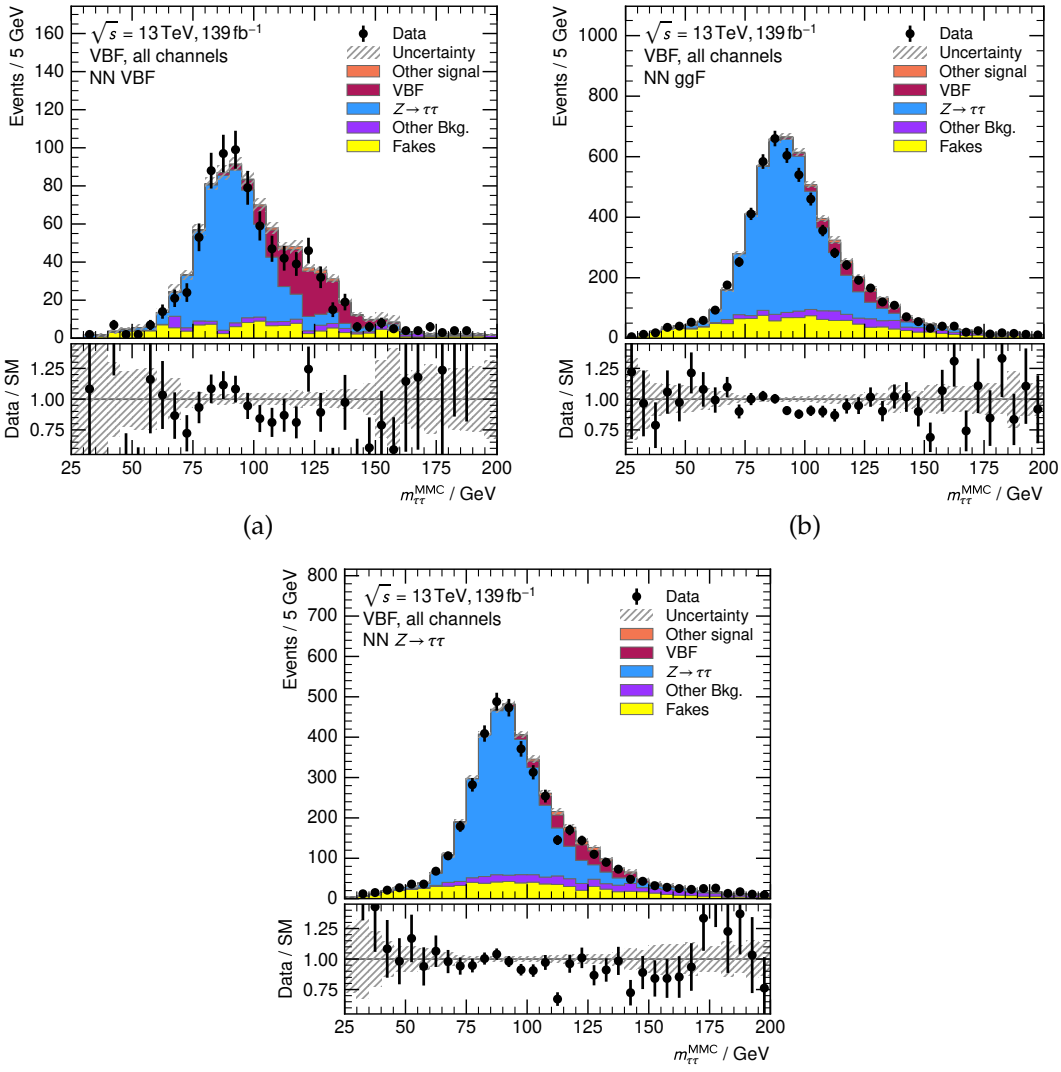


Figure 5.12.: Exclusive VBF signal region categories as defined by the NN output for all analysis channels combined. By definition, the categories are designed to increase the fraction of events from VBF (a), ggF (b), and $Z \to \tau\tau$ processes (c). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

One set of systematic uncertainties affects only the event weight. To implement these uncertainties in the analysis, the set of events in each signal category is considered again with the alternative event weights applied. The distributions constructed in this way can differ from the nominal distributions in the overall normalization, in the shape of the distribution, and the relative composition of processes since the varied weights can depend on kinematic properties and the physics process.

The second group of systematic uncertainties affects kinematic properties of the event, such as reconstructed 4-momenta or whether particles are correctly identified. These variations require reprocessing the whole dataset. Derived quantities such as $E_{\mathrm{T}}^{\mathrm{miss}}$ or $m_{\tau\tau}^{\mathrm{MMC}}$ are recomputed based on the varied event. Events can migrate into the signal categories, fall out of acceptance, or migrate between signal categories. Furthermore, since kinematic quantities are changed, these variations can modify the shape of distributions in the signal category and their relative process composition.

Besides this categorization from the implementation side, one can group the uncertainties in three groups based on their physics origin: experimental uncertainties, theoretical uncertainties on the background, and theoretical uncertainties on the signal. The three groups are discussed in detail in the following sections. In most cases, the systematic uncertainties are parametrized as a $\pm 1\,\sigma$ variation on the initial source and propagated to the final distribution used in the likelihood fit. As described in Section 5.7, each systematic uncertainty is handled by a nuisance parameter with a constraint term.

5.6.1. Experimental uncertainties

The dominant experimental uncertainties for the analysis are the jet energy scale, tau identification and its energy scale, and misidentified taus. For the data-driven fake methods, specific uncertainties have been assigned to assess the uncertainty introduced by the estimation technique.

The uncertainty of the jet energy scale (JES) originates from the uncertainty of the in situ calibration, pile-up jets, the extrapolation of the calibration to higher jet momenta, and differences in the detector response between quark and gluon-initiated jets. The uncertainty on the jet energy scale is 1 % in the central region of the detector and over a wide momentum range $250\,\mathrm{GeV} < p_{\mathrm{T}} < 2000\,\mathrm{GeV}$. The largest impact on the analysis from jet energy resolution (JER) comes indirectly from its effect on $E_{\mathrm{T}}^{\mathrm{miss}}$, which enters the computation of the invariant mass $m_{\tau\tau}^{\mathrm{MMC}}$. A worse energy resolution broadens the $Z \to \tau\tau$ resonance peak in the $m_{\tau\tau}^{\mathrm{MMC}}$ spectrum, also affecting the mass range most sensitive to the Higgs boson signal. The relative JER depends on the jet momentum and is between (24 ± 5) % at 20 GeV and (6.0 ± 0.5) % at 300 GeV [70].

The tau identification efficiency uncertainty is between 2 % and 6 %. The tau trigger efficiency uncertainty is between 1 % and 1.5 %. The tau energy scale (TES) is of the order of 1 % to 4 %. The uncertainties depend on the tau momentum and track multiplicity. The efficiency uncertainty of the electron-BDT used to reject electrons misidentified as taus is between 1 % and 2 % and depend on tau momentum and on decay mode, which includes information on the neutral tau decay products (mainly neutral π^0 mesons). The tau reconstruction efficiency and its uncertainty are typically measured using MC samples [164].

The luminosity used to scale the MC expectation to the recorded data events was obtained from a measurement of the LUCID-2 detector [165]. The uncertainty of the measurement is 1.7 % [24]. This uncertainty is not considered for background samples whose normalization is measured in control regions.

The data-driven fake modeling methods are affected by systematic uncertainties. In the $\tau_e\tau_\mu$ channel, dedicated uncertainties are introduced to account for the statistical uncertainty of the efficiency ϵ_f and ϵ_r measurement, the dependency of the efficiencies on the number b-tagged jets and the process in which they are measured, and the normalization of processes with a real tau.

For the fake factor method in the $\tau_\mathrm{lep}\tau_\mathrm{had}$ channel, dedicated uncertainties account for the statistical uncertainty of the fake factors and R_QCD, uncertainties from the flavor composition, and the uncertainty from processes with a real tau. In the $\tau_\mathrm{had}\tau_\mathrm{had}$ channel, the fake factor method received dedicated uncertainties from the statistical uncertainty of the fake factor measurement, the uncertainty in the composition of the background, and an uncertainty originating from the choice of parametrization. The sizes of the additional uncertainties in all three channels are between 5 % and 15 %.

A 1 % uncertainty has been assigned to the extrapolation of the Z+jets normalization measured in the control region to the signal region. The small extrapolation uncertainty is justified by the kinematic embedding approach that eliminates phase space mismatches. The size is based on discrepancies observed between Data and MC. The uncertainty is assumed to be uncorrelated between analysis regions [84].

5.6.2. Background theory uncertainties

Systematic uncertainties from the theoretical prediction of the background processes are only considered for the two main background processes Z+jets and $t\bar{t}$. The contribution from other backgrounds in the signal region is small, and their theory uncertainty is negligible.

For the Z+jets sample, uncertainties of the matrix element are evaluated by six variations of the factorization scale μ_F and the renormalization scale μ_R [166]. The scales are each multiplied by 0.5, 1, and 2, not considering the combination where both scales go in opposite directions. The uncertainty of the strong coupling constant is handled by varying it by approximately ± 0.85 %. An uncertainty is assigned to the resummation by varying the resummation scale by a factor of 0.5 and 2. Furthermore, uncertainties related to the matching between partons and jets are introduced based on the CKKW parameters of Sherpa.

For the $t\bar{t}$ sample, the uncertainties of the matrix element are estimated by the comparison with an MC sample generated with MadGraph5_aMC. The uncertainty of the parton shower is estimated by the comparison with an MC sample showered with Herwig 7.04 [167]. The uncertainty from initial state radiation (ISR) is obtained by variation of generator parameters, the renormalization and factorization scales, μ_R and μ_F, and the propagation of the uncertainty of α_S [168]. To account for the uncertainty from final state radiation (FSR), the renormalization scale for emissions is varied by a factor 0.5 and 2.

For both samples, the uncertainty of the PDF set is estimated from the 100 replicas of the PDF, measuring the standard deviation of the spread. Additionally, the PDFs are

compared to the CT14NNLO [169] and MMHT2014NNLO [170] PDF sets.

5.6.3. Signal theory uncertainties

Theory uncertainties in the signal prediction represent the largest limitation of the analysis. In principle, one would expect the impact of signal theory uncertainties to be small in a cross section measurement compared to a signal strength measurement since the former does not depend on the predicted signal yield. However, in this analysis, the uncertainties introduced by the parton shower play a large role and have the largest impact on some of the measured cross sections.

To assess the uncertainty of the parton shower and hadronization model for the main signal process, the nominal signal samples created with Powheg Box v2 interfaced to Pythia 8 are compared to the expected signal from a Powheg Box v2 sample interfaced to Herwig 7. The uncertainty of the PDF4LHC15NLO PDF set is parametrized by their eigenvectors and propagated to a varied MC sample.

The effects of missing higher orders in the matrix element computation are considered, and dedicated uncertainties are assigned. Especially the ggF contamination in the VBF enriched category is subject to uncertainties following the Steward-Tackmann prescription [171] handling uncertainties introduced by the BDT or NN that act as a third-jet veto in the VBF enriched category.

Additional variations are introduced to account for uncertainties on the Higgs boson momentum in the case of no additional jets and the uncertainty of the top quark mass in loop corrections. The comparison of the nominal sample generated with Powheg Box v2 with a sample generated with MadGraph5_aMC using the FxFx prescription [172] gives rise to uncertainties on ggF in the VBF and VH phase space.

5.7. Statistical analysis

The statistical analysis is performed using a binned likelihood fit of the prediction model to the observed data. The prediction is given as a binned template in the $m_{\tau\tau}^{\text{MMC}}$ observable for each signal region and category. For the control regions, the template consists of a single bin. The model is parametrized as a function of a set of normalization factors $\boldsymbol{\phi}$ that scale the expected signal or background yields. Normalization factors for the signal can be interpreted as parameters of interest (POI) to perform a measurement and estimate confidence intervals of the parameter. Furthermore, the model is given as a function of nuisance parameters $\boldsymbol{\theta}$, that account for systematic variations. Formally, the expected number of events in a bin b is denoted $\nu_b(\boldsymbol{\phi}, \boldsymbol{\theta})$. The nuisance parameters are constrained by auxiliary measurements. Given a true value θ_j of a nuisance parameter, the probability of observing the value a_j in the auxiliary measurement is denoted $A_j(a_j|\theta_j)$. Often, a Gaussian distribution is used for the auxiliary constraint, and their parameter θ_j is scaled to be centered around zero with unit standard deviation. Technically, the predicted number of events $\nu_b(\boldsymbol{\phi}, \boldsymbol{\theta})$ for arbitrary θ_j are interpolated from the $\pm 1\,\sigma$ variations obtained by varying the systematic uncertainties.

Following the approach outlined in [173], the full statistical model is given by

$$P(\boldsymbol{n}, \boldsymbol{a}|\,\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \left(\prod_{b \in \text{bins}} \text{Pois}\left(n_b|\gamma_b \nu_b(\boldsymbol{\phi}, \boldsymbol{\theta})\right) \text{Pois}(m_b|\gamma_b \tau_b) \right) \cdot \left(\prod_j A_j(a_j|\theta_j) \right) \quad (5.8)$$

where n_b denotes the number of observed events in bin b. The Poisson term $\text{Pois}(m_b|\,\gamma_b \tau_b)$ accounts for the statistical uncertainty of the MC samples. The creation of m_b unweighted MC events in bin b is treated as a random experiment. The term quantifies the probability to observe m_b MC events when expecting $\gamma_b \tau_b$ events. The observed number of events is used as the best estimate for the number of expected MC events $\tau_b = m_b$. The term is to be understood as a function of γ_b, whose most likely value is $\gamma = 1$. In practice, the number of MC events is estimated by considering the uncertainty σ on the event yield $\nu_b = \sum_{\text{i events}}^{m_b} w_i$ given by

$$\sigma = \frac{\sqrt{m_b}}{m_b} \cdot \nu_b = \frac{\nu_b}{\sqrt{m_b}}. \quad (5.9)$$

Instead of treating the uncertainty for every sample independently, only the combined uncertainty of all processes and samples in each bin is considered.

The dataset recorded with the ATLAS detector provides the number of observed events $\boldsymbol{n}$. Since $\boldsymbol{n}$ and also $\boldsymbol{a}$ are fixed after recording and analyzing the dataset, the probability $P(\boldsymbol{n}, \boldsymbol{a}|\,\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma})$ can be regarded as a function of its unknown conditional parameters and is then termed the likelihood function

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = P(\boldsymbol{n}, \boldsymbol{a}|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}). \quad (5.10)$$

Maximizing the likelihood function with respect to its parameters yields the most likely values for the parameter of interest and, therefore, the result of the measurement. The test statistic of the analysis is the likelihood ratio. The confidence intervals on the parameters of interest are derived using the asymptotic approximation. A more detailed description can be found in References [174, 175].

5.8. Results

The analysis has been performed in a *blinded* way to avoid biasing analysis decisions on statistical fluctuations in data. The data points of signal-sensitive bins in the $m_{\tau\tau}^{\text{MMC}}$ distribution between 100 GeV and 150 GeV were removed until the analysis was finalized. For the same reasons, the fit model has been tested first with an artificial *Asimov* dataset built to perfectly match the expected yields in every bin. As a next step, the signal prediction was scaled by a secret random but reproducible factor. These so-called random-μ fits allow assessing the behavior of nuisance parameters and systematic uncertainties without being able to infer the potential effect of decisions on the signal measurement.

This section summarizes the results of the analysis. First, blinded comparisons between different machine learning classifiers are presented. Secondly, the results of the partially blinded random-μ fits are discussed. Finally, the chapter concludes with the presentation of the unblinded cross section measurement.

5.8.1. BDT and NN comparison

Section 5.5 described the training of a BDT and an alternative multiclass NN to classify events in the VBF category. To get a realistic comparison of the two methods, both have been evaluated with systematic uncertainties to produce the input histograms for the likelihood fit. In the case of the BDT, two VBF categories have been used, while for the NN, the VBF category is split into three subcategories. The comparison is intended to show the advantages of the different methods. For this approach, a fully blinded Asimov fit is sufficient. Rerunning the kinematic rescaling procedure in the $Z \to \ell\ell$ control regions, including its systematic uncertainties with the NN classifier, is a major effort and requires considerable computing power. Therefore, the complexity of this setup is reduced by removing the $Z \to \ell\ell$ control regions. By construction, measuring the Z normalization in the control region on an Asimov dataset would yield unity. The impact on the statistical uncertainty of the normalization is negligible. Furthermore, the $t\bar{t}H$ categories are pruned from the setup. It is expected that these changes do not impact the conclusions drawn from a blinded fit. The comparison is performed with a 3-POI fit, with r_{VBF}, r_{ggF}, and r_{VH} as parameters of interest.

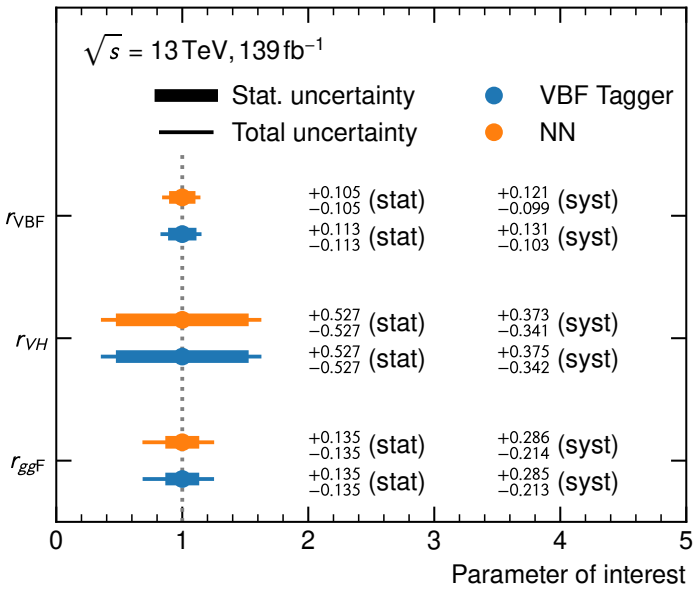
Figure 5.13 summarizes the expected sensitivity on the POI for the BDT and the NN. Since there is no change to the ggF or VH signal category, the uncertainty on the POI is approximately equal between the two methods, as expected. On the other hand, the NN classifier is able to produce approximately 8 % smaller uncertainties on the r_{VBF} parameter resulting in a more precise expected cross section measurement. The improvement affects both the statistical and the systematic part of the uncertainty as would be required for better signal discrimination. The reduction of the statistical uncertainty is assumed to originate from the increased number of bins in case of the multiclass NN with different signal purities.

5.8.2. Partially unblinded fit

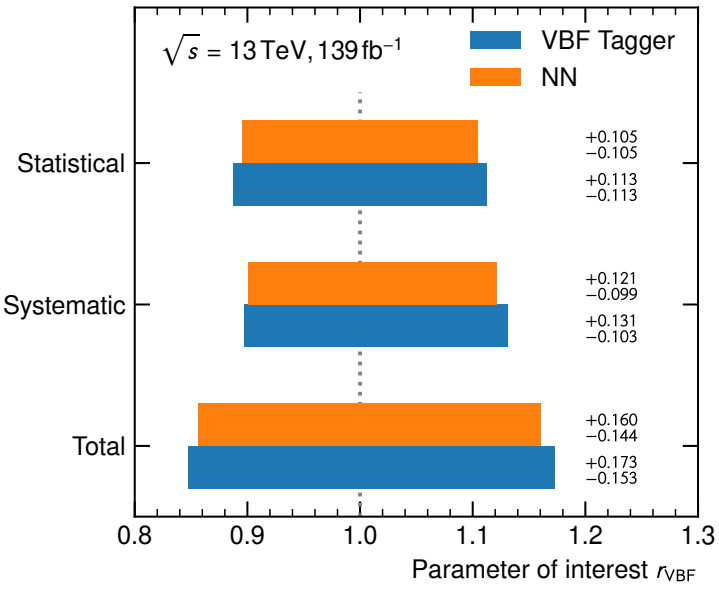
A so-called random-μ fit is performed to validate the fit model and study constraints or pulls of nuisance parameters. Fits performed on an Asimov dataset do not show nuisance parameters or norm factors deviate from its nominal value since, by construction, the dataset matches the nominal prediction. However, once the data is unblinded, it is expected to show small, random deviations from the nominal prediction. To study these deviations and the effect of normalization factors applied in control regions and signal regions, one needs to include the measured dataset in the fit.

With a random-μ fit, it is possible to study pulls and constraints while keeping the analysis blinded. Before running the fit on the measured dataset, the signal prediction is scaled by a random factor u. The effect of the random scaling on all other parameters and normalization factors of the fit is expected to be small. The values of nuisance parameters that maximize the likelihood and their uncertainties are expected to be similar to the value obtained from a fully unblinded fit on measured data. The procedure is illustrated in Figure 5.14.

The scaling is performed by introducing an additional constant parameter in the fit model. The random-μ parameter is applied to the signal samples analog to the POI. Before maximizing the likelihood, the signal prediction is scaled by setting the random-μ parameter to a random value u and making the parameter a constant vari-



(a)



(b)

Figure 5.13.: Comparison of the sensitivity between VBF Tagger and NN for each parameter of interest in a 3-POI fit (a) and a decomposition of the uncertainties for the r_{VBF} parameter (b). The NN-based analysis achieves an approximately 8 % smaller expected uncertainty on the VBF cross section measurement. The fit is performed on an Asimov dataset, and the central value is exactly unity.

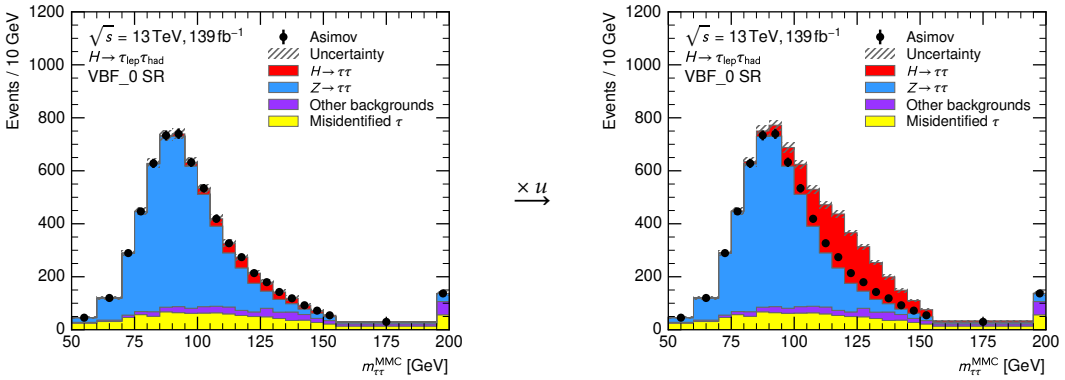


Figure 5.14.: Illustration of the random scaling of the signal prediction. The signal part of the initial MC distribution (left) is scaled by the random factor u to obtain the new MC template (right) in the random fit. The black points represent the Asimov dataset built from MC to match the prediction before scaling the signal.

able. The value of u is not revealed during the procedure. Before using the method, its implementation was tested and verified on an Asimov dataset with the unscaled signal expectation.

The random value u is drawn from a Gaussian distribution with a mean value of 1.5 and a standard deviation of 0.3. The range of the random distribution has been chosen to avoid convergence issues observed for fits with large random scales, e.g., $u = 10$. The random value is drawn from an unbounded random distribution (as opposed to, e.g., a uniform interval) to void the potential leakage of information about the real POI value. For example, consider a random value u drawn from the inclusive interval $[2, 10]$ and a fitted POI of $\mu = 0.8$, one could infer that the real POI satisfies $r_{\text{signal}} \geq 2 \times 0.8$.

The value u is generated using a pseudo-random generator with a known seed value. Reusing the same seed reproduces previous results. Switching the seed value was rarely done to avoid inferring the signal strength from the average of multiple fits with different seed values.

Figure 5.15 shows the values of the nuisance parameters that maximize the likelihood and their uncertainty. The full fit model consists of approximately 1000 parameters. The diagram is limited to those parameters whose central value is pulled by more than one unit of their post-fit uncertainty. It is not expected to observe 31.7 % of the assigned error bars to exclude the initial value. The auxiliary measurements characterized by A_j in Equation (5.8) represent the best measurement or calibration performed with the ATLAS detector. The analysis itself cannot be seen as an independent experiment, and the assigned error bars are expected to contain the initial value far more frequently than 68.3 %.

The pulls shown in Figure 5.15 can be grouped into three categories: related to the estimation of misidentified taus or electrons, related to the jet energy resolution, and related to the tau reconstruction efficiency. The pulls of the nuisance parameters re-

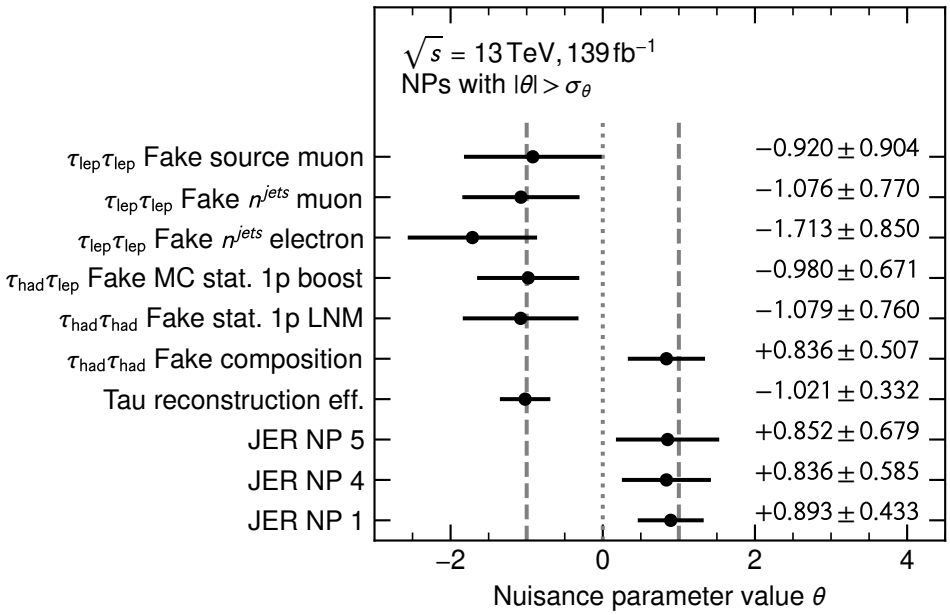


Figure 5.15.: Observed pulls and constraints in the random-μ fit. Out of the approximately 1000 parameters, only those are shown whose pull is larger than the post-fit uncertainty. The parameters can be grouped into related to the fake-estimate, related to the jet energy resolution, and related to the tau reconstruction efficiency.

lated to fake modeling are not surprising. The uncertainties are in part assigned to the fake modeling method because discrepancies between the modeling and data were observed, and the uncertainty is a parametrization of unknown complexity.

The jet energy resolution is parametrized by twelve effective nuisance parameters obtained as statistically independent uncertainties from a principal component analysis. Three of these parameters appear in Figure 5.15. The sensitivity of the analysis to JER is known since the Run 2 analysis on the partial dataset [83]. The dependence is introduced via the mass reconstruction $m_{\tau\tau}^{\text{MMC}}$. The jet energy resolution affects the missing transverse mass $E_{\text{T}}^{\text{miss}}$ via the momenta of the physics objects and the soft terms. A worse jet energy resolution leads to a broader Z resonance in the $m_{\tau\tau}^{\text{MMC}}$ spectrum. A careful analysis of this effect in collaboration with the relevant groups within the ATLAS Collaboration led to the conclusion that the analysis is sensitive to the jet energy resolution, and the observed pulls of approximately 1σ in a small fraction of the nuisance parameters are not a sign of an issue in the fit model.

Lastly, the pulls and constraints show a significant difference from the nominal value for the τ reconstruction nuisance parameters. The issue can be traced to the initial estimation of the τ reconstruction efficiency and its uncertainty. The initial τ reconstruction uncertainty is based on MC simulation. However, similar to misidentified τ leptons, inefficiencies in the τ reconstruction are difficult to assess with MC events only.

Neglecting the effect of the $Z\ell\ell$ control regions for now, a lower-than-expected τ reconstruction efficiency could be absorbed by the free parameter scaling the total Z+jets normalization as discussed in Section 5.4.1. However, the kinematic rescaling method employed in the $Z\ell\ell$ control regions allows measuring the Z+jets normalization in data. The observed pull of the reconstruction efficiency is, therefore, an indication of a deficit in the τ reconstruction efficiency. Based on these insights, the fit model was modified. The Gaussian constraint of the reconstruction efficiency nuisance parameter has the tendency to shift a large pull to other correlated nuisance parameters, such as the identification efficiency nuisance parameter, which results in an overall smaller penalty term A_j in the likelihood function in Equation (5.8). To avoid any bias of the identification efficiency or other correlated nuisance parameters, the uncertainty of the reconstruction efficiency was increased, making A_j a flat function and effectively removing the constraint from the model. This change reflects the insight that with this analysis, the reconstruction efficiency can be measured more precisely than the initial estimate only based on MC simulation.

Figure 5.16 illustrates the effect of the tau reconstruction efficiency in the analysis categories. The top panel shows the process composition in the signal categories in a window around the Z boson resonance. The width of the window is adjusted per category to match the bin boundaries used in the fit setup. Each bin in the figure corresponds to a single signal category. The filled stacks are the pre-fit expected yields, while the green outline is the post-fit result. The post-fit yields reflect the best-fit configuration of all normalization and nuisance parameters that maximize the likelihood. As by construction, the post-fit outline agrees with observed data points, given the uncertainties in the bottom panel. The green outline includes the effect of the reduced tau reconstruction efficiency as determined by the fit.

The orange line demonstrates the effect of the tau reconstruction efficiency nuisance parameter when manually setting its post-fit value back to the nominal expectation. This is identical to fixing the reconstruction efficiency to its pre-fit value. Comparing the effect in the different signal categories, it is reassuring that resetting the reconstruction efficiency has no impact on the $\tau_e\tau_\mu$ channel. It is difficult to draw quantitative conclusions from this representation since the effect of the reconstruction efficiency on the fake modeling is not easy to assess.

The effect of the tau reconstruction efficiency becomes clearer by looking at a signal-only window around the Higgs boson peak. Again, the window width is adjusted per category to match the bin boundaries used in the fit setup. Figure 5.17 shows the expected yields in green per signal category after applying the best-fit nuisance parameters. The figure does not show the number of observed data events nor the number of expected events, as this study was performed when the analysis was still blinded. Similar to the plot around the Z mass window, the orange outline in Figure 5.17 shows the expected yields after manually setting the tau reconstruction efficiency to its pre-fit value. The x-axis categories are sorted by channel. The emerging steps can be understood by considering the tau multiplicity in each channel. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the event yield is $\Delta_1 = (5.20 \pm 0.02)\,\%$ higher if we assume the initial reconstruction efficiency, while in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the event yield is $\Delta_2 = (10.70 \pm 0.08)\,\%$ higher if we assume the initial reconstruction efficiency. Due to the tau multiplicity, one expects the change of the single tau channel to affect the $\tau_{\text{had}}\tau_{\text{had}}$ channel as its squared

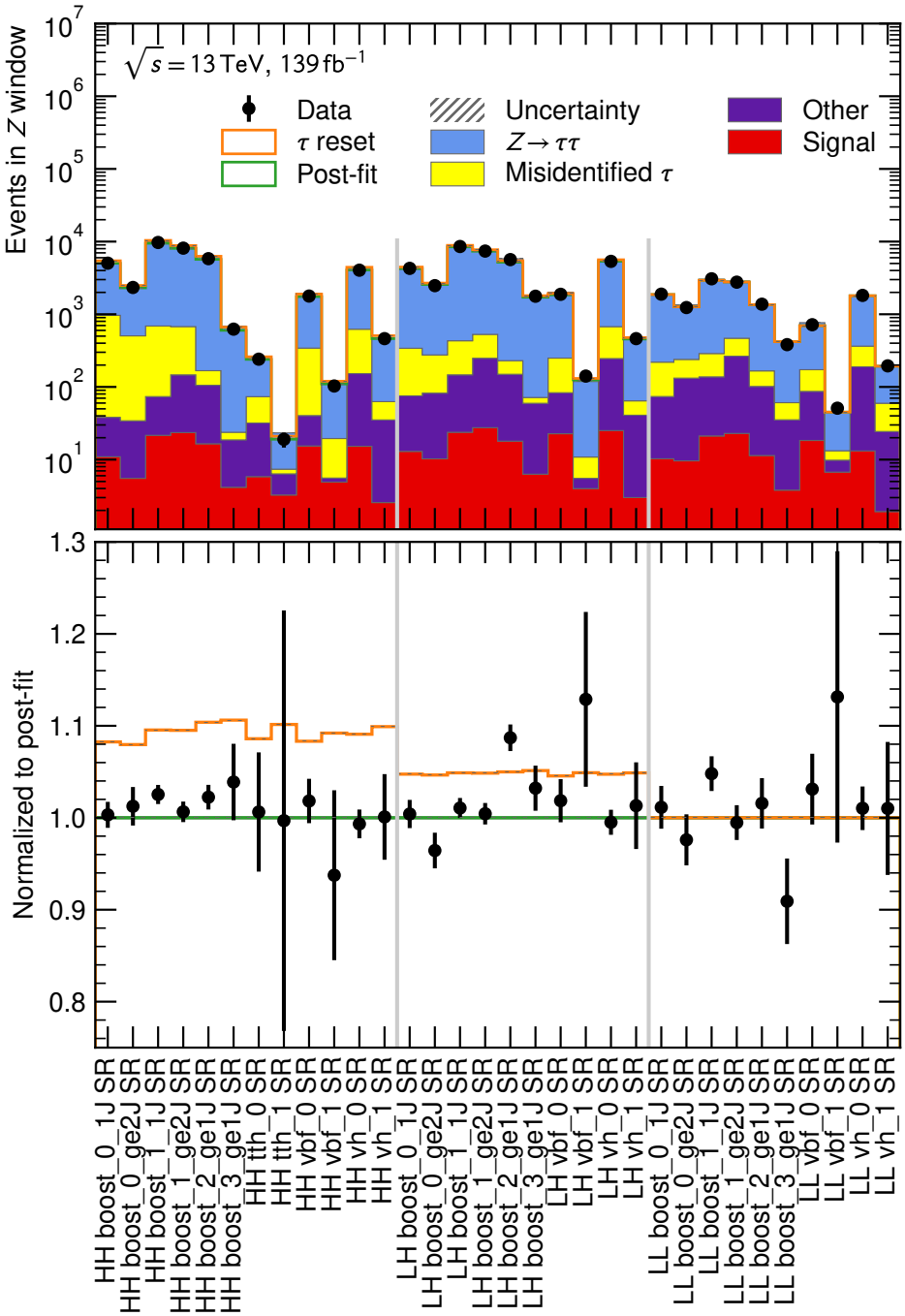


Figure 5.16.: Post-fit effect of the tau reconstruction efficiency in all signal regions in the mass window around the Z boson. The top panel shows the absolute yield comparision between data, and pre-fit, post-fit and τ-reset expectations. The bottom panel represents the comparison normalized to the post-fit expectation.

effect,

$$(1+\Delta_1)^2 = 1 + \underbrace{2\Delta_1 + \Delta_1^2}_{=\Delta_2} = 1 + (10.67 \pm 0.04)\,\% \tag{5.11}$$

which matches the observed value of Δ_2 within the uncertainty.

The observation that the τ reconstruction efficiency can be measured with this analysis has influenced the fit setup and promoted the nuisance parameter to a free parameter in the likelihood fit. On the other hand, the fact that the kinematic rescaling technique allows measuring the reconstruction efficiency in data means that the efficiency could be determined in data instead of MC. Such a global measurement is not yet in place but could be a potential improvement for a wide range of analyses with τ leptons.

5.8.3. Fully unblinded fit

The fully unblinded fit yields the measurement results of the $pp \to H \to \tau\tau$ analysis using the full Run 2 dataset with an integrated luminosity of $139\,\text{fb}^{-1}$. The results are published in Reference [84]. The measured quantity is the product of the Higgs boson production cross section and the branching ratio to tau leptons. Figure 5.18 summarizes the measurement of the total cross section and the measurement for each production mode. The total measured cross section

$$\sigma_{H\to\tau\tau} = 2.94 \pm 0.21(\text{stat})\,^{+0.37}_{-0.32}(\text{syst})\,\text{pb} \tag{5.12}$$

is in agreement with the SM expectation of $(3.17 \pm 0.09)\,\text{pb}$ given the statistical and systematic uncertainties. For the total cross section measurement, the p-value is 0.58, which illustrates its compatibility with the SM expectation. Figure 5.19 shows the $m_{\tau\tau}^{\text{MMC}}$ distributions with the best-fit value of all nuisance parameters applied and the post-fit uncertainty of the background and signal model. Figures B.9–B.11 show the same distributions split by analysis channel.

The observed pulls and constraints of nuisance parameters have been found to agree with the results from the partially unblinded random-μ fits. The analysis and the fit setup have been compared to fits split by production mode and by decay channel. No significant discrepancy between the channels or between the production modes have been found. The results show good compatibility with the SM prediction with p-values between 0.30 and 0.98.

Besides the total cross section measurement, the cross section has been measured for each production mode individually in a simultaneous fit. The most precise measurement is obtained for the VBF production mode. This is enabled by the signal enriched VBF category that, in turn, reduces the impact of theory uncertainties from other processes. The observed cross section is

$$\sigma_{H\to\tau\tau}^{\text{VBF}} = 0.197 \pm 0.028(\text{stat})\,^{+0.032}_{-0.026}(\text{syst})\,\text{pb}. \tag{5.13}$$

The VBF production mode was observed at a significance of $5.3\,\sigma$ ($6.2\,\sigma$ expected). The second most precisely measured production mode was ggF. The measured cross section is

$$\sigma_{H\to\tau\tau}^{gg\text{F}} = 2.7 \pm 0.4(\text{stat})\,^{+0.9}_{-0.6}(\text{syst})\,\text{pb}. \tag{5.14}$$

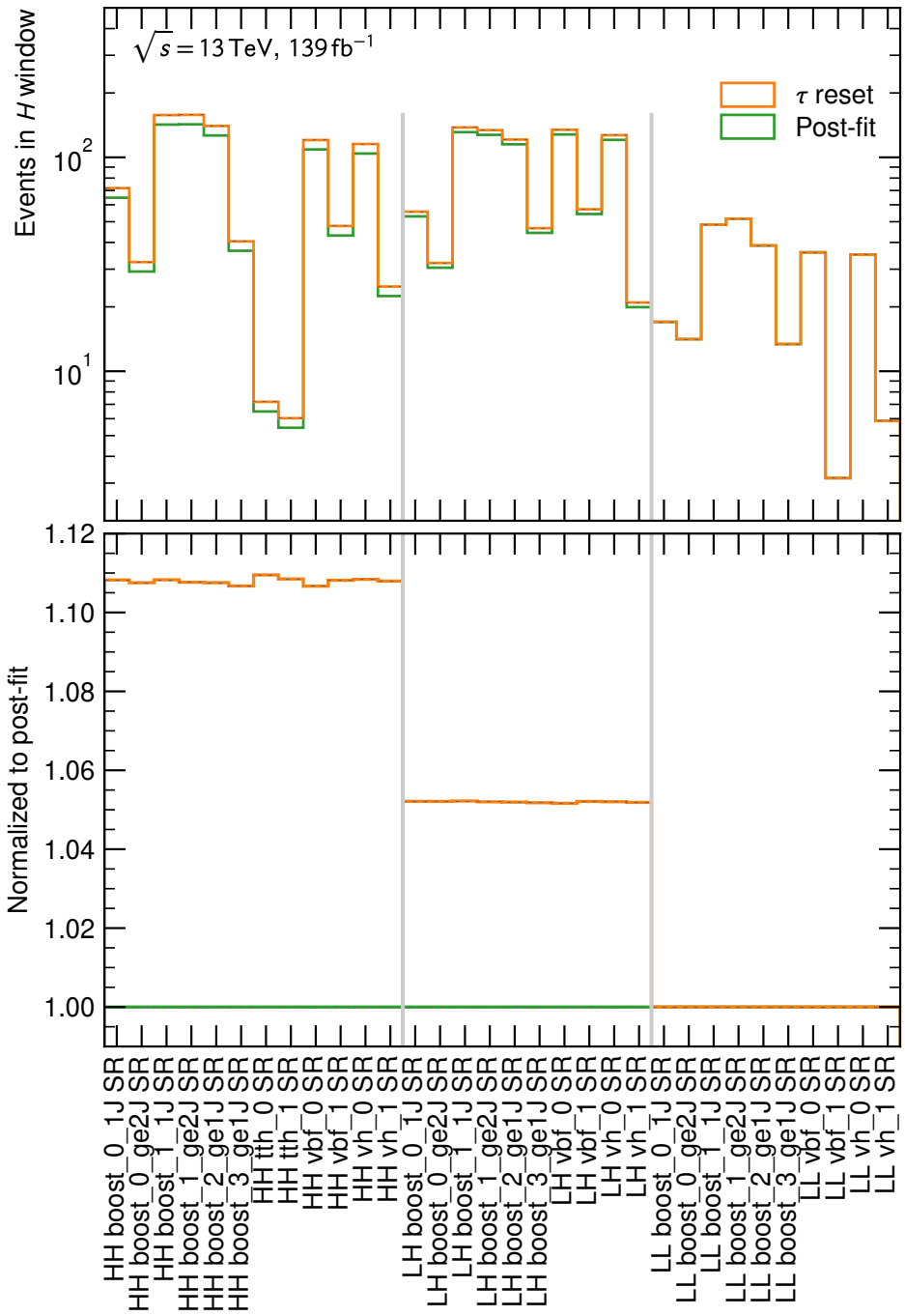


Figure 5.17.: Post-fit effect of the tau reconstruction efficiency in all signal regions on the signal in the mass window around the Higgs boson. The top panel shows absolute yield comparision between post-fit and τ-reset expectations. The bottom panel represents the comparison normalized to the post-fit expectation.

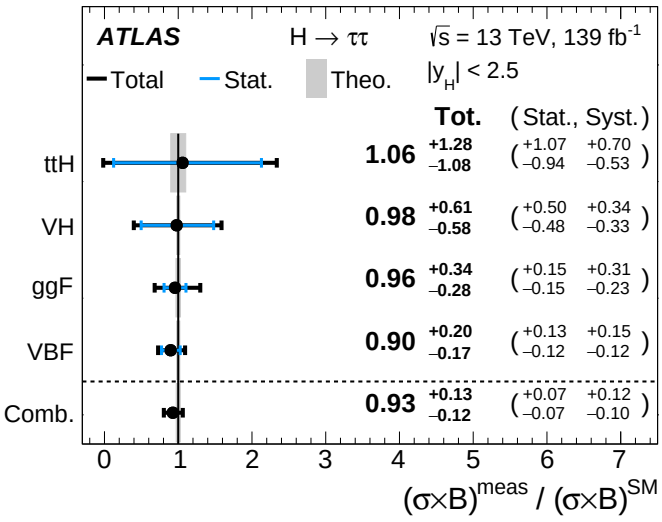


Figure 5.18.: Observed cross sections normalized to the SM prediction. The top part shows the cross section measurement for each Higgs boson production mode. The bottom part shows the total combined cross section [84].

Finally, the measured cross sections for VH and $t\bar{t}H$ are $0.12 \pm 0.6(\text{stat}) \pm 0.4(\text{syst})$ pb and $0.033^{+0.033}_{-0.029}(\text{stat})\,^{+0.022}_{-0.017}(\text{syst})$ pb, respectively, which both are limited by statistical uncertainties. Additionally, cross sections have been measured in nine bins of the STXS framework. The measured cross section for each STXS bin is shown in Figure 5.20.

The measurements for VBF and ggF, and since these two production modes drive the combined cross section precision, also the total cross section measurement, are limited by systematic uncertainties. The relative impact of different sources of uncertainty on the measured total cross section is listed in Table B.2. The ranking of the individual nuisance parameters is shown in Figure 5.21. Theoretical uncertainties on the signal prediction have the largest impact. In contrast to a signal strength measurement that depends on the theory prediction, one does not expect theory uncertainties on the signal cross section to affect a cross section measurement. However, the large impact of theory uncertainties is due to the parton shower model, especially in the VBF and ggF signal regions. As described in Section 5.6.3, the parton shower uncertainty is estimated by the comparison of Pythia 8 and Herwig 7. The generator comparison is an estimate of the uncertainty in the absence of a better understanding of the parton shower. The discrepancy between the two generators might improve in the future. The largest experimental uncertainty is related to jets and $E_{\text{T}}^{\text{miss}}$.

Comparing these results and the first Run 2 measurement [83], this analysis significantly improves the measurement precision in the ggF and VBF production cross section. Measurements for the VH and $t\bar{t}H$ were not performed in the previous publication. The greater statistical power makes it possible to perform an STXS measurement in nine bins compared to only three in the previous analysis. Additionally, the impact of systematic uncertainties has been significantly reduced. The uncertainty from signal theory predictions, those related to jets and $E_{\text{T}}^{\text{miss}}$, and the MC statistical uncertainty

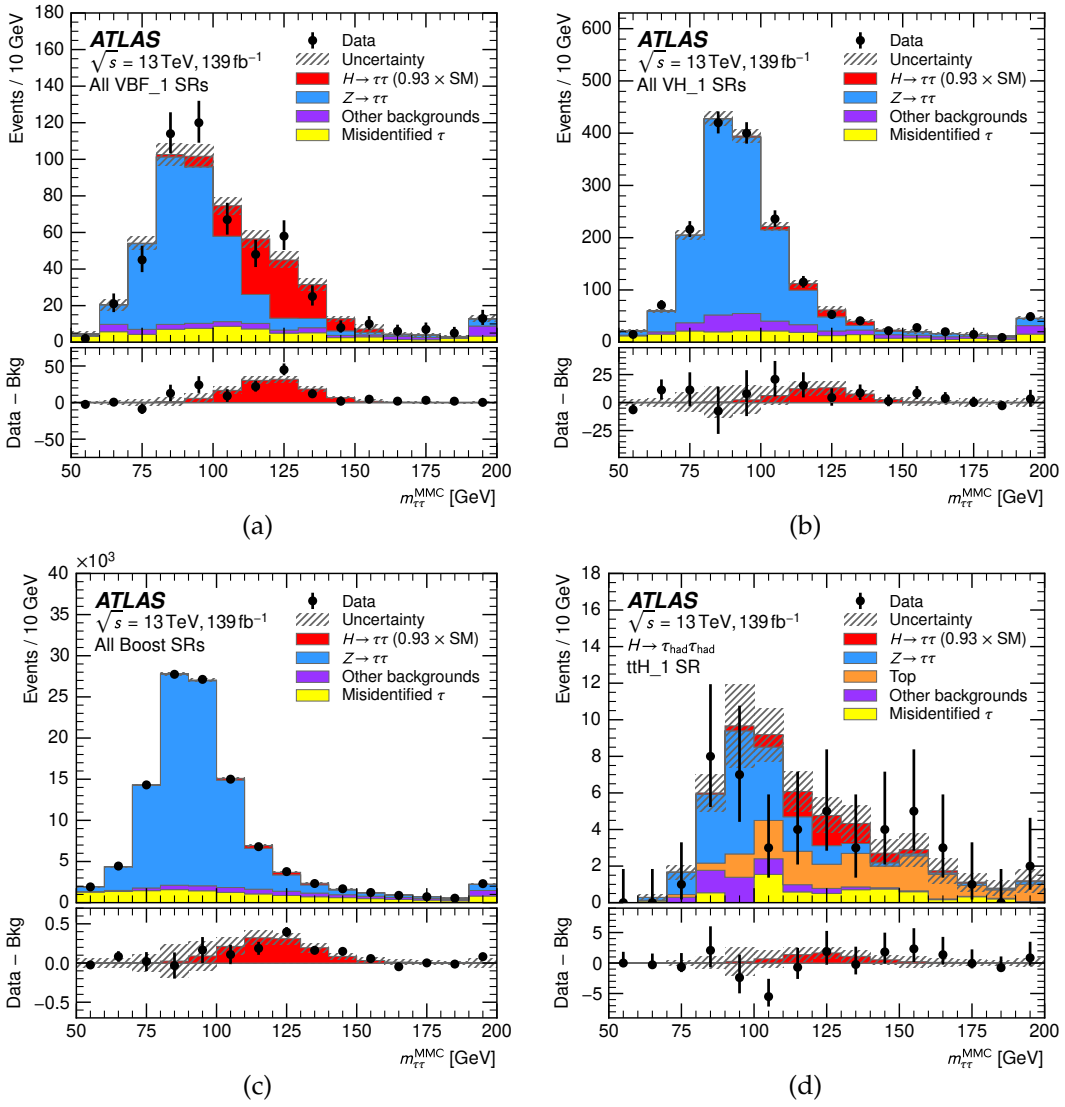


Figure 5.19.: Invariant mass $m_{\tau\tau}^{\mathrm{MMC}}$ distributions in selected signal regions with the all nuisance parameters set to the value that maximizes the likelihood function [84]. The hatched band and the error bars indicate total uncertaintainty of the background expectation and statistical unceratinty measured data, respectively.

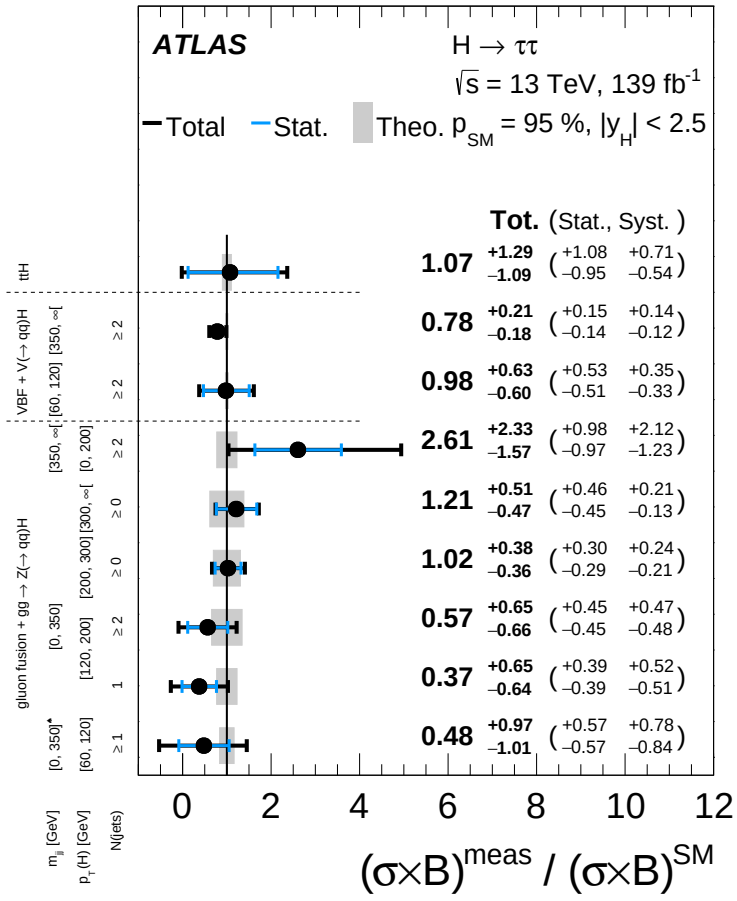


Figure 5.20.: Observed cross sections in each measured STXS bins normalized to the SM prediction [84].

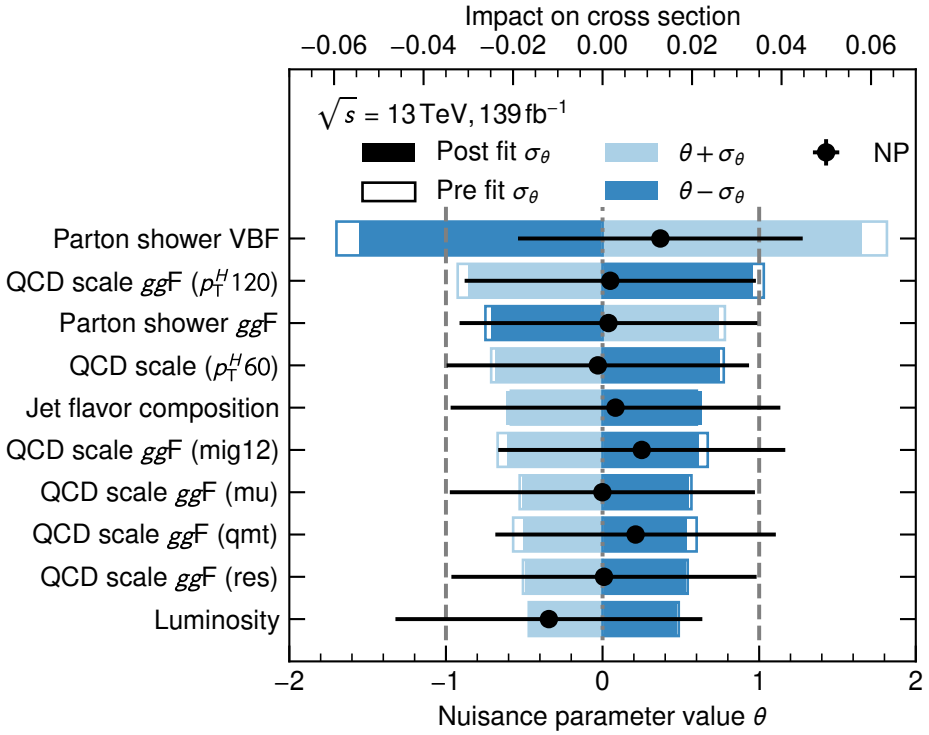


Figure 5.21.: Pull of the nuisance parameters in the 1-POI fit and their relative impact on the measured cross section. The nuisance parameters are sorted by their impact. Only the first ten most significant parameters are shown.

have each been reduced by a factor of 1.5–3.

CHAPTER 6

Combination of multiple Higgs boson channels

We are only as strong as we are united, as weak as we are divided.

— J. K. Rowling, *Harry Potter and the Goblet of Fire*, 2000

As introduced in Chapter 2, the SM is incomplete and not the ultimate description of matter and its interactions. New physics beyond the SM might manifest itself as a deviation from the SM expectation in a single Higgs boson decay channel. However, as we've seen in Chapter 5, the measurements in the $H \to \tau\tau$ decay channel are in agreement with the predictions, supporting the SM. Signs of new physics, if detectable in the Higgs sector, seem to be harder to observe and might require a more detailed analysis of multiple Higgs boson production and decay channels to probe more complex models than would be possible with a single decay mode.

The κ-framework, as introduced in Section 2.3.2, can be used to interpret the combined analysis of multiple Higgs boson channels within the ATLAS Collaboration. The framework introduces global coupling modifiers κ_i that scale the Higgs boson coupling factor for particle i. In order to achieve consistent results, the total Higgs boson decay width is adjusted to account for the changes obtained with $\kappa_i \neq 1$.

Physics not described by the SM could be detected by individual κ_i modifiers that deviate from unity or a pattern of deviations across the multiple coupling modifiers. Potential BSM contributions to loop-induced processes affect the measurement with effective coupling modifiers that modify indirect Higgs boson couplings. Additionally, the framework includes the two branching ratios B_{inv} and B_{undet} to account for BSM decays to invisible particles and BSM decays to which non of the included analyses are sensitive. The SM is reproduced for $\kappa_i = 1$ and $B_{\text{inv}} = B_{\text{undet}} = 0$.

The results presented in the κ-framework represent different models built with different assumptions. Each model treats a different set of coupling modifiers κ_i as independent parameters, determines their values, and thus probes different manifestations of new physics. The assumptions for each model are stated together with the measured result. Most models have some assumption on the total decay width of the Higgs bo-

son. The last model presented in Section 6.4.5 measures ratios of κ modifiers, which are independent of the total decay width. All models assume the Lorentz structure of the Higgs boson to be SM-like, resulting in the same kinematic distributions as expected by the SM. This limitation of the κ-framework means that the framework can only scale the kinematic distribution but is not sensitive to more generic changes in the shape of the distribution. A summary of various models of new physics and their effect on the κ modifiers can be found in Reference [176].

6.1. Effective and resolved couplings

Loop-induced processes can be treated in two different ways in the κ-framework: as an effective coupling or as a resolved coupling. For example, in the Higgs boson production mode $gg \to H$, the gluon does not couple directly to Higgs boson as gluons are massless. In the SM, the production is facilitated at leading-order via a fermion loop. Depending on the model under consideration, one can assign an effective coupling modifier κ_g that scales the effective gluon-to-Higgs-boson vertex factor. With this kind of setup, the κ parameters are sensitive to particles not described in the SM if they contribute to a loop diagram. The particles in the loop can appear as virtual particles whose rest mass could be too large to be produced directly in an interaction at the LHC.

On the other hand, assuming only SM particles in loop-induced processes, one can parametrize the effective coupling between gluons and the Higgs boson in terms of the fundamental coupling modifiers used in a model. For example, for the ggF Higgs boson production mode, the relevant contributions stem from the top and bottom quarks as well as the interference between different quark flavors. If an effective coupling is expressed in terms of their direct coupling modifiers, the coupling is said to be *resolved*. The resolved and effective coupling modifiers for $H \to \gamma\gamma, Z\gamma$ decays and $gg \to H$ production are illustrated in Figure 6.1.

The methodology to derive the functional form of the resolved coupling is presented in References [28, 177]. Table 6.1 summarizes the relation between effective and resolved coupling modifiers for loop-induced processes. The table includes an expression for the total Higgs decay width in terms of its fundamental couplings. The expression is used to achieve consistent results by scaling the total Higgs decay under the assumption of no invisible or undetected decays $B_{\text{inv}} = B_{\text{undet}} = 0$.

The terminology for some processes refers to more than one diagram with a different coupling structure. For example, the VBF production mode has contributions from Z boson and W boson fusion. These processes are expressed by their fundamental couplings. The relations are listed in Table 6.2.

6.2. Input channels

The interpretation in the κ-framework requires sensitivity to multiple production and decay modes. Most of the analyses used in this combination target one specific Higgs boson decay channel and multiple production modes. All input channels use a consistent set of MC generators, detector simulation, and event reconstruction algorithms.

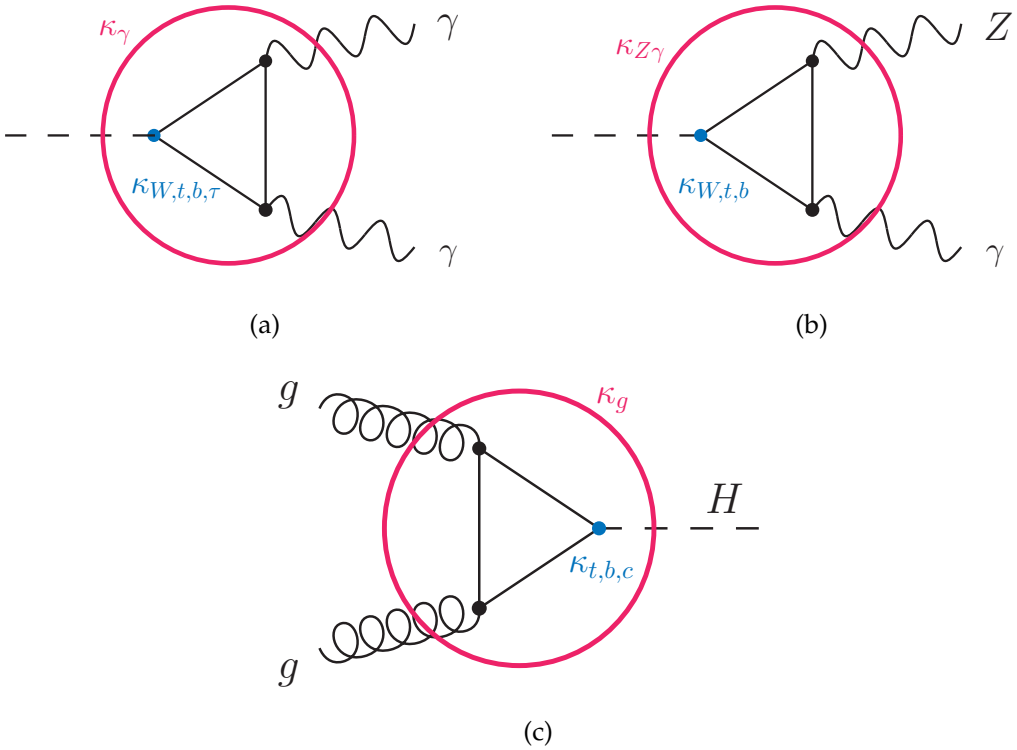


Figure 6.1.: Higgs boson decay and production Feynman diagrams with resolved and effective coupling modifiers κ_γ (a), $\kappa_{Z\gamma}$ (b), and κ_g (c).

Table 6.1.: Summary of effective and resolved coupling modifiers for loop-induced production (σ_i) or decay (Γ_f) processes [8, 28, 177]. The process $gg \to ZH$ is never treated as an effective coupling. The expression for the total Higgs decay width Γ_H does not consider the effects from invisible and undetected decays.

Process	Effective modifier	Resolved modifier
$\sigma_{gg\mathrm{F}}$	κ_g^2	$1.040\,\kappa_t^2 + 0.002\,\kappa_b^2 - 0.038\,\kappa_t\kappa_b - 0.005\,\kappa_t\kappa_c$
$\sigma_{gg\to ZH}$	—	$2.456\,\kappa_Z^2 + 0.456\,\kappa_t^2 - 1.903\,\kappa_Z\kappa_t - 0.011\,\kappa_Z\kappa_b + 0.003\,\kappa_t\kappa_b$
Γ_{gg}	κ_g^2	$1.111\,\kappa_t^2 + 0.012\,\kappa_b^2 - 0.123\,\kappa_t\kappa_b$
$\Gamma_{\gamma\gamma}$	κ_γ^2	$1.589\,\kappa_W^2 + 0.072\,\kappa_t^2$ $-0.674\,\kappa_W\kappa_t + 0.009\,\kappa_W\kappa_\tau + 0.008\,\kappa_W\kappa_b$ $-0.002\,\kappa_t\kappa_b - 0.002\,\kappa_t\kappa_\tau$
$\Gamma_{Z\gamma}$	$\kappa_{Z\gamma}^2$	$1.118\,\kappa_W^2 - 0.125\,\kappa_W\kappa_t + 0.004\,\kappa_t^2 + 0.003\,\kappa_W\kappa_b$
Γ_H	κ_H^2	$0.581\,\kappa_b^2 + 0.215\,\kappa_W^2 + 0.082\,\kappa_g^2$ $+0.063\,\kappa_\tau^2 + 0.026\,\kappa_Z^2 + 0.029\kappa_c^2$ $+0.0023\,\kappa_\gamma^2 + 0.0015\,\kappa_{Z\gamma}^2 + 0.0004\kappa_s^2$ $+0.00022\,\kappa_\mu^2$

Table 6.2.: Summary of coupling modifiers for composite processes that correspond to more than one fundamental process. [8, 28, 177].

Process	Coupling modifier
σ_{VBF}	$0.733\,\kappa_W^2 + 0.267\,\kappa_Z^2$
σ_{tHW}	$2.909\,\kappa_t^2 + 2.310\,\kappa_W^2 - 4.220\,\kappa_t\kappa_W$
σ_{tHq}	$2.633\,\kappa_t^2 + 3.578\,\kappa_W^2 - 5.211\,\kappa_t\kappa_W$

Table 6.3.: Summary of the input channels and analyses used in the combination.

Production	Decay	L_{int}/fb	Reference
ggF, VBF, VH, $t\bar{t}H$, tH	$H \to \gamma\gamma$	139	[179]
ggF, VBF, VH, $t\bar{t}H$	$H \to ZZ^* \to 4\ell$	139	[180]
ggF, VBF	$H \to WW^*$	139	[154]
ggF, VBF, VH, $t\bar{t}H$	$H \to \tau\tau$	139	[84]
VH	$H \to bb$	139	[181–183]
$t\bar{t}H$	$H \to bb$	139	[184]
VBF	$H \to bb$	126	[185]
ggF, VBF, VH, $t\bar{t}H$	$H \to \mu\mu$	139	[186]
ggF, VBF, VH, $t\bar{t}H$	$H \to Z\gamma$	139	[187]
VBF	$H \to$ inv	139	[188]
$t\bar{t}H$	$H \to ZZ^*, WW^*, \tau\tau$	36.1	[86]

The combination presented in this chapter is an improvement over the previous Higgs boson combination [178] within the ATLAS Collaboration. The improvement is mainly driven by improved and updated input measurements used in the combination. The input measurements and their changes with respect to the previous interpretation in the κ-framework are listed in the following. All input analyses are summarized in Table 6.3.

The combination includes the measurement of the Higgs boson decay to photons $H \to \gamma\gamma$ [179]. The analysis provides particle-level measurements in phase spaces defined by the STXS framework for the ggF, VBF, VH, $t\bar{t}H$, and tH production modes using the full Run 2 dataset corresponding to an integrated luminosity of 139 fb^{-1}.

The interpretation in the κ-framework profits from the inclusion of the measurement of $H \to ZZ^* \to 4\ell$ [180]. Cross section measurements are provided in STXS bins for the ggF, VBF, VH, $t\bar{t}H$ production mode. This analysis comprises the full Run 2 dataset with an integrated luminosity of 139 fb^{-1}.

Cross section measurements at particle-level for $H \to WW^*$ in phase space regions as defined by the STXS framework are included in the combination. The analysis [154] provides measurements for the ggF and VBF production modes using the full Run 2 dataset. Compared to the previous κ measurement, this analysis was updated to use the full dataset corresponding to an integrated luminosity of 139 fb^{-1}. Additionally, a new analysis category for the ggF production mode with two or more reconstructed jets was added.

The analysis presented in Chapter 5 measuring the decay of a Higgs boson to a pair of tau leptons in the ggF, VBF, VH, and $t\bar{t}H$ production modes on the full Run 2 dataset [84] is used in the combination. The measurements for VH are restricted to fully hadronic decays of the vector boson. The measurements for $t\bar{t}H$ are restricted to fully hadronic decays of tau leptons and fully hadronic $t\bar{t}$ decays. Both production modes, VH and $t\bar{t}H$, are included for the first time. Compared to the previous round, the granularity of

the STXS splitting is improved for the ggF production mode. Furthermore, the signal sensitivity is increased for all production modes, particularly in the VBF production mode.

The measurements for the decay of a Higgs boson to a pair of b quarks are performed in separate analyses depending on the production mode. Particle-level cross sections, as defined by the STXS framework, are measured for the VH production mode [181–183] and the $t\bar{t}H$ production mode [184], both on the full Run 2 dataset corresponding to an integrated luminosity of 139 fb^{-1}. Measurements for the VBF production mode [185] are included based on a dataset corresponding to an integrated luminosity of 126 fb^{-1}. The integrated luminosity is smaller than for the other analyses because the required trigger was not available in 2015. Compared to the previous combination, the measurement for $t\bar{t}H$ was updated to use the full Run 2 dataset and provide the measurement as a function of transverse Higgs boson momentum p_{T}^{H}. Additionally, a new measurement for VH production with boosted Higgs bosons was added.

As a measurement of a Higgs boson decaying to second-generation particles, the analysis of $H \to \mu\mu$ [186] is included in the combination. The measurement uses the full Run 2 dataset corresponding to an integrated luminosity of 139 fb^{-1}. The analysis provides measurements for the ggF, VBF, VH, and $t\bar{t}H$ production modes.

For the first time, the measurement of $H \to Z\gamma$ [187] is included in the combination. This is of particular interest for the loop model presented in Section 6.4.3 using effective coupling modifiers probing the presence of new heavy particles in loops. The measurement uses the full Run 2 dataset corresponding to an integrated luminosity of 139 fb^{-1} and provides measurements for the ggF, VBF, VH, and $t\bar{t}H$ production modes.

A dedicated analysis of Higgs boson decays to invisible particles $H \to$ inv [188] is only included in the κ models that are sensitive to it with B_{inv} as a free parameter. Specifically, the measurement is included in Section 6.4.3 and Section 6.4.4. The measurement targets the VBF production mode and uses the full Run 2 dataset corresponding to an integrated luminosity of 139 fb^{-1}.

Lastly, a measurement targeting the $t\bar{t}H$ multilepton production mode [86] is included. The measurement is complementary to the $t\bar{t}H$ measurement in the $H \to \tau\tau$ and $H \to ZZ^{*} \to 4\ell$ analyses. The measurement uses a partial Run 2 dataset corresponding to an integrated luminosity of 36.1 fb^{-1}.

6.3. Statistical combination

The statistical analysis of the combination is facilitated by a likelihood fit. The combined likelihood function is the product of the likelihood functions of all input analyses. The likelihood functions in the input analyses are formulated as a function of expected event yields in exclusive analysis categories c. The number of expected signal events per category is expressed as a function of the integrated luminosity L_{int}^{c} of the corresponding analyses, the product of cross section and branching ratio $(\sigma \cdot B)_{if}$ for the specific production i and decay mode f, and the product of acceptance and reconstruction efficiency $(A \cdot \epsilon)_{if}^{c}$ in the analysis category. The relation is summarized

in the master equation [8] for the combination

$$n^{c,\text{signal}} = L^{c}_{\text{int}} \sum_{i} \sum_{f} (\sigma \cdot B)_{if} \, (A \cdot \epsilon)^{c}_{if}, \tag{6.1}$$

where the sums run over all available production and decay mode combinations listed in Table 6.3. For the interpretation in the κ-framework, the products $(\sigma \cdot B)_{if}$ are factorized and expressed in terms of the SM prediction and the couplings modifiers according to Equation (2.25).

The systematic uncertainties in the statistical model for all input analyses are parametrized using nuisance parameters as described for the $H \to \tau\tau$ analysis in Section 5.7. Uncertainties that affect all analysis categories in the same way are shared between the input analyses, correctly representing the correlation between them. For example, the systematic uncertainty on the electron reconstruction efficiency is correlated across all analyses since they share the same detector and the same reconstruction algorithm. The uncertainties on the background models are usually treated as uncorrelated. The statistical uncertainty from the limited number of MC events is incorporated as defined in Equation (5.8).

Theory uncertainties on the signal prediction as described in Section 5.6.3 affect the signal yield via changes to the cross section and branching ratio $(\sigma \cdot B)_{if}$ but also via changes to the acceptance A^{c}_{if}. The interpretations of the combination in the κ-framework take both contributions into account. The combination considers correlations between the theoretical uncertainties on the branching ratio [28]. Modifications to the correlation model from BSM contributions are deemed negligible. The uncertainties are identically parametrized across the channels and correlated in the fit model.

Limits and uncertainties on parameters of interest are derived from a likelihood ratio in the asymptotic limit similar to the procedure outlined in Section 5.7. A more detailed description can be found in References [174, 175].

6.4. Results

This section presents the results of the combination of multiple Higgs boson production and decay channels and its interpretation in the κ-framework with global coupling modifiers. Each subsection describes the model, its constraints, and assumptions used for the fit, alongside the obtained results. The results were made public in Reference [8].

6.4.1. Fermion and weak boson couplings

The first model uses only two free parameters, the coupling of the Higgs boson to fermions $\kappa_F \equiv \kappa_\tau = \kappa_\mu = \kappa_t = \kappa_b$ and to the weak bosons $\kappa_V \equiv \kappa_W = \kappa_Z$. As stated in Section 2.3.2, the coupling to second-generation quarks is scaled with κ_F. The effect of coupling modifier for first-generation fermions is negligible. The total Higgs boson width is adjusted according to Table 6.1 to account for changes to the coupling modifiers expressed by κ_F and κ_V. The model is useful to probe differences in the couplings between the gauge bosons that receive their interaction in the spontaneous symmetry breaking and the fermions for which Yukawa couplings are added in an ad hoc

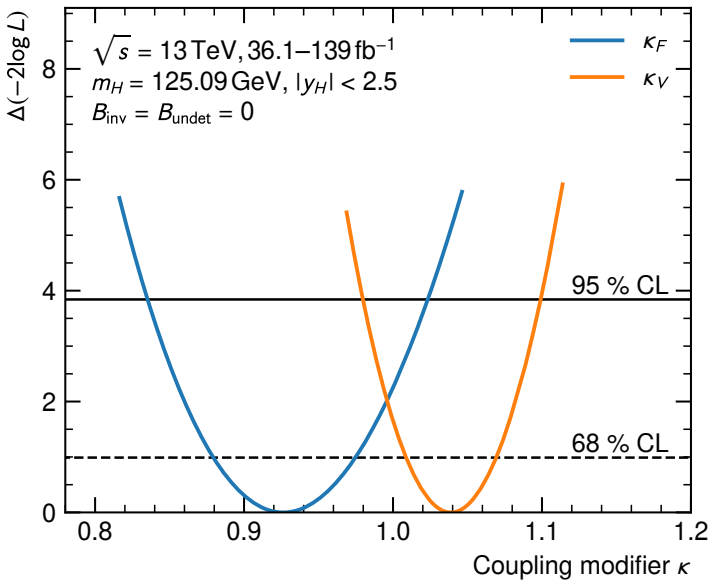


Figure 6.2.: Scan of the NLL for the coupling modifiers κ_F and κ_V independently to determine the best-fit values and their uncertainties.

fashion. Loop-induced couplings are resolved and expressed as defined in Table 6.1. Invisible and undetected decays are assumed to not exist, i.e., $B_{\text{inv}} = B_{\text{undet}} = 0$. From the interference terms in the resolved coupling modifiers, it is only possible to measure the relative sign between κ_F and κ_V. The hypothesis of a negative sign between the two coupling modifiers was rejected with high confidence in a previous measurement [82]. Therefore, without loss of generality, only positive values are considered in this analysis. The measured coupling modifiers and their uncertainties are

$$\kappa_V = 1.039^{+0.031}_{-0.030} \quad \text{and} \quad \kappa_F = 0.93 \pm 0.05. \tag{6.2}$$

Both values are compatible with the SM expectation of unity. Figure 6.2 shows the negative logarithmic likelihood (NLL) scans used to determine the above results. Figure 6.3 shows a representation of the fit result in the (κ_V, κ_F) plane and highlights the correlation between the uncertainties of the couplings. The correlation reduces the compatibility with the SM to $p = 2.8\,\%$. Reference [8] attributes this to the updates in the $H \to b\bar{b}$ and $H \to \tau\tau$ measurements. However, the largest discrepancies in decay-channel-specific cross sections are observed for $t\bar{t}H$+tH, $H \to b\bar{b}$, and VBF, $H \to \gamma\gamma$ with the observations at $0.35^{+0.34}_{-0.33}$ and $1.47^{+0.27}_{-0.24}$ of its expectation, respectively. No other measurement of cross section times branching ratio deviates from the expectation by significantly more than one standard deviation. The couplings in the $t\bar{t}H$+tH, $H \to b\bar{b}$ measurement is governed by κ_F, so a smaller observed cross section is consistent with a smaller κ_F value. Similarly, according to Table 6.1 and Table 6.2, the couplings in the VBF, $H \to \gamma\gamma$ measurement is dominated by κ_V. A larger observed cross section is consistent with a larger κ_V value.

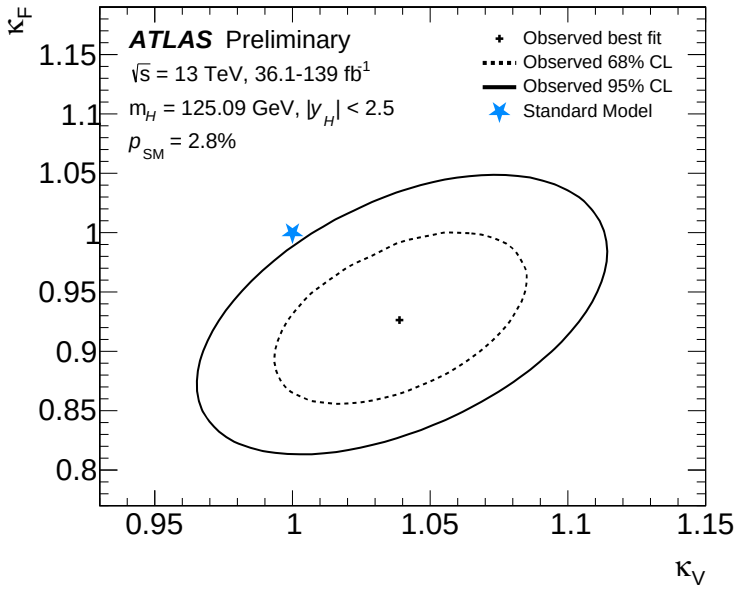


Figure 6.3.: Visualization of the 95 % and 68 % confidence intervals in the (κ_V, κ_F) plane derived from the likelihood fit assuming no contribution from invisible and undetected decays. The compatibility with the SM is $p = 2.8\,\%$. The linear correlation coefficient between the two κ modifiers is 43 % [8].

6.4.2. Generic parametrization with resolved loops

The generic parametrization in the κ-framework without effective loop couplings introduces the coupling modifiers for κ_t, κ_b, κ_τ, and κ_μ for fermions, and κ_Z and κ_W for bosons. The coupling to second-generation quarks scale as their third-generation partners. The contribution from first-generation fermions is negligible and is assumed to be SM-like. The total Higgs boson width is adjusted according to Table 6.1 to account for changes to the coupling modifiers. Loop-induced couplings use the resolved parametrization as listed in Table 6.1. All coupling modifiers are assumed to be positive. BSM contributions to invisible or undetected decays are assumed to not exist. The best-fit values and their uncertainties for the coupling parameters are

$$\begin{aligned} \kappa_t &= 0.92 \pm 0.06, \\ \kappa_b &= 0.88 \pm 0.11, \\ \kappa_\tau &= 0.92 \pm 0.07, \\ \kappa_\mu &= 1.07^{+0.25}_{-0.31}, \\ \kappa_Z &= 0.99 \pm 0.06, \quad \text{and} \\ \kappa_W &= 1.03 \pm 0.05. \end{aligned} \tag{6.3}$$

The NLL scan for the coupling modifier κ_τ is shown in Figure 6.4. Given the number of parameters and their uncertainties, the results are in agreement with the SM.

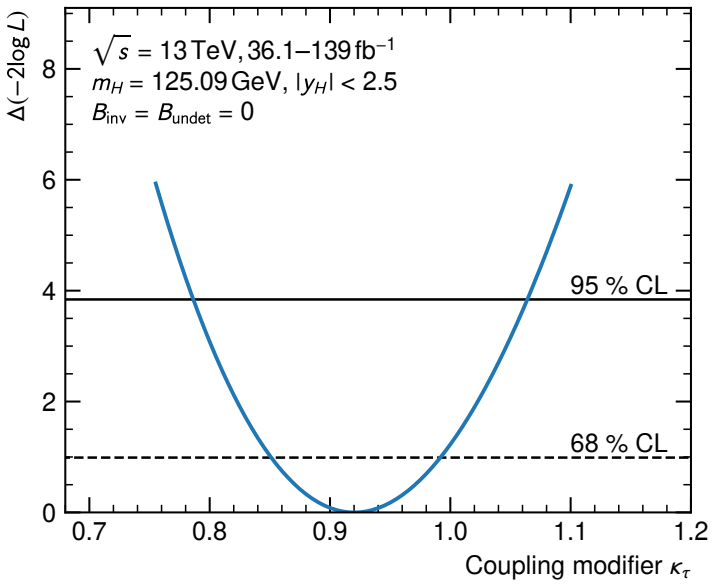


Figure 6.4.: Scan of the NLL for the coupling modifiers κ_τ to determine the best-fit value and its uncertainties.

The modifiers κ_τ, κ_t, and κ_b show an observed value smaller than the expectation by approximately one standard deviation. A similar trend was observed in the (κ_F, κ_V) model. The observation of smaller values in the generic model supports the arguments stated in Section 6.4.1. However, given the number of parameters, the compatibility with the SM is $p = 19\,\%$.

The previous combination [178] performed a measurement for the same model with a smaller set of input analyses and on a smaller dataset. The previous interpretation yielded a value of $\kappa_\tau = 1.06^{+0.15}_{-0.14}$. The current combination reduces the uncertainty on κ_τ by a factor of approximately 2. This is a clear indication of the gain achieved by using the updated input analysis for $H \to \tau\tau$ presented in Section 5. The size of the improvement seen here in the combination is similar to the size of the improvement achieved on the updated cross section measurement with the $H \to \tau\tau$ analysis.

The result of the measurement can be visualized by translating the measured coupling modifiers into reduced coupling-strength scale factors s_F for fermions and s_V for vector bosons. For fermions with mass m_F, the scale factor

$$s_F = \kappa_F \frac{Y_F}{\sqrt{2}} = \kappa_F \frac{m_F}{v} \tag{6.4}$$

is related to the Yukawa couplings Y_F defined in Equation (2.18), where v is the vacuum expectation value of approximately 246 GeV. Analogously, for gauge bosons with mass m_V, the scale factor is defined as

$$s_V = \sqrt{\kappa_V \frac{g_V}{2v}} = \sqrt{\kappa_V} \frac{m_V}{v}. \tag{6.5}$$

where g_V is the coupling strength to vector bosons, see Equation (2.16) and (2.15).

The two scale factors illustrate the relationship between coupling strength and mass as imposed by the SM. The Yukawa coupling of fermions to the Higgs boson is introduced ad hoc in the SM while the mass of the bosons is fundamentally built into the symmetry breaking. The fermionic coupling depends on the mass of particle. The bosonic coupling depends on the square of the coupling. The scale factors s_F for fermions and s_V for bosons are defined separately such that both are linearly dependent on the particle mass. Figure 6.5 shows the expected and observed linear relationship between coupling-strength scale factors s_F and s_V as a function of particle masses m_F and m_V. All measurements are in agreement with the SM. The observed linear relationship demonstrates that the Brout-Englert-Higgs mechanism is responsible for the masses of vector bosons and the fermions.

6.4.3. Effective loop couplings

The model considered in this section measures possible deviations from the SM in loop-induced Higgs couplings. This is an interesting measurement since the model is sensitive to potentially heavy BSM particles contributing as virtual particles to the coupling to the Higgs boson. The model uses effective coupling modifiers for κ_g, κ_γ, and $\kappa_{Z\gamma}$. The κ modifiers appear only as squared terms without any interference. Therefore information about relative or absolute signs between the modifiers cannot be inferred.

New physics is only assumed to occur in the loop diagrams. Therefore, the branching ratios B_{inv} and B_{undet} are set to zero. The only free parameters in the fit are the coupling modifiers for κ_g, κ_γ, and $\kappa_{Z\gamma}$. Alternatively, the contributions from BSM processes to the total width are measured by adding B_{inv} and B_{undet} as free parameters to the fit. In this case, the effects from B_{inv} and B_{undet} on the total Higgs boson decay width are taken into account according to Equation (2.27).

With the assumption $B_{\text{inv}} = B_{\text{undet}} = 0$, the measured effective coupling modifiers are

$$\begin{aligned} \kappa_g &= 1.00 \pm 0.05, \\ \kappa_\gamma &= 1.06 \pm 0.05, \quad \text{and} \\ \kappa_{Z\gamma} &= 1.43^{+0.31}_{-0.38}. \end{aligned} \tag{6.6}$$

The measured values are in agreement with the SM. The measured values and their uncertainties are visualized in Figure 6.6. The null hypothesis that the $Z\gamma$ coupling is non-existent could be rejected with $2.2\,\sigma$ significance ($1.1\,\sigma$ expected).

Adding B_{inv} and B_{undet} as free parameters in the model allows accounting for contributions to the total decay width of the Higgs boson. The analysis of VBF, $H \to$ inv [188], is added as an input to this model in this case. By construction, the branching ratios are non-negative. Limits on the branching ratios are expressed in terms of the 95 %

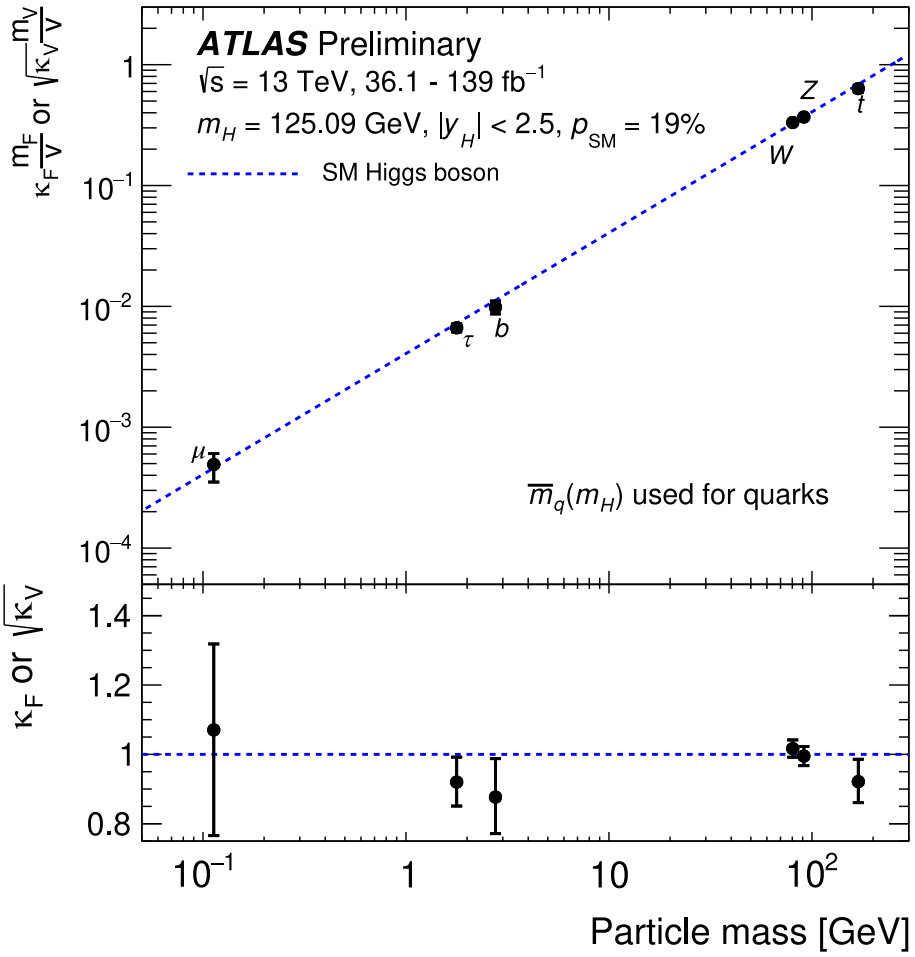


Figure 6.5.: Visualization of the linear expected and observed relationship between coupling-strength scale factors Y_F and Y_V as a function of the particle's mass. The measurement is performed for third-generation quarks, charged second and third-generation leptons, and the two heavy vector bosons. Quark masses are evaulated in the $\overline{MS}$ scheme at the mass of the Higgs boson m_H [8]. The same figure is shown in Figure 1.1 as an introduction.

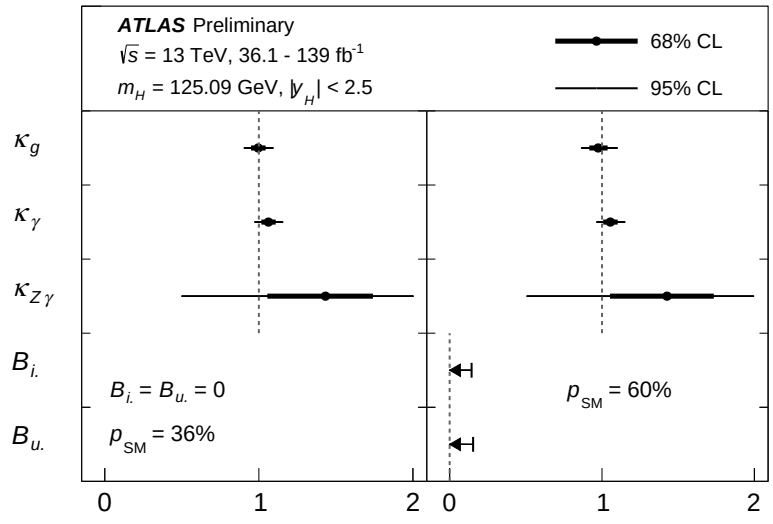


Figure 6.6.: Visualization of the 95 % and 68 % confidence intervals on κ_g, κ_γ, and $\kappa_{Z\gamma}$ under the assumption $B_{\text{inv}} = B_{\text{undet}} = 0$ (left), and with B_{inv} and B_{undet} as free parameters (right) [8].

confidence level (CL). The measured coupling modifiers and branching ratios are

$$\begin{aligned} \kappa_g &= 0.98 \pm 0.06, \\ \kappa_\gamma &= 1.06 \pm 0.05, \\ \kappa_{Z\gamma} &= 1.43^{+0.31}_{-0.37}, \\ B_{\text{inv}} &< 0.14 \text{ at } 95\,\% \text{ CL}, \quad \text{and} \\ B_{\text{undet}} &< 0.15 \text{ at } 95\,\% \text{ CL}. \end{aligned} \tag{6.7}$$

This result is in agreement with the SM and compatible with the measurement summarized in Equation (6.6).

6.4.4. Generic parametrization

The generic model with resolved couplings presented in Section 6.4.2 assumes no contribution from BSM processes as invisible and undetected decays and expresses the loop processes in terms of resolved couplings. These restrictions can be dropped with the generic parametrization. In addition to the parameters measured in the generic model with resolved couplings, this model uses the effective couplings κ_g, κ_γ, and $\kappa_{Z\gamma}$ for the respective loop-induced processes. Without loss of generality, the coupling modifier κ_Z is assumed to be positive. Through the interference terms in $gg \to HZ$, the measurement is sensitive to the relative sign between κ_Z and κ_t. Therefore κ_t is allowed to be negative. All other parameters are assumed to be positive. The coupling to second-generation quarks scale as their third-generation partners. The contribution from first-generation fermions is negligible and is assumed to be SM-like.

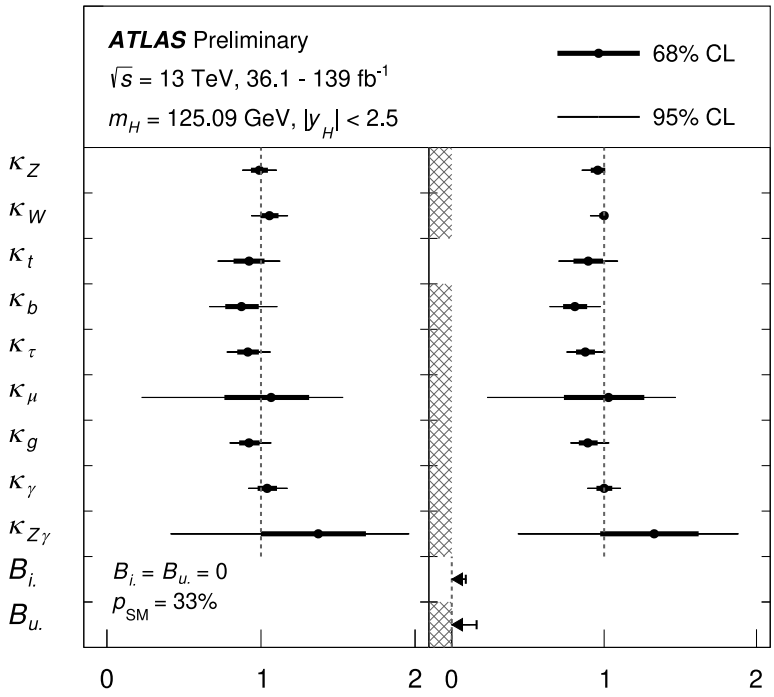


Figure 6.7.: Summary of the measured κ modifiers and branching ratios to invisible and undetected particles in the generic model with effective coupling modifiers [8].

The model is first fit to data assuming no BSM contribution to the decay to invisible and undetected particles, i.e., $B_{\text{inv}} = B_{\text{undet}} = 0$. Afterward, the model is evaluated while allowing BSM decays to invisible and undetected particles. For the latter approach, the VBF, $H \rightarrow$ inv analysis [188] is included to provide constraints on B_{inv}. Since none of the input analyses are sensitive to B_{undet} by definition, additional conditions are required to perform a measurement in this model. If B_{undet} is used as a free parameter, the conditions $\kappa_W \leq 1$ and $\kappa_Z \leq 1$ are added. Furthermore, B_{undet} must be positive. The effects from B_{inv} and B_{undet} on the total Higgs boson decay width are taken into account according to Equation (2.27).

The result for both alternatives regarding B_{inv} and B_{undet} are summarized in Figure 6.7 and Table 6.4. All observed κ modifiers, B_{inv}, and B_{undet} are in agreement with the SM. The measurement under the assumption of no BSM contributions to the decay excludes a negative sign for κ_t with a significance of 4.3σ (3.8σ expected).

6.4.5. Ratios of coupling modifiers

Previous models required assumptions to account for BSM contributions to the total decay width of the Higgs boson via B_{inv} and B_{undet}. A more general measurement with fewer assumptions can be obtained by measuring ratios of coupling modifiers since the total width cancels in such an expression. All ratios are assumed to be positive in the fit. The definition of the ratios and their measured values are summarized in Table 6.5.

Table 6.4.: Summary of measured κ modifers and braching ratios to invisible and undetected particles in the generic model with effective coupling modifers. The model is evaluated under the assumption of $B_{\text{inv}} = B_{\text{undet}} = 0$ and with $B_{\text{undet}} \geq 0,\ \kappa_{W,Z} \leq 1$ [8].

Parameter	$B_{\text{inv}} = B_{\text{undet}} = 0$	$B_{\text{undet}} \geq 0$ and $\kappa_{W,Z} \leq 1$
κ_Z	0.99 ± 0.06	$0.96^{+0.04}_{-0.05}$
κ_W	1.06 ± 0.06	$1.00^{+0.00}_{-0.03}$
κ_t	0.92 ± 0.10	0.90 ± 0.10
κ_b	0.87 ± 0.11	0.81 ± 0.08
κ_τ	0.92 ± 0.07	0.88 ± 0.06
κ_μ	$1.07^{+0.25}_{-0.30}$	$1.03^{+0.23}_{-0.29}$
κ_g	$0.92^{+0.07}_{-0.06}$	$0.89^{+0.07}_{-0.06}$
κ_γ	1.04 ± 0.06	1.00 ± 0.05
$\kappa_{Z\gamma}$	$1.37^{+0.31}_{-0.37}$	$1.33^{+0.29}_{-0.35}$
B_{inv}		< 0.09 at 95 % CL
B_{undet}		< 0.16 at 95 % CL

The coupling to second-generation quarks scale as their third-generation partners. The contribution from first-generation fermions is negligible and is assumed to be SM-like.

The parameter κ_{gZ} measures the $gg \to H \to ZZ^*$ process and serves as the overall normalization. Except for $\lambda_{\mu\tau}$, all other ratios are either normalized to κ_Z or κ_g. The ratios are sensitive to a variety of different extensions of the SM.

After spontaneous symmetry breaking, the SM exhibits an approximate global $SU(2)$ symmetry referred to as the custodial symmetry of the SM [11, 189–191]. The symmetry and experimental measurements at LEP and Tevatron [192] impose strict limits on

$$\rho = \frac{m_W^2}{m_Z^2 \cos\theta_w} = 1. \tag{6.8}$$

Radiative correction terms to ρ break the custodial symmetry. The parameter λ_{WZ} is of particular interest due to its sensitivity to the relative couplings of W and Z bosons, making λ_{WZ} a direct probe of the custodial symmetry.

Furthermore, in contrast to $H \to ZZ^*$ decays, the loop processes with κ_γ and $\kappa_{Z\gamma}$ are sensitive to new electrically charged particles coupling to the Higgs boson. The ratios $\lambda_{\gamma Z}$ and $\lambda_{Z\gamma Z}$ are therefore sensitive to these contributions without assumptions on the total Higgs decay width. New colored particles directly affect ggF loop processes while keeping the structure of $t\bar{t}H$ unchanged. The ratio λ_{tg} is therefore sensitive to colored BSM particles. Finally, the parameter $\lambda_{\mu\tau}$ probes differences between second and third-generation lepton couplings to the Higgs boson.

All observed ratios are in agreement with the SM expectation. The results are summarized in Figure 6.8.

Table 6.5.: Defintion of ratios of coupling modifers and κ_{gZ} as well as their measured values [8].

Parameter	Definition	Measured value
κ_{gZ}	$\kappa_g\kappa_Z/\kappa_H$	0.98 ± 0.05
λ_{tg}	κ_t/κ_g	1.00 ± 0.11
λ_{Zg}	κ_Z/κ_g	1.07 ± 0.09
λ_{WZ}	κ_W/κ_Z	1.07 ± 0.06
λ_{bZ}	κ_b/κ_Z	$0.89^{+0.10}_{-0.09}$
$\lambda_{\gamma Z}$	κ_γ/κ_Z	1.05 ± 0.06
$\lambda_{Z\gamma Z}$	$\kappa_{Z\gamma}/\kappa_Z$	$1.39^{+0.31}_{-0.37}$
$\lambda_{\tau Z}$	κ_τ/κ_Z	0.93 ± 0.07
$\lambda_{\mu\tau}$	κ_μ/κ_τ	$1.16^{+0.28}_{-0.33}$

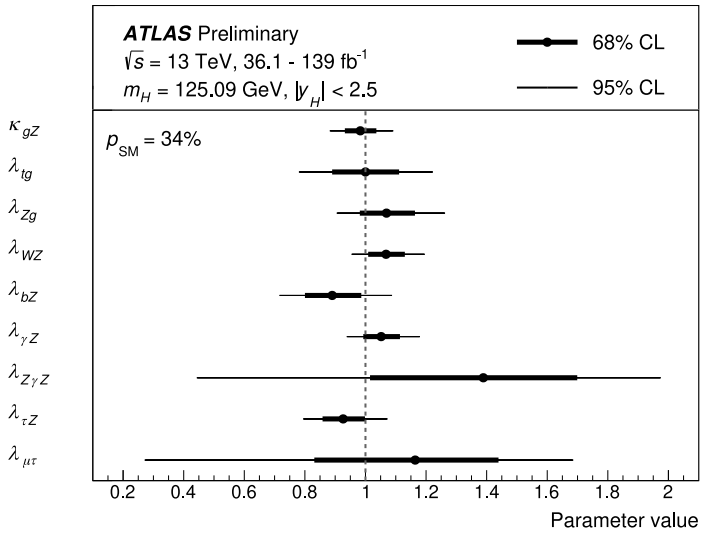


Figure 6.8.: Summary of the measured ratios of modifiers and κ_{gZ} [8].

CHAPTER 7

Conclusion

Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve.

— Karl Popper, *Objective Knowledge*, 1972

In this thesis, studies of Higgs boson production and decays in tau tau final states have been studied based on the complete dataset collected by the ATLAS Collaboration at the Large Hadron Collider (LHC). The results are a significant improvement over previous measurements in Run 1 and Run 2. Compared to previous studies, detailed cross section measurements have been achieved for all major Higgs boson production modes at the LHC in fiducial regions of phase space as defined by the Simplified Template Cross Section (STXS) framework. Additionally, the results were combined with other Higgs boson analyses within the ATLAS Collaboration to interpret the measurement as global coupling parameters in the κ-framework. The obtained limits on the coupling modifiers are a significant improvement over previous combinations. The work encompasses as well investigations of the noise behavior of the SCT detector.

The study described in Chapter 4 assessed different thresholds for the detection of noisy SCT strips. Thermal or electrical noise causes strips to exceed the discrimination threshold leading to noise hits. Single event upsets can change the discrimination threshold, potentially causing a strip to become noisy continuously. The effective suppression of noisy strips improves the tracking performance needed for high precision physics measurements. With increasing radiation damage to the SCT modules and higher particle multiplicities in Run 2 and Run 3, the detection of noisy strips may require adjustments.

The first optimization studied the impact of different noisy strip identification thresholds. A higher threshold implies that fewer strips are marked as being noisy, fewer artificial holes are introduced, but more noise hits need to be considered during track fitting. The study showed that the performance deteriorates with lower thresholds. A

small improvement was observed for higher thresholds. However, other effects have a larger impact on the figures of merit.

Marking strips as being noisy introduces artificial holes if a masked strip is hit by a charged particle leading to a track in the SCT. To prevent the degradation of tracking performance, an alternative treatment of noisy strips was considered. Instead of removing the hit information from noisy strips, the strips are marked as dead detector components, such that the track fit is not penalized by missing hits. The reevaluation of measured data with this alternative procedure showed no improvement over the baseline treatment.

The study did not find parameter settings that significantly improve both tracking figures of merit over the baseline approach used since Run 1. As a result, the settings were not changed for the remainder of Run 2. The number of pile-up interactions and the amount of radiation damage to the SCT are expected to increase for Run 3. The current configuration might lead to a large number of strips marked as noisy, significantly impacting the tracking performance. The study described in Chapter 4 shows that increasing the noise threshold does not negatively impact the tracking performance, thus suggesting a higher threshold as a viable option for Run 3.

The measurement of the Higgs boson production cross section times branching ratio in the $H \to \tau\tau$ channel described in Chapter 5 is a significant improvement over the Run 1 analysis [80] and the first Run 2 analysis [83] performed by the ATLAS Collaboration. The more precise cross section measurement is driven by a combination of a larger dataset, smaller statistical uncertainties in the background prediction, more refined analysis techniques, and more advanced object reconstruction techniques with smaller systematic uncertainties.

Tau leptons can decay fully leptonically $\tau \to \ell\bar{\nu}_\ell\nu_\tau$ with $\ell = e, \mu$ or hadronically $\tau \to \text{hadrons}\,\nu_\tau$. The analysis is split into three analysis channels depending on the number of leptonically decaying tau leptons: $\tau_e\tau_\mu$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_{\text{had}}\tau_{\text{had}}$. The analysis is harmonized between the channels while accounting for different background compositions in each channel. The event selection defines signal region categories that target the measurement in various Higgs boson production modes: ggF, VBF, VH, and $t\bar{t}H$. The $t\bar{t}H$ categories exist only in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

The dominant background in the analysis is $Z \to \tau\tau$. The dependency on MC generators for the total normalization of $Z \to \tau\tau$ is avoided with the kinematic rescaling technique. Events from $Z \to \ell\ell$ which can be selected with high purity in control regions are used to estimate the normalization of $Z \to \tau\tau$ in the signal region categories. The mismatch between ℓ and τ_{vis} (the visible decay production of a tau) kinematics is compensated by rescaling the momentum p^ℓ of the light lepton ℓ.

Another important background constitutes jet production via QCD processes, where jets are misidentified as tau leptons or electrons. For all analysis channels, a dedicated data-driven method is implemented to estimate the kinematic distributions of events with misidentified objects.

The signal sensitivity is improved by using machine-learning techniques such as BDTs or NNs. The analysis published in Reference [84] uses BDTs in the VBF, VH, and $t\bar{t}H$ categories to define a signal and a background enriched subcategory. Section 5.5.2 describes the training and optimization of a multiclass NN in the VBF categories to

define three subcategories that are enriched in VBF signal, ggF signal, and $Z \to \tau\tau$. The motivation for the multiclass classification is that ggF signal events with a VBF topology are subject to large systematic uncertainties. Likewise, the contamination from the $Z \to \tau\tau$ background reduces the measurement sensitivity. Therefore, separating VBF from ggF and $Z \to \tau\tau$ reduces the impact of their respective systematic uncertainties and thus improves the precision of the measurement. The NN approach achieved an 8 % reduction of uncertainties compared to the BDT approach on an Asimov dataset considering all relevant systematic uncertainties.

The complexity of the analysis is reflected by the complexity of the fit model. Especially with the new kinematic rescaling approach, it is necessary to validate the fit model and its assumptions on the correlation between nuisance parameters and uncertainties in partially unblinded fits. The so-called random-μ fit procedure is described in Section 5.8.2 and revealed insight that could not be obtained from a fit on an Asimov dataset.

The random-μ fit in combination with the kinematic rescaling showed that the tau reconstruction efficiency deviates from the initial assumption derived on MC by approximately 5 %. This observation led to changes to the fit model to prevent the compensation of this effect by other systematic uncertainties, potentially biasing the output.

The product of total cross section and $H \to \tau\tau$ branching ratio within $|y_H| < 2.5$ was extracted from a likelihood fit of the background and signal model to data and is measured to be

$$\sigma_{H\to\tau\tau} = 2.94 \pm 0.21(\text{stat}) \,^{+0.37}_{-0.32}(\text{syst})\,\text{pb} \tag{7.1}$$

which is $0.93^{+0.13}_{-0.12}$ of its SM expectation. Additionally, the product of cross section and branching ratio was measured for each of the four main production modes and in nine bins defined by the Simplified Template Cross Section STXS framework. All measurements are in agreement with the SM expectation.

As described in Chapter 6, the results from the $H \to \tau\tau$ measurements were combined with similar measurements from other Higgs boson production modes and decay channels. The combined model was interpreted in the κ-framework. In this framework, the coupling parameters of each process are scaled by coupling modifiers to have a consistent parametrization of deviations from the SM with a straightforward interpretation. Various models for resolved and effective coupling modifiers were fit to the data. Contributions from BSM decays to undetected or invisible processes are either assumed not to exist or taken into account. Additionally, ratios of coupling modifiers are measured, which are independent of assumptions on the total Higgs boson decay width.

The interpretation in the generic model with resolved loops provides a test of the SM and its predicted relation between Higgs boson couplings and particle mass. The expected proportionality between coupling strength and particle mass could be verified experimentally, supporting the Brout-Englert-Higgs mechanism and electroweak symmetry breaking. The measurement of the coupling modifier κ_τ that scales the interaction vertex of the Higgs boson and τ leptons yielded

$$\kappa_\tau = 0.92 \pm 0.07 \tag{7.2}$$

which is, at the time of writing, the most precise measurement of this parameter.

Run 3 is planned to start in 2022. The evaluation of the tracking performance will show if increasing the noisy strip threshold is required in order to ensure the proper operation of the SCT and a high tracking performance of the combined ID. With the expected larger dataset of Run 3, the even larger Run 2 and Run 3 combined dataset, more precise detector calibrations, and predictions with smaller theoretical uncertainties, it is expected to measure the Higs boson production cross section in the $H \to \tau\tau$ channel even more precisely, in more granular bins of the STXS framework, or differentially in kinematic variables. Similar improvements for other Higgs boson analyses are expected to reduce the uncertainties on a combined measurement in the κ-framework, in an STXS combination, and in interpretations in terms of effective field theories. Reducing the uncertainties narrows down the space where new physics could hide.

To paraphrase David Griffiths [7], now is the time to study the effects in the Higgs sector empirically. Overall, all measurements are in agreement with the SM. No significant deviations from the SM prediction are observed supporting the current version of the SM with the Brout-Englert-Higgs mechanism and Yukawa couplings for fermions.

APPENDIX A

Noise optimization

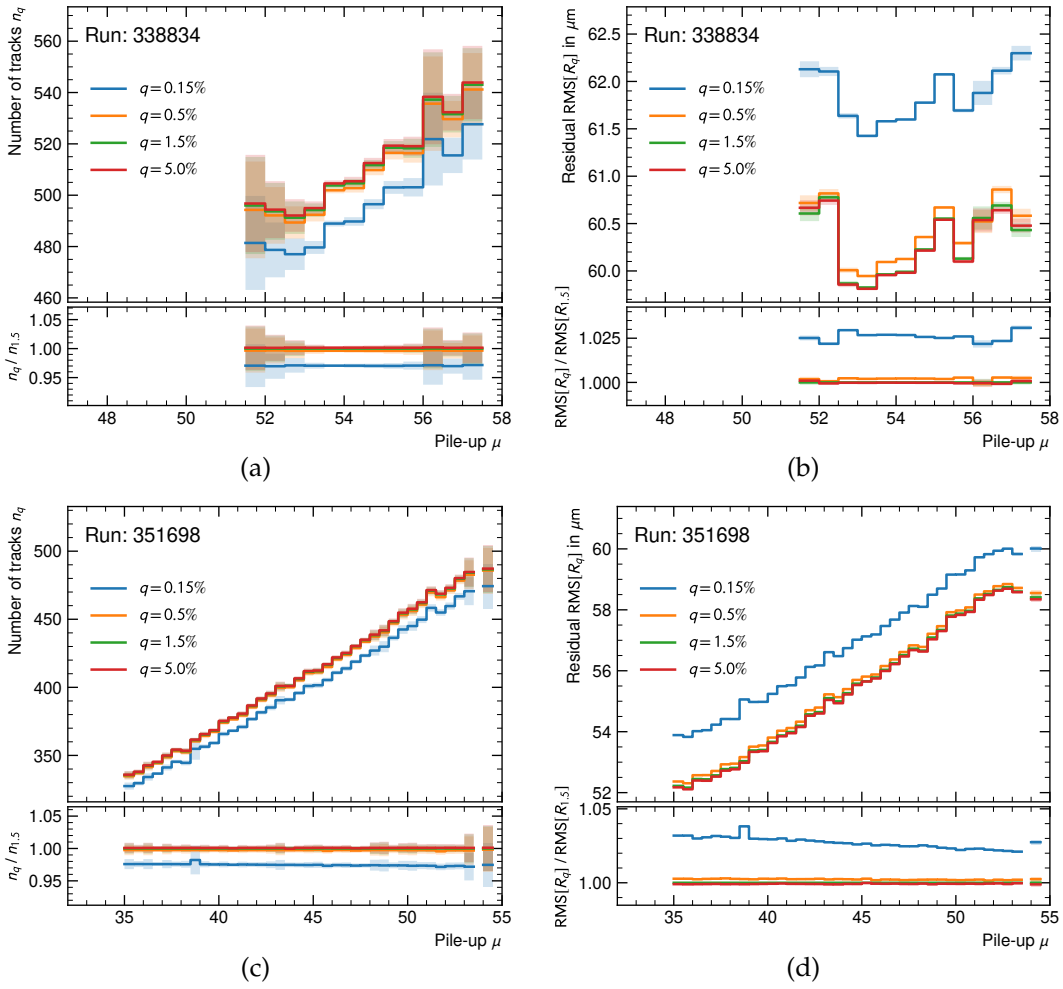


Figure A.1.: Number of reconstructed tight tracks (left) and residual in the barrel region (right) for run 338834 (top) and run 351698 (bottom) for different noise thresholds as a function of number of pile-up interactions. The shaded bands indicate the statistical uncertainty.

APPENDIX B

Cross section measurement

B.1. Fakes

Table B.1.: Number of expected events with misidentifed objects for each channel and category. The last column states the fraction of fake events over the total number of expected events.

Channel	Category	Fake events	Fraction of total / %
$\tau_e\tau_\mu$	VBF_0	184	13.5
	VBF_1	14	13.9
	VH_0	453	13.2
	VH_1	83	26.6
	boost_0_1J	263	8.7
	boost_0_ge2J	194	9.4
	boost_1_1J	291	6.4
	boost_1_ge2J	447	9.3
	boost_2	187	7.6
	boost_3	85	9.7
$\tau_{\text{lep}}\tau_{\text{had}}$	VBF_0	559	15.5
	VBF_1	27	10.3
	VH_0	1496	16.3
	VH_1	90	12.3
	boost_0_1J	1017	14.8
	boost_0_ge2J	756	16.5
	boost_1_1J	1185	9.3
	boost_1_ge2J	1286	10.7
	boost_2	470	5.6
	boost_3	121	4.2
$\tau_{\text{had}}\tau_{\text{had}}$	VBF_0	961	27.4
	VBF_1	38	17.1
	VH_0	1529	21.7
	VH_1	103	13.8
	boost_0_1J	3208	35.0
	boost_0_ge2J	1625	35.4
	boost_1_1J	2135	14.4
	boost_1_ge2J	1888	14.4
	boost_2	264	3.1
	boost_3	70	5.4
	tth_0	179	24.2
	tth_1	6	11.3

B.2. Input variables

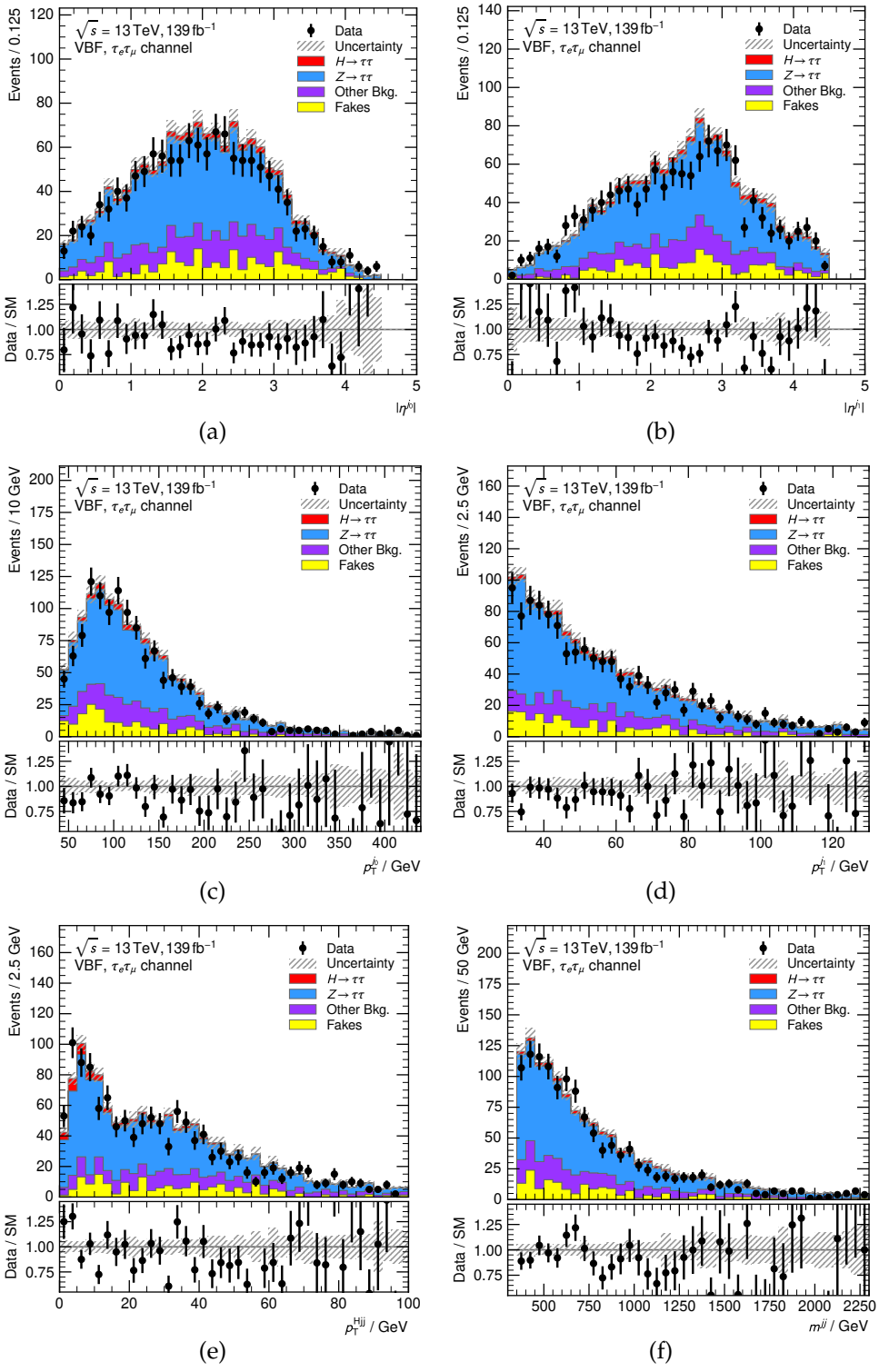


Figure B.1.: Selected input variables of the neural network in the VBF inclusive signal region for the $\tau_e\tau_\mu$ channel (a)–(e), and the invariant mass of the two jet system m^{jj} (f). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

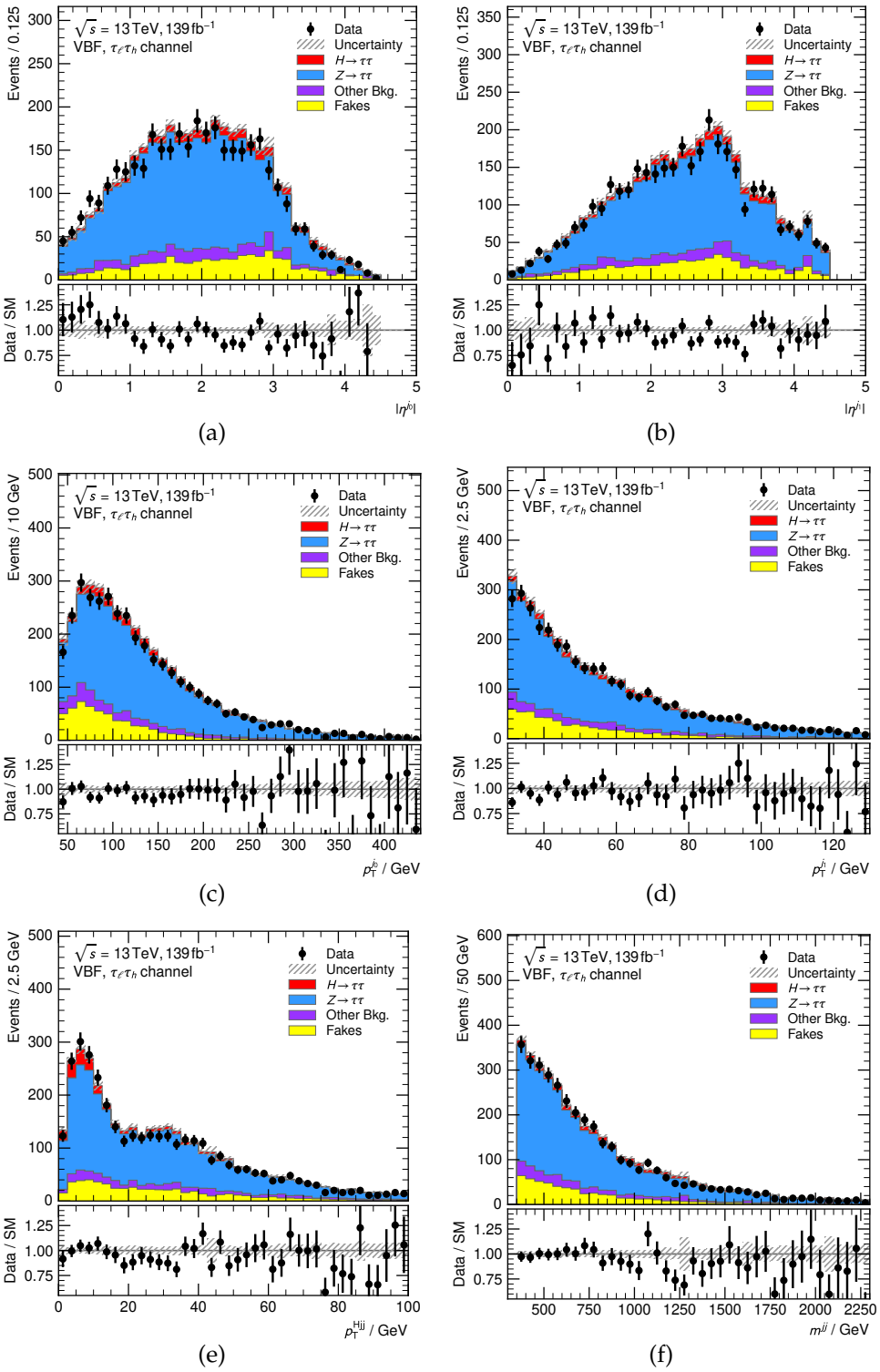


Figure B.2.: Selected input variables of the neural network in the VBF inclusive signal region for the $\tau_{\text{lep}}\tau_{\text{had}}$ channel (a)–(e), and the invariant mass of the two jet system m^{jj} (f). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

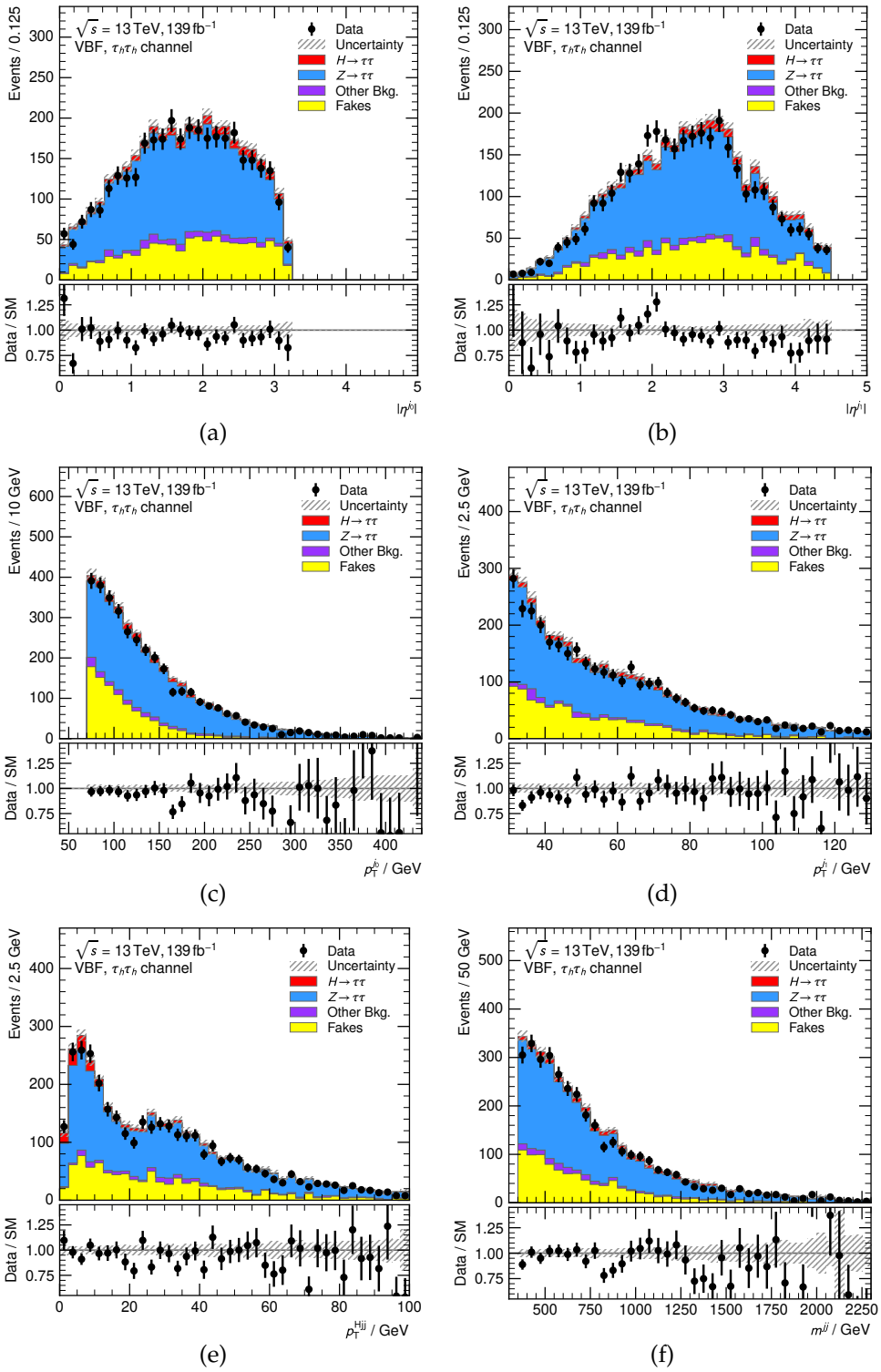


Figure B.3.: Selected input variables of the neural network in the VBF inclusive signal region for the $\tau_{\text{had}}\tau_{\text{had}}$ channel (a)–(e), and the invariant mass of the two jet system m^{jj} (f). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

B.3. Neural network training

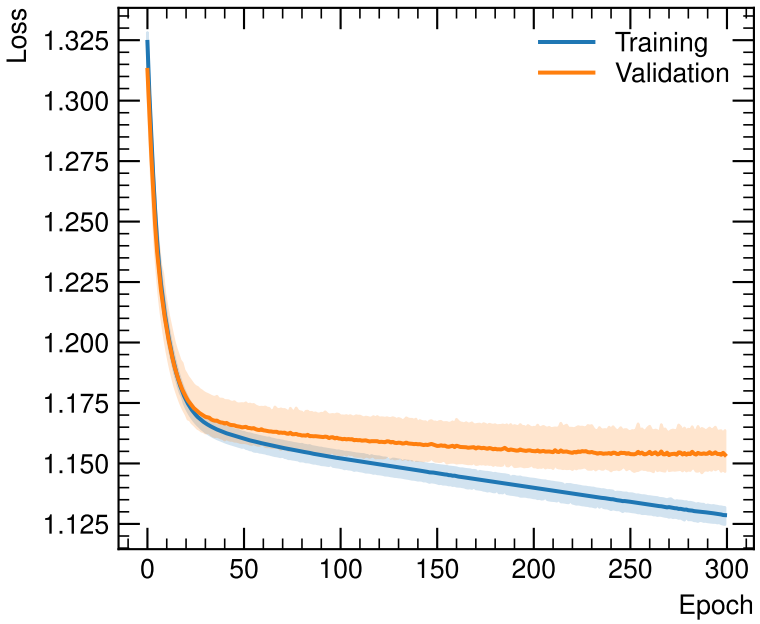


Figure B.4.: The loss function on the training and validation set as a function of the training epochs. The uncertainty band represents the spread between the $k = 4$ cross validation folds.

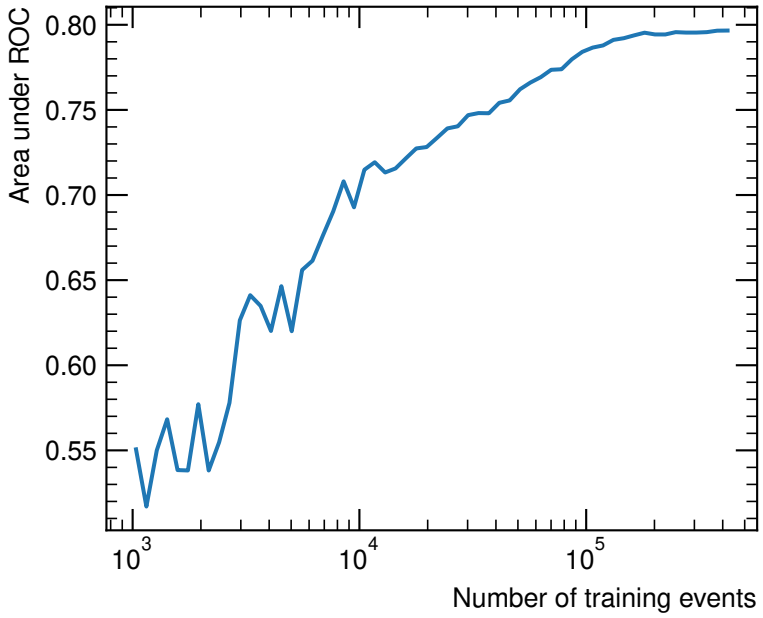


Figure B.5.: Area under the ROC curve obtained from networks trained with the same hyperparameter configuration but artificially reduced training set sizes.

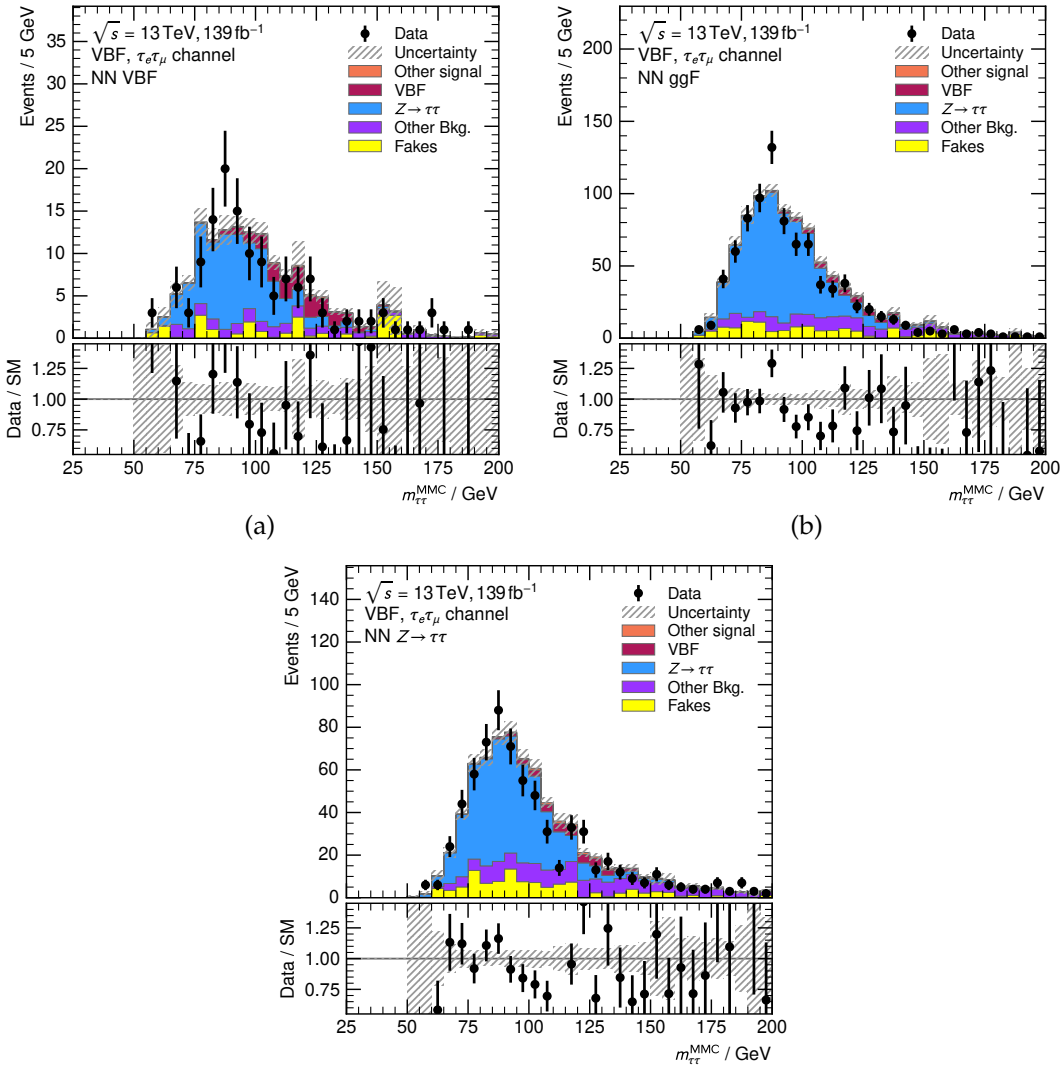


Figure B.6.: Exclusive VBF signal region categories as defined by the NN output for the $\tau_e\tau_\mu$ channel. By definition, the categories are designed to increase the fraction of events from VBF (a), ggF (b), and $Z \to \tau\tau$ processes (c). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

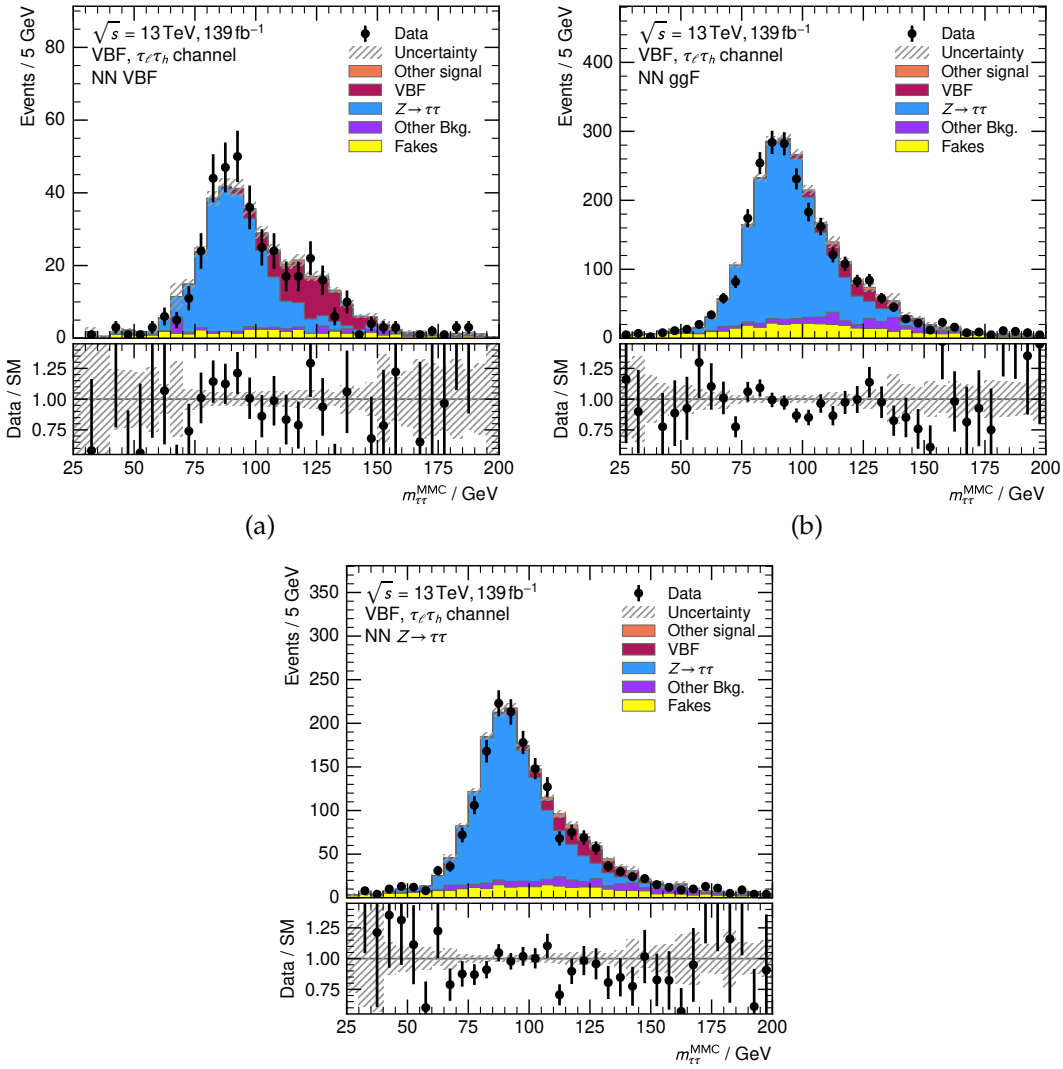


Figure B.7.: Exclusive VBF signal region categories as defined by the NN output for the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel. By definition, the categories are designed to increase the fraction of events from VBF (a), ggF (b), and $Z \to \tau\tau$ processes (c). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

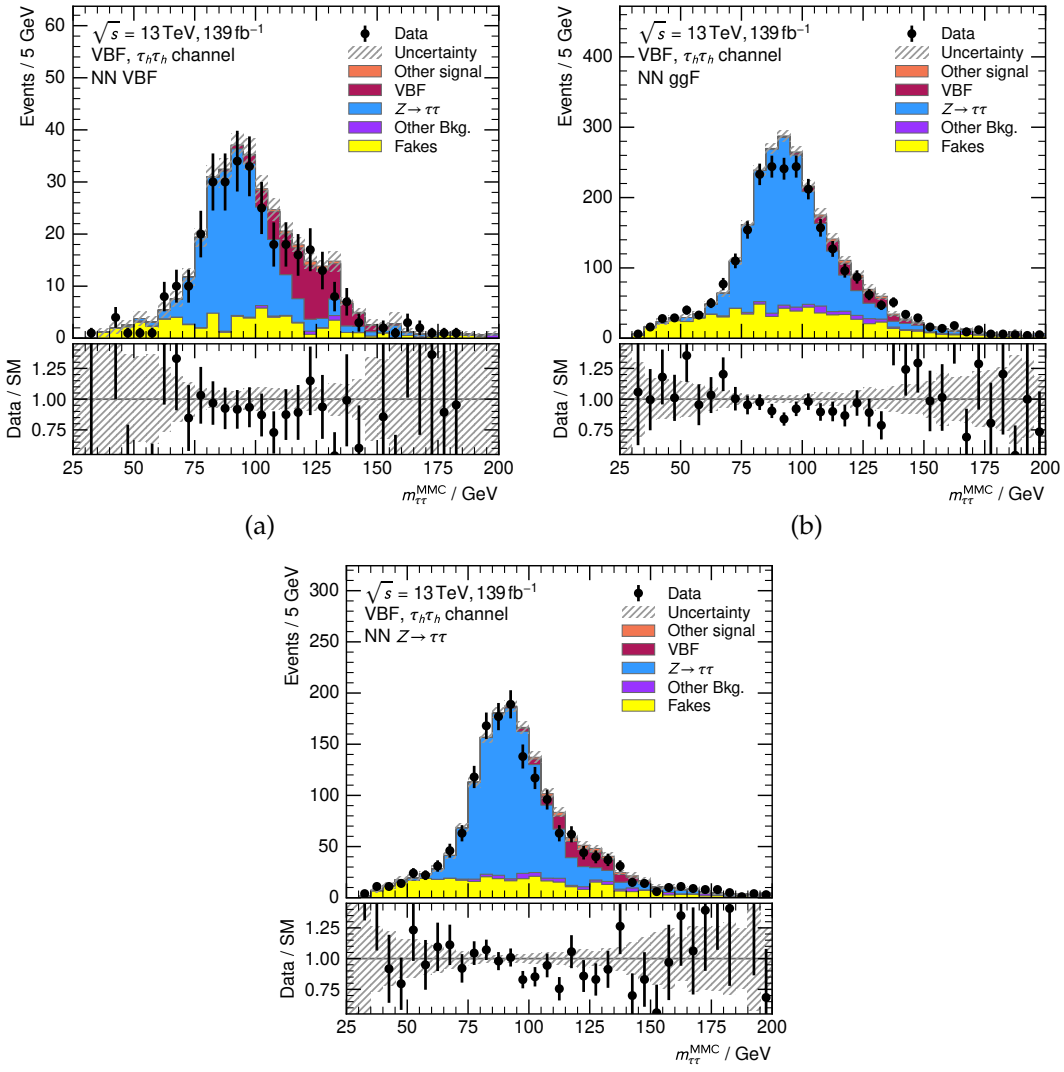


Figure B.8.: Exclusive VBF signal region categories as defined by the NN output for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. By definition, the categories are designed to increase the fraction of events from VBF (a), ggF (b), and $Z \to \tau\tau$ processes (c). The hatched band and the error bars indicate statistical uncertainties of the background expectation and measured data, respectively.

B.4. Fit results

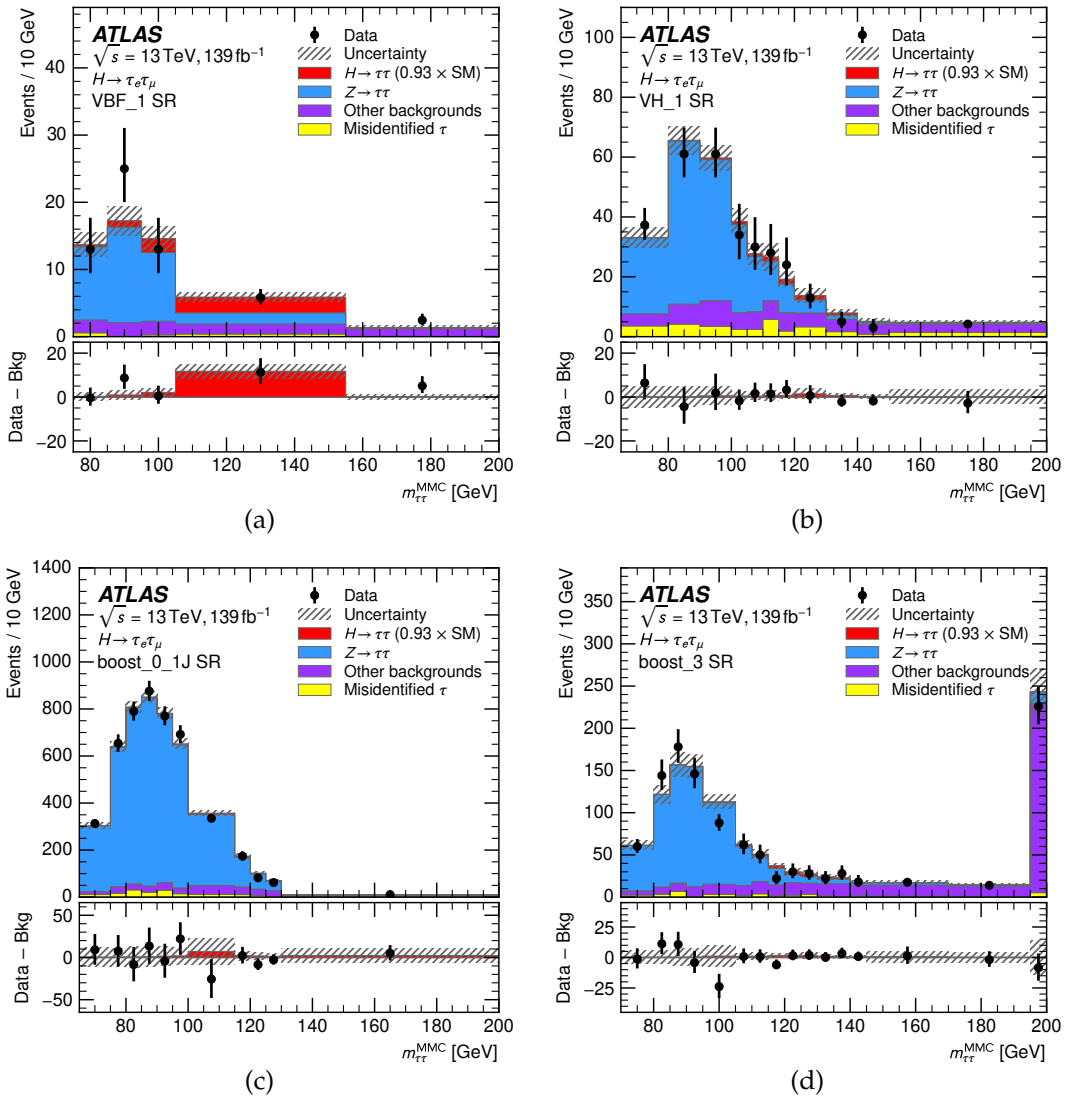


Figure B.9.: Mass $m_{\tau\tau}^{\mathrm{MMC}}$ distributions in selected signal regions of the $\tau_e\tau_\mu$ channel with the all nuisance parameters set to the value that maximizes the likelihood function [84]. The hatched band and the error bars indicate total uncertaintainty of the background expectation and statistical unceratinty measured data, respectively.

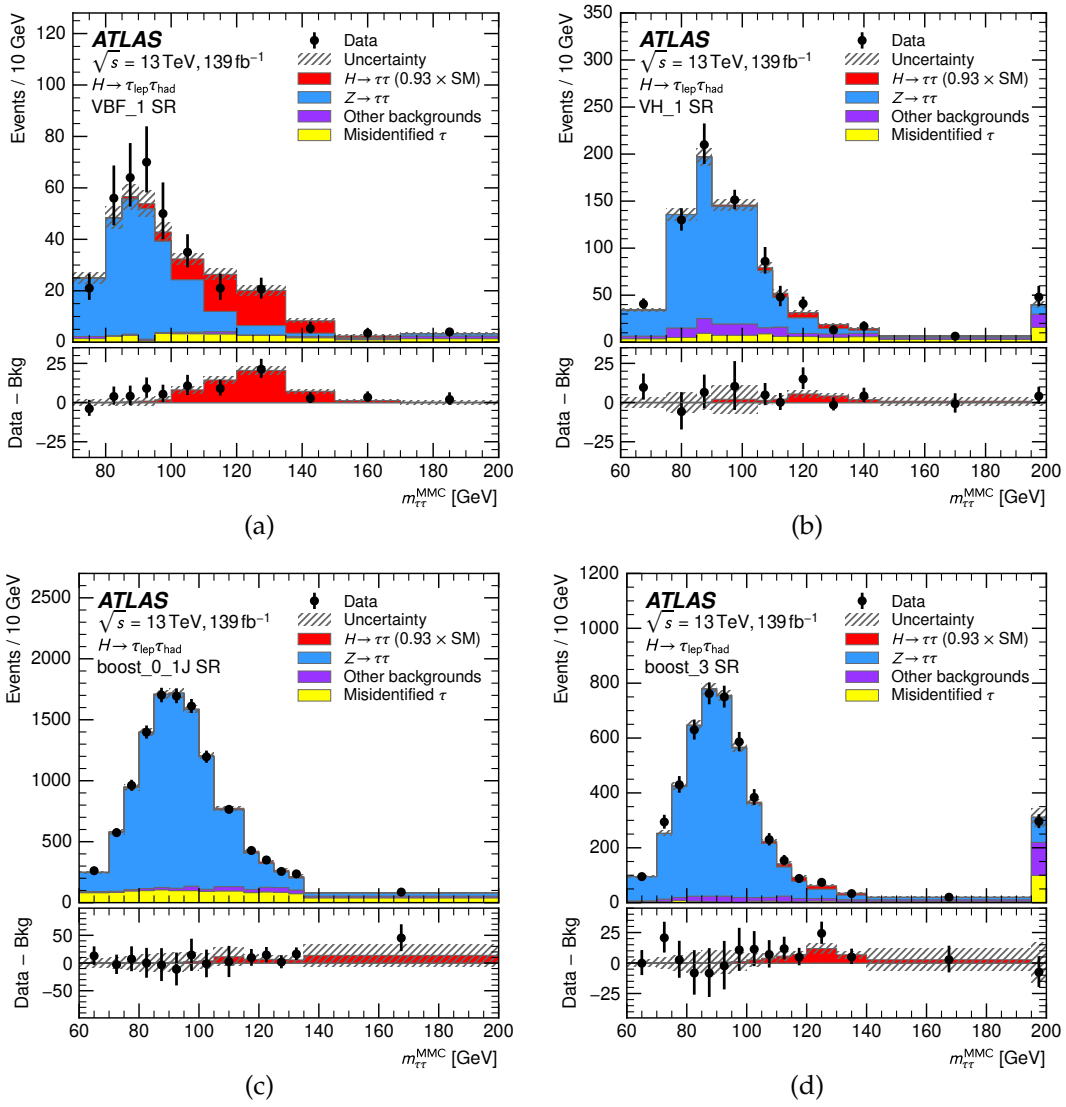


Figure B.10.: Mass $m_{\tau\tau}^{\text{MMC}}$ distributions in selected signal regions of the $\tau_{\text{lep}}\tau_{\text{had}}$ channel with the all nuisance parameters set to the value that maximizes the likelihood function [84]. The hatched band and the error bars indicate total uncertaintainty of the background expectation and statistical uncertainty measured data, respectively.

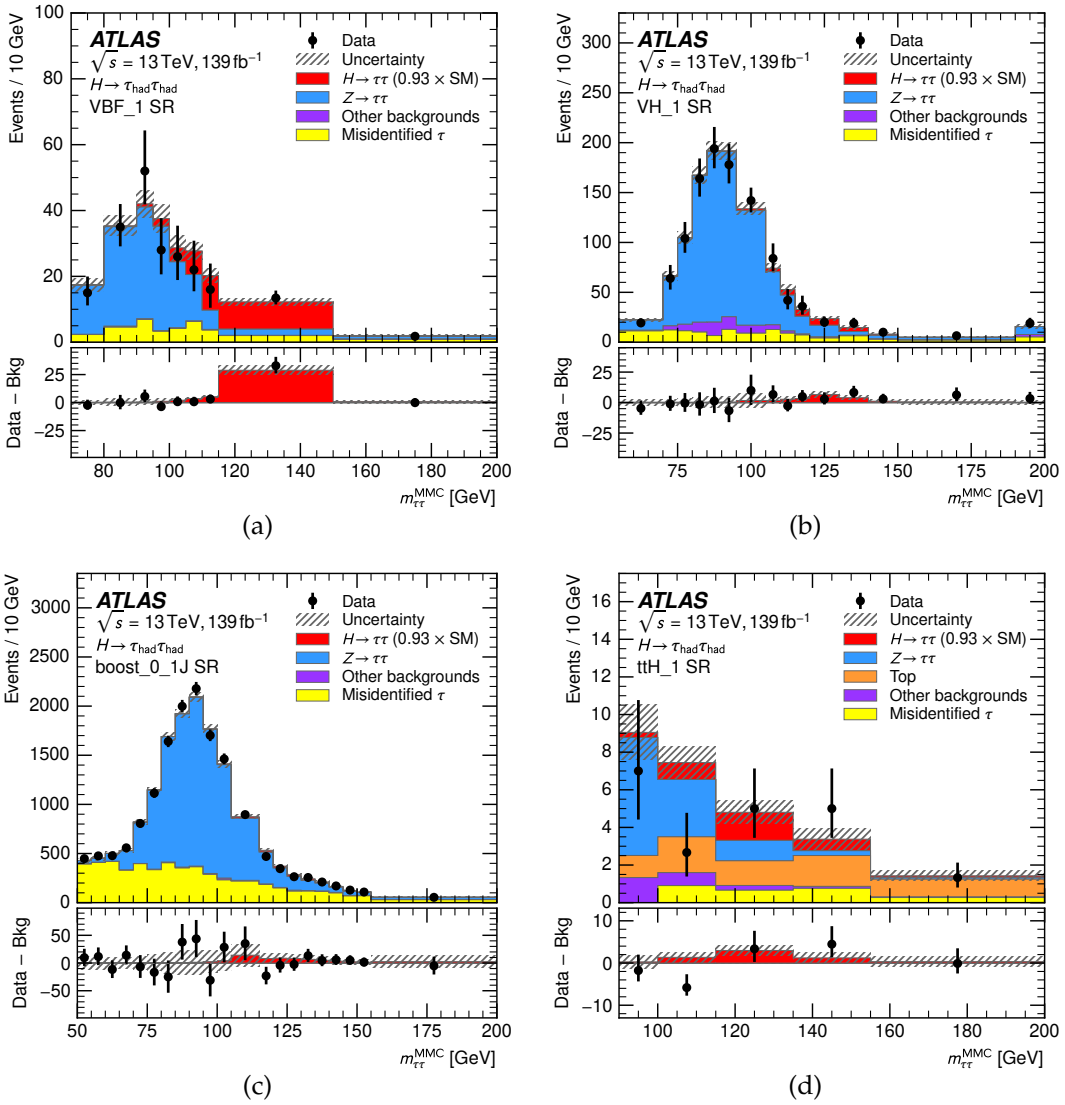


Figure B.11.: Mass $m_{\tau\tau}^{\mathrm{MMC}}$ distributions in selected signal regions of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel with the all nuisance parameters set to the value that maximizes the likelihood function [84]. The hatched band and the error bars indicate total uncertaintainty of the background expectation and statistical uncertainty measured data, respectively.

Table B.2.: Observed and from an Asimov fit expected impact of different sources of uncertainty [84]. The impact is measured as the relative change in cross section $\Delta\sigma/\sigma$ in percent (%).

Uncertainty	Observed impact	Expected impact
Signal theory uncertainty	8.7	8.5
Jet and $E_{\mathrm{T}}^{\mathrm{miss}}$	4.5	4.2
MC statistical uncertainty	4.0	3.7
Hadronic taus	2.1	2.1
Misidentificaiton estimation	2.0	2.0
Luminosity	1.8	1.8
Theory uncertainty in Z+jets processes	1.7	1.2
Theory uncertainty in top processes	1.1	1.1
Flavor tagging	0.4	0.5
Light leptons	0.4	0.4
Total systematic uncertainty	12.0	11.4
Data statisical uncertainty	7.2	6.7
Total	13.9	13.2

Acknowledgement

The completion of this thesis marks a significant milestone in my life. I would like to take a moment and thank all the people who helped me and supported my work to achieve this goal. It was an honor and a pleasure to work with so many brilliant people in an international collaboration on a truly inspiring topic, to expand the frontier of human understanding, and to put it boldly, to unravel the mysteries of the universe.

I am very grateful and would like to thank Karl Jakobs and Christian Weiser for their supervision, the opportunity to pursue a deeper understanding of Higgs physics, and the chance to achieve a doctoral degree with this research. I'd like to take the opportunity to thank Karsten Köneke for his continuous and sustained support, the excellent feedback, and the time he took to provide supervision despite other important obligations. Throughout the years, Karsten has been a constant source of advice.

Since the beginning of my master thesis, I have had the opportunity to learn from experienced postdocs at the University of Freiburg and at CERN. In chronological order, I'd like to thank Lei Zhang, Minoru Hirose, Shigeki Hirose, Spyros Argyropoulos, Daniele Zanzi, and Christopher Young for the opportunity to learn from your experience.

I'd also like to express my gratitude to Théo Megy with whom I had the pleasure of sharing an office for most of the time. Furthermore, I'd like to express my appreciation to all current and former group members at the University of Freiburg for creating a nice and productive environment. Lastly, I'd like to thank my family for their continued support and encouragement in mastering this achievement.

Ever since I had the opportunity to participate a guided tour at CERN as a high school student, I have been intrigued to participate in this outstanding scientific endeavor. Working within the ATLAS Collaboration at CERN and operating the ATLAS detector in the control room was literally a dream coming true. I hope I was able to ignite the same fire of enthusiasm in the next generation of aspiring students during my work as a CERN guide, within the Netzwerk Teilchenwelt, as a tutor with physics students, and as a supervisor of bachelor students and research assistants.

List of Figures

List of Tables

Bibliography

[1] Bureau international des poids et mesures, *The International System of Units*, tech. rep., 2019 (cit. on p. xiii).

[2] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 7).

[3] Serguei Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 7).

[4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on pp. 1, 7).

[5] Peter W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132 (cit. on pp. 1, 7).

[6] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, ed. by J. C. Taylor (cit. on pp. 1, 7).

[7] David Griffiths, *Introduction to elementary particles*, Wiley-VCH, 2008, ISBN: 978-3-527-40601-2 (cit. on pp. 2, 5, 12, 128).

[8] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to* 139 fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ *TeV collected with the ATLAS experiment*, tech. rep., CERN, 2021, URL: https://cds.cern.ch/record/2789544 (cit. on pp. 3, 112, 115–117, 120–124).

[9] F. Halzen and Alan D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, Wiley, 1984, ISBN: 978-0-471-88741-6 (cit. on pp. 5, 9).

[10] Mark Thomson, *Modern particle physics*, New York: Cambridge University Press, 2013, ISBN: 978-1-107-03426-6 (cit. on pp. 5, 24).

[11] M.E. Peskin, *An Introduction To Quantum Field Theory*, CRC Press, 2018, ISBN: 9780429972102 (cit. on pp. 5, 123).

[12] Michael Kachelriess, *Quantum Fields: From the Hubble to the Planck Scale*, Oxford Graduate Texts, Oxford University Press, 2017, ISBN: 978-0-19-880287-7 (cit. on pp. 5, 9).

[13] A. Zee, *Quantum field theory in a nutshell*, Princeton University Press, 2003, ISBN: 978-0-691-14034-6 (cit. on pp. 5, 9).

[14] P.A. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01 (cit. on pp. 7, 20, 46, 63, 64).

[15] Philip W. Anderson, *Plasmons, Gauge Invariance, and Mass*, Phys. Rev. **130** (1963) 439, ed. by J. C. Taylor (cit. on p. 7).

[16] Peter W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156 (cit. on p. 7).

[17] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (cit. on p. 7).

[18] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (5 1967) 1554, URL: https://link.aps.org/doi/10.1103/PhysRev.155.1554 (cit. on p. 7).

[19] G.'t Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*, Nuclear Physics B **35** (1971) 167, ISSN: 0550-3213, URL: https://www.sciencedirect.com/science/article/pii/0550321371901398 (cit. on p. 9).

[20] Steven Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264 (cit. on p. 13).

[21] Lyndon Evans and Philip Bryant, *LHC Machine*, JINST **3** (2008) S08001 (cit. on pp. 13, 16, 35, 43).

[22] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv: 0901.0002 [hep-ph] (cit. on p. 15).

[23] W. J. Stirling and G. Watt, *private communication*, (2022), accessed 2022-04-27, URL: https://mstwpdf.hepforge.org/plots/plots.html (cit. on p. 15).

[24] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, tech. rep., CERN, 2019, URL: http://cds.cern.ch/record/2677054 (cit. on pp. 16, 36, 37, 92).

[25] ATLAS Collaboration, *internal communication*, (2019), accessed 2022-04-27, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (cit. on pp. 16, 37).

[26] Frank Sauerburger, *LHC cross section plot*, (2022), accessed 2022-04-27, URL: https://lhc-xsecs.org/ (cit. on p. 18).

[27] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex] (cit. on pp. 20, 21).

[28] LHC Higgs Cross Section Working Group, D. de Florian, et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: 1610.07922 [hep-ph] (cit. on pp. 21, 67, 110, 112, 115).

[29] J. R. Andersen et al., "Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report", *9th Les Houches Workshop on Physics at TeV Colliders*, 2016, arXiv: 1605.04692 [hep-ph] (cit. on p. 23).

[30] Nicolas Berger et al., *Simplified Template Cross Sections - Stage 1.1*, (2019), arXiv: 1906.02754 [hep-ph] (cit. on pp. 23–25).

[31] S. Amoroso et al., "Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report", *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 2020, arXiv: 2003.01700 [hep-ph] (cit. on pp. 23, 26).

[32] M. Boggia et al., *The HiggsTools handbook: a beginners guide to decoding the Higgs sector*, J. Phys. G **45** (2018) 065004, arXiv: 1711.09875 [hep-ph] (cit. on p. 23).

[33] Edvige Corbelli and Paolo Salucci, *The extended rotation curve and the dark matter halo of M33*, Monthly Notices of the Royal Astronomical Society **311** (2000) 441, ISSN: 0035-8711, URL: https://doi.org/10.1046/j.1365-8711.2000.03075.x (cit. on p. 26).

[34] Jürgen Renn, Tilman Sauer, and John Stachel, *The Origin of Gravitational Lensing: A Postscript to Einstein's 1936 Science Paper*, Science **275** (1997) 184 (cit. on p. 26).

[35] Planck Collaboration, *Planck 2015 results - XIII. Cosmological parameters*, A&A **594** (2016) A13, URL: https://doi.org/10.1051/0004-6361/201525830 (cit. on p. 26).

[36] S. Fukuda et al., *Solar B-8 and hep neutrino measurements from 1258 days of Super-Kamiokande data*, Phys. Rev. Lett. **86** (2001) 5651, arXiv: hep-ex/0103032 (cit. on p. 26).

[37] Q. R. Ahmad et al., *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301, arXiv: nucl-ex/0106015 (cit. on p. 26).

[38] Q. R. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, arXiv: nucl-ex/0204008 (cit. on p. 26).

[39] M. Aker et al., *Direct neutrino-mass measurement with sub-electronvolt sensitivity*, Nature Phys. **18** (2022) 160, arXiv: 2105.08533 [hep-ex] (cit. on p. 26).

[40] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016 (cit. on pp. 27, 46, 47).

[41] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, New York, NY, USA: Springer New York Inc., 2001 (cit. on pp. 27, 49).

[42] C.M. Bishop, *Pattern Recognition and Machine Learning*, Information science and statistics, Springer, 2006 (cit. on p. 27).

[43] Enrico Bothmann et al., *Event Generation with Sherpa 2.2*, SciPost Phys. **7** (3 2019) 34, URL: https://scipost.org/10.21468/SciPostPhys.7.3.034 (cit. on p. 29).

[44] Kurt Hornik, Maxwell Stinchcombe, and Halbert White, *Multilayer feedforward networks are universal approximators*, Neural Networks **2** (1989) 359, ISSN: 0893-6080 (cit. on p. 32).

[45] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Mathematics of Control, Signals, and Systems (MCSS) **2** (1989) 303, ISSN: 0932-4194, URL: http://dx.doi.org/10.1007/BF02551274 (cit. on p. 32).

[46] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, *Learning Representations by Back-propagating Errors*, Nature **323** (1986) 533, URL: http://www.nature.com/articles/323533a0 (cit. on p. 32).

[47] Federico Girosi, Michael Jones, and Tomaso Poggio, *Regularization Theory and Neural Networks Architectures*, Neural Computation **7** (1995) 219, ISSN: 0899-7667, URL: https://doi.org/10.1162/neco.1995.7.2.219 (cit. on p. 32).

[48] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, 2012, URL: https://arxiv.org/abs/1207.0580 (cit. on p. 32).

[49] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) S08002, URL: https://doi.org/10.1088/1748-0221/3/08/s08002 (cit. on p. 36).

[50] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 36, 37, 39–43, 52).

[51] CMS Collaboration, *The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment*, JINST **3** (2008) S08004. 361 p, URL: https://cds.cern.ch/record/1129810 (cit. on p. 36).

[52] LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) S08005, URL: https://doi.org/10.1088/1748-0221/3/08/s08005 (cit. on p. 36).

[53] B. Muratori and T. Pieloni, *Luminosity levelling techniques for the LHC*, (2014) 177, arXiv: 1410.5646 [physics.acc-ph] (cit. on p. 36).

[54] CERN and J. Pequenao, *Computer generated image of the whole ATLAS detector*, (2008), accessed 2022-04-27, URL: https://cds.cern.ch/record/1095924 (cit. on p. 38).

[55] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 39).

[56] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 39), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[57] Hermann Kolanoski and Norbert Wermes, *Teilchendetektoren: Grundlagen und Anwendungen*, Springer, 2016, ISBN: 978-3-662-45349-0, 978-3-662-45350-6 (cit. on pp. 40, 51, 52).

[58] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, JINST **12** (2017) P05002, arXiv: 1702.06473 [hep-ex] (cit. on p. 40).

[59] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 43).

[60] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8\,TeV$ using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on p. 43).

[61] T Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119** (2008) 032014, URL: https://doi.org/10.1088/1742-6596/119/3/032014 (cit. on p. 44).

[62] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 44).

[63] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009, arXiv: 1406.7690 [hep-ex] (cit. on p. 44).

[64] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 45, 70).

[65] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13\,TeV$*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on p. 45).

[66] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on p. 45).

[67] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 45).

[68] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 45).

[69] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 45).

[70] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\,TeV$ with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: 2007.02645 [hep-ex] (cit. on pp. 45, 91).

[71] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13\,TeV$*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 46).

[72] Georges Aad et al., *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, Eur. Phys. J. C **76** (2016) 295, arXiv: 1512.05955 [hep-ex] (cit. on p. 46).

[73] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, tech. rep., CERN, 2015, URL: https://cds.cern.ch/record/2064383 (cit. on p. 46).

[74] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2688062 (cit. on pp. 46, 47, 70).

[75] 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., CERN, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on p. 46).

[76] R. Keith Ellis, I. Hinchliffe, M. Soldate, and J. J. van der Bij, *Higgs Decay to tau+ tau-: A Possible Signature of Intermediate Mass Higgs Bosons at the SSC*, Nucl. Phys. B **297** (1988) 221 (cit. on p. 48).

[77] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. A **654** (2011) 481, arXiv: 1012.4686 [hep-ex] (cit. on p. 49).

[78] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, Texts in statistical science series, Chapman & Hall/CRC, 2004, ISBN: 9780203594292 (cit. on p. 49).

[79] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, tech. rep., CERN, 2015, URL: https://cds.cern.ch/record/2037683 (cit. on p. 54).

[80] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117, arXiv: 1501.04943 [hep-ex] (cit. on pp. 63, 126).

[81] Serguei Chatrchyan et al., *Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons*, JHEP **05** (2014) 104, arXiv: 1401.5041 [hep-ex] (cit. on p. 63).

[82] ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) 045, arXiv: 1606.02266 [hep-ex] (cit. on pp. 63, 116).

[83] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072001, arXiv: 1811.08856 [hep-ex] (cit. on pp. 63, 98, 103, 126).

[84] Georges Aad et al., *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2022) 175, arXiv: 2201.08269 [hep-ex] (cit. on pp. 63, 68, 76, 80, 83, 92, 101, 103–105, 113, 126, 142–145).

[85] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with a vector boson and decaying into a tau pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Rev. D **93** (2016) 092005, arXiv: 1511.08352 [hep-ex] (cit. on p. 64).

[86] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D **97** (2018) 072003, arXiv: 1712.08891 [hep-ex] (cit. on pp. 64, 113, 114).

[87] ATLAS Collaboration, *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*, tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2693930 (cit. on p. 64).

[88] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002 (cit. on p. 67).

[89] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph] (cit. on p. 67).

[90] Paolo Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 67).

[91] Stefano Frixione, Paolo Nason, and Carlo Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 67).

[92] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 67).

[93] Heribertus B. Hartanto, Barbara Jager, Laura Reina, and Doreen Wackeroth, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D **91** (2015) 094003, arXiv: 1501.04498 [hep-ph] (cit. on p. 67).

[94] Keith Hamilton, Paolo Nason, and Giulia Zanderighi, *MINLO: Multi-Scale Improved NLO*, JHEP **10** (2012) 155, arXiv: 1206.3572 [hep-ph] (cit. on p. 67).

[95] John M. Campbell et al., *NLO Higgs Boson Production Plus One and Two Jets Using the POWHEG BOX, MadGraph4 and MCFM*, JHEP **07** (2012) 092, arXiv: 1202.5475 [hep-ph] (cit. on p. 67).

[96] Keith Hamilton, Paolo Nason, Carlo Oleari, and Giulia Zanderighi, *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, JHEP **05** (2013) 082, arXiv: 1212.4504 [hep-ph] (cit. on p. 67).

[97] Stefano Catani and Massimiliano Grazzini, *An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC*, Phys. Rev. Lett. **98** (2007) 222002, arXiv: hep-ph/0703012 (cit. on p. 67).

[98] Torbjörn Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 67).

[99] Georges Aad et al., *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex] (cit. on p. 67).

[100] Jon Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 67).

[101] Marco Bonetti, Kirill Melnikov, and Lorenzo Tancredi, *Higher order corrections to mixed QCD-EW contributions to Higgs boson production in gluon fusion*, Phys. Rev. D **97** (2018) 056017, [Erratum: Phys.Rev.D 97, 099906 (2018)], arXiv: 1801.10403 [hep-ph] (cit. on p. 67).

[102] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati, *NNLO Computational Techniques: The Cases H → gamma gamma and H → g g*, Nucl. Phys. B **811** (2009) 182, arXiv: 0809.3667 [hep-ph] (cit. on p. 67).

[103] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati, *NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders*, Phys. Lett. B **670** (2008) 12, arXiv: 0809.1301 [hep-ph] (cit. on p. 67).

[104] Alexey Pak, Mikhail Rogal, and Matthias Steinhauser, *Finite top quark mass effects in NNLO Higgs boson production at LHC*, JHEP **02** (2010) 025, arXiv: 0911.4662 [hep-ph] (cit. on p. 67).

[105] Robert V. Harlander, Hendrik Mantler, Simone Marzani, and Kemal J. Ozeren, *Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass*, Eur. Phys. J. C **66** (2010) 359, arXiv: 0912.2104 [hep-ph] (cit. on p. 67).

[106] Robert V. Harlander and Kemal J. Ozeren, *Top mass effects in Higgs production at next-to-next-to-leading order QCD: Virtual corrections*, Phys. Lett. B **679** (2009) 467, arXiv: 0907.2997 [hep-ph] (cit. on p. 67).

[107] Robert V. Harlander and Kemal J. Ozeren, *Finite top mass effects for hadronic Higgs production at next-to-next-to-leading order*, JHEP **11** (2009) 088, arXiv: 0909.3420 [hep-ph] (cit. on p. 67).

[108] Falko Dulat, Achilleas Lazopoulos, and Bernhard Mistlberger, *iHixs 2 — Inclusive Higgs cross sections*, Comput. Phys. Commun. **233** (2018) 243, arXiv: 1802.00827 [hep-ph] (cit. on p. 67).

[109] Charalampos Anastasiou et al., *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, Phys. Rev. Lett. **114** (2015) 212001, arXiv: 1503.06056 [hep-ph] (cit. on p. 67).

[110] Charalampos Anastasiou et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, JHEP **05** (2016) 058, arXiv: 1602.00695 [hep-ph] (cit. on p. 67).

[111] M. Ciccolini, Ansgar Denner, and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs + 2jets via weak interactions at the LHC*, Phys. Rev. Lett. **99** (2007) 161803, arXiv: 0707.0381 [hep-ph] (cit. on p. 67).

[112] Mariano Ciccolini, Ansgar Denner, and Stefan Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, Phys. Rev. D **77** (2008) 013002, arXiv: 0710.4749 [hep-ph] (cit. on p. 67).

[113] Paolo Bolzoni, Fabio Maltoni, Sven-Olaf Moch, and Marco Zaro, *Higgs production via vector-boson fusion at NNLO in QCD*, Phys. Rev. Lett. **105** (2010) 011801, arXiv: 1003.4451 [hep-ph] (cit. on p. 67).

[114] M. L. Ciccolini, S. Dittmaier, and M. Kramer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. D **68** (2003) 073003, arXiv: hep-ph/0306234 (cit. on p. 67).

[115] Oliver Brein, Abdelhak Djouadi, and Robert Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. B **579** (2004) 149, arXiv: hep-ph/0307206 (cit. on p. 67).

[116] Oliver Brein, Robert Harlander, Marius Wiesemann, and Tom Zirke, *Top-Quark Mediated Effects in Hadronic Higgs-Strahlung*, Eur. Phys. J. C **72** (2012) 1868, arXiv: 1111.0761 [hep-ph] (cit. on p. 67).

[117] Ansgar Denner, Stefan Dittmaier, Stefan Kallweit, and Alexander Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, Comput. Phys. Commun. **195** (2015) 161, arXiv: 1412.5390 [hep-ph] (cit. on p. 67).

[118] Oliver Brein, Robert V. Harlander, and Tom J. E. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*, Comput. Phys. Commun. **184** (2013) 998, arXiv: 1210.5347 [hep-ph] (cit. on p. 67).

[119] Robert V. Harlander, Anna Kulesza, Vincent Theeuwes, and Tom Zirke, *Soft gluon resummation for gluon-induced Higgs Strahlung*, JHEP **11** (2014) 082, arXiv: 1410.0217 [hep-ph] (cit. on p. 67).

[120] David J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002 (cit. on p. 67).

[121] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, tech. rep., CERN, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 67).

[122] Richard D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 67).

[123] W. Beenakker et al., *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. B **653** (2003) 151, arXiv: hep-ph/0211352 (cit. on p. 67).

[124] S. Dawson et al., *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. D **68** (2003) 034022, arXiv: hep-ph/0305087 (cit. on p. 67).

[125] Yu Zhang et al., *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Phys. Lett. B **738** (2014) 1, arXiv: 1407.1110 [hep-ph] (cit. on p. 67).

[126] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *Weak corrections to Higgs hadroproduction in association with a top-quark pair*, JHEP **09** (2014) 065, arXiv: 1407.0823 [hep-ph] (cit. on p. 67).

[127] Barbara Jager, Laura Reina, and Doreen Wackeroth, *Higgs boson production in association with b jets in the POWHEG BOX*, Phys. Rev. D **93** (2016) 014030, arXiv: 1509.05843 [hep-ph] (cit. on p. 67).

[128] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 68).

[129] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro, *Higgs production in association with a single top quark at the LHC*, Eur. Phys. J. C **75** (2015) 267, arXiv: 1504.00611 [hep-ph] (cit. on p. 68).

[130] Federico Demartin et al., *tWH associated production at the LHC*, Eur. Phys. J. C **77** (2017) 34, arXiv: 1607.05862 [hep-ph] (cit. on p. 68).

[131] Enrico Bothmann et al., *Event Generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on p. 68).

[132] Tanju Gleisberg and Stefan Hoeche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 68).

[133] Federico Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph] (cit. on p. 68).

[134] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 68).

[135] Ansgar Denner, Stefan Dittmaier, and Lars Hofer, *Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations*, Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph] (cit. on p. 68).

[136] A. Bredenstein, Ansgar Denner, S. Dittmaier, and M. M. Weber, *Precision calculations for the Higgs decays $H \to ZZ/WW \to 4$ leptons*, Nucl. Phys. B Proc. Suppl. **160** (2006) 131, ed. by J. Blumlein, S. Moch, and T. Riemann, arXiv: hep-ph/0607060 (cit. on p. 68).

[137] Stefan Hoeche, Frank Krauss, Marek Schonherr, and Frank Siegert, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049, arXiv: 1111.1220 [hep-ph] (cit. on p. 68).

[138] Stefan Hoeche, Frank Krauss, Marek Schonherr, and Frank Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 68).

[139] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD matrix elements + parton showers*, JHEP **11** (2001) 063, arXiv: hep-ph/0109231 (cit. on p. 68).

[140] Stefan Hoeche, Frank Krauss, Steffen Schumann, and Frank Siegert, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv: 0903.1219 [hep-ph] (cit. on p. 68).

[141] Charalampos Anastasiou, Lance J. Dixon, Kirill Melnikov, and Frank Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266 (cit. on p. 68).

[142] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl. Phys. B **855** (2012) 695, arXiv: 1109.1536 [hep-ph] (cit. on p. 68).

[143] Matteo Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612, arXiv: 1111.5869 [hep-ph] (cit. on p. 68).

[144] Peter Bärnreuther, Michal Czakon, and Alexander Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph] (cit. on p. 68).

[145] Michal Czakon and Alexander Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv: 1207.0236 [hep-ph] (cit. on p. 68).

[146] Michal Czakon and Alexander Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv: 1210.6832 [hep-ph] (cit. on p. 68).

[147] Michał Czakon, Paul Fiedler, and Alexander Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph] (cit. on p. 68).

[148] Michal Czakon and Alexander Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Com. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 68).

[149] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2020) 578, arXiv: 2012.00578 [hep-ex] (cit. on p. 70).

[150] Georges Aad et al., *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 70).

[151] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex] (cit. on p. 70), Erratum: Eur. Phys. J. C **77** (2017) 712.

[152] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, tech. rep., CERN, 2017, URL: https://cds.cern.ch/record/2273281 (cit. on p. 70).

[153] Georges Aad et al., *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on p. 70).

[154] ATLAS Collaboration, *Measurements of gluon fusion and vector-boson-fusion production of the Higgs boson in $H \to WW^* \to e\nu\mu\nu$ decays using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., CERN, 2021, URL: https://cds.cern.ch/record/2759651 (cit. on pp. 71, 113).

[155] Martina Laura Ojeda, *Measurement of Higgs couplings to top quarks and τ leptons with the ATLAS detector*, (2021), accessed 2022-04-27, URL: https://hdl.handle.net/1807/109253 (cit. on p. 76).

[156] ATLAS Collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **74** (2014) 3023, arXiv: 1405.6583 [hep-ex] (cit. on p. 77).

[157] Andreas Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an] (cit. on p. 81).

[158] Dan Guest, Kyle Cranmer, and Daniel Whiteson, *Deep Learning and Its Application to LHC Physics*, Annual Review of Nuclear and Particle Science **68** (2018) 161, eprint: https://doi.org/10.1146/annurev-nucl-101917-021019 (cit. on p. 82).

[159] Takuya Akiba et al., "Optuna: A Next-generation Hyperparameter Optimization Framework", *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, URL: https://doi.org/10.1145/3292500.3330701 (cit. on p. 84).

[160] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl, "Algorithms for Hyper-Parameter Optimization", *Advances in Neural Information Processing Systems*, ed. by J. Shawe-Taylor et al., vol. 24, Curran Associates, Inc., 2011, URL: https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf (cit. on p. 84).

[161] François Chollet et al., *Keras*, https://keras.io, 2015 (cit. on p. 86).

[162] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, URL: https://www.tensorflow.org/ (cit. on p. 86).

[163] Dirk Merkel, *Docker: lightweight linux containers for consistent development and deployment*, Linux journal **2014** (2014) 2 (cit. on p. 86).

[164] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 91).

[165] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** (2018) P07017, URL: https://doi.org/10.1088/1748-0221/13/07/p07017 (cit. on p. 92).

[166] Enrico Bothmann, Marek Schönherr, and Steffen Schumann, *Reweighting QCD matrix-element and parton-shower calculations*, Eur. Phys. J. C **76** (2016) 590, arXiv: 1606.08753 [hep-ph] (cit. on p. 92).

[167] Johannes Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on p. 92).

[168] *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*, tech. rep., CERN, 2017, URL: https://cds.cern.ch/record/2261938 (cit. on p. 92).

[169] Sayipjamal Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. D **93** (2016) 033006, arXiv: 1506.07443 [hep-ph] (cit. on p. 93).

[170] L.A. Harland-Lang, A.D. Martin, P. Motylinski, and R.S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75** (2015) 204, arXiv: 1412.3989 [hep-ph] (cit. on p. 93).

[171] Iain W. Stewart and Frank J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, Phys. Rev. D **85** (2012) 034011, arXiv: 1107.2117 [hep-ph] (cit. on p. 93).

[172] Rikkert Frederix and Stefano Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012) 061, arXiv: 1209.6215 [hep-ph] (cit. on p. 93).

[173] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012, URL: http://cds.cern.ch/record/1456844 (cit. on p. 94).

[174] G. Cowan, *Statistical Data Analysis*, Oxford science publications, Clarendon Press, 1998, ISBN: 9780198501558 (cit. on pp. 94, 115).

[175] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, Wiley, 2013, ISBN: 9783527653430 (cit. on pp. 94, 115).

[176] Tim Barklow et al., *Improved Formalism for Precision Higgs Coupling Fits*, Phys. Rev. D **97** (2018) 053003, arXiv: 1708.08912 [hep-ph] (cit. on p. 110).

[177] LHC Higgs Cross Section Working Group, J R Andersen, et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, (2013), ed. by S Heinemeyer, C Mariotti, G Passarino, and R Tanaka, arXiv: 1307.1347 [hep-ph] (cit. on pp. 110, 112).

[178] ATLAS Collaboration, *A combination of measurements of Higgs boson production and decay using up to* 139 fb^{-1} *of proton–proton collision data at* $\sqrt{s}$ *= 13 TeV collected with the ATLAS experiment*, tech. rep., CERN, 2020, URL: https://cds.cern.ch/record/2725733 (cit. on pp. 113, 118).

[179] ATLAS Collaboration, *Measurement of the properties of Higgs boson production at* $\sqrt{s}$*=13 TeV in the* $H \to \gamma\gamma$ *channel using 139* fb^{-1} *of pp collision data with the ATLAS experiment*, tech. rep., CERN, 2020, URL: https://cds.cern.ch/record/2725727 (cit. on p. 113).

[180] ATLAS Collaboration, *Higgs boson production cross-section measurements and their EFT interpretation in the* 4ℓ *decay channel at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 957, arXiv: 2004.03447 [hep-ex] (cit. on p. 113), Erratum: Eur. Phys. J. C **81** (2021) 29.

[181] ATLAS Collaboration, *Measurements of WH and ZH production in the* $H \to b\bar{b}$ *decay channel in pp collisions at* 13 *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 178, arXiv: 2007.02873 [hep-ex] (cit. on pp. 113, 114).

[182] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **816** (2021) 136204, arXiv: 2008.02508 [hep-ex] (cit. on pp. 113, 114).

[183] ATLAS Collaboration, *Combination of measurements of Higgs boson production in association with a W or Z boson in the* $b\bar{b}$ *decay channel with the ATLAS experiment at* $\sqrt{s} = 13$ *TeV*, tech. rep., CERN, 2021, URL: https://cds.cern.ch/record/2782535 (cit. on pp. 113, 114).

[184] Georges Aad et al., *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, (2021), arXiv: 2111.06712 [hep-ex] (cit. on pp. 113, 114).

[185] ATLAS Collaboration, *Measurements of Higgs Bosons Decaying to Bottom Quarks from Vector Boson Fusion Production with the ATLAS Experiment at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81** (2020) 537, arXiv: 2011.08280 [hep-ex] (cit. on pp. 113, 114).

[186] ATLAS Collaboration, *A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector*, Phys. Lett. B **812** (2021) 135980, arXiv: 2007.07830 [hep-ex] (cit. on pp. 113, 114).

[187] ATLAS Collaboration, *A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **809** (2020) 135754, arXiv: 2005.05382 [hep-ex] (cit. on pp. 113, 114).

[188] ATLAS Collaboration, *Search for invisible Higgs boson decays with vector boson fusion signatures with the ATLAS detector using an integrated luminosity of 139 fb^{-1}*, tech. rep., CERN, 2020, URL: https://cds.cern.ch/record/2715447 (cit. on pp. 113, 114, 119, 122).

[189] C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. C **40** (2016) 100001 (cit. on p. 123).

[190] P. Sikivie, Leonard Susskind, Mikhail B. Voloshin, and Valentin I. Zakharov, *Isospin Breaking in Technicolor Models*, Nucl. Phys. B **173** (1980) 189 (cit. on p. 123).

[191] H. Georgi, *Effective field theory*, Ann. Rev. Nucl. Part. Sci. **43** (1993) 209 (cit. on p. 123).

[192] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, Tevatron Electroweak Working Group, SLD Electroweak, Heavy Flavour Groups, *Precision Electroweak Measurements and Constraints on the Standard Model*, (2010), arXiv: 1012.2367 [hep-ex] (cit. on p. 123).