Search for Higgs boson pair production in the bbWW final state with leptonic W boson decays at the CMS Experiment

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Manfred Peter Fackeldey

M. Sc. RWTH

aus Düren

Berichter: Univ.-Prof. Dr. rer. nat. Martin Erdmann
Univ.-Prof. Dr. rer. nat. Alexander Schmidt

Tag der mündlichen Prüfung: 16.02.2024

Diese Dissertation ist auf den Internetseiten der Universitatsbibliothek verfügbar.

Manfred Peter Fackeldey
Search for Higgs boson pair production in the bbWW final state with leptonic W boson decays at the CMS Experiment

Dissertation in Physik
Rheinisch-Westfälische Technische Hochschule Aachen
III. Physikalisches Institut A

Berichter:
Univ.-Prof. Dr. rer. nat. Martin Erdmann
Univ.-Prof. Dr. rer. nat. Alexander Schmidt

Abstract

This thesis presents a search for Higgs boson pair production in the $HH \rightarrow b\bar{b}W^+W^-$ decay channel where both W bosons decay leptonically. The search uses data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ at the LHC. In total, data collected by the CMS experiment corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$ are analyzed.

The biggest challenge of the presented analysis is the small amount of expected signal events compared to the large amount of background events. This imbalance of expected events requires a well understood modelling of the background processes. A dedicated deep learning based analysis strategy is used to tackle these challenges. In addition, novel computing technologies are leveraged to enable a sustainable and efficient analysis.

No evidence for the production of Higgs boson pairs is found. Upper exclusion limits on the inclusive signal strength modifier are observed (expected) to be 19 (27), and 385 (424) on the HH(VBF) signal strength modifier at 95% CL_s. The Higgs boson self-coupling (κ_λ) and the direct coupling between two Higgs bosons and two vector bosons (κ_{2V}) are measured to be $\kappa_\lambda = 3.17^{+6.19}_{-6.15}$ and $\kappa_{2V} = 1.04^{+1.56}_{-1.53}$. In addition, both couplings are constrained with observed (expected) exclusion limits in the range of $-8.64 \leq \kappa_\lambda \leq 15.21$ ($-11.43 \leq \kappa_\lambda \leq 17.83$) and $-1.71 \leq \kappa_{2V} \leq 3.84$ ($-2.22 \leq \kappa_{2V} \leq 4.37$). These measurements do not deviate from the expectation of the SM, and mark the strictest results in the $HH \rightarrow b\bar{b}W^+W^-$ decay channel to date.

Kurzfassung

In dieser Arbeit wird eine Suche nach der Higgs Boson Paarproduktion im $HH \rightarrow b\bar{b}W^+W^-$ Zerfallskanal vorgestellt, bei dem beide W-Bosonen leptonisch zerfallen. Die Suche verwendet Daten aus Proton-Proton Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ am LHC. Insgesamt werden Daten, gesammelt vom CMS Experiment, analysiert, die einer integrierten Luminsosität von $138\,\text{fb}^{-1}$ entsprechen.

Die größte Herausforderung der vorgestellten Analyse ist die geringe Menge an erwarteten Signalereignissen im Vergleich zur großen Menge an Untergrundereignissen. Dieses Ungleichgewicht der erwarteten Ereigniszahlen erfordert eine gut verstandene Modellierung der Untergrundprozesse. Eine dedizierte deep learning basierte Analysestrategie wird benutzt, um diese Herausforderungen zu bewältigen. Außerdem werden neuartige Rechentechnologien benutzt, um eine nachhaltige und effiziente Analyse zu ermöglichen.

Es sind keine Hinweise auf die Produktion von Higgs Boson Paaren gefunden worden. Die oberen Ausschlussgrenzen der Signalstärkemodifikatoren für den inklusiven und den HH(VBF) Produktionsmodus werden bei 19(27) und 385(424) bei 95% CL_s beobachtet (erwartet). Die Selbstkopplung des Higgs Bosons (κ_λ) und die direkte Kopplung zwischen zwei Higgs Bosonen und zwei Vektorbosonen (κ_{2V}) werden gemessen zu $\kappa_\lambda = 3.17^{+6.19}_{-6.15}$ und $\kappa_{2V} = 1.04^{+1.56}_{-1.53}$. Darüber hinaus sind die beide Kopplungen mit beobachteten (erwarteten) Ausschlussgrenzen in den Bereichen $-8.64 \leq \kappa_\lambda \leq 15.21$ ($-11.43 \leq \kappa_\lambda \leq 17.83$) und $-1.71 \leq \kappa_{2V} \leq 3.84$ ($-2.22 \leq \kappa_{2V} \leq 4.37$) eingeschränkt. Diese Messungen weichen nicht von den Erwartungen des SM ab und repräsentieren die bisher stricktesten Ergebnisse im $HH \rightarrow b\bar{b}W^+W^-$ Zerfallskanal.

Contents

Chapter 1

Introduction

Particle physics is the science of the smallest building blocks of nature, so-called elementary particles. The research of these particles and their interactions pushes the boundaries of our understanding of the universe. In the last century, more and more particles have been discovered with experiments that accelerate and collide particles at the highest energies. These experiments are designed, built, and operated by international collaborations of thousands of scientists.

A theory, called the standard model of particle physics (SM), has been developed by physicists to describe elementary particles and their interactions. The SM is a quantum field theory based on the concept of gauge invariance. Gauge invariance demands Lagrange densities describing the dynamics of particles to remain unchanged under local transformations of fields. The SM predicts processes in the field of particle physics with high precision.

Initially the SM could not explain the origin of particle masses. This problem has been solved for bosons in 1964 by introducing a mechanism theorized by Peter Higgs [1, 2], Robert Brout, and François Englert [3], Gerald S. Guralnik, Carl R. Hagen, and Tom W. B. Kibble [4]. The mechanism is known as the Higgs mechanism. An extension of the mechanism to matter particles was developed by including Yukawa interactions [5]. The Higgs mechanism introduces the scalar Higgs field and its energy potential (Higgs potential). The Higgs field entails the existence of a new particle, the Higgs boson, which has been discovered 48 years after its prediction in 2012 at the Large Hadron Collider (LHC) by the CMS and ATLAS collaborations [6, 7]. This discovery marked a first confirmation of the Higgs mechanism and thus a milestone in particle physics. It was awarded with the Nobel Prize in physics in 2013 for François Englert and Peter Higgs.

Although the SM is a success story, it is not a complete theory, e.g., it cannot explain the existence of dark matter [8] and dark energy [9]. Therefore, it is of utmost importance to test the SM to its limits and search for new physics beyond the SM. The production of Higgs boson pairs is a fundamental property of the Higgs mechanism, and thus of the SM, that has not been observed before. Its observation tests the SM to its limits, marks another important validation of the Higgs mechanism, and in case of deviations from theory predictions hints the existence of new physics. Additionally, it tests fundamental parameters of the mechanism, such as the Higgs boson self-coupling κ_λ, that may help to unveil the structure of the Higgs potential.

The analysis presented in this thesis searches for Higgs boson pair production in the HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ decay channel where both W bosons decay leptonically (W $\rightarrow$ eν_e and W $\rightarrow$ $\mu\nu_\mu$). It uses proton-proton collision data at a center-of-mass energy of $\sqrt{s} = 13$ TeV recorded by the CMS experiment at the LHC in 2016, 2017, and 2018 corresponding to an integrated

luminosity of around $138\,\mathrm{fb}^{-1}$. The two most dominant production modes of Higgs boson pair production are considered, namely the gluon fusion (HH(GGF)) and the vector boson fusion (HH(VBF)) production. While the HH(GGF) production mode is particularly sensitive to the Higgs boson self-coupling κ_λ, the HH(VBF) production mode is additionally sensitive to the direct coupling between two Higgs bosons and two vector bosons, referred to as κ_{2V} coupling.

The presented analysis has been published in Ref. [10]. The publication shows results of the combination of the dileptonic and semi-leptonic $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$ analysis. The semi-leptonic analysis has been developed in parallel to the presented analysis; both analyses were designed from the beginning to be combined. The presented analysis will contribute to the joined effort of the CMS collaboration aiming to combine all Higgs boson pair production analyses.

The dileptonic $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$ analysis benefits from the large branching fractions of the $\mathrm{H} \to \mathrm{b\bar{b}}$ process, and a clean signature of two leptons from the W boson decays. Spin correlations of the W vector bosons result in narrow opening angles between both leptons (and neutrinos) in the final state. This effect benefits the sensitivity of this analysis. Several challenges appear in the search for the dileptonic $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$ process. Background processes, especially the irreducible $t\bar{t}$ + Jets process, overwhelm the small number of expected signal events by several orders of magnitude. While 112 HH(GGF) events and 6 HH(VBF) events are expected to be produced at the interaction point of the CMS experiment in the Run 2 data-taking period, roughly $\mathcal{O}(10^8)$ are expected for the $t\bar{t}$ + Jets background alone. Furthermore, the two undetected neutrinos and the broad invariant mass spectrum of the $\mathrm{H} \to \mathrm{b\bar{b}}$ process are a challenge for the event reconstruction. These challenges are tackled by a dedicated analysis strategy. At first, a loose event selection is applied to retain as many signal events as possible. Two data-driven background estimation methods are used to improve the modelling of the Drell-Yan background and the background of jets misidentified as leptons. Then, a deep learning model is used to isolate physics processes from each other, especially signal processes from background processes. Evaluating the deep learning model creates dedicated analysis regions mutually enriched by particular physics processes, helping to create signal sensitive regions and regions to control the main background processes. Finally, these regions are used to measure signal strength modifiers affecting the rate of the HH(GGF) and HH(VBF) production modes and coupling modifiers ($\kappa_\lambda, \kappa_{2V}$).

These analysis steps demand vast computational efforts. Novel computing techniques are applied and partly developed to cope with the large amount of data ($\mathcal{O}(10)$ TB) in this analysis. First of all, events are analyzed in a vectorized manner, i.e., multiple events are processed simultaneously on a single thread. Furthermore, I/O latencies are diminished by leveraging a caching system backed by solid-state drives (SSDs) at the VISPA computing cluster [11, 12]. This caching system has been developed during the thesis and published in Ref. [13]. Classically, analyses are scaled up by submitting "jobs" to distributed computing clusters. This paradigm is replaced by analyzing events with the "MapReduce" concept using Dask [14–17]. Here, resource-intensive tasks, such as histogramming and evaluating deep learning models, are externalized to server-like services. Finally, the measurement of signal strength modifiers and coupling modifiers requires their implementation into the likelihood function, a complex pipelining system of fitting routines, and visualization methods of the results. For this, the HH INFERENCE TOOLS [18, 19] software package has been developed and maintained during this thesis.

This thesis is structured as follows. First, the theoretical foundation of the SM, the Higgs mechanism, physics at high energy proton-proton collisions, and Higgs boson pair production are explained in chapter 2. Then, the experimental setup, including the LHC and the CMS detector, is described in the next chapter. Here, central reconstruction steps of the CMS experiment and the used computing infrastructures are described. This is followed by chapter 4, which explains the used computing tools, the mathematical foundation of deep learning, and the statistical methods used for the measurements. Afterwards, the analysis strategy is outlined in chapter 5. Chapter 6 describes the recorded and simulated datasets, and the event selection. Then, the signal extraction based on a deep learning model is explained in chapter 7. The final measurement procedure and results are presented in chapter 8. Chapter 9 concludes the thesis with results of the combination of the dileptonic and semi-leptonic $HH \rightarrow b\bar{b}W^+W^-$ analyses [10].

Chapter 2

Theoretical Foundation

The standard model of particle physics (SM) describes the elementary particles and their interactions. In particular, its particle content comprises integer and half-integer spin particles, called *bosons* and *fermions* respectively. While fermions form matter, bosons mediate the interactions between those. Out of four known forces, the electromagnetic, the weak, the strong, and the gravitational force, the SM describes the first three of them, while the latter is described by general relativity. In order to explain the origin of particle masses, the SM is extended through the so-called "Higgs mechanism".

The following chapter describes the particles and interactions of the SM. In addition, properties of the Higgs pair production and decay in the proton-proton collisions are described.

2.1 The Standard Model of Particle Physics

The standard model of particle physics is a quantum field theory represented by the following gauge group:

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

where the $SU(3)_C$ group describes the strong interaction, and the $SU(2)_L \times U(1)_Y$ group the unified theory of the electromagnetic and weak interaction, with the corresponding color charge (C), weak isospin (L), and the weak hypercharge (Y). The SM obeys symmetries of special relativity, i.e., translation, rotation, and Lorentz invariance, conservation laws of the Noether theorem, i.e., conservation of energy, momentum, and angular momentum, and global and local gauge invariance. Especially, local gauge invariance characterizes the SM as it gives rise to the mediators of each interaction.

The elementary particles of the SM are represented by different types of fields depending on their spin. Scalar fields represent particles with spin 0, Dirac spinor fields particles with spin $\frac{1}{2}$, and vector fields particles with spin 1. The dynamics of a physical system containing these fields is described by a *Lagrangian*. The *free*, i.e., including no interaction, Lagrangians for the previously mentioned fields are given in the following equations:

$$\text{Spin } 0: \quad \mathcal{L} = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{m_\phi^2}{2}\phi, \tag{2.2}$$

$$\text{Spin } \frac{1}{2}: \quad \mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m_\psi \bar{\psi}\psi, \tag{2.3}$$

$$\text{Spin } 1: \quad \mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{m_A^2}{2}A_\mu A^\mu, \tag{2.4}$$

where ϕ describes a scalar field, ψ the Dirac spinor field, A the vector field, γ^μ the gamma matrices, $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ the field strength tensor and m the mass of the particle. Interactions between fields are also described by a Lagrangian, leading to the full Lagrangian of the form:

$$\mathcal{L} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{interaction}}. \tag{2.5}$$

For instance, the Lagrangian of a fermion with electromagnetic interactions that are mediated through (massless) photons, leads to the Lagrangian of the Quantum Electrodynamics (QED) and reads:

$$\mathcal{L}_{\text{QED}} = \underbrace{i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi}_{\mathcal{L}_{\text{fermion}}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{0^2}{2}A_\mu A^\mu}_{\mathcal{L}_{\text{photon}}} + \underbrace{e\bar{\psi}\gamma^\mu A_\mu\psi}_{\mathcal{L}_{\text{interaction}}} \tag{2.6}$$

$$= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + e\bar{\psi}\gamma^\mu A_\mu\psi \tag{2.7}$$

$$= \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.8}$$

where $D_\mu = \partial_\mu - ieA_\mu$ denotes the covariant derivative and e the electric charge. Here, the motivation of local gauge symmetry within the $U(1)$ symmetry group becomes visible. While $\mathcal{L}_{\text{fermion}}$ is not gauge invariant under local transformations, i.e.,

$$\mathcal{L} \to \mathcal{L}' \qquad \text{with} \qquad \psi \to \psi' = e^{i\alpha(x)}\psi, \tag{2.9}$$

$\mathcal{L}_{\text{QED}}$ is invariant by introducing the interaction between fermions mediated through photons. Hereby, the mediator of the electromagnetic interaction - the photon - needs to be massless.

These interactions can be arbitrary complex in a physics process, e.g., involving multiple emissions in the initial or final state, prohibiting exact solutions for the field equations. Treating these (additional) interactions as small perturbations allows to obtain probability amplitudes of the involved physics processes with a perturbative theory using Fermi's golden rule. This perturbative ansatz allows to extend calculations of scattering amplitudes of leading order (LO) processes to higher orders (NLO, NNLO, ...). These calculations involve matrix elements, which can be associated to cross sections and decay widths using Fermi's golden rule. In addition, Feynman diagrams can be used to visualize different orders of physics processes.

2.1.1 Particles

The standard model of particle physics describes the fundamental building blocks of matter and their interactions as point-like particles. The mediators of the three forces in the SM are bosons with a spin of 1. Matter forming particles are called fermions and have a spin of $\frac{1}{2}$. For every fermion there is an anti-fermion with the same mass but opposite charge quantum numbers. Fermions are further subdivided into *leptons* and *quarks*, each grouped, based on their electrical charge, weak isospin, and mass, into three generations. Additionally, quarks with a positive charge are called *up-type* quarks, and quarks with a negative charge are called *down-type* quarks. While quarks are subject to the strong, weak, and electromagnetic force, leptons are only affected by the weak and electromagnetic forces, as they do not carry any color charge. Part of the lepton family are neutrinos, which only participate in the weak interaction, as they do not carry any electrical charge contrary to all other leptons. In addition, neutrinos are massless in the SM; however, the recent discovery [20, 21] of neutrino

oscillations [22, 23] implies non-zero masses. Table 2.1 lists all fermions and their properties. Ordinary matter is only composed of fermions of the first generation due to the increasing mass hierarchy within the three fermion generations. This "instability" of heavy particles means that they can often only be detected by their decay products.

Table 2.1: The elementary fermions, leptons (neutrinos and charged leptons) and quarks, of the SM grouped into the three generations and their properties.

Fermions	Generation			El. Charge	Weak Isospin	Color
	I	II	III			
Leptons	ν_e	ν_μ	ν_τ	0	$+\frac{1}{2}$	0
	e	μ	τ	$-e$	$-\frac{1}{2}$	
Quarks	u	c	t	$+\frac{2}{3}e$	$+\frac{1}{2}$	r,g,b
	d	s	b	$-\frac{1}{3}e$	$-\frac{1}{2}$	

The bosons of the SM are summarized in Tab. 2.2 together with the interaction they mediate, their spin quantum number, and their mass. The Higgs boson does not mediate any interaction; instead, it is a consequence of the Higgs mechanism, which explains the masses of the W and Z bosons, and the masses of fermions, excluding neutrinos, through Yukawa interactions (cf. Sec. 2.1.4).

Table 2.2: The elementary bosons of the SM including their spin, mass, and the interaction they mediate.

Boson	Interaction	Spin	Mass [GeV]
Photon γ	electromagnetic	1	0
Z^0	weak	1	91.1876 ± 0.0021 [24]
$W^\pm$		1	80.379 ± 0.012 [24]
8 Gluons g	strong	1	0
Higgs boson H	-	0	125.25 ± 0.17 [24]

In the following sections the individual interactions are described in more detail including the role of their corresponding bosons.

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) describes the strong interaction between quarks mediated by gluons. It is based on the non-abelian $SU(3)_C$ symmetry group, introduces the color charge, and describes interactions between color charge carriers. The color charge, "(anti-)red", "(anti-)green", and "(anti-)blue", was introduced to explain the observation of the Δ^{++} baryon [25–27]. The measurement suggested a bound state of three up quarks (uuu), which would violate Pauli's principle. The color charge, different for each up quark, introduces a new degree of freedom for the wave function, which restores its antisymmetry.

The Lagrangian for a free quark is given by Eq. 2.2 for a fermion respecting three color

charges:

$$\mathcal{L}_{\text{free quark}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi, \tag{2.10}$$

$$= \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi, \tag{2.11}$$

$$\text{with the color triplet} \quad \psi = \begin{pmatrix} q_a \\ q_b \\ q_c \end{pmatrix}. \tag{2.12}$$

This Lagrangian possesses global symmetry but not local symmetry. Demanding local gauge invariance yields the following Lagrangian:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}\left(i\gamma^{\mu}D_{\mu} - m\right)\psi - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}, \tag{2.13}$$

$$\text{with} \quad D_{\mu} = \partial_{\mu} - ig_s t_a G^{a}_{\mu}, \tag{2.14}$$

$$G^{a}_{\mu\nu} = \partial_{\mu}G^{a}_{\nu} - \partial_{\nu}G^{a}_{\mu} + g_s f^{abc}G^{b}_{\mu}G^{c}_{\nu}. \tag{2.15}$$

Here, g_s denotes a dimensionless real parameter, G^{a}_{μ} ($a = 1\dots 8$) the eight massless gluon fields, t_a the eight generators of the $SU(3)_C$ symmetry group, D_{μ} the covariant derivative, and $G^{a}_{\mu\nu}$ the gluon field strength tensor. The generators t_a do not commute, hence their commutation is defined with the help of the antisymmetric structure constants f^{abc} as follows:

$$[t_a, t_b] = if^{abc}t_c. \tag{2.16}$$

The QCD Lagrangian (Eq. 2.13), looks very similar to the QED Lagrangian (Eq. 2.6). In addition to the quark-gluon interaction, the non-abelian structure of the $SU(3)_C$ symmetry group gives rise to trilinear and quartic self-interaction vertices for gluons. These do not exist for photons in QED. The corresponding interaction Feynman diagrams are shown in Fig. 2.1 including the relevant terms of the QCD Lagrangian (Eq. 2.13).

At a typical momentum transfer scale Q^2 of the LHC, in the GeV^2 range, corrections to the quark and gluon propagators, exemplarily depicted in Fig. 2.2, are considered within perturbation theory in order to calculate physical quantities, which depend on the probability amplitudes of the involved physics processes. In these calculations, ultraviolet divergences occur, requiring a procedure called "renormalization" [28] to be applied, which introduces an arbitrary cutoff parameter μ_R. In order to avoid dependencies of physical observables on such arbitrary parameters, the coupling constant has to be "running", which means that it depends on Q^2 and is defined as follows for 1-loop corrections:

$$\alpha_s\left(Q^2\right) = \frac{4\pi}{\beta_0 \ln\left(Q^2/\Lambda^2\right)}, \tag{2.17}$$

$$\text{with} \quad \beta_0 = 11 - \frac{2}{3}n_f, \tag{2.18}$$

where Λ denotes the QCD scale parameter, and n_f is the number of quark flavors active at the considered energy scale. The most commonly used renormalization procedure is the minimal subtraction scheme ($\overline{\text{MS}}$), which defines the QCD scale parameter to be $\Lambda = 0.34\,\text{GeV}$. The dependence on the energy scale (Eq. 2.17) has a peculiar feature, where the coupling constant vanishes for large Q^2, called "asymptotic freedom": $\alpha_s\left(Q^2\right) \to 0$ for $Q^2 \to \infty$.

(a) ggg $\propto g_s f^{abc} G_\mu^a G_\nu^b \partial^\mu G^{c,\nu}$.

(b) gggg $\propto g_s^2 f^{abc} f^{ade} G_\mu^b G_\nu^c G^{d,\mu} G^{e,\nu}$.

(c) qgq $\propto g_s \bar{\psi} t_a \gamma^\mu G_\mu^a \psi$
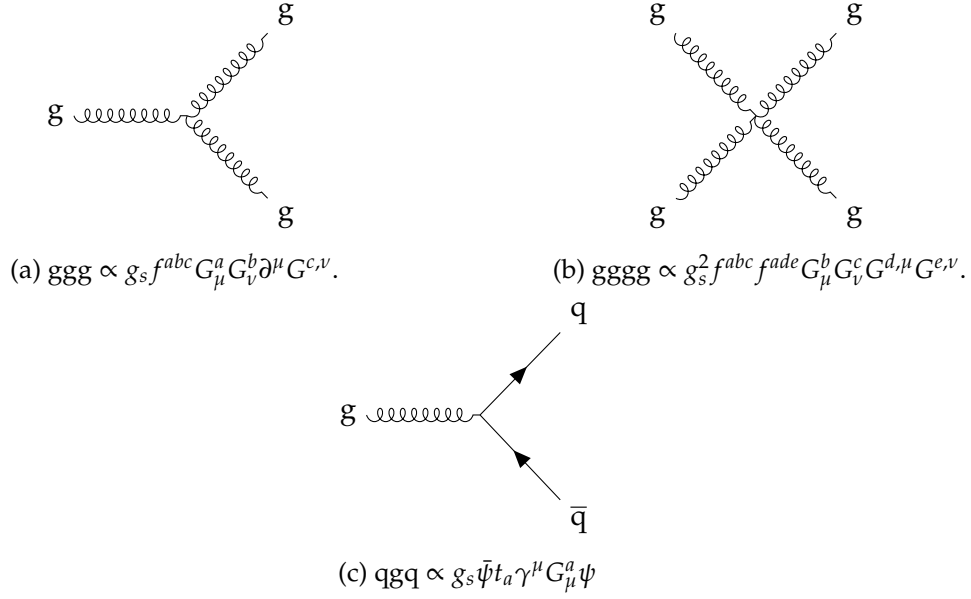
Figure 2.1: Feynman diagrams for trilinear (2.1a) and quartic (2.1b) gluon self interactions, and the quark gluon (2.1c) interaction.

(a) 1-loop gluon propagator correction.

(b) 1-loop quark propagator correction.
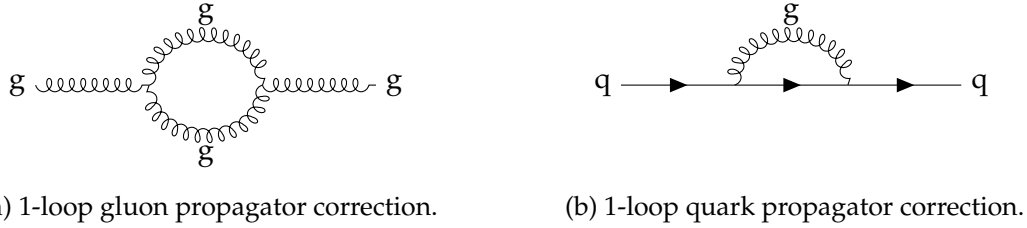
Figure 2.2: Exemplary Feynman diagrams for 1-loop corrections to the gluon and quark propagators used in perturbation theory.

Vice versa the coupling constant increases for small Q^2: $\alpha_s \left(Q^2 \right) \to \infty$ for $Q^2 \to \Lambda^2$. Close to the value of the QCD scale Λ, the coupling α_s diverges; at the same time, the QCD potential has a linear dependency on the distance:

$$V_{\text{QCD}}(r) = -\frac{a}{r} + kr, \tag{2.19}$$

$$\Rightarrow V_{\text{QCD}}(r) \to \infty \quad \text{for} \quad r \to \infty. \tag{2.20}$$

Thus, the energy increases when separating two quarks, leading to a phenomenon called "color confinement", where it becomes energetically more favorable, for large r, to create new quark anti-quark pairs. This phenomenon prohibits the existence of free quarks in nature, instead they exist in color neutral bound states, called "hadrons".

At high energy accelerators, such as the LHC, quarks are initially produced with large Q^2, leading to "free" quarks for a certain distance of flight. After that distance, color confinement prevails. This transition is called "hadronization" and leads to the formation of multiple collimated bound states in a so-called "jet".

2.1.3 Electroweak Theory

The electromagnetic and weak interaction (EWK) has been successfully unified by Glashow, Salam, and Weinberg [5, 29, 30]. It is based on the $SU(2)_L \times U(1)_Y$ symmetry group. While the electromagnetic interaction describes the interaction between electrically charged particles through photons, the weak interaction describes the transition of fermions through the exchange of W and Z bosons. The third component of the weak isospin I_3, a generator of the $SU(2)_L$ symmetry group, is the quantum number describing these transitions. Here, the chirality of the fermions is essential since right-handed fermions do not carry any weak isospin and thus do not couple to the W boson. An example of this transition, including the I_3 quantum number, is shown in a Feynman diagram of a muon decay in Fig. 2.3.
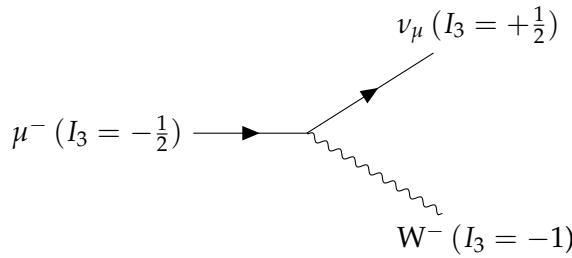
Figure 2.3: Feynman diagram of a muon decay through a W boson. The corresponding third component of the weak isospin (I_3) is annotated for each particle.

The generator of the $U(1)_Y$ group - the weak hypercharge Y - and I_3 can be used to construct a relation to the electrical charge Q:

$$Q = I_3 + \frac{1}{2}Y. \tag{2.21}$$

As described before, right-handed fermions do not couple to the W boson, which allows for arranging the fermions fields as doublets (left-handed, L) and singlets (right-handed, R) depending on their chirality:

$$\psi_{L,\text{leptons}} = \left\{ \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \right\}; \psi_{L,\text{quarks}} = \left\{ \begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \right\}, \tag{2.22}$$

$$\psi_{R,\text{leptons}} = \{e, \mu, \tau\} \qquad ; \ \psi_{R,\text{quarks}} = \{u, d, c, s, b, t\}. \tag{2.23}$$

In the following, the left-handed fermion spinor field doublets will be denoted as ψ_L, and the right-handed singlets as ψ_R:

$$\psi_L^j \quad \text{with} \quad j = 1 \ldots \left| \psi_{L,\text{leptons}} \cup \psi_{L,\text{quarks}} \right|, \tag{2.24}$$

$$\psi_R^k \quad \text{with} \quad k = 1 \ldots \left| \psi_{R,\text{leptons}} \cup \psi_{R,\text{quarks}} \right|, \tag{2.25}$$

where $|A \cup B|$ describes the length of the union of the finite sets A and B. Then, the Lagrangian of the EWK theory is given by the following equation assuming the intrinsic sum-

mation over all participating fermion fields (j, k):

$$\mathcal{L}_{\text{EWK}} = \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{gauge}} \tag{2.26}$$

$$= \underbrace{i\bar{\psi}_L^j \gamma^\mu D_\mu^L \psi_L^j}_{j\,\text{doublets}} + \underbrace{i\bar{\psi}_R^k \gamma^\mu D_\mu^R \psi_R^k}_{k\,\text{singlets}} + \mathcal{L}_{\text{gauge}} \tag{2.27}$$

$$= \underbrace{i\bar{\psi}_L^j \gamma^\mu D_\mu^L \psi_L^j + i\bar{\psi}_R^k \gamma^\mu D_\mu^R \psi_R^k}_{\mathcal{L}_{\text{fermion}}} \underbrace{- \frac{1}{4}W^{i\mu\nu}W_{i\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}}_{\mathcal{L}_{\text{gauge}}}, \tag{2.28}$$

$$\text{with} \quad D_\mu^{L,R} = \partial_\mu + ig I_i^{L,R} W_\mu^i + ig' \frac{Y}{2} B_\mu, \tag{2.29}$$

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g\epsilon^{ijk} W_\mu^j W_\nu^k, \tag{2.30}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.31}$$

where γ^μ are the Dirac matrices, $D_\mu^{L,R}$ is the covariant derivative, W_μ^i, B_μ are the gauge fields, g, g' are real dimensionless parameters, I is the isospin ($I_i^R = 0$ for right-handed singlets), and Y the hypercharge.

The physical fields ($W^\pm_\mu$, A_μ, Z_μ) of the W_μ^i and B_μ gauge fields are obtained through a linear combination and a rotation using the weak mixing angle (θ_W):

$$W^\pm_\mu = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp W_\mu^2 \right), \tag{2.32}$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \cdot \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}. \tag{2.33}$$

The weak mixing angle, also called the "Weinberg" angle, can be expressed through the two real dimensionless parameters g and g':

$$\sin^2\theta_W = \frac{g^2}{g^2 + g'^2}, \tag{2.34}$$

which has been measured by multiple collaborations [31] at the Large Electron-Positron collider (LEP) to be $\sin^2\theta_W = 0.23153 \pm 0.00016$.

The Lagrangian in Eq. 2.26 does not contain mass terms, as they would violate the local gauge symmetry. Instead, the missing mass terms are recovered by the Higgs mechanism, which is explained in the following section.

2.1.4 Higgs Sector

In order to retain mass terms for the W and Z bosons, the Higgs mechanism has been introduced by Peter Higgs [1, 2], Robert Brout, and François Englert [3], Gerald S. Guralnik, Carl R. Hagen, and Tom W. B. Kibble [4]. It extends the EWK Lagrangian (Eq. 2.26) by an additional gauge invariant term, namely the Higgs Lagrangian, which introduces four new real scalar fields ϕ_i:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu\phi)^\dagger (D_\mu\phi) - V(\phi), \tag{2.35}$$

$$\text{with} \quad \phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.36}$$

where D_μ corresponds to the covariant derivative introduced in Eq. 2.26 with hypercharge $Y = 1$ and weak isospin $I = \frac{1}{2}$. The Higgs potential $V(\phi)$ is defined as:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi \right)^2, \tag{2.37}$$

where μ^2 and λ are real parameters. These parameters are chosen to satisfy the following conditions, which leads to an infinite number of non-trivial minima ($|\phi_{\min}|$) of the potential $V(\phi)$:

$$\mu^2 < 0 \quad \text{and} \quad \lambda > 0, \tag{2.38}$$

$$\frac{\partial V}{\partial \phi^\dagger \phi} = \mu^2 + 2\lambda \left(\phi^\dagger \phi \right) \overset{!}{=} 0, \tag{2.39}$$

$$\text{with} \quad \left(\phi^\dagger \phi \right) = |\phi_{\min}|^2 \quad \text{follows} \tag{2.40}$$

$$\to |\phi_{\min}| = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}}, \tag{2.41}$$

where v denotes the vacuum expectation value of the scalar field. In unitary gauge three of the four fields ϕ_i can be chosen to be:

$$\phi_1 = \phi_2 = \phi_4 = 0 \tag{2.42}$$

in order to reduce the degrees of freedom. The third field ϕ_3 is chosen to be a linear expansion by the real Higgs field (H) around the minimum of the potential in radial direction. The Higgs potential can then be written as follows:

$$\phi_3 = v + \text{H} \tag{2.43}$$

$$\to \phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \text{H} \end{pmatrix}, \tag{2.44}$$

which leads to the Higgs potential:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi \right)^2 \tag{2.45}$$

$$= \mu^2 \left(\frac{1}{\sqrt{2}} (v + \text{H}) \right)^2 + \lambda \left(\frac{1}{\sqrt{2}} (v + \text{H}) \right)^4 \tag{2.46}$$

$$\text{with} \quad \mu^2 = -\lambda v^2 \quad \text{follows} \tag{2.47}$$

$$\Rightarrow V(\phi) = -\frac{\lambda}{2} v^2 \left(v^2 + 2v\text{H} + \text{H}^2 \right) + \frac{\lambda}{4} \left(v^4 + 4v^3\text{H} + 6v^2\text{H}^2 + 4v\text{H}^3 + \text{H}^4 \right) \tag{2.48}$$

$$= -\frac{\lambda}{2} v^4 - \lambda v^3 \text{H} - \frac{\lambda}{2} v^2 \text{H}^2 + \frac{\lambda}{4} v^4 + \lambda v^3 \text{H} + \frac{3\lambda}{2} v^2 \text{H}^2 + \lambda v \text{H}^3 + \frac{\lambda}{4} \text{H}^4 \tag{2.49}$$

$$= -\frac{\lambda}{2} v^4 + \lambda v^2 \text{H}^2 + \lambda v \text{H}^3 + \frac{\lambda}{4} \text{H}^4. \tag{2.50}$$

Figure 2.4 shows the potential $V(\phi)$, including the position of the minimum and thus the meaning of the vacuum expectation value for this choice.

The expanded Higgs potential (Eq. 2.50) consecutes interesting aspects of the Higgs field. First of all, a mass term arises for the Higgs boson of the form $m_h = \sqrt{2\lambda} v$. Second, self-interaction terms arise for trilinear ($\propto \text{H}^3$) and quartic ($\propto \text{H}^4$) interactions of the Higgs field.
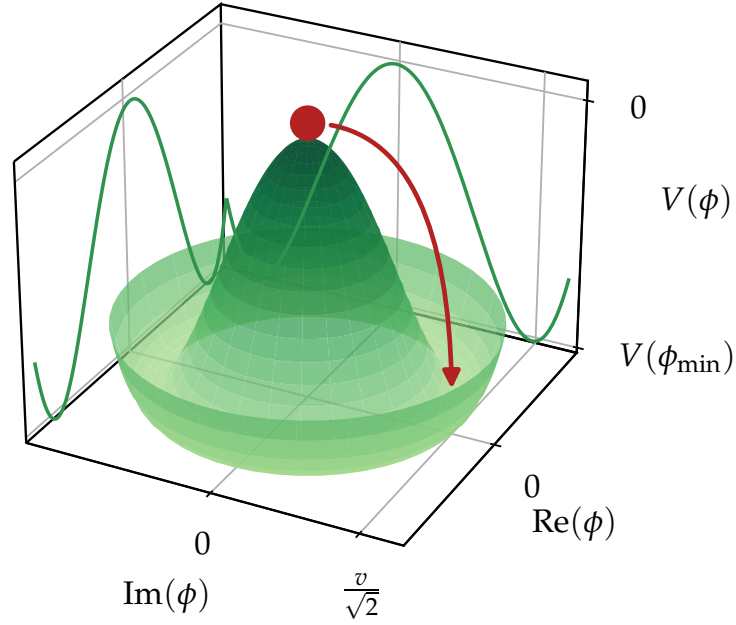
Figure 2.4: A sketch of the Higgs potential $V(\phi)$, including the new groundstate of ϕ at the vacuum expectation value v. The spontaneous symmetry breaking is depicted with the red arrow, which shows the state (red sphere) transition to a new ground state.

(a) Trilinear Higgs self-interaction.

(b) Quartic Higgs self-interaction.
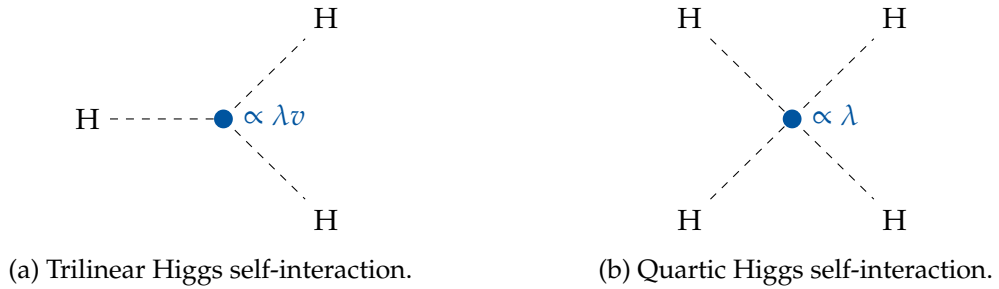
Figure 2.5: Feynman diagrams of the trilinear and quartic Higgs self-interactions including their coupling strength.

Figure 2.5 depicts the two corresponding Feynman diagrams with their corresponding coupling strength. In addition, the mass terms of the W and Z bosons arise from the kinematic term of the Higgs lagrangian (Eq. 2.35), by inserting the Higgs field (Eq. 2.43), and the defi-

nition of the covariant derivative (Eq. 2.26):

$$\mathcal{L}_{\text{Higgs, kin.}} = \left(D^\mu \phi\right)^\dagger \left(D_\mu \phi\right) = |D^\mu \phi|^2 \tag{2.51}$$

$$= \left| D^\mu \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \mathrm{H} \end{pmatrix} \right|^2 \tag{2.52}$$

$$= \left| \left(\partial_\mu + ig\frac{\sigma_i}{2}W_\mu^i + ig'\frac{1}{2}B_\mu\right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \mathrm{H} \end{pmatrix} \right|^2 \tag{2.53}$$

$$= \left| \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \partial_\mu \mathrm{H} \end{pmatrix} + \frac{i}{\sqrt{8}} \begin{pmatrix} gW_\mu^1 - igW_\mu^2 \\ -gW_\mu^3 + g'B_\mu \end{pmatrix} (v + \mathrm{H}) \right|^2 \tag{2.54}$$

$$= \frac{1}{2} \left(\partial_\mu \mathrm{H}\right)^2 + (v + \mathrm{H})^2 \left(\frac{g^2}{8} \left(W_\mu^1 W^{1,\mu} + W_\mu^2 W^{2,\mu}\right) + \frac{1}{8} \left(-gW_\mu^3 + g'B_\mu\right)^2 \right). \tag{2.55}$$

The definitions of the physical W and Z bosons (Eq. 2.32 and Eq. 2.33) replace the abstract fields $(W_\mu^1, W_\mu^2, W_\mu^3, B_\mu)$ in the last two terms of $\mathcal{L}_{\text{Higgs, kin.}}$ (Eq. 2.55) as follows:

W bosons:

$$\mathrm{W}^\pm{}_\mu = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp W_\mu^2\right) \tag{2.56}$$

$$\Rightarrow W_\mu^1 W^{1,\mu} + W_\mu^2 W^{2,\mu} = \mathrm{W}^+{}_\mu \mathrm{W}^{+\mu} + \mathrm{W}^-{}_\mu \mathrm{W}^{-\mu} \tag{2.57}$$

Z boson and photon:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \cdot \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{2.58}$$

$$\text{with} \quad \sin^2\theta_W = \frac{g^2}{g^2 + g'^2} \quad \text{and} \tag{2.59}$$

$$\cos^2\theta_W = \frac{g'^2}{g^2 + g'^2} \quad \text{follows} \tag{2.60}$$

$$\Rightarrow -gW_\mu^3 + g'B_\mu = -\sqrt{g^2 + g'^2}Z_\mu + 0 \cdot A_\mu \tag{2.61}$$

Applying these substitutions, the kinematic term of the Higgs lagrangian can be rewritten as

follows:

$$\mathcal{L}_{\text{Higgs, kin.}} = \frac{1}{2}\left(\partial_\mu H\right)^2 + \frac{1}{2}\underbrace{\left(\frac{1}{2}gv\right)^2}_{m_W^2}\left(W^+{}_\mu W^{+\mu} + W^-{}_\mu W^{-\mu}\right) \tag{2.62}$$

$$+ \frac{1}{2}\underbrace{\left(\frac{1}{2}\sqrt{g^2 + g'^2}v\right)^2}_{m_Z^2} Z_\mu Z^\mu + 0 \cdot A_\mu A^\mu \tag{2.63}$$

$$+ \underbrace{\frac{m_Z^2}{v}H Z_\mu Z^\mu + 2\frac{m_W^2}{v}H W^+{}_\mu W^{-\mu}}_{\text{HVV}} \tag{2.64}$$

$$+ \underbrace{\frac{m_Z^2}{2v^2}H^2 Z_\mu Z^\mu + \frac{m_W^2}{v^2}H^2 W^+{}_\mu W^{-\mu}}_{\text{HHVV}}. \tag{2.65}$$

The mass terms for the W and Z bosons are retained, while the Higgs mechanism generates no mass term for the photon. In addition, interaction terms for the W and Z bosons with the Higgs boson are obtained as depicted in Fig. 2.6.

(a) HVV interaction.   (b) HHVV interaction.

Figure 2.6: Feynman diagrams of the HVV and HHVV interactions.

The value of the vacuum expectation value can be indirectly calculated from the mass of the W boson [32], which can also be expressed in terms of the Fermi constant G_F:

$$\text{Higgs mechanism:}\quad m_W^2 = \left(\frac{1}{2}gv\right)^2 \tag{2.66}$$

$$\text{Fermi theory:}\quad m_W^2 = \frac{\sqrt{2}g^2}{8G_F} \tag{2.67}$$

$$\Rightarrow v = \left(\sqrt{2}G_F\right)^{-\frac{1}{2}} \simeq 246.22\,\text{GeV}, \tag{2.68}$$

where G_F can be precisely inferred from muon lifetime measurements [33].

Fermion masses are not described so far. An additional lagrangian describing Yukawa interactions, i.e., interactions between the fermion fields and the Higgs field, is introduced [5]:

$$\mathcal{L}_{\text{fermion}} = -\lambda_f\left(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L\right), \tag{2.69}$$

where λ_f denotes the Yukawa coupling for a fermion. Using the previously defined scalar field, including the real Higgs field (Eq. 2.43) and exemplary the four spinor fields for bottom

quarks (left-handed doublets and right-handed singlets), the Yukawa interaction to bottom quarks reads:

$$\mathcal{L}_{\mathrm{b}} = -\frac{1}{\sqrt{2}}\lambda_{\mathrm{b}}\left[\left(\bar{\mathrm{t}}\ \ \bar{\mathrm{b}}\right)_L \begin{pmatrix} 0 \\ v+\mathrm{H} \end{pmatrix} \mathrm{b}_R + \bar{\mathrm{b}}_R \left(0\ \ v+\mathrm{H}\right)\begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix}_L\right] \tag{2.70}$$

$$= -\frac{1}{\sqrt{2}}\lambda_{\mathrm{b}}\left(v+\mathrm{H}\right)\underbrace{\left[\bar{\mathrm{b}}_L\mathrm{b}_R + \bar{\mathrm{b}}_R\mathrm{b}_L\right]}_{\bar{\mathrm{b}}\mathrm{b}} \tag{2.71}$$

$$= -\underbrace{\frac{1}{\sqrt{2}}\lambda_{\mathrm{b}}v\,\bar{\mathrm{b}}\mathrm{b}}_{m_{\mathrm{b}}} - \underbrace{\frac{1}{\sqrt{2}}\lambda_{\mathrm{b}}}_{m_{\mathrm{b}}/v\,\propto\,m_{\mathrm{b}}}\ \mathrm{H}\bar{\mathrm{b}}\mathrm{b}. \tag{2.72}$$

Here, the mass term of the bottom quark arises, which is proportional to the Yukawa coupling for bottom quarks λ_{b}. In addition, an interaction term is introduced with a coupling proportional to the bottom quark mass. This mechanism is applied in the same way for all fermions (except neutrinos), although for up-type fermions the conjugate of the Higgs doublet ($\tilde{\phi}$) has to be used:

$$\tilde{\phi} = \frac{1}{\sqrt{2}}\begin{pmatrix} v+\mathrm{H} \\ 0 \end{pmatrix} \quad \text{and} \quad \tilde{\phi}^{\dagger} = \frac{1}{\sqrt{2}}\left(v+\mathrm{H}\ \ 0\right). \tag{2.73}$$

Finally, the lagrangian for all fermions, except neutrinos, can be written in the following form (assuming summation over all fermions):

$$\mathcal{L}_{\mathrm{fermion}} = -m_{\mathrm{f}}\bar{\psi}\psi - \frac{m_{\mathrm{f}}}{v}\mathrm{H}\bar{\psi}\psi. \tag{2.74}$$

The coupling of the Higgs boson to two fermions is depicted in the Feynman diagram in Fig. 2.7. While the coupling between the Higgs boson and the Z and W bosons depends
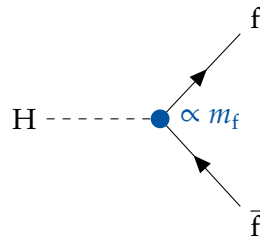
Figure 2.7: Feynman diagram of the coupling between the Higgs boson and two fermions including the coupling strength.

quadratically on the mass of the vector boson, the coupling to fermions has a linear dependency on the mass of the fermion.

2.2 Higgs Boson Pair Production

The study of non-resonant Higgs boson pair production gives access to the self-coupling λ through the Higgs boson self-interactions, which arise after the EWK symmetry breaking. A measurement of this self-coupling helps to unveil the structure of the Higgs potential,

which may have multiple theoretical implications. First and foremost, this measurement is a stress test of the Higgs mechanism and interactions at the electroweak scale. In addition, the structure of the Higgs potential may have cosmological consequences, such as the vacuum stability of the universe [34].

Three different states are possible for a post-inflationary (late) universe: a *stable*, a *metastable*, or an *unstable* universe. Figure 2.8 shows the stability diagram, including the three different states, parametrized by the top quark and Higgs boson masses. Currently, a *metastable* universe is favored, although a *stable* universe is still possible within the experimental and theoretical uncertainties on the masses (68%, 95%, and 99% contour ellipsis). An *unstable* universe is excluded, given the existence of our universe.
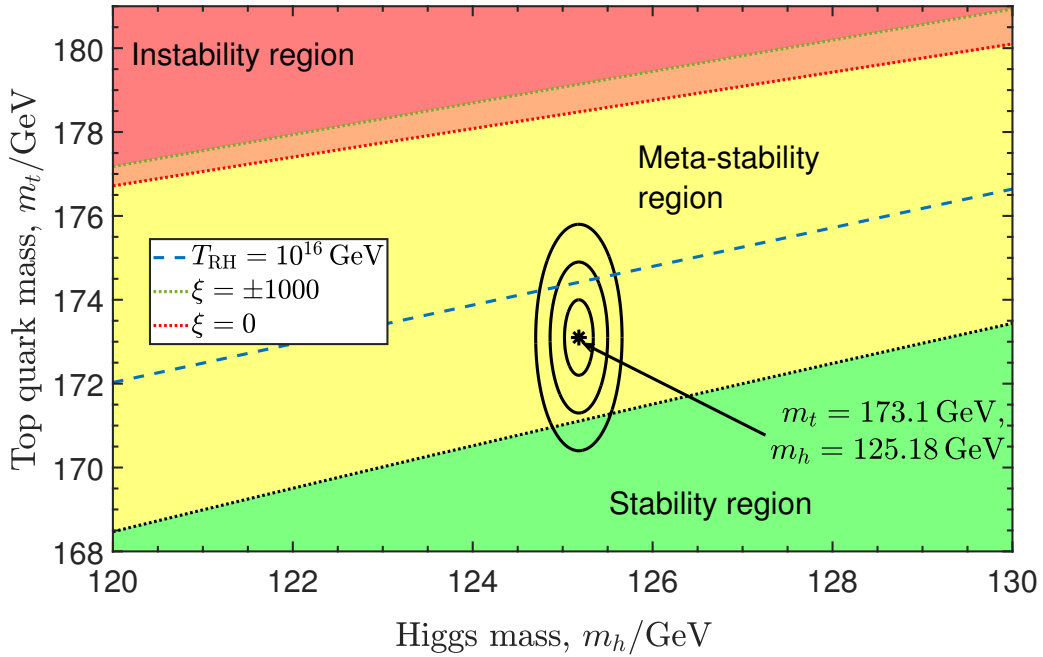
Figure 2.8: Stability diagram of vacuum stability of the standard model [34] parametrized by the top quark and Higgs boson pole masses. The black star marks the state of the universe, including the 68%, 95%, and 99% uncertainty contours. The instability boundary is shown for different curvature coupling parameters (red and green dotted line), and for a reheat temperature of $T_{\mathrm{RH}} = 10^{16}\,\mathrm{GeV}$ (blue dashed line).

The corresponding structures of the Higgs potential for these three stability categories are shown in Fig. 2.9. Additional higher order terms in the Higgs potential, such as a dimension-6 term of the form $\propto \left(\phi^\dagger \phi\right)^3$, allow for *metastable* and *unstable* scenarios. These terms occur in effective theories, which incorporate new physics at a high mass scale Λ (ultraviolet completion of the SM) [35, 36].

For the currently favored *metastable* state, a global (true) minimum exists in the shape of the Higgs potential into which the electroweak vacuum may decay. This so-called *false vacuum decay* may happen through quantum tunneling or classically if there is sufficient energy in the system. A new global ground state likely imposes new laws of physics.

In addition, the measurement of the Higgs boson pair production and its self-coupling may
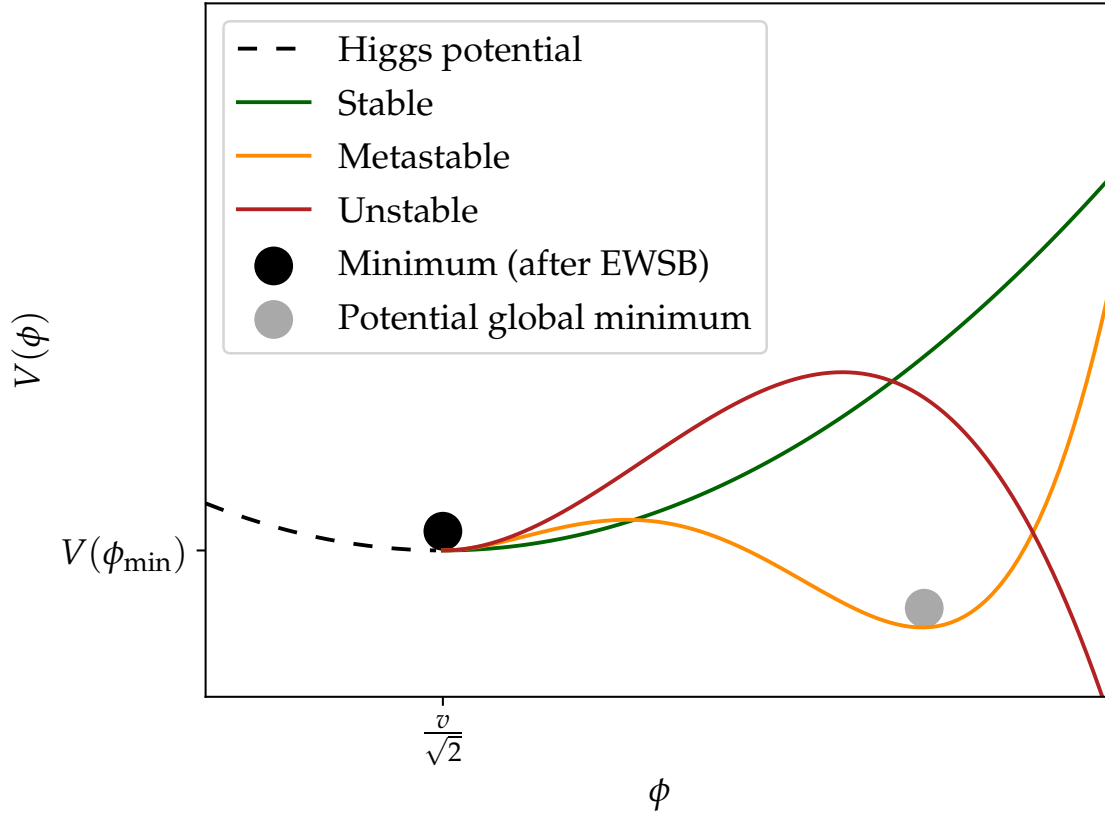
Figure 2.9: A sketch of the three different stability categories (*stable*, *metastable*, and *unstable*) of Higgs potential in the two dimensional plane. The black ball marks the current ground state, which has been reached after electroweak symmetry breaking (EWSB). The dark grey ball shows the ground state for a potential global minimum in the *metastable* scenario.

shed light on the dynamics of the electroweak phase transition [37], the production of primordial black holes [38], and the formation of new particles through a resonant Higgs boson pair production with a radion or a Kaluza-Klein graviton [39].

This thesis focuses on the non-resonant production of Higgs boson pairs, the measurement of the Higgs boson self-coupling, and the coupling between two vector bosons and two Higgs bosons.

2.2.1 Proton Proton Collisions at High Energies

Circular hadron colliders, such as the LHC, are designed to accelerate particle beams in the opposite direction until a collision occurs. In contrast to a linear collider, a circular collider has the advantage that the energy can be ramped up continuously until the desired center-of-mass energy is reached. However, the energy loss from Bremsstrahlung has to be taken into account, thus making heavy particles a preferred choice as the energy loss is proportional to $(E/m)^4$.

At the LHC, protons and ions are accelerated, allowing for a much higher center-of-mass energy than its predecessor, the Large Electron-Positron Collider (LEP), which accelerated

electrons and positrons. The latter is suitable for performing high-precision measurements since electrons and positrons are elementary particles. However, the energy losses due to synchroton radiation limited the beam energy heavily with a peak beam energy of around 104 GeV [40]. Protons are composite particles of two up valence quarks, a down valence quark, and a sea of quark anti-quark pairs and gluons. When two protons collide, two constituents (partons) interact with each other. These partons carry only a fraction of the total proton momentum, the so-called Bjorken scaling variable x, described by parton distribution functions (PDFs). Figure 2.10 shows the parton distribution functions of the NNPDF3.1 [41] set for two different momentum transfer scales Q^2.
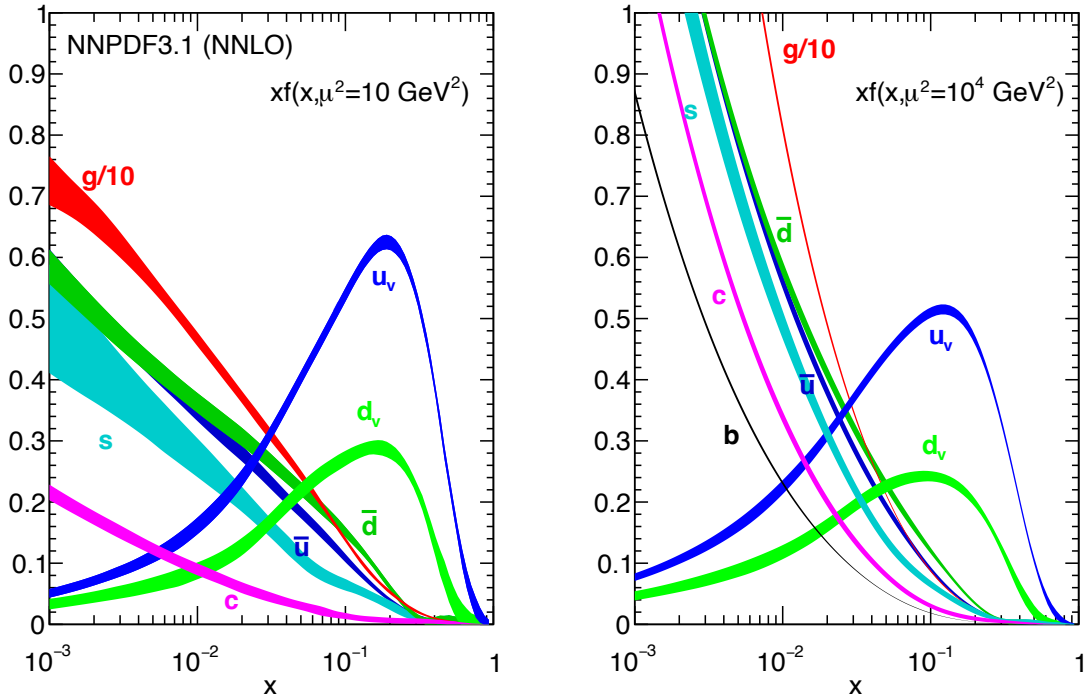
Figure 2.10: Parton distribution function for the NNPDF3.1 [41] set as a function of the momentum fraction x. The PDF is shown for two different momentum transfers: low momentum transfer of $Q^2 = 10 \, \mathrm{GeV}^2$ (left) and high momentum transfer of $Q^2 = 10^4 \, \mathrm{GeV}^2$. The gluon distribution is scaled down by a factor of 10 for visibility.

The PDF is also significantly different for various parton flavors, especially for the u_v and d_v valence quarks compared to the other partons from the sea. For high momentum transfer scales, Q^2, the PDFs for sea quarks and gluons become more relevant. Especially gluons, exceeding the PDFs of the valence quarks, become the dominant parton to participate in the interaction. In addition, at the LHC, the center-of-mass energy exceeds the typical energies of EWK physics (e.g., the mass of the Z boson), leading to most interactions involving low momentum fractions x of the partons. Thus, gluon-gluon interactions are the dominant source of most SM processes, including the Higgs bosons production at the LHC. PDFs cannot be calculated analytically; they have to be measured in experiments, such as the HERA experiment [42, 43].

The PDFs are essential to calculate cross sections at the LHC. The calculation of cross sections can be factorized according to the QCD factorization theorem [44] at the factorization scale μ_F into short (asymptotic freedom) and long (confinement) distance parton interactions. The pp $\rightarrow$ ab production cross section is calculated as follows:

$$\sigma_{\text{pp}\rightarrow\text{ab}} = \sum_{i,j} \underbrace{f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2)}_{\text{long dist.}} \cdot \underbrace{\hat{\sigma}_{ij\rightarrow\text{ab}}(\mu_F^2, \mu_R^2 \, \hat{s}, \alpha_s)}_{\text{short dist.}} dx_1 dx_2, \tag{2.75}$$

where $\hat{s}$ denotes the center-of-mass energy of both partons, and μ_R the scale of the renormalization. The functions $f_{i,j}$ are the PDFs of both protons, which describe the parton dynamics for long distance parton interactions. The partonic cross section $\hat{\sigma}_{ij\rightarrow\text{ab}}$ describes the asymptotic freedom of partons. The partonic cross sections can be calculated directly with matrix elements. Additional QCD multijet vertices (higher order corrections) add terms proportional to α_s, which can be calculated with the help of perturbation theory as the value of α_s is sufficiently small. However, for small values of Q^2 ("soft interaction") the value of α_s diverges, which makes the application of perturbation theory invalid. In these cases, PDFs are used to describe the parton dynamics.

2.2.2 Production Modes in the Standard Model

At the LHC, Higgs boson pairs are mainly produced - similarly to the single Higgs production - by four different production modes: gluon-gluon fusion (HH(GGF)), vector boson fusion (HH(VBF)), double Higgsstrahlung (VHH), and top quark associated double Higgs boson production (t$\bar{\text{t}}$HH). This analysis considers the first two production modes, the predominant ones at the LHC. The leading-order Feynman diagrams of the HH(GGF) and HH(VBF) production modes are shown in Fig. 2.11. Here, three couplings are accessible in the two production modes: κ_λ and $\kappa_{\text{t,b}}$ in the HH(GGF) production mode, and κ_λ, κ_{2V}, and κ_V in the HH(VBF) production mode. These vertices have been previously shown in Fig 2.5 (κ_λ), Fig. 2.7 ($\kappa_{\text{t,b}}$), and Fig. 2.6 (κ_{2V} and κ_V). Each coupling parameter κ_i is defined by the so-called "kappa framework" [45–47] as the coupling strength value of the Higgs bosons to particles i divided by their corresponding SM values. The SM scenario is defined as $\kappa_i = 1$.

The cross section of these production modes is shown in Fig. 2.12 as a function of the self-coupling for a center-of-mass energy of 14 TeV at leading and next-to-leading order.

At the SM scenario ($\kappa_i = 1$) the production cross sections for the HH(GGF) and HH(VBF) production modes at next-to-next-to-leading order amount to [49]:

$$\sigma_{\text{HH(GGF)}}^{\text{SM}} = 31.05^{+6\%}_{-23\%} \quad (\text{QCD scale} + m_t) \pm 3.0\% \, (\text{PDF} + \alpha_s) \text{ fb}, \tag{2.76}$$

$$\sigma_{\text{HH(VBF)}}^{\text{SM}} = 1.726^{+0.03\%}_{-0.04\%} (\text{QCD scale}) \qquad \pm 2.1\% \, (\text{PDF} + \alpha_s) \text{ fb}. \tag{2.77}$$

These cross sections are about three orders of magnitude smaller than the corresponding single Higgs boson production ones. Figure 2.12 shows a quadratic dependency of the cross section (and its uncertainty bands) on the self-coupling (black lines and bands). Thus, the cross section and its scale variations can be parametrized in the following way for the HH(GGF)
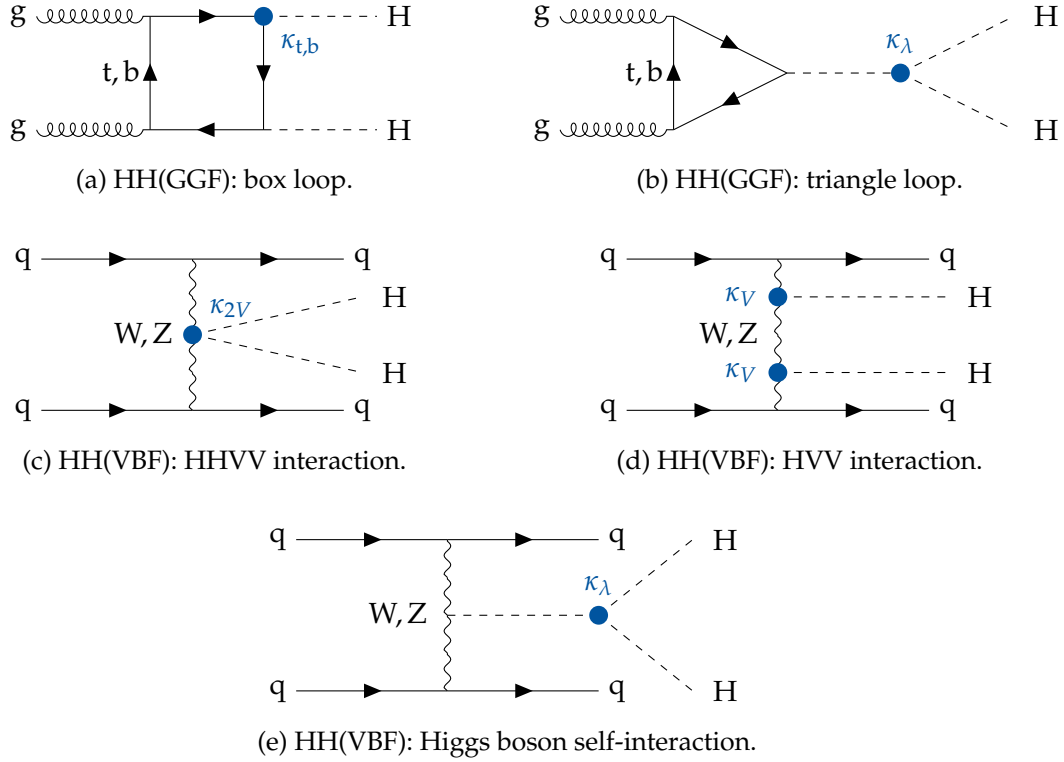
(a) HH(GGF): box loop.

(b) HH(GGF): triangle loop.

(c) HH(VBF): HHVV interaction.

(d) HH(VBF): HVV interaction.

(e) HH(VBF): Higgs boson self-interaction.

Figure 2.11: Leading-order Feynman diagrams for the two dominant Higgs pair production modes at the LHC. Here, multiple couplings ($\kappa_{t,b}$, κ_λ, κ_{2V}, κ_V) are accessible denoted by the blue dots.

production mode at next-to-next-to-leading order [49]:

$$\sigma_{\text{HH(GGF)}}(\kappa_\lambda) = 70.3874 - 50.4111 \cdot \kappa_\lambda + 11.0595 \cdot \kappa_\lambda^2 \text{ fb,} \tag{2.78}$$

$$\sigma_{\text{HH(GGF)}}^{+1\sigma}(\kappa_\lambda) = \max(76.6075 - 56.4818 \cdot \kappa_\lambda + 12.635 \cdot \kappa_\lambda^2, \tag{2.79}$$

$$75.4617 - 56.3164 \cdot \kappa_\lambda + 12.7135 \cdot \kappa_\lambda^2) \text{ fb,} \tag{2.80}$$

$$\sigma_{\text{HH(GGF)}}^{-1\sigma}(\kappa_\lambda) = \min(57.6809 - 42.9905 \cdot \kappa_\lambda + 9.58474 \cdot \kappa_\lambda^2, \tag{2.81}$$

$$58.3769 - 43.9657 \cdot \kappa_\lambda + 9.87094 \cdot \kappa_\lambda^2) \text{ fb.} \tag{2.82}$$

Variations of the PDF uncertainties are found to be independent of the self-coupling.

The triangle (Fig. 2.11b) and box (Fig. 2.11a) diagrams of the HH(GGF) production mode interfere destructively with each other. This effect is visible in Fig. 2.12, where a minimum of the black line is located at $\kappa_\lambda = 2.45$ - the maximal destructive interference. The SM scenario is close to this minimum and thus leads to a strongly reduced cross section. The interference effect on the total cross section depends on the strength of the self-coupling as it dictates the relevance of the triangle diagram. The effect of the interference is shown in Fig. 2.13 [50].

Here, the differential cross section at leading order is shown as a function of the invariant mass of the Higgs boson pair. The black line denotes the total distribution, which is the sum of the triangle diagram (red), the box diagram (blue), and their interference (green). The interference reduces the total cross section by about 50% compared to the box-only contribution.

In addition, the interference of the triangle and box diagrams of the HH(GGF) production
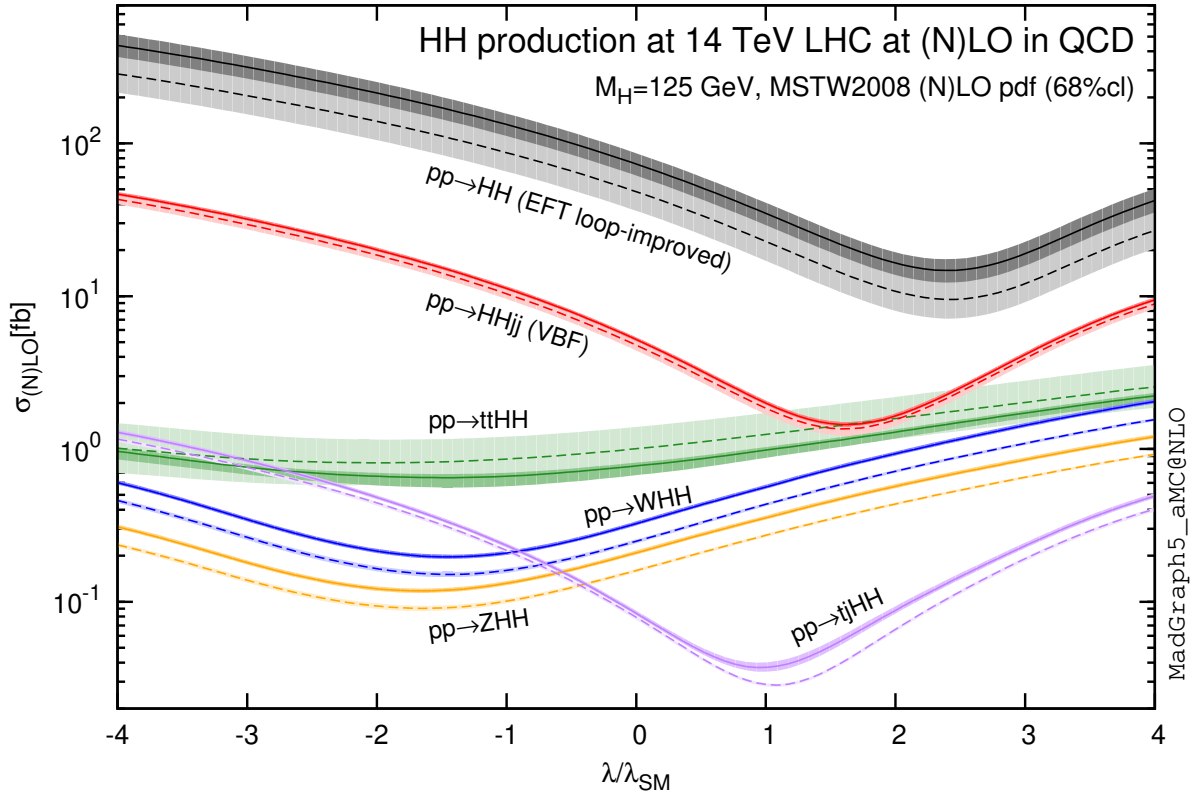
Figure 2.12: Higgs pair production cross section as a function of the self-coupling for a center-of-mass energy of 14 TeV at next-to-leading order [48]. The dashed (solid) lines and light (dark) color bands correspond to the leading order (next-to-leading order) cross sections and to the scale and PDF uncertainties added linearly. The SM scenario can be seen at $\kappa_\lambda = \lambda/\lambda_{SM} = 1$.

mode has kinematic consequences. The triangle diagram involves two Higgs bosons originating, produced at rest from the same mother particle: a strongly off-shell produced Higgs boson. This diminishes the contribution of the triangle diagram significantly since the width of the Higgs boson is very small, and thus the production of a strongly off-shell Higgs boson quite improbable. On the other hand, the box diagram does not share this characteristic feature, as both Higgs boson pairs are produced through the top quark loop. This leads to a broad spectrum, often involving Higgs bosons with large momenta. Thus, the interference effect on the kinematic distribution also depends on the self-coupling's strength. This effect is shown in Fig. 2.14 [50].

It can be seen that a characteristic minimum is present for $\kappa_\lambda = 2.45$ (Fig. 2.14a) at the maximal destructive interference. In addition, the more the triangle diagram contribution is enhanced ($\kappa_\lambda < -1$ and $\kappa_\lambda > 5$) the more pronounced the low-m_{HH} region is.

Varying κ_λ also has effects on the single Higgs boson production and branching fractions through electroweak corrections [51, 52]. These effects are taken into account in the measurements of this analysis.
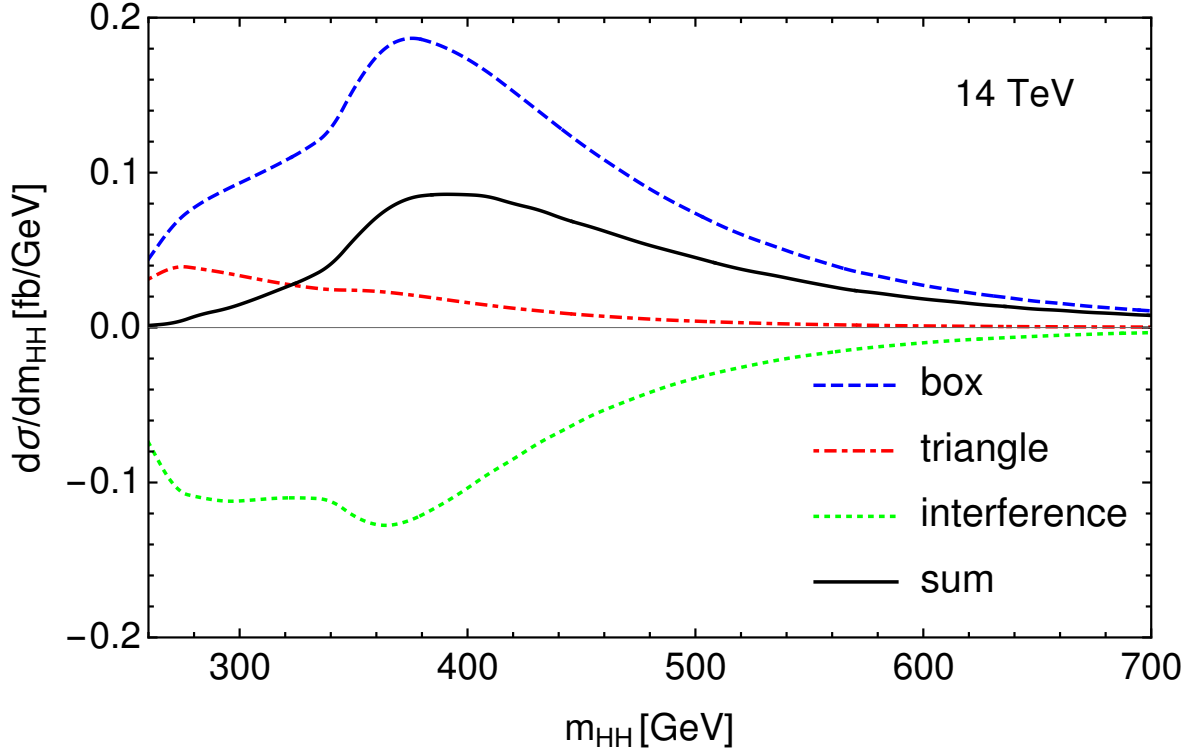
Figure 2.13: Interference of the triangle and box diagrams of the HH(GGF) production mode at leading order [50] shown in the invariant mass spectrum of the Higgs boson pair.
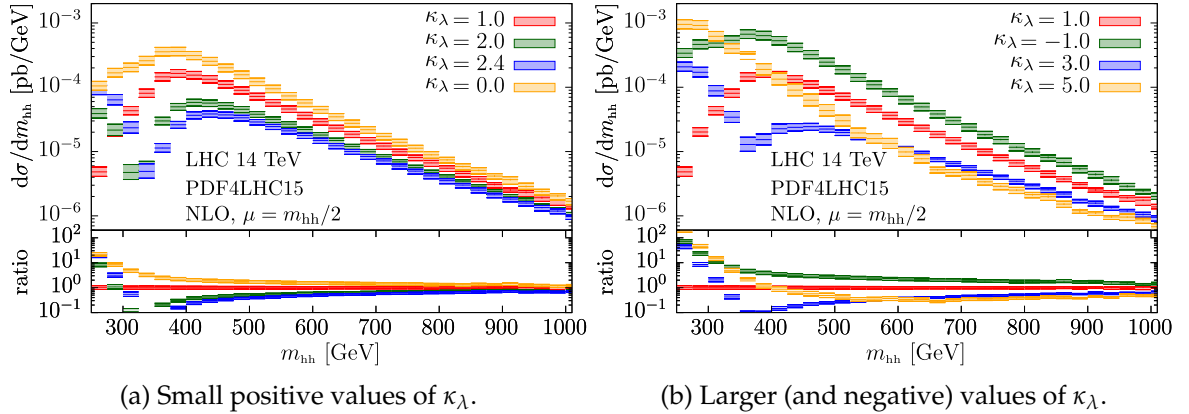
(a) Small positive values of κ_λ.

(b) Larger (and negative) values of κ_λ.

Figure 2.14: Invariant mass spectrum of two Higgs bosons at 14 TeV shown for different values of the self-coupling [50].

2.2.3 Characteristics of the Dileptonic $HH \rightarrow b\bar{b}W^+W^-$ Process

The $HH \rightarrow b\bar{b}W^+W^-$ process has several unique characteristics. First and foremost, it has a relatively large branching fraction especially due to one Higgs boson decaying into a pair of b-quarks. In addition, the Higgs boson decays via two W bosons fully leptonically yielding a clean event signature. The fully leptonic decay of $H \rightarrow W^+W^-$ shares another unique feature: the spin correlation of both W bosons resulting in a narrow opening angle between the two leptons. These characteristics are described in the first part of this section.

Afterwards, all background processes are discussed. In addition, their cross sections are given.

Finally, an overview of all physics processes considered in this analysis is shown.

Higgs Boson Decay Modes

The signal process of the analysis presented in this thesis comprises two Higgs boson decays. One into a pair of bottom quarks, and the other into two W bosons, which further decay leptonically. The branching fractions ($\mathcal{B}$) of all Higgs boson pair decay modes are visualized in Fig. 2.15.
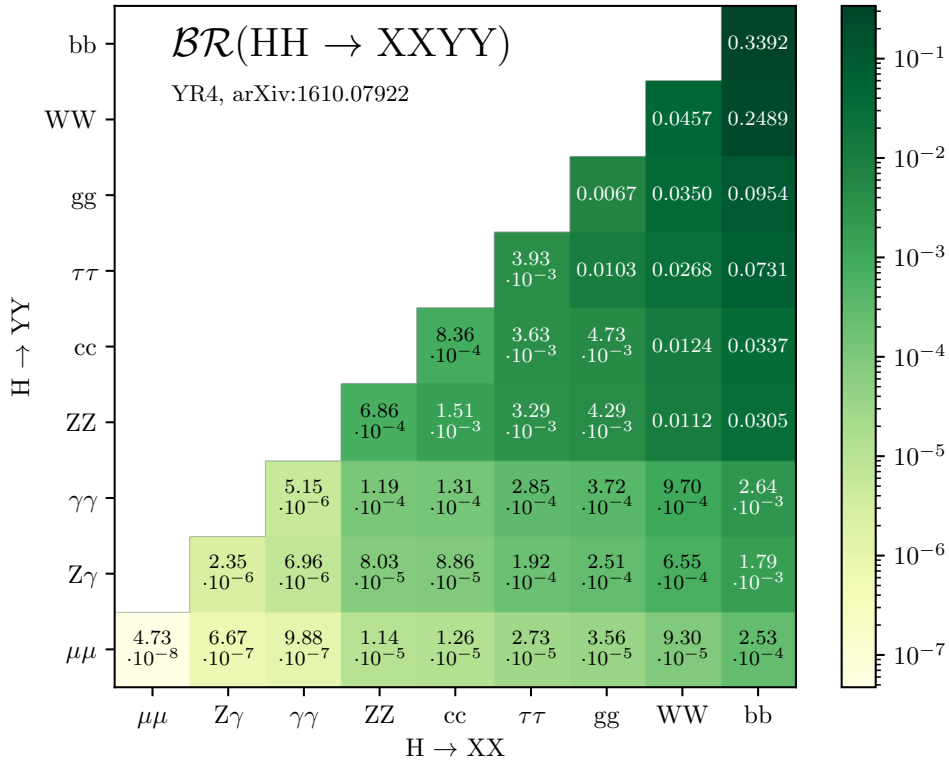
Figure 2.15: Higgs boson pair decay modes.

The decay into two bottom quarks has the largest branching fraction of about 58.24%. The second highest branching fraction has the decay to two W bosons with around 21.37%. Each W boson further decays to a lepton and neutrino, labeled W_ℓ from now on, with a branching fraction of around 32.59%. These branching fractions finally result in a total branching fraction of approximately 2.64% for the $\mathrm{HH} \to \mathrm{b\bar{b}W^+_\ell W^-_\ell}$ signal process. For comparison, the branching fractions of the inclusive $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$ process, i.e., not only leptonic W boson decays, is shown in Fig. 2.15. With a branching fraction of 24.89% it is almost ten times larger than the fully leptonic decay mode.

The fully leptonic $\mathrm{H} \to \mathrm{W^+W^-}$ decay has the benefit of a narrow opening angle between the direction of flight of the two leptons. This effect is shown in Fig. 2.16.
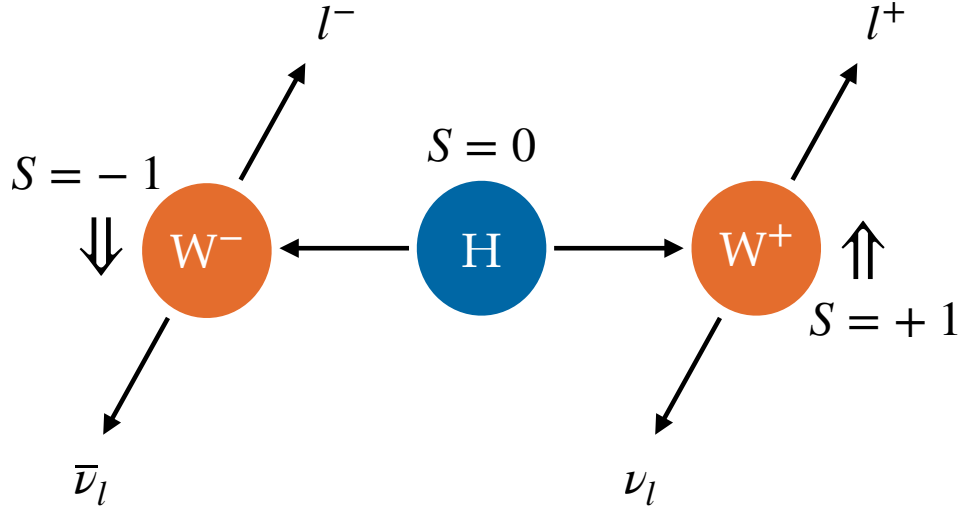
Figure 2.16: Sketch of the spin correlation in the fully leptonic $H \rightarrow W^+W^-$ process.

The W bosons have a spin in opposite direction ($S = \pm 1$) as the Higgs boson has a spin of zero. In the left decay branch of Fig. 2.16 it can be seen that the W^- decays into a charged lepton and a anti-neutrino. Both decay products carry a spin of $-\frac{1}{2}$, since the W^- boson has a spin of -1. The weak interaction enforces the coupling to a left-handed charged lepton, i.e., its spin points into the opposite direction of the direction of flight, and a right-handed anti-neutrino, i.e., its spin points into the same direction of flight. The same holds for the other (W^+) decay branch, but vice-versa. Therefore, both leptons are emitted into the same direction of flight resulting in a narrow opening angle between them. Likewise, the argument applies to both neutrinos. This effect is characteristic for the leptonic $H \rightarrow W^+W^-$ decay mode and is used to distinguish it from other physics processes in this thesis.

The $H \rightarrow b\bar{b}$ decay mode has the benefit of a large branching fraction. Combined with the characteristic signature of the decay to two W_ℓ bosons, the $H \rightarrow b\bar{b}W^+_\ell W^-_\ell$ decay mode is a competitive channel for the search for Higgs boson pair production at the LHC. Finally, events from $HH \rightarrow b\bar{b}ZZ$ and $HH \rightarrow b\bar{b}\tau^+\tau^-$ decays are considered as well although their contribution is significantly smaller.

Background Processes

Background processes in the presented analysis arise from physics processes that mimic the signal process, and thus survive the event selection. Often they comprise the same or a similar final state as the signal process. This section lists all relevant background processes

including their cross sections to this analysis. Unless uncertainties on the cross sections are given in this section, they are discussed in Sec. 8.1.2 and given in Tab. 8.3 or are negligible.

The most dominant background processes are the $t\bar{t}$ + Jets and single t-quark production. The t-quark decays via a W boson and a b-quark, and thus results for the $t\bar{t}$ + Jets process in the identical final state as the signal process. The predicted $t\bar{t}$ + Jets production cross section is $\sigma_{t\bar{t}\,+\,\text{Jets}} = 831.76^{+19.77}_{-29.20}(\text{scale})^{+35.06}_{-35.06}(\text{PDF} + \alpha_s)$ pb as calculated with the Top++2.0 program to NNLO order in perturbative QCD, including soft-gluon resummation to NNLL order (see [53] and references therein), and assuming a top-quark mass $m_\tau = 172.5\,\text{GeV}$. The first uncertainty comes from the independent variation of the factorization and renormalization scales, μ_F and μ_R, while the second one is associated to variations in the PDF and α_s, following the PDF4LHC prescription with the MSTW2008 68% CL NNLO order, CT10 NNLO order and NNPDF2.3 5f FFN PDF sets (see [54] and references therein, and [55–57]). The cross section of single t-quark production is split into three channels: t-channel, s-channel, and tW-channel. For the first two channels the cross section is further split by the charge of the t-quark. Their cross sections are given in Tab. 2.3 for a center-of-mass energy of 13 TeV.

Table 2.3: Single t-quark production cross sections at 13 TeV center-of-mass energy and $m_\tau = 172.5\,\text{GeV}$.

Process	σ (t-quark) $[\text{pb}]$	σ ($\bar{\text{t}}$-quark) $[\text{pb}]$
t-channel	$136.02^{+4.09}_{-2.92}(\text{scale}) \pm 3.52(\text{PDF} + \alpha_s)$	$80.95^{+2.53}_{-1.71}(\text{scale}) \pm 3.18(\text{PDF} + \alpha_s)$
s-channel	$6.35^{+0.18}_{-0.15}(\text{scale}) \pm 0.14(\text{PDF} + \alpha_s)$	$3.97^{+0.11}_{-0.09}(\text{scale}) \pm 0.15(\text{PDF} + \alpha_s)$
tW-channel	$71.7 \pm 1.80(\text{scale}) \pm 3.40(\text{PDF})$	

The single t-quark production cross sections are calculated for a top quark mass of 172.5 GeV at next-to-leading order (NLO) in QCD with Hathor v2.1 [58, 59]. The PDF and α_s uncertainties are calculated using the PDF4LHC prescription [60] with the MSTW2008 68% CL NLO [61, 62], CT10 NLO [63] and NNPDF2.3 [64] PDF sets, added in quadrature to the scale uncertainty. The single t-quark events may pass the event selection in presence of additional emissions and event misreconstructions.

The next largest background is the Drell-Yan process. It has two leptons in its final state and may look similar to the signal process in presence of two additional jets. This process is estimated with a data-driven procedure as explained in Sec. 6.6.

Backgrounds arising from misreconstructed jets as leptons are called Misid. leptons. They are estimated with a data-driven procedure as explained in Sec. 6.5. The misreconstruction of a jet as a lepton may occur in any physics process. All events, that are subject to such misreconstructions, constitute the Misid. leptons background.

Physics processes involving the production of multiple bosons are called "VV(V)" (V = $\{W, Z, \gamma\}$). They contribute as background in case the bosons decay to two leptons and two b-quarks. An overview of all relevant VV(V) processes is given in Tab. 2.4.

Standard model single Higgs boson processes are another set of background processes. Although their cross sections are significantly smaller compared to the other background processes, they are still several orders of magnitude larger than the one of the inclusive signal process. The cross sections [66] of all relevant single Higgs boson processes are given in Tab. 2.5 for a center-of-mass energy of 13 TeV and $m_H = 125\,\text{GeV}$.

Table 2.4: Relevant VV(V) processes with their production cross sections [65] at 13 TeV center-of-mass energy. Uncertainties on these cross sections are given in Tab. 8.3.

Process	$\sigma\,[\mathrm{pb}]$	Accuracy
$W\gamma$	656.40	NLO
$Z\gamma$	179.49	NLO
WW	118.7	NNLO
WZ ($2\ell2q$, $3\ell\nu$)	10.03	NLO
ZZ ($2\ell2q$)	5.52	NLO
VVV	0.48456	NLO

Table 2.5: Single Higgs boson production cross sections [66] at 13 TeV center-of-mass energy and considering $m_H = 125\,\mathrm{GeV}$. In addition, the accuracy in QCD and EWK corrections are given.

Process	$\sigma\,[\mathrm{pb}]$ with relative uncertainties	QCD Acc.	EWK Acc.
H(GGF)	$48.58 \pm 3.9\%(\text{th. Gauss}) \pm 1.9\%(\text{PDF}) \pm 2.6\%(\alpha_s)$	N3LO	NLO
H(VBF)	$3.782^{+0.4\%}_{-0.3\%}(\text{scale}) \pm 2.1\%(\text{PDF}) \pm 0.5\%(\alpha_s)$	NNLO	NLO
WH	$1.373^{+0.5\%}_{-0.7\%}(\text{scale}) \pm 1.7\%(\text{PDF}) \pm 0.9\%(\alpha_s)$	NNLO	NLO
ZH	$0.8839^{+3.8\%}_{-3.1\%}(\text{scale}) \pm 1.3\%(\text{PDF}) \pm 0.9\%(\alpha_s)$	NNLO	NLO
$t\bar{t}$H	$0.5071^{+5.8\%}_{-9.2\%}(\text{scale}) \pm 3.0\%(\text{PDF}) \pm 2.0\%(\alpha_s)$	NLO	NLO
tHq	$0.07425^{+6.5\%}_{-14.9\%}(\text{scale}) \pm 3.5\%(\text{PDF}) \pm 1.2\%(\alpha_s)$	NLO	NLO
tHW	$0.01517^{+4.9\%}_{-6.7\%}(\text{scale}) \pm 6.1\%(\text{PDF}) \pm 1.5\%(\alpha_s)$	NLO	NLO

Even though the single Higgs boson processes are considered as background processes, they are sensitive through higher order EWK corrections to the Higgs boson self-coupling κ_λ. Exemplary Feynman diagrams for these corrections are shown in Fig. 2.17.
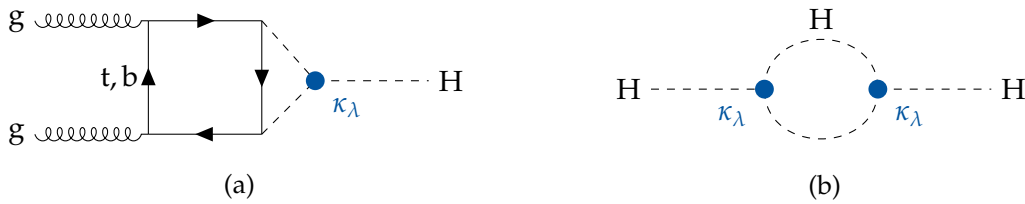
Figure 2.17: Feynman diagrams of exemplary EWK corrections to single Higgs boson processes involving the self-coupling κ_λ.

The effect of EWK corrections of single Higgs boson processes on the self-coupling κ_λ is taken into account in the presented analysis.

Finally, several other small background processes are considered and grouped together in this analysis. These include among others $t\bar{t}$ + Jets production in association with one or two more bosons V = $\{W, Z, \gamma\}$ and the production of two $t\bar{t}$ pairs. A full list of these background processes are listed in Tab. 2.6.

Table 2.6: Relevant remaining background processes with their production cross sections [65] at 13 TeV center-of-mass energy. Uncertainties are negligible in the scope of this analysis if none are given.

Process	$\sigma\,[\mathrm{pb}]$ with relative uncertainties	Accuracy
$t\bar{t}W$ ($1\ell\nu$, 2q)	$0.6001^{+12.9\%}_{-11.5\%}(\mathrm{scale}) \pm 2.0\%(\mathrm{PDF}) \pm 2.70(\alpha_s)$	NLO
$t\bar{t}Z$ (2ℓ, 2q)	$0.9504^{+9.6\%}_{-11.3\%}(\mathrm{scale}) \pm 2.8\%(\mathrm{PDF}) \pm 2.78(\alpha_s)$	NLO
$t\bar{t}W^+W^-$	0.006981	NLO
$t\bar{t}\gamma$	4.215	NLO
$t\gamma$	1.018	NLO
tZq	0.0735	NLO
$t\bar{t}t\bar{t}$	0.008213	NLO
W^+W^+ + Jets	0.04926	NLO

Overview of All Relevant Physics Processes

A summary of all physics processes considered for this analysis is given in Tab. 2.7.

Table 2.7: All physics processes considered for this analysis, including their estimation method and color scheme.

Process	Sub-process(es)	Estimation method	Color
HH(GGF)	$b\bar{b}W^+W^-$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}ZZ$	simulation (sim.)	
HH(VBF)	$b\bar{b}W^+W^-$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}ZZ$	simulation	
Top	$t\bar{t}$ + Jets, single t-quark	simulation	
Drell-Yan + Multiboson	Drell-Yan, VV(V)	data-driven, sim.	
Misid. leptons	Misid. leptons	data-driven	
Other	single H, remaining bkgs.	simulation	

The processes are partly grouped together based on similarities, such as their rate, event topologies, and kinematic distributions. Additionally, Tab. 2.7 lists their estimation method and the color scheme used in the figures throughout this thesis. The first two processes are the two different signal production modes. Afterwards, all background processes are listed sorted by their contribution to the presented analysis.

2.2.4 Overview of Related Publications

Three publications from the ATLAS and CMS collaborations using proton-proton collisions at the LHC at $\sqrt{s} = 13$ TeV are related to the presented analysis. Two publications are searches for Higgs boson pair production in the HH $\rightarrow$ $b\bar{b}W^+W^-$ decay mode. One is the predecessor to the presented analysis in this thesis by the CMS collaboration [67] using data recorded in the 2016 data-taking period. This analysis analyzes the dileptonic final state of the HH $\rightarrow$ $b\bar{b}W^+W^-$ decay mode. The other publication is from the ATLAS collaboration [68] using data of the Run 2 data-taking period. It considers one Higgs boson decay via a pair of bottom quarks and one via a pair of W/Z bosons or a pair of tau leptons to two leptons (e, μ). The third publication summarizes and combines the current Higgs boson pair production analysis results from the CMS collaboration [69] using data recorded in the Run

2 data-taking period. It does not include the results of the $HH \to b\bar{b}W^+W^-$ decay channel. An overview of all results is given in Tab. 2.8.

Table 2.8: Overview of results from publications of the CMS and ATLAS collaborations.

Collaboration	$\mathcal{L}\,[\,\mathrm{fb}^{-1}]$	Analysis	Obs. (Exp.) Limit on μ	Ref.
CMS	35.9	$HH \to b\bar{b}\ell\nu\ell\nu$	79 (89)	[67]
ATLAS	139	$HH \to b\bar{b}\ell\nu\ell\nu$	40 (29)	[68]
		$HH \to b\bar{b}ZZ$	32 (40)	[69]
		$HH \to \mathrm{Multilepton}$	21 (19)	[69]
CMS	137	$HH \to b\bar{b}\gamma\gamma$	8.4 (5.5)	[69]
		$HH \to b\bar{b}\tau\tau$	3.3 (5.2)	[69]
		$HH \to b\bar{b}b\bar{b}$	6.4 (4.0)	[69]
CMS	137	Combined	3.4 (2.5)	[69]

Chapter 3

Experimental Setup

This chapter describes the experimental setup needed for this analysis. First, an overview of the Large Hadron Collider is given with its different collider and fixed-target experiments. Then, the Compact Muon Solenoid detector and its subdetector systems are described in greater detail. Afterwards, this chapter explains how data is recorded and filtered through a multi-stage trigger system. Next, a description of the reconstruction algorithms for different particle types, such as leptons and jets, follows. Finally, the NanoAOD data format and the computing infrastructure used for this analysis are described.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the most powerful human-built particle collider located and operated by the Conseil Européen pour la Recherche Nucléaire (CERN) in Switzerland. It is built inside the tunnel previously used for the Large Electron-Positron collider (LEP) and lies between 45 m and 170 m under the earth's surface. A chain of pre-accelerators exist for the LHC, which have been used for the LEP accelerator as well. Multiple experiments are served by the LHC, of which the four largest are all collider experiments with dedicated entry points. Namely, these are the ALICE [70] (A Large Ion Collider Experiment), ATLAS [71] (A Toroidal LHC Apparatus), CMS [72] (Compact Muon Solenoid), and the LHCb [73] (LHC-beauty) experiments. Additionally, the LHC and its pre-accelerator accelerate particles for multiple fixed-target experiments. A sketch of the accelerator chain and the associated experiments is shown in Fig. 3.1.

At first, protons are obtained from a H_2 source by applying a strong electric field separating protons and electrons. Afterwards, these protons are accelerated by the LINAC2 to the energy of 50 MeV. From there on, the Proton Synchrotron Booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS) increase the energy of the protons to 450 GeV in three stages. Finally, the protons are injected into the LHC ring, accelerating them to the beam energy of 7 TeV with a bunch spacing of 25 ns using radiofrequency cavities. The bunches are kept on their trajectory with 1232 superconducting magnets. Each of these magnets provides a magnetic field of 8.33 T and is operated at less than 2 K. Orbit corrections and beam focusing shortly before the interaction points are achieved using higher-order multipole magnets. At design operation, the LHC stores 2808 proton-proton bunches in each beam at a center-of-mass energy of 14 TeV and can deliver an instantaneous peak luminosity of 10^{34} cm^{-2} s^{-1} for the ATLAS and CMS experiments. The two lower luminosity experiments, LHCb and TOTal Elastic and diffractive cross section Measurement (TOTEM) [75], are served with 10^{32} cm^{-2} s^{-1} and $2 \cdot 10^{29}$ cm^{-2} s^{-1} with 156 bunches respectively. The ALICE experiment is a dedicated ion experiment aiming at an instantaneous peak luminosity of
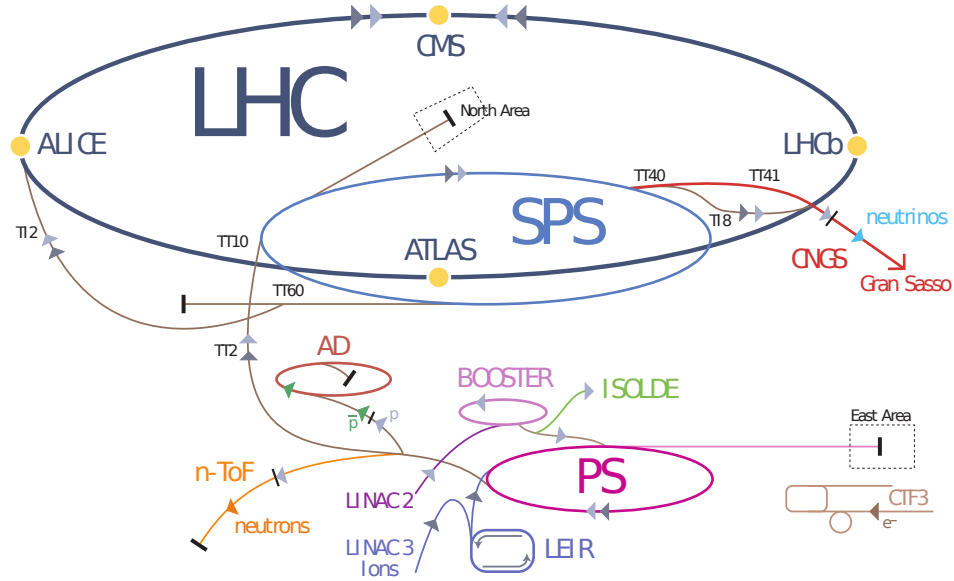
Figure 3.1: Sketch of the complete accelerator chain of the LHC experiment and its associated collider and fixed-target experiments [74].

$10^{27}\,\mathrm{cm^{-2}\,s^{-1}}$ for lead-lead ion collisions and recording collisions at a center-of-mass energy of $5.52\,\mathrm{TeV}$ for lead ions delivered by the LHC. The instantaneous luminosity $\mathcal{L}_{\mathrm{inst.}}$ allows to infer the event rate of a certain physics process with the associated cross section σ:

$$\frac{dN}{dt} = \sigma \mathcal{L}_{\mathrm{inst.}}, \tag{3.1}$$

$$\text{with} \quad \mathcal{L}_{\mathrm{inst.}} = \frac{N_b^2 n_b f_{\mathrm{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} \cdot F \quad \text{for a Gaussian beam distribution.} \tag{3.2}$$

Here, N_b denotes the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta function at the collision point, and F the geometric reduction factor due to the beam inclination at the interaction point. The number of total events can be inferred by integrating formula 3.1 over the measurement time:

$$N = \int \frac{dN}{dt} dt = \sigma \int \mathcal{L}_{\mathrm{inst.}} dt = \sigma \mathcal{L}, \tag{3.3}$$

with the integrated luminosity $\mathcal{L}$.

The LHC has been in operation since 2008 and successfully recorded events containing Higgs bosons leading to its discovery in 2012. The two discovery experiments ATLAS and CMS are multi-purpose experiments, mainly built to discover the Higgs boson and new physics beyond the standard model of particle physics. In particular, both experiments push the frontiers of the Higgs sector, high precision SM measurements, new physics in the dark matter, extra dimensions, and supersymmetric sector. Both experiments provide results with similar competitive precision and thus serve as cross-validation of scientific results for each other. The primary purpose of the LHCb experiment is to explain CP violation by analyzing events

containing b- and c-quarks with a dedicated tracking system. The ALICE experiment focuses on quark-gluon plasma studies in ion collisions only.

During the Run 2 data-taking period (2015-2018) the LHC operated at 13 TeV delivering in total an integrated luminosity of around $164\,\mathrm{fb}^{-1}$, of which the CMS experiment recorded around $150\,\mathrm{fb}^{-1}$. Fig. 3.2 shows the integrated luminosity as a function of the measurement time.
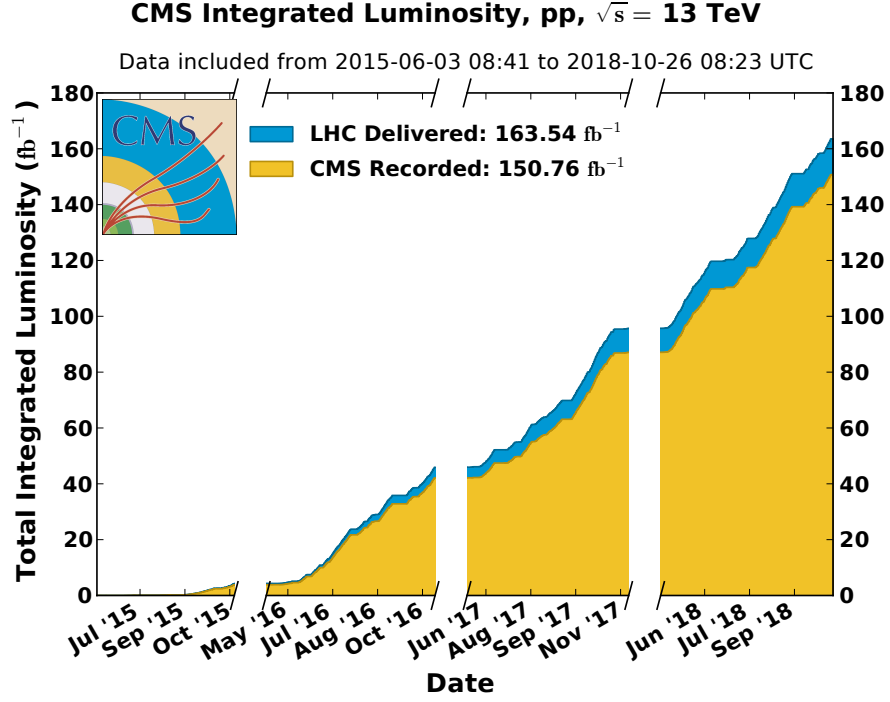
Figure 3.2: Integrated luminosity delivered by the LHC (blue) and recorded by the CMS experiment (yellow) for the Run 2 data-taking period [76].

The analysis presented in this thesis uses data taken between 2016 and 2018. The small integrated luminosity of $4\,\mathrm{fb}^{-1}$ recorded in 2015 is omitted.

3.2 The CMS Experiment

The multi-purpose Compact Muon Solenoid (CMS) experiment has been built near the french city Cessy approximately 100 m underground to investigate physics phenomena at the TeV scale. In particular, the nature of electroweak symmetry breaking induced by the Higgs mechanism, multiple new theories, such as supersymmetry (SUSY) or extra dimensions, and high precision tests of the SM are studied at the CMS experiment. This comprehensive physics program requires a versatile detector setup. Key features of this setup are the ability to accurately measure particle momenta and energies, identify particle types, such as muons or b-quark jets, and reconstruct trajectories of charged particles. These features allowed to observe the Higgs boson in 2012 mainly in the $H \to \gamma\gamma$ and the $H \to ZZ \to 4\ell$ decay. The $H \to \gamma\gamma$ channel heavily profited from the high mass resolution of the di-photon system and an efficient π^0 rejection, while the $H \to ZZ \to 4\ell$ channel mainly benefited from the high mass resolution of the Z boson and the precise identification of e and μ leptons.

A sketch of the CMS detector is depicted in Fig. 3.3. It has an onion-like multi-layer subdetector structure in a cylindrical shape. Its center is located at the interaction point, surrounded by the pixel and strip tracking subdetectors. The subsequent layers are the electromagnetic (ECAL) and hadronic (HCAL) calorimeters, followed by a superconducting magnet of 3.8 T with several iron return yokes. Finally, the muon chambers are installed as the outermost subdetector between the iron return yokes. The cylindrical detector is encapsulated at both sides with two endcaps holding additional muon chambers and a forward calorimeter.
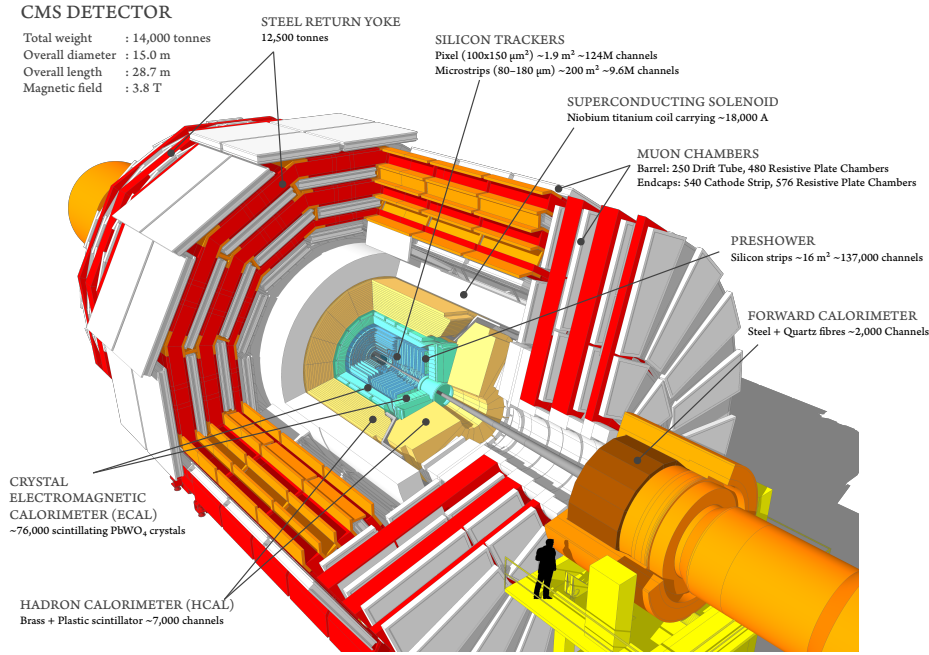
Figure 3.3: Cutaway view of the CMS detector including its subdetector systems [77].

The detector design is described in more detail in the following subsections, starting with its coordinate system, followed by the subdetector systems from the inside to outside based on references [72, 78–81].

3.2.1 Coordinate Conventions

The origin of the coordinate system of the CMS experiment is located at the nominal collision point of the two beams. The x-axis points towards the center of the LHC ring, while the y-axis points towards the earth's surface. The z-axis is aligned counterclockwise along the beam direction. The angular orientation in the x-y plane is described by the azimuthal angle ϕ, whereas the polar angle θ is measured in the z-x plane. In order to retain lorentz invariance for differences in the z-x plane, the pseudorapidity is introduced:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{3.4}$$

The angular separation of two particles can now be computed using the azimuthal angle ϕ and the pseudorapidity η:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{3.5}$$

Due to the energy conservation in the x-y plane, many variables are calculated only with respect to their transverse component, such as the imbalance of energy in the x-y plane called $E_{\mathrm{T}}^{\mathrm{miss}}$.

3.2.2 Tracking System

The innermost subdetector of the CMS detector is the tracker installed closely around the interaction point. It comprises two systems: the silicon pixel and the silicon strip tracking systems. Both systems cover together a region of $|\eta| \leq 2.5$. Within this region, charged particles traverse depletion zones induced by bias voltages, as both tracking systems utilize semiconductor technology. Traversing these induces electron-hole pairs, allowing to measure ionization currents and thus the position of the charged particle for a single pixel or at least two strip modules. Multiple positions can be aggregated and used to reconstruct the helix-shaped trajectory of the charged particle. The radius of curvature of these trajectories is then used to calculate the particle's full three-momentum.

Initially, the silicon pixel tracker consisted of three layers and two disks until the end of the data-taking period in 2016. During the technical stop at the end of 2016, the pixel tracker was upgraded to four layers and three disks in the scope of the phase-1 upgrade plan. The upgraded layout is shown in Fig. 3.4a and Fig. 3.4b. This upgrade allowed maintaining the same (or even better) tracking efficiencies for high luminosities and thus higher pile-up rates up to and exceeding 50. The four layers are distanced from the center at radii 3.0 cm, 6.8 cm, 10.9 cm and 16.0 cm and the three disks at longitudinal positions of ±29.1 cm, ±39.6 cm ±51.6 cm with a radial coverage from 4.5 cm to 16.1 cm. They provide a four-hit coverage for trajectories up to $|\eta| < 2.5$. The tracking efficiency is noticeably reduced at the crossing point of the layers and disks near $|\eta| = 1.5$. In comparison to the original layout the number of pixels has been increased by a factor of 1.6 from 48×10^6 to 79×10^6, each of size $100 \times 150\,\mu\mathrm{m}^2$.

Outside of the silicon pixel tracking system, the silicon strip tracking system is located. The layout of this system is shown in Fig. 3.5. It consists of 15 400 modules divided into two parts in the barrel region, namely the TIB (Tracker Inner Barrel) and the TOB (Tracker Outer Barrel), and two parts in the endcap region, called the TEC (Tracker End Cap) and the TID (Tracker Inner Disks). The TIB and TOB cover in total ten silicon strip layers between radii of 20 cm and 110 cm, the TEC comprises nine layers in the region $120\,\mathrm{cm} < |z| < 280\,\mathrm{cm}$ and the TID three disks, which close the gap between the TIB and TEC. The first two layers of the TIB and TOB are made of modules, which are tilted to each other at an angle of 100 mrad in order to operate in the $r - \phi$ and $r - z$ planes. This results in a single-point resolution for the barrel region of $23 - 34\,\mu\mathrm{m}$ ($35 - 52\,\mu\mathrm{m}$) in the $r - \phi$ direction and $23\,\mu\mathrm{m}$ ($52\,\mu\mathrm{m}$) in z direction for the TIB (TOB). In the endcap region, the first two ring layers of the TIC and TEC and the fifth of the TEC are also "stereo" modules.

3.2.3 Electromagnetic Calorimeter

The homogeneous electromagnetic calorimeter (ECAL) is designed to measure electromagnetic showers, primarily induced by photons and electrons. A cross section view of the ECAL can be found in Fig 3.6. The ECAL consists of lead tungstate (PbWO$_4$) crystals of which 61 200 are installed in the barrel region and 7324 in each endcap. Lead tungstate provides several material benefits, such as a short radiation (Moliere) length of $X_0 = 0.89\,\mathrm{cm}$ (2.2 cm), fast light emittance and radiation hardness. This results in a compact ECAL lay-
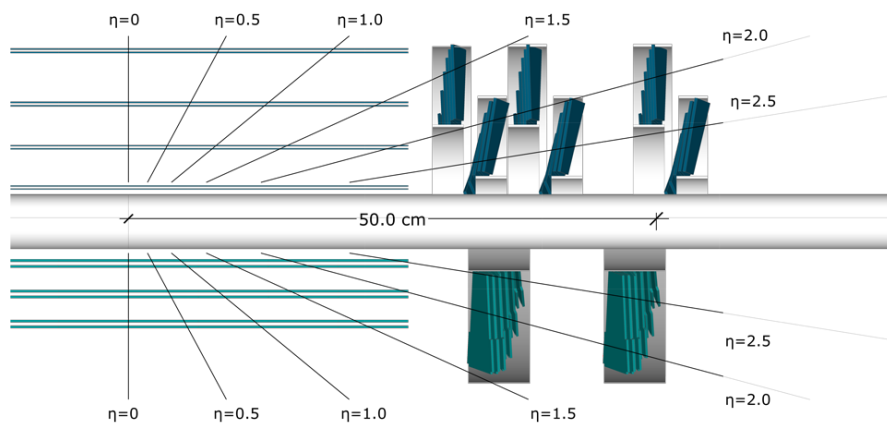
(a) Sketch of the original (bottom) and upgraded (top) pixel tracker [80].
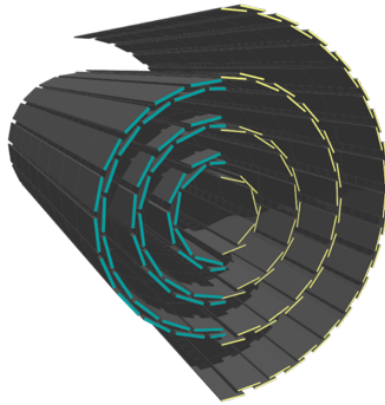
(b) Sketch of the original (left) and upgraded (right) pixel tracker [80].
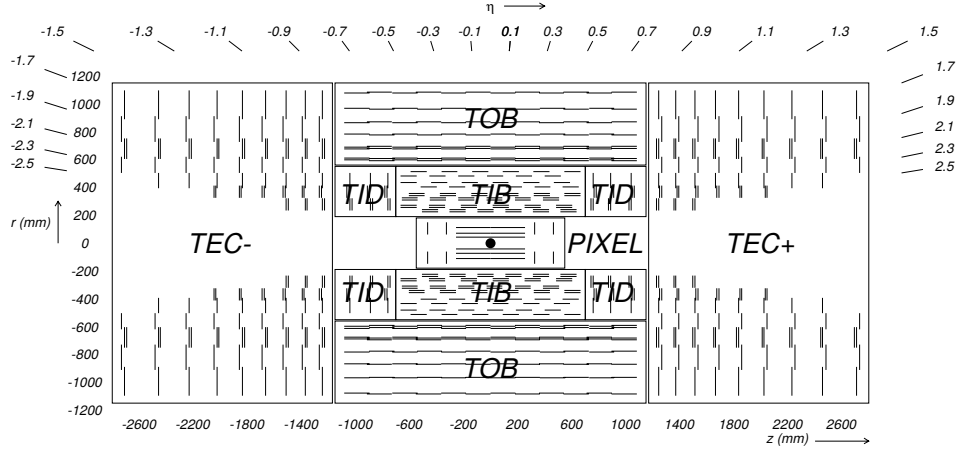
Figure 3.5: Layout of the silicon strip tracking system, composed of the TIB and TOB subsystems in the barrel region and the TID and TEC subsystems in the endcap regions [78].
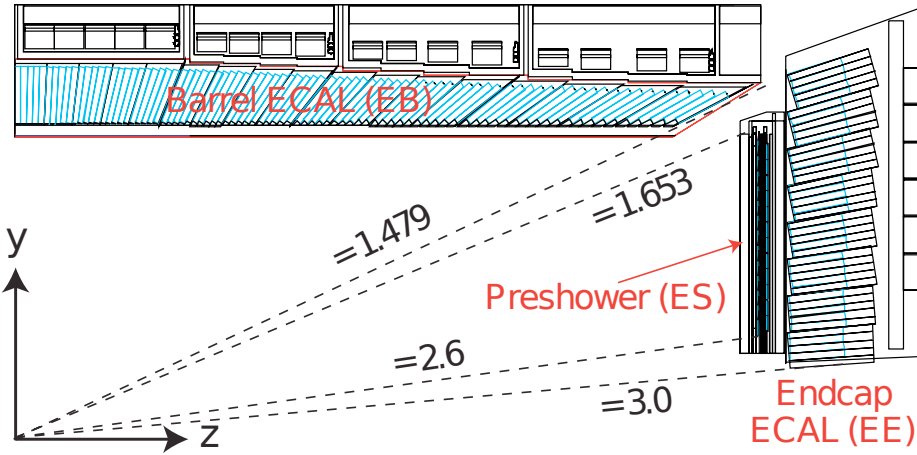
Figure 3.6: Cross section sketch of the electromagnetic calorimeter, consisting of the EB system in the barrel region and the ES and EE systems in the endcap region [78].

out, which is fast and radiation-resistant. When a photon or electron traverses a crystal, it is absorbed by the dense scintillator material. The primary processes of this energy loss are bremsstrahlung and photon pair production, which result in an electromagnetic shower and thus excites the scintillator material. When the material returns to its ground state, a certain amount of scintillation light is emitted, which is a measure of the deposited energy. The low light yield of lead tungstate is compensated by amplifying it with avalanche photodiodes in the EB and vacuum phototriodes in the EC.

The barrel section of the ECAL, called EB, covers a region of up to a radius of 129 cm and $|\eta|$ of 1.479. Each crystal in this region has a granularity of 0.0174 in $\Delta\phi$ and $\Delta\eta$ with a front face cross section of $\approx 22 \times 22 \, \text{mm}^2$ and a length of 230 mm. This corresponds to a radiation length of $25.8 X_0$. The endcap section of the ECAL, called EE, is located at $z = \pm 314$ cm and covers a range of $1.479 < |\eta| < 3.0$. These crystals have a front face cross section of $\approx 28.6 \times 28.6 \, \text{mm}^2$ and a length of 220 mm, resulting in a radiation length of $24.7 X_0$. Additionally a preshower detector (ES) exists in the region of $1.653 < |\eta| < 2.6$. The ES

comprises two planes of silicon strip detectors, behind two absorption layers at depths $2X_0$ and $3X_0$. This detector system allows to distinguish neutral pions decaying into two nearly parallel photons from single photons.

The energy resolution of the ECAL can be parametrized by a function with three terms. The first is a stochastic term, taking into account fluctuations of photon and electron counts, the second one is a noise term, arising from noise in readout electronics and other collision events (pile-up). Finally, a constant term exists, respecting imperfections of the detector material and energy losses in dead detector materials [72]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.30\%)^2, \tag{3.6}$$

where E is given in GeV. This energy reconstruction is achieved by summing 3×3 crystals.

3.2.4 Hadronic Calorimeter

The hadronic calorimeter (HCAL) is a sampling calorimeter designed to measure the energy of strongly interacting particles. It consists of alternating layers of brass (absorber) and plastic scintillators. Brass has a short interaction length and is non-magnetic. Hadronic showers are created when particles traverse the brass layers. These showers are more complex as they consist of a hadronic, mainly consisting of charged pions, and an electromagnetic component, arising from neutral pion to photon decays, which limits the energy resolution. Figure 3.7 shows the layout of the HCAL detector, consisting of four subsystems, namely the HB (barrel region), the HE (endcap region), the HO (outer barrel region), and the HF (forward region).
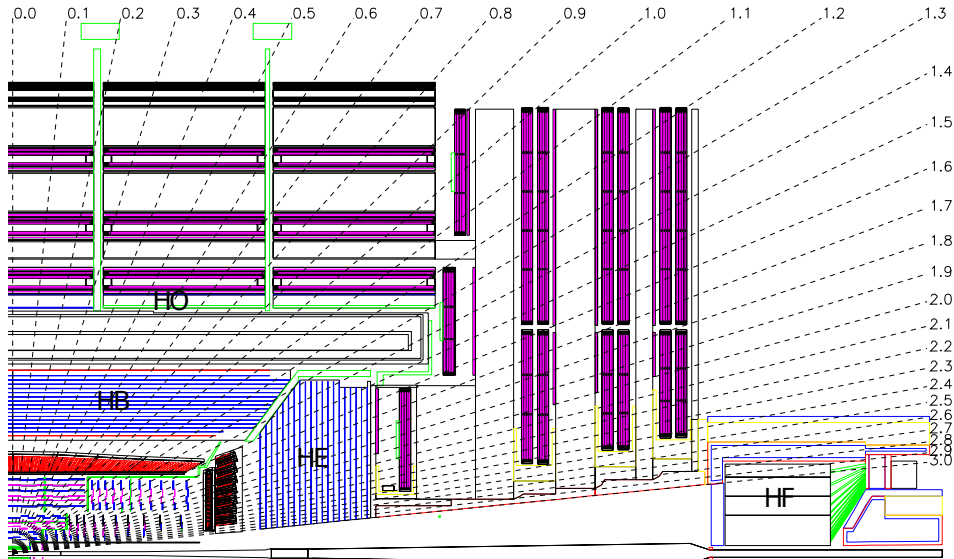
Figure 3.7: Layout of the calorimetry detector systems [72]. The hadronic calorimeter consists of the HB and HO systems in the barrel region, the HE system in the endcap region and the dedicated HF system in the forward region for large values of $|\eta|$.

The HB covers a region of $|\eta| < 1.4$, while in the endcap region the HE covers a region of $1.3 < |\eta| < 3.0$. Both consist of 2304 towers each with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ for $1.6 < |\eta| < 3.0$. In addition, the HO exists

to measure late starting showers and shower tails in the region of $|\eta| < 1.26$ as the combined stopping power of the EB and HB is sometimes not sufficient enough in the barrel region. It is located in the barrel muon system outside of the magnetic coil and uses the latter one as an additional absorber layer extending the absorption power of the HB to ten interaction lengths. The steel/quartz fibre HF extends the HCAL to the forward region $3.0 < |\eta| < 5.0$. Its front face is 11.2 m distanced from the interaction point. In total, 900 towers exist in the HF with a granularity of $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$. The choice of steel and quartz fibre for this part of the HCAL is mainly motivated by the harsh radiation environment in the forward region, allowing survivability for up to ten years of LHC operation. The energy resolution (in GeV) of the HCAL has been measured to be [82]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{85\%}{\sqrt{E}}\right)^2 + (7.00\%)^2 . \tag{3.7}$$

3.2.5 Solenoid

The superconducting solenoid creates a magnetic field of up to 4 T to bend the trajectories of charged particles. This allows the momentum measurement using the trajectory's radius of curvature and thus its charge. The sign of muons decaying from narrow resonances with an energy of ≈ 1 TeV can be determined with this solenoid at a momentum resolution of $\approx 10\%$. This precision is required by the muon system, ensuring a good performance of the muon system. Its inner bore has a diameter of 6.3 m, a total length of 12.5 m, and weighs 220 tons. The flux is returned by an iron yoke of 10 000 tons. The solenoid encapsulates the tracking systems and the calorimeters and is operated at ≈ 4.5 K in order to make the coil superconducting. Its strong magnetic field causes large mechanical deformations resulting in several millimeters displacement of yoke disks in the endcap region.

3.2.6 Muon Chambers

The outermost subdetector system are the muon chambers. Up to these chambers, almost all particles are absorbed already, except for muons and neutrinos. While the latter ones leave the CMS detector without a trace, the muon chambers allow precise measurements of the momentum and energy of muons. In combination with the reconstructed trajectory of the tracking subdetector system, the resolution of these measurements can be significantly improved. In total, three types of gaseous detector parts are used as shown in Fig. 3.8, namely the drift tube (DT) chambers, the resistive plate chambers (RPC), and the cathode strip chambers (CSC).

The gaseous detectors measure electrical currents when muons traverse them by ionizing the gas. The released charge carrier creates an avalanche while drifting through an electric field, which leads to a measurable current at anode wires and cathode strips. Spatial information can be inferred from the corresponding cathode strip, allowing the reconstruction of the muon trajectory and thus the momentum from its radius of curvature.

In total, the muon chambers cover a region of $|\eta| < 2.4$. In the barrel region ($|\eta| < 1.2$) there are 250 DT chambers organized in four layers together with six RPCs. They are interleaved with the iron return yoke and located at radii of approximately 4.0 m, 4.9 m, 5.9 m and 7.0 m. The DT chambers are designed to provide a global resolution of $\approx 100\,\mu$m in the $r - \phi$ direction. The endcap region comprises 486 CSCs complemented by three RPC chambers (until $|\eta| < 1.6$) to cover a region of up to $|\eta| < 2.4$. The CSCs provide a spatial resolution of
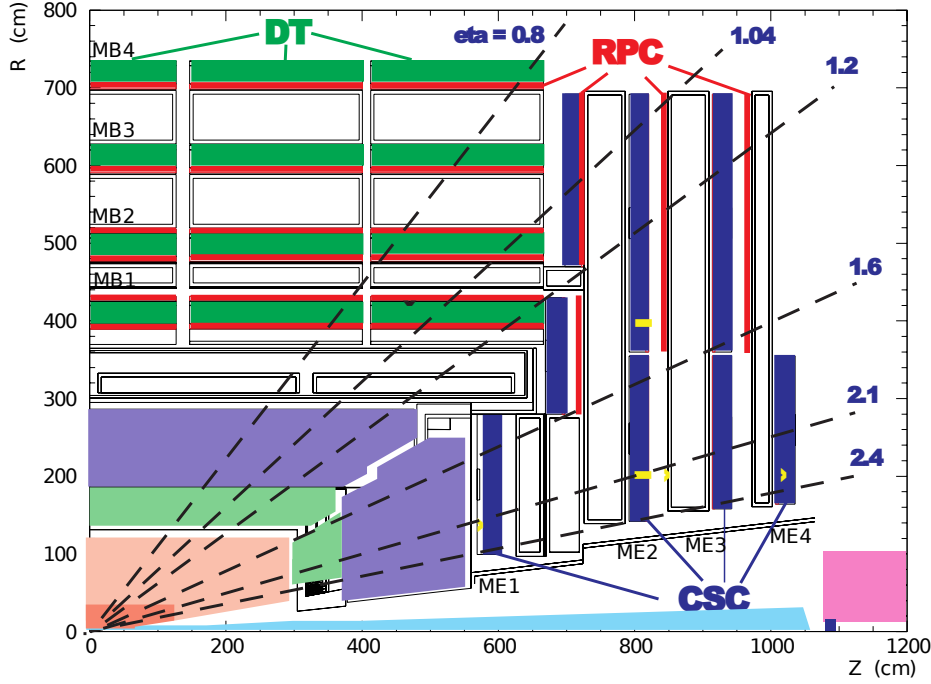
Figure 3.8: Layout of the muon chambers, comprising the three subdetector parts. In
the barrel region the DT and RPC chambers are located, while in the endcap
region the RPC and CSC chambers are installed [78].

$\approx 200\,\mu$m. While the RPC can barely compete with the spatial resolution of the DT and CSC, it has a time resolution, which is significantly smaller than the time difference of two bunch crossings (25 ns), helping to identify the corresponding bunch crossing.

The track reconstruction efficiency has been measured to be $95 - 98\%$ for the year 2010 ($\sqrt{s} = 7\,$TeV) [83] and the years $2015 - 2016$ ($\sqrt{s} = 13\,$TeV) [84].

3.2.7 Data Acquisition and Trigger System

Given the high collision rate, a multi-stage trigger system [85] is deployed at the CMS experiment. A data rate of approximately 40 TB per second is obtained for a typical event size of approximately one mega byte [81] and a bunch crossing every 25 ns. This rate corresponds to a collision rate of 40 MHz. Out of those collisions, only a small fraction is interesting for the CMS physics program. In addition, the data acquisition system is only capable of storing a data stream of approximately 100 MB per second [86] on disk.

The first stage of this trigger system is the so-called Level-1 trigger. It accepts a maximum data rate of 100 kHz, spending at most $3\,\mu$s for a decision per event. Events are held in memory buffers on frontend electronics for this period, overcoming the time between two consecutive bunch crossings. Of these $3\,\mu$s, only $1\,\mu$s is spent on trigger decision calculations, while the rest of the time is spent on signal transit between the detector and the trigger hardware. This limits the available amount of information to make the trigger decision. In the end, only information from the calorimeters, the muon chambers, and their correlations can be used. From these "trigger primitives" objects are built, such as photons, electrons, muons, and jets, but also event level objects, such as H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$. The trigger decision is then based on the presence of these objects and their transverse momentum and energy.

Dedicated hardware, i.e., Field Programmable Gate Arrays (FPGAs), are used at the Level-1 trigger to minimize latencies.

The next stage is the HLT trigger, which is a pure software trigger. Here, a total processing time of approximately 1 s is spent for the trigger decision. At this stage, more event information and more precise calibrations are used, close to the one used by offline physics analyses. Multiple trigger decisions are grouped into "paths". After each step in such a path, it is decided whether the event is kept or not. These decisions are sorted such that timely calculations are performed later, minimizing the total processing time. The HLT trigger farm is designed to output an event rate of approximately 1000 events per second [87].

3.2.8 Physics Object Reconstruction

The previously described CMS detector collects signals from up to $\mathcal{O}(10^8)$ readout channels during one bunch crossing. However, these signals need to be transformed into interpretable outputs for physics analyses, such as hits in the tracking system and clusters from energy deposits in calorimeters. Finally, particle hypotheses, or "physics objects", can be reconstructed from these outputs. The primary dependence of the physics objects on the collected signals from the CMS subdetectors is described in the following:

Muons are identified and reconstructed mainly based on the response of the muon chambers.

Electrons (and photons) primarily use the energy deposits in the ECAL for identification and reconstruction. Electrons are further reconstructed with their corresponding track in the tracking system.

Jets are made of hadrons and photons. Their energy reconstruction is solely based on calorimeter information. In addition, the tagging of jets, e.g., originating from b-quark hadronization, and hadronic τ decays, uses tracker information of the jet constituents.

Neutrinos cannot be measured by the CMS detector. However, they create an energy imbalance in the transverse plane. This imbalance is called E_T^{miss} and is used to estimate the neutrino presence in the event.

In addition, the correlation or "linking" of several subdetector systems provide great improvements for the reconstruction and identification of the above listed *physics objects*. The reconstruction software for this linking approach is called PARTICLE-FLOW (PF) [88]. An example of its benefits is shown in Fig. 3.9.

The figure shows the event display of a simulated dijet event. The reference jets (Ref jet) are known from the simulation. Two jet reconstructions are shown, which aim to reconstruct the reference jets: one using only calorimeter information (Calo jet), and one using the PARTICLE-FLOW algorithm (PF jet). Here, the jets reconstructed by the PARTICLE-FLOW algorithm are much closer to the reference jets in energy and direction compared to the jets reconstructed using only the calorimeter information highlighting the benefit of exploiting information of multiple subdetector systems simultaneously. The following subsections describe first the PARTICLE-FLOW algorithm and then the reconstruction of the aforementioned *physics objects* in detail.
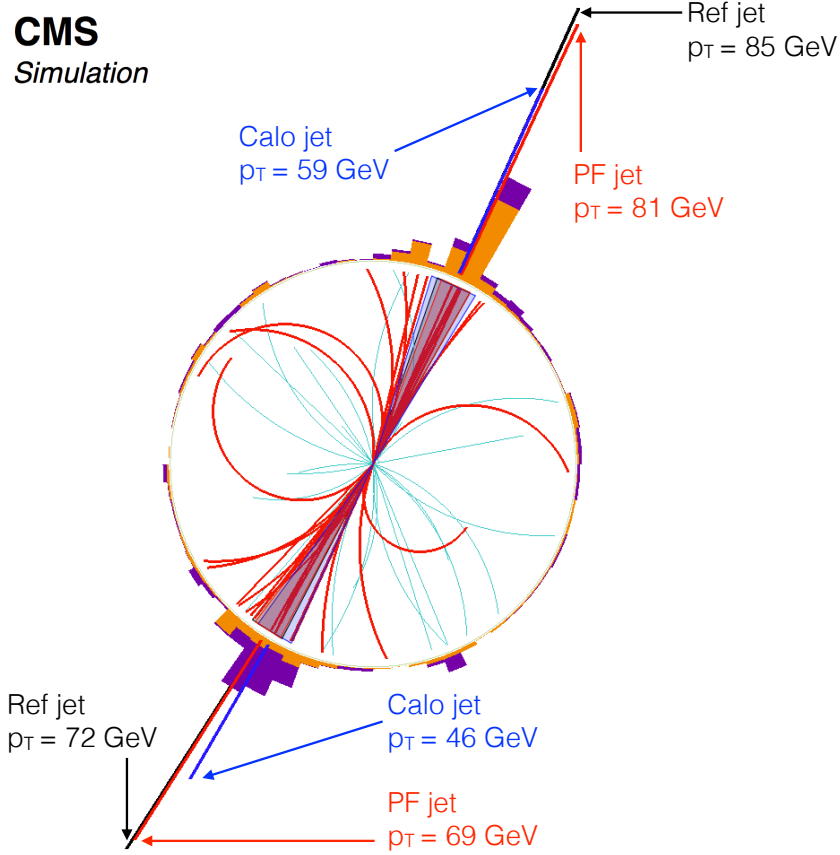
Figure 3.9: Event display of a simulated dijet event [88]. In black, the reference jets are depicted; in red, the reconstructed jets with the PARTICLE-FLOW (PF) algorithm, and in blue, the reconstructed jets with only the calorimeter information (Calo) are shown. Both reconstructions aim to reconstruct the reference jets.

3.2.8.1 The PARTICLE-FLOW Algorithm

The PARTICLE-FLOW algorithm, part of the CMS software framework "CMSSW", is designed to identify and reconstruct all particles within an event from detector signals. By linking these detector signals, particle candidates are found. Trajectories of charged particles, clusters of energy deposits in the calorimeters, and trajectories of muons serve as baseline input for the algorithm. These inputs are the smallest unit, called "PF elements", of which multiple are linked together. The algorithm quantifies the distance between linked elements as a quality measure, establishing a qualitative correlation of different detector information. The linked PF elements lead to a complete event description, in which muons, electrons, photons, jets originating from light and heavy quarks, hadronic tau lepton decays, and derived quantities are identified and reconstructed. A detailed description of the PARTICLE-FLOW algorithm is given in Ref. [88].

Reconstruction of Charged Particle Tracks and Energy Clusters

The reconstruction of tracks in the inner tracking systems is based on an iterative fitting approach [89] using the combinatorial track finder [90] (CTF), which is an extended version of

the combinatorial Kalman filter [91–94]. The fitting procedure starts with seeds from pairs or triplets of hits and extrapolates new candidates of hits to outer tracking layers, which fit the expected helix-like trajectory. The procedure ends once the end of the inner tracking system is reached or no more hits are available. Tracks are required to be reconstructed with at least eight hits, have at most one missing hit, originate close to the beam axis, and have transverse momentum of more than 0.9 GeV. This fitting procedure is repeated multiple times using different initial seeds with relaxed quality requirements in each consecutive iteration, increasing the total tracking efficiency. Hits of tracks, which are kept, are masked for next iterations in order to reduce computing time.

The clustering of energy deposits is performed separately for the ECAL and HCAL in the barrel and endcap regions and the two preshower layers. Here, cluster seeds are identified as the cells with the highest energy compared to their neighbors and fulfilling a certain seed threshold. Neighboring cells are aggregated into a topological cluster if their energy is above a certain threshold of E_{T}. Cell energies are not exclusive to one cluster and can be shared among different clusters. The final clusters are built within these topological clusters using an expectation-maximization algorithm based on a Gaussian-mixture model. This model can identify, e.g., two energy clusters deposited by two collimated photons originating from a neutral pion decay within the same topological cluster.

Linking between Tracks and Energy Clusters

The PARTICLE-FLOW algorithm links the previously described charged particle tracks and energy clusters. The linking is performed pairwise, considering only the nearest neighbors in the $\eta - \phi$ plane. This consideration minimizes computing time significantly due to the $\mathcal{O}(n^2)$ complexity. Multiple PF elements linked directly or indirectly are called *PF blocks*. Starting from a track, its last hit is extrapolated to the energy clusters in the ECAL, HCAL, and the two preshower layers. While the extrapolation depth for the HCAL is defined as one interaction length, for the ECAL and the two preshower layers, it corresponds to the expected depth for a typical longitudinal shower profile. A link is established if the track extrapolation is within the energy cluster in the $\eta - \phi$ plane for the ECAL and HCAL barrel region and in the $x - y$ plane for the ECAL endcaps and the preshower layers. In case of ambiguities of track links to multiple energy clusters, only the one with the smallest distance between the extrapolated track position and the cluster position in the $\eta - \phi$ plane is kept. The same procedure is performed to connect HCAL clusters to ECAL clusters and ECAL clusters to preshower clusters. In addition, tracks are also linked together if they share a common secondary vertex. Finally, the algorithm links tracks to the information of the muon chambers.

Reconstruction and Identification of Particles

The PARTICLE-FLOW algorithm iteratively identifies particles using the aforementioned established links. First, muon candidates are identified and reconstructed (cf. Sec. 3.2.8.2). Their tracks and energy clusters are removed from the initial PF block. Next, electrons (and energetic or isolated photons) are reconstructed (cf. Sec. 3.2.8.3). Their energy loss due to bremsstrahlung is estimated by extrapolating photons as a tangent of the electron track to the calorimeter. Again, their associated tracks and energy deposits in the ECAL and preshower systems are removed from the corresponding PF blocks. Finally, the remaining elements in the PF blocks are used to reconstructed hadrons and photons: first charged hadrons, taking into account information of the inner tracking and HCAL system, then neutral hadrons and

non-isolated photons only relying on energy clusters in the HCAL and ECAL (cf. Sec. 3.2.8.4).

In the end, muons, electrons, photons, charged and neutral hadrons are identified and re-constructed. However, in rare cases, misreconstructions of especially muons can lead to artificially large p_T^{miss} values. Sources are, e.g., muons reconstructed from cosmic muons with trajectories close to the beam axis, misreconstructions of muon momenta, or energetic charged hadrons reaching the muon chambers ("punch-throughs"). A post-processing step takes care of these events without affecting events with genuine p_T^{miss}, such as $t\bar{t}$ + Jets events.

3.2.8.2 Muons

The reconstruction of muons [78, 88] is based on hits in the muon chambers and region-compatible hits in the silicon trackers. The region compatibility requirement reduces com-puting time significantly as the muon hit information in the tracker constitutes only a few percent of the total available tracker information, allowing the reconstruction to run during time critical HLT trigger decisions (online). Especially the fact that only muons (and neutri-nos) can reach the muon chambers enables a high purity of the muon reconstruction. Three different muon collections are reconstructed depending on the detector information, with a reconstruction efficiency of 99%.

Standalone muons are reconstructed solely with the information of the muon chambers. Track segments are built based on hits in the DT and CSC independently. Afterwards, a fit is performed using the determined DT and CSC track segments as input seeds. The fit is iteratively repeated and stopped once the outermost detector layer of the muon system is reached. Finally, hits in the DT, CSC, and RPC chambers and the nominal interaction point are combined to the trajectory of a *standalone muon*.

Global muons use the information of the muon chambers and the inner tracking system. First, a *standalone muon* trajectory is extrapolated based on its expected path of flight to the silicon tracker. Then, compatible tracks in the muon chambers and the silicon tracker are matched. Here, only tracks in the silicon tracker are considered in a specific region of interest defined by the extrapolated *standalone muon* trajectory. Both tracks are fitted together to the trajectory of a *global muon*, improving the momentum resolution for muons with large transverse momenta ($p_T \gtrsim 200\,\text{GeV}$).

Tracker muons are defined by tracks in the silicon tracker, which can be extrapolated to the muon chambers. These tracks need to have at least a momentum of 2.5 GeV and a transverse momentum of 0.5 GeV. At least one match of a muon segment to the track from the silicon tracker is required for a muon candidate to qualify as a *tracker muon*.

While *global muons* have a precise momentum resolution for muons with large transverse momenta, tracker muons are a more suitable choice for muons with small momenta ($p_T \lesssim 10\,\text{GeV}$). These muons often do not reach the outermost layer of the muon system due to absorption resulting in fewer hits in the muon chambers. The *tracker muons* and *global muons*, merged in case they share the same trajectory in the silicon tracker, serve as input to the PARTICLE-FLOW algorithm. The PARTICLE-FLOW algorithm considers additionally energy deposits in the calorimeters, which are linked to the reconstructed track, yielding an im-proved identification of muons. Muons, which are misreconstructed from punch-throughs, are taken care of using additional selection criteria.

3.2.8.3 Electrons

The reconstruction of electrons [78, 88] is based on their energy deposition in the ECAL calorimeter, their trajectory in the silicon tracker, and the linking between both. The material thickness of the detector causes a challenge for the reconstruction, whereby electrons experience a significant energy loss due to bremsstrahlung. This energy loss manifests itself through photon emission, which may create new electron-positron pairs, emitting photons due to bremsstrahlung again. Thus, it is essential to collect the energy of bremsstrahlung photons to ensure a precise electron reconstruction. Due to the solenoidal magnetic field, the energy deposited by the bremsstrahlung photons is spread in the ϕ direction. The low electron mass yields significant bremsstrahlung effects such that in 10% of the cases more than 95% of the initial energy is lost. In addition, other particles than electrons create signals in the ECAL and the silicon tracker, unlike muons in the muon chambers, resulting in a reduced identification efficiency and a higher misreconstruction rate.

The reconstruction starts first by building a cluster of energy clusters, called *supercluster*. Two clustering algorithms [86, 95] take care of the *supercluster* reconstruction. Their main purpose is to deal with the previously described energy spread in the ϕ direction. The clustering procedure is seeded by energy clusters above a certain threshold of E_T. The resulting *superclusters* then drive the seed finding for the track reconstruction. The *supercluster* allows exploiting the fact that its energy weighted average describes an electron, which did not experience any energy loss. This initial position is extrapolated to the tracking system using the expected flight path for both charge hypotheses. The first compatible hit in the innermost pixel layer within a certain region requirement is used as a starting point. From there on, a second hit is searched for in the next layer. Once both hits are found, they serve as seeds for the track reconstruction. The trajectory is fitted similarly to the standard track reconstruction, described in 3.2.8.1, taking into account energy losses due to bremsstrahlung. A dedicated extension of the Kalman filter algorithm is used incorporating a non-linear Gaussian Sum Filter (GSF) [96] accounting for kinks in the trajectory due to sudden energy losses.

In the end, more than ten variables are used, combined in a boosted decision tree, to identify isolated and non-isolated electrons in the barrel and endcap region separately. These variables comprise information of the trajectory, such as its energy loss, information of the energy deposits in the calorimeters, e.g., the energy ratio deposited in the ECAL and HCAL calorimeters, and combined information, such as the position offset between the supercluster seed and the extrapolated trajectory into the ECAL calorimeter. The identification performance has been measured using a tag and probe method in $Z/\gamma^* \rightarrow e^+e^-$ events for the Run 2 data-taking period [97] to about 60% identification efficiency for electrons with small transverse momentum ($p_T \leq 20\,\text{GeV}$) and $70 - 80\%$ identification efficiency above.

3.2.8.4 Jets

Quark and gluons produced in proton-proton collisions are subject to the hadronization process due to QCD confinement. This process occurs when two color charge carriers are separated in spacetime. The distance-dependent linear component of the QCD potential rises and eventually exceeds an energy threshold, where the production of new quark anti-quark pairs becomes energetically favorable. Color neutral baryons and mesons are created together with leptons and photons, originating from decays, bremsstrahlung, or photon conversions. A collection of different stable particles originating from the same initial quark or gluon traverse the CMS detector in a collimated cone - a so-called "jet". The original quark

or gluon energy and momentum can be inferred from the reconstruction of such a jet.

Clustering algorithms are used to merge regional and energy compatible stable particles into jets. Here, the anti-k_T algorithm [98], implemented in the FASTJET package [99], is used with a distance parameter of $R = 0.4\,(0.8)$ to cluster small radius jets (large radius jets). The clustering algorithm is robust against the presence of additional soft emissions or collinear particle splittings, called infrared and collinear safety (IRC). Leptons and photons, which have already been identified by the PARTICLE-FLOW algorithm and qualify as isolated, do not participate in the jet clustering. All other particles undergo an iterative clustering, which consecutively accumulates particles within the defined distance parameter R into jets. The clustering is based on a distance measure for each particle (d_i) and each particle pair (d_{ij}):

$$d_i = p_{\mathrm{T},i}^{-2}, \tag{3.8}$$

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}\right) \cdot \frac{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2}{R^2}. \tag{3.9}$$

In case d_{ij} is smaller than d_i, both particles are merged into one object. Both are excluded from further clustering iterations; instead, the new merged object is added. A jet is found once the condition $d_i < d_{ij}$ is fulfilled. The clustering procedure finishes when all particles are accumulated into jets. Particles with large transverse momenta (hard) are expected to be closer to the center of the jet cone, thus clustered first. In contrast, particles with small momenta (soft) are expected to be distanced further away from the cone center and therefore merged in later iterations. This characterizes the strength of the anti-k_T algorithm: it prioritizes hard particles and is less dependent on soft particles, which usually come from soft emissions (IRC).

The energy calibration of jets (JEC) is performed in a factorized approach [100]. It comprises a manifold of energy scale calibrations (JES) and an energy resolution calibration (JER). Consecutively energy corrections are applied to correct for offsets due to pileup effects, detector responses, and differences between data and simulation.

Pileup offset corrections are applied to account for tracks and energy deposits from additional proton-proton collisions during bunch-crossings. Here, pileup interactions from the same bunch crossing (in-time pileup) and previous and subsequent bunch crossings (out-of-time pileup) are considered. In-time pileup is mitigated by removing charged hadrons, which tracks match with pileup vertices. This procedure is called "charged hadron subtraction" (CHS). Out-of-time pileup effects are reduced by minimizing the time window in which signal is collected in different subdetector systems. Additional (overlapping) soft jets originating from pileup interactions are taken care of by the multivariate pileup jet identification algorithm [101, 102] (PUJETID). The remaining energy diffusion due to neutral hadrons and out-of-time pileup is estimated by an extended hybrid jet area method [100]. Finally, particle-level offsets between data and simulation are estimated using zero-bias events [100] collected by a heavily prescaled random trigger.

Detector simulations corrections are used to reduce differences induced by the detector simulation. True (particle-level) jets are matched to reconstructed jets with a simulated QCD multijet dataset. A detector response correction is calculated based on the arithmetic mean between the matched particle-level jet and the reconstructed jet momen-

tum. Correction factors are derived in bins of p_T and η for different jet sizes. Especially in regions of large $|\eta|$ ($|\eta| \gtrsim 4.0$), and the transitions regions at $|\eta| \approx 1.3$ and $|\eta| \approx 3.0$, the detector response correction rises due to an increasing amount of detector material.

Residual differences are corrected in two steps. First, η-dependent relative JES differences are estimated using the missing transverse momentum fraction method (MPF) [103] in dijet events. In the second step, p_T-dependent absolute differences are derived from a global fit using $Z + $ jets, $\gamma + $ jets, and multijet events. The resulting differences are a few percent in the barrel region and up to 10% in the endcap region. Only in the transition region of $|\eta| \approx 3.0$ the difference rises to approximately 20%. A significant difference to the other corrections is that the residual differences are explicitly applied to data.

Jet flavor corrections are account for differences between reconstructed and particle-level jet momenta ratios for different jet flavors. The differences are estimated by calculating this ratio for different MC generators. Following the recommendations of the CMS collaboration, this correction is not included in the analysis presented in this thesis; however, dedicated uncertainties are included.

These corrections are applied in strict order, as listed above. In each consecutive step, the previous corrections are already incorporated. Figure 3.10 shows the effect of uncertainties from the different correction sources.
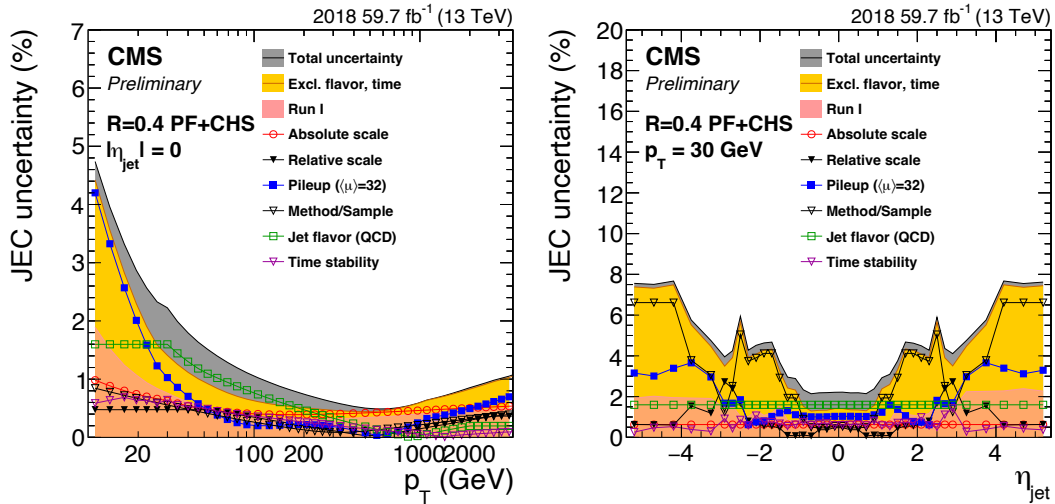
Figure 3.10: Uncertainties arising from jet energy scale corrections as a function of p_T (left) and η (right) for the data-taking period of 2018 [104]. The jets are reconstructed using the PARTICLE-FLOW algorithm including the charged hadron subtraction. The jet clustering is performed with the anti-k_T algorithm using a distance parameter of $R = 0.4$. For comparison, the uncertainty using data recorded during Run 1 are shown.

For jets in the barrel region (Fig. 3.10, left), the total uncertainty is in the range of $1 - 5\%$, rising towards low transverse momenta. In this low transverse momentum region, the effect of pileup interactions is the most dominated one. The η-dependence is shown for a typical jet with a transverse momentum of $p_T = 30 \, \text{GeV}$ in Fig. 3.10 on the right side. The total uncertainty varies between 2% in the barrel region and up to 8% in the forward regions.

3.2.8.5 Missing Transverse Energy

The presence of particles leaving the CMS detector undetected, e.g., neutrinos or particles predicted by BSM theories, can be estimated by the event-level energy imbalance. An energy imbalance along the longitudinal axis (beam pipe) is not measurable as only partons of the proton collide, which involve rather large uncertainties on their a priori unknown initial energy; instead, an imbalance can be measured with sufficient accuracy in the transverse plane as momentum conservation applies here. The p_T^{miss} (E_T^{miss}) quantity serves as a measure of this transverse momentum (energy) imbalance. In the past, this quantity lead to the observation of the W boson [105] at the UA1 experiment [106] with only six events, including its mass measurement to $m_W = 81 \pm 5\,\text{GeV}$.

The E_T^{miss} observable is calculated by summing the transverse momenta of all particles identified by the PARTICLE-FLOW algorithm:

$$\vec{p}_T^{\,miss} = - \sum_{i}^{N_{particles}} \vec{p}_{T,i}, \tag{3.10}$$

$$E_T^{miss} = \left| \vec{p}_T^{\,miss} \right|. \tag{3.11}$$

As described in the previous Sec. 3.2.8.4, jets are subject to several correction steps. The impact of these corrections is propagated to the E_T^{miss} quantity by subtracting the difference of the transverse jet momentum between the "raw" PARTICLE-FLOW jet and the corrected one:

$$\vec{p}_T^{\,miss} = - \sum_{i}^{N_{particles}} \vec{p}_{T,i} - \sum_{j}^{N_{jets}} \left(\vec{p}_{T,j}^{\,corr.} - \vec{p}_{T,j} \right). \tag{3.12}$$

The resolution of the E_T^{miss} observable is relatively coarse as all particles are considered in the calculation, including their reconstruction inefficiencies. In addition, the presence of pileup interactions complicates the reconstruction of E_T^{miss}. Thus, the E_T^{miss} resolution has been found to be in the range of $10 - 30\%$ using a simulated $t\bar{t}$ + Jets dataset [88].

In the presented analysis, two neutrinos are part of the final state of the signal process, yielding a non-zero contribution of E_T^{miss}. The major backgrounds do not share this topology, except for the dileptonic $t\bar{t}$ + Jets process, which also incorporates two neutrinos in the final state; however, the spin-correlation of the H $\rightarrow$ W$^+$W$^-$ process (part of the signal process) favors the collimated path of flight of both neutrinos, whereas a nearly uniform distribution is expected for the dileptonic $t\bar{t}$ + Jets process. Thus, not only the E_T^{miss} quantity itself is powerful in the presented analysis, but also angular correlations between the two neutrinos (E_T^{miss}) and other particles in the event.

3.2.8.6 Identification of b-quark Jets

Identifying b-quark jets is essential for the presented analysis as one of the two Higgs bosons decays to two b-quarks. These jets contain long-lived b-hadrons, which have a relatively sizeable decay length of approximately $c\tau \approx 0.45\,\text{mm}$ [78], leading to measurable secondary vertices in the tracking systems. In addition, they are characterized by high track multiplicities and an increased branching fraction to decays containing leptons. Jets originating from

b-quarks are characterized by the collective of these features, which lead to the use of multi-variate techniques - especially deep learning techniques - to build powerful discriminators. Requiring jets to fulfill certain thresholds of such a discriminator allows identifying b-quark initiated jets, also called "b-tagging", at a particular selection efficiency and misidentification probability.

In the presented analysis, the DEEPJET algorithm [107] is leveraged. It is a deep neural network that uses approximately 650 input features comprising information of charged and neutral jet constituents, secondary vertices, and the whole event. Compared to preceding b-tagging algorithms, the DEEPJET algorithm focuses on as many low-level features of the jet constituents as possible. This immense feature space, except for the global event features, is fed into 1×1 convolutions [108], which creates a constituent-level feature embedding. Next, the embedded features are passed into recurrent layers (LSTM) [109], which are suitable to deal with varying numbers of constituents in a jet. Finally, the outputs of the LSTM layers and the global event features are forwarded to a fully-connected layer, which outputs jet identification probabilities for heavy quark initiated jets (b-tagging or c-tagging) and light quark or gluon initiated jets (u,d,s,g).

The performance of the DEEPJET algorithm has been measured using data collected in data-taking period of 2017 [110]. Figure 3.11 shows the misidentification probability as a function the b-tagging efficiency for the DEEPJET algorithm and its predecessor, the DEEPCSV algorithm [111]. Due to performance differences of these algorithms between data and simulations, dedicated scale factors (SF) are applied to simulation provided by the CMS collaboration [112].

Ideal performance is achieved at low misidentification probabilities and high b-tagging efficiencies. The CMS collaboration provides three different working points for misidentification probabilities of approximately 10% (loose), 1% (medium), and 0.1% (tight) respectively, marked with circles and triangles in Fig. 3.11. It can be seen that the DEEPJET algorithm outperforms its predecessor for every possible working point.

Semi-leptonic decays in b-quark jets are characterized by the presence of a neutrino, which escapes the CMS detector and remains unmeasured. Thus, the energy of the b-quark jet is misreconstructed, ultimately worsening the resolution of invariant mass of the $H \rightarrow b\bar{b}$ system. A neural network based regression algorithm [113] is used in this analysis to correct the b-quark jet energy spectrum.

The signal process of the presented analysis consists of two b-quark jets in the final state. Choosing exemplary the medium working point of the DEEPJET algorithm, Fig. 3.11 suggests a misidentification probability of approximately 1%, while the b-tagging efficiency is at approximately 80%. Thus, this results in a total tagging efficiency for two b-quark jets of nearly 65%. In addition, b-tagging also helps to reject background processes, which do not incorporate b-quarks in their final state, leading to a significant reduction of the Drell-Yan background contamination for this analysis.

3.2.9 The NanoAOD Data Format

This analysis leverages a new data format [114] released by the CMS collaboration, namely NANOAOD. As described in Sec. 3.2.7 the typical event size is one megabyte initially; however, the NANOAOD data format has a reduced event content of $1 - 2 \, \text{kB}$. This significantly diminishes the amount of storage needed, providing a sustainable solution for the current
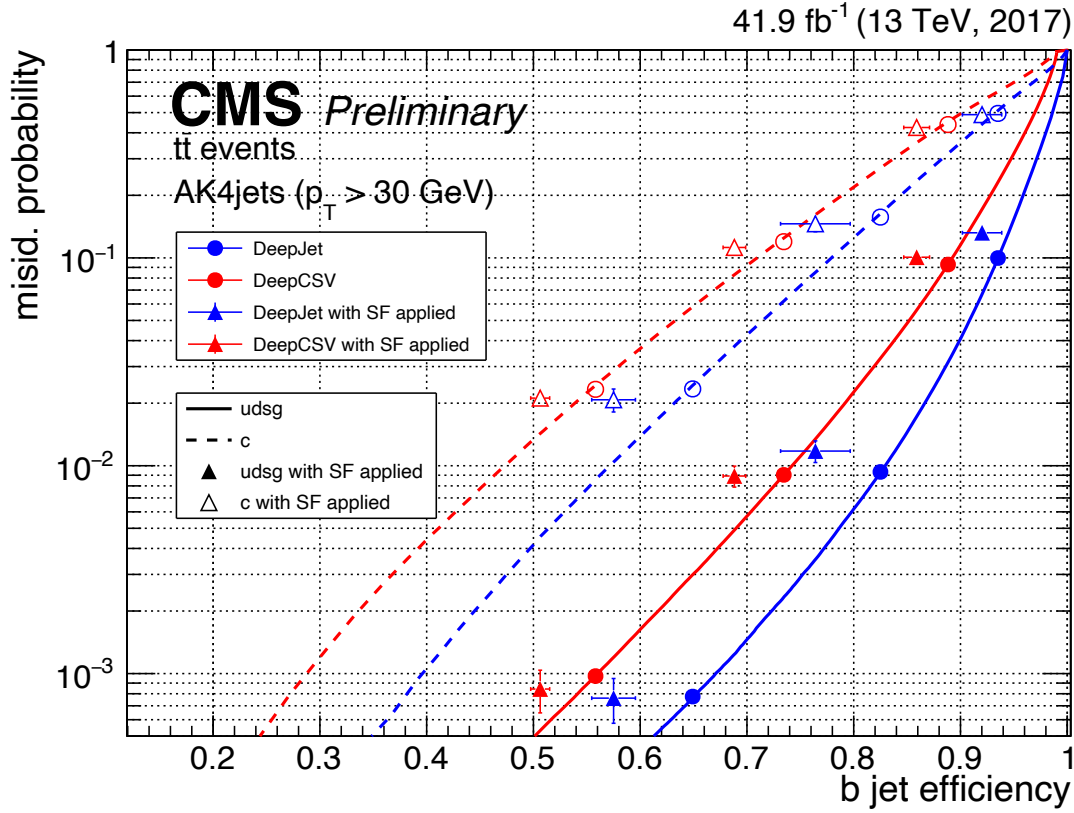
Figure 3.11: Performance of the DEEPJET (blue) and DEEPCSV (red) b-tagging algorithms measured with $t\bar{t}$ + Jets events for jets with a transverse momentum larger than 30 GeV [110]. Three different working points are shown (indicated by the colored markers), which are provided by the CMS collaboration to select b-quark jets at different misidentification probabilities and b-tagging efficiencies.

and the upcoming high luminosity data-taking periods.

The event size reduction is achieved by storing only the most common event-level variables and high-level *physics objects* shared by the CMS physics program and by limiting the float precision of their features, i.e., zeroing insignificant mantissa bits. Such a feature, e.g., the transverse momentum of a muon, is stored as a flat array. In addition, the number of muons per event is stored, allowing to infer the relevant transverse muon momenta per event. Cross-references between *physics objects*, such as generator level particles and reconstructed particles, are stored with indices. All branches are stored using basic datatypes: signed or unsigned integers, floats, or booleans. Figure 3.12 shows the relative size of the different branch collections in a NANOAOD file averaged over all events using the number of compressed bytes per branch collection for a typical dileptonic $t\bar{t}$ + Jets simulation dataset in 2017. Due to the high jet multiplicities in $t\bar{t}$ + Jets events, the *Jet* and *Generator Particle* collection occupy more than half of the total event size. The *Other* collection makes up the next largest portion; it comprises global event information, trigger relevant information, and track and vertex information.

In contrast to previous data formats, such as AOD [81], where physics objects are stored as

Generator Particles (72)

31.0%

Photons (31)

3.7%

Muons (53)

5.1%

Taus (47)

6.9%

Jets (109) — 20.9%

7.3%

Trigger Information (1027)

17.7% 7.4%

Electrons (63)

Other (184)
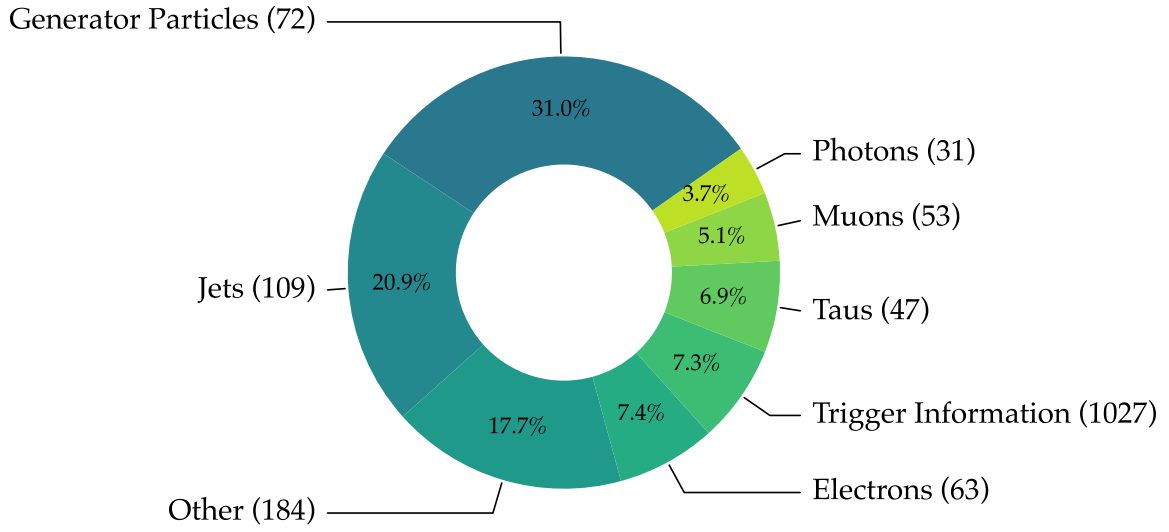
Figure 3.12: The relative contribution of different branch collections, including the number of branches in brackets, for an (averaged) event within a typical dileptonic $t\bar{t}$ + Jets simulated NANOAOD dataset (2017).

high-level C classes, no additional dependence on CMSSW is needed to interpret and analyze these objects. The NANOAOD data format uses the LZMA compression algorithm; however, in this analysis it has been recompressed with the ZSTD [115] compression algorithm in order to speed up the decompression step.

3.3 Computing Infrastructures

Complex reconstruction algorithms, e.g., as described in Sec. 3.2.8, massive amounts of data recorded during the Run 2 data-taking period as depicted in Fig. 3.2, and corresponding simulations require dedicated computing infrastructure. Over the last decades the computing demands of LHC experiments have gradually risen, which finally led to the foundation of the Worldwide LHC Computing Grid (WLCG) [81]. The WLCG connects multiple computing sites organized in four different tiers with (primarily) devoted duties: data acquisition and storage (tape archiving), event reconstruction, calibration and skimming, simulation and physics analysis.

The presented analysis further leverages the Visual Physics Analysis (VISPA) computing cluster [11], which provides software libraries and hardware accelerators for high-performance computing needed for fast analysis cycles and computing-demanding tasks, such as deep learning.

In the following two sections, the WLCG and VISPA computing infrastructures and their use for the presented analysis are described in order.

3.3.1 Worldwide LHC Computing Grid

The WLCG has a tiered architecture comprising approximately 170 geographically distributed computing sites across 40 different countries. It is designed to archive, distribute, and analyze tens of petabytes of data recorded by different LHC experiments per year. Four different

tier levels are operated with dedicated duties [81].

The Tier-0 computing site is located directly at CERN and shared by multiple LHC experiments. Here, data is gathered from the CMS experiment, processed in a first reconstruction step, and then archived to tape storage. From there, the data is distributed to the Tier-1 sites.

The Tier-1 computing centers provide additional archiving capacities, which leads to secure and robust data storage through site redundancy. In addition, offline reconstruction and calibration steps, as outlined in Sec. 3.2.8, are performed at these sites.

Tier-2 and Tier-3 computing centers are usually the smallest in terms of resources individually; however, there are significant more sites ($\mathcal{O}(100)$) of this type in comparison to the Tier-1 sites ($\mathcal{O}(10)$). Here, Monte Carlo simulation - only at Tier-2 sites - and physics analysis are performed and thus used by the majority of CMS users.

While the CERN IT department centrally manages the Tier-0 computing site, the others are operated, developed, and maintained by local universities and institutes. Typically, CMS users only interact with the Tier-2 and Tier-3 sites and can transparently address those using the CRAB workload management system [116], which can handle tens of thousands user jobs simultaneously.

This analysis uses simulated and recorded events, which have been produced, reconstructed, and calibrated with the WLCG resources. In addition, the conversion to the NANOAOD data format (cf. Sec. 3.2.9) has been centrally performed and provided by the CMS collaboration using the WLCG.

3.3.2 VISPA

The VISPA computing cluster currently comprises a few hundred logical CPU cores, at least 2 GB of random access memory (RAM) per core, about 30 GPUs, and approximately 100 TB of local file storage. These resources are split between 13 individual machines, each backed by SSD caches, enabling an enormous acceleration of repetitive access of data from the local file storage. Even though the total CPU power is significantly smaller than typical Tier-2 and Tier-3 computing centers, the resources are more than sufficient to perform the analysis presented in this thesis. Latencies of I/O operations are even diminished through the SSD caches mentioned earlier. In addition, the VISPA cluster enables significant acceleration through GPUs for the deep learning techniques used in the presented analysis. Finally, the NANOAOD data format significantly reduces the storage and CPU needs, enabling small-scale computing clusters as a viable alternative to the Tier-2 and Tier-3 computing centers.

The VISPA project is developed and maintained by the III. Physics Institute A at the RWTH Aachen University. A web-fronted [12], allowing users to connect and interact with the computing cluster transparently through a browser, facilitates the frequent use of VISPA for research, education, and outreach [117, 118]. In order to improve the web-based usage and thus deepen the possibility to share knowledge [119], the VISPA project is currently migrated towards a JUPYTERLAB [120] based web-frontend [121].

Chapter 4

Analysis Techniques

The analysis presented in this thesis leverages several state-of-the-art analysis techniques. These techniques are grouped into three different application fields: the processing of events, a physics process classification with deep learning, and the final measurement. The tools and techniques related to these three topics are described in this chapter.

The processing of events leverages a multitude of different software packages from the Scikit-HEP [122] and NUMPY [123] ecosystems. These packages are described in the first section of this chapter. They allow vectorization of event processing, i.e., simultaneous analysis of multiple events in a single thread, and scaling of this processing to a distributed computing cluster. This leads to a significant analysis speedup, which unveils a new bottleneck: the time spent on delivering data to the processing units (I/O latencies). Section 4.2 describes a caching mechanism to reduce these latencies. This mechanism has been published in Ref. [13], and was developed during the analysis presented in this thesis.

The analysis presented in this thesis uses deep learning to classify collision events into different physics processes. The mathematical foundation of deep learning, the training procedure of neural networks, and more advanced deep learning techniques are explained in Sec. 4.3.

Methods for the statistical inference are described in Sec. 4.4 to extract upper exclusion limits, best fit values with uncertainties, goodness-of-fit tests, and more. Finally, a software package, called HH INFERENCE TOOLS, is used to apply these methods and is described in Sec. 4.4.3. This package has been developed and maintained during the work of this thesis. The HH INFERENCE TOOLS have been used to produce the results of the majority of Run 2 HH analyses of the CMS experiment, including results presented in Ref. [69]. In this thesis they have been used to produce the results presented in chapter 8 and chapter 9.

4.1 Scientific Software Environment

This analysis utilizes software packages from the Scikit-HEP and NUMPY ecosystems. Each serves its own purpose following a highly polylithic software design. In this section all software packages and their purpose, which have been used in this analysis, are described in detail. These packages are provisioned on the VISPA cluster using the CONDA [124] package manager.

Workflow management: LAW

The software package LAW [125, 126] is used to automate each analysis step with full flexibility. It ensures reproducibility and reduces the number of human errors significantly. Each step is isolated in a so-called "task", which defines its inputs through

requirements of other tasks, its outputs, and how they are produced. In this way, a directed acyclic graph (DAG) of tasks is created, which is then executed in order by the LUIGI [127] package. Each task defines its own parameters, that are passed upstream through the DAG, enabling a highly interactive and customizable behavior of the analysis. In addition, the LAW package provides a simple interface to store and access remote data, which is used in this analysis to facilitate file transfers between the VISPA and the CERN computing clusters.

ROOT file reading and writing: UPROOT

The UPROOT [128] package solely serves the purpose of reading and writing ROOT [129] files. This is facilitated by casting ROOT's TSTREAMERINFO from a ROOT file to NUMPY arrays. UPROOT is then able to interpret these arrays as different type of objects through rather generic python classes (lightweight re-implementations of the corresponding ROOT classes). In addition, UPROOT is able to handle different compression algorithms, remote ROOT files (streaming), and to export data to multiple scientific libraries in Python, such as PANDAS [130, 131] or CUPY [132]. In this work, UPROOT is used to read the NANOAOD datasets, calibration files, and write one-dimensional histograms for the statistical inference.

Jagged array manipulation: AWKWARD-ARRAY

AWKWARD-ARRAY [133] is a software package to manipulate nested and variable-sized data with NUMPY-like idioms. This is particularly useful for LHC event data, since the amount of particles (per type) varies per event resulting in a so-called "jagged" data structure: *awkward-arrays*. The AWKWARD-ARRAY package allows to manipulate these data structures with compiled operations, allowing the usage of vectorization. In addition, custom behaviors are supported for *awkward-arrays*, e.g., four momenta properties. Besides, the previously mentioned UPROOT package can natively stream data from NANOAOD datasets into *awkward-arrays*.

Columnar HEP tools: COFFEA

The COFFEA [134] package uses the UPROOT and AWKWARD-ARRAY packages to provide a NUMPY-like interface to HEP collision events. While the package also includes many standard tools for physicists, such as histogramming and plotting, this analysis uses only the high-level data structure for NANOAOD files. This data structure consists of nested records of jagged *awkward-arrays* that automatically resolve complex relationships between collections in the NANOAOD files, such as mother-daughter relationships of generator particles. These underlying records feature four momenta properties and behaviors, which allow, e.g., the calculation of the invariant mass of two particles or the calculation of the ΔR between two particles. The interface for processing event blocks, scaling to computing clusters, and the tool for extracting jet energy corrections have been adapted and optimized to the needs of this analysis.

Computational heavy lifting: DASK and DASK-JOBQUEUE

The processing of chunks of events is scaled to the VISPA cluster with DASK [14, 15] and DASK-JOBQUEUE [16, 17]. In particular DASK-JOBQUEUE allows to deploy a DASK cluster on the HTCONDOR [135] cluster of VISPA. In a first step, the number of events of each file is read through local multiprocessing steered by DASK. Then, the full workflow is split into chunks of 10×10^5 events, that are submitted and processed in parallel by multiple DASK workers, each with certain memory and CPU requirements, running on the HTCONDOR cluster. Finally, their outputs are collected and merged by dedi-

cated DASK workers ("mergers"), which have no memory limitations. These steps are orchestrated by a DASK scheduler responsible for the coordination of all DASK workers.

Histogramming: BOOST-HISTOGRAM

The BOOST-HISTOGRAM [136, 137] library provides an object-oriented interface to multi-dimensional histograms. It is designed to operate memory efficient storage types in order to minimize CPU cache misses. In addition, BOOST-HISTOGRAM supports weighted input and a variety of axis types, such as categorical axes for, e.g., systematic variations or analysis categories. In this analysis workflow, histograms are the output of the main processing step, where analyzed events are binned into multi-dimensional histograms to reduce the initial dimensionality (number of events) significantly.

Visualization: MATPLOTLIB and MPLHEP

MATPLOTLIB [138] is a Python software package for visualizing data. It is highly integrated with other scientific libraries, that provide NUMPY-like data structures. The MPLHEP [139] package extends MATPLOTLIB to create publication-ready figures for the HEP community. This analysis uses both packages for all visualizations except for the ones related to the statistical inference.

Deep Learning: TENSORFLOW and TENSORFLOW SERVING

The TENSORFLOW [140, 141] software library with a NUMPY-like API for high performance numeric calculations and automatic differentiation. Operations are compiled into a static computing graph optimized with respect to hardware accelerators, such as GPUs. Thus, it is highly suitable for modern computing-demanding deep learning applications. TENSORFLOW provides a rich toolset for a vast variety of deep learning architectures, layers, activation functions, optimizers, etc. In this work, TENSORFLOW is utilized to train a physics process classification neural network. The trained network is served on a single GPU for evaluation to the DASK cluster with the TENSORFLOW SERVING [142] package.

4.2 Cache-Aware DASK Cluster

Parts of this section are published in Ref. "Fast Columnar Physics Analyses of Terabyte-Scale LHC Data on a Cache-Aware Dask Cluster" [13].

When developing a physics analysis, many different studies need to be performed. Examples include data-driven background estimations, efficiency measurements, training and evaluation of multivariate methods, and determination of systematic uncertainties. All of these studies typically require multiple processing of at least a significant portion of the data and simulations. In addition, analyses are subjected to an experiment-internal peer review process, which requires numerous further consolidation studies. Consequently, every data analysis is inevitably subjected to a large number of iterations.

The duration of an analysis cycle has considerably increased due to the very successful LHC operation and the associated growth of recorded data. Typical analyzed data volumes are in the order of terabytes (TB). Without further developments in analysis technology, the prospect of the LHC upgrade for high luminosities will again significantly prolong analysis cycles. Three key concepts are exploited in this work to compensate this increase and

improve the runtime of an analysis cycle.

The first concept tackles the way of processing events. While classically collision events are analyzed one after another, vectorized array operations can process multiple events simultaneously. The scientific Python ecosystem NumPy provides these vectorized array operations using the processor-specific "single instruction multiple data" (SIMD) instruction sets.

Secondly, the programming paradigm "MapReduce" [143] is a key concept for this project. Operations, such as selection and reconstruction, are mapped to subsets of collision events. Their partial output is then accumulated (reduced) to a single output. Dask orchestrates this paradigm on any computing cluster.

The third key concept is caching. Caching increases the efficiency of repeated data access. Here, a caching mechanism caches collision data on processor-near solid-state disks (SSDs). Subsequent analysis cycles benefit from this and show a substantial reduction in cycle time. The caching is facilitated for each worker by the FSCache available within the Linux kernel [144]. Once enabled for a particular network file system (NFS) mount-point, it operates transparently upon all I/O requests for files therein. In particular, data is cached at a page-size granularity (4 KB) which enables selective caching, i.e., of only the accessed branches of a `.root` file. Since all I/O operations (read and write) fill the cache, the occurrence of cache-trashing is minimized by only caching a certain volume, which is predominantly used for write-once read-often data. The cache is configured to store its contents on the SSDs of the workers, thus profiting from their superior data transfer rates.

Since each cache will only contain the contents of data requested by its worker at some point prior, it is of utmost importance to route such requests - or rather the jobs that cause these particular requests - in a cache-hit maximizing manner. This is done through a worker – job affinity mechanism, where each worker and job is identified in a reproducible manner. The identifier consists of 1) the worker by its hostname, 2) the job by the input file UUID, and 3) the range of the event numbers.

These identifiers are then uniformly mapped into a high dimensional bounded space by interpreting their cryptographic hash value (i.e., SHA512 [145]) as a vector of integers (i.e., $[0..255]^{64}$). For any pair of such values a distance D can be calculated as such: $D(\vec{a},\vec{b}) = \sum_i d(|a_i - b_i|)$ where $d(x) = \min(x, 256 - x)$. Each job is then assigned to the worker it has the smallest distance to, ensuring a reasonably even distribution. The assignment is not strict, allowing idle workers to steal jobs from busy workers. This concept is referred to as work-stealing. It avoids trailing jobs due to heterogeneous job runtimes, thus improving the overall runtime. Especially in the case of the addition or removal of workers, the affected jobs are redistributed homogeneously while avoiding the reallocation of all other jobs. Additionally, the allocation ratio of jobs between workers can be changed smoothly by including a worker-specific distance factor - which is used to equalize the workload despite the varying processing power of all the workers. Multiple users can participate and profit from the data caching when using the same files and affinity mechanism.

The performance of the VISPA computing cluster with on-worker SSD caching is measured for a subset of simulated datasets in the NanoAOD data format. In total, the read data amounts to 1439 GB, which corresponds to the event information of 1.05×10^9 events in the scope of a realistic Higgs pair production analysis. All datasets are compressed with the level ten Z-standard compression algorithm [115], which has been changed from NanoAOD's default compression in order to reduce the decompression time. Our benchmark consists of

multiple consecutive cycles. Throughout each cycle, 221 Dask workers carry out the processing with one thread and 1.5 GB RAM each. Fig. 4.1 shows the performance benchmark for ten cycles.
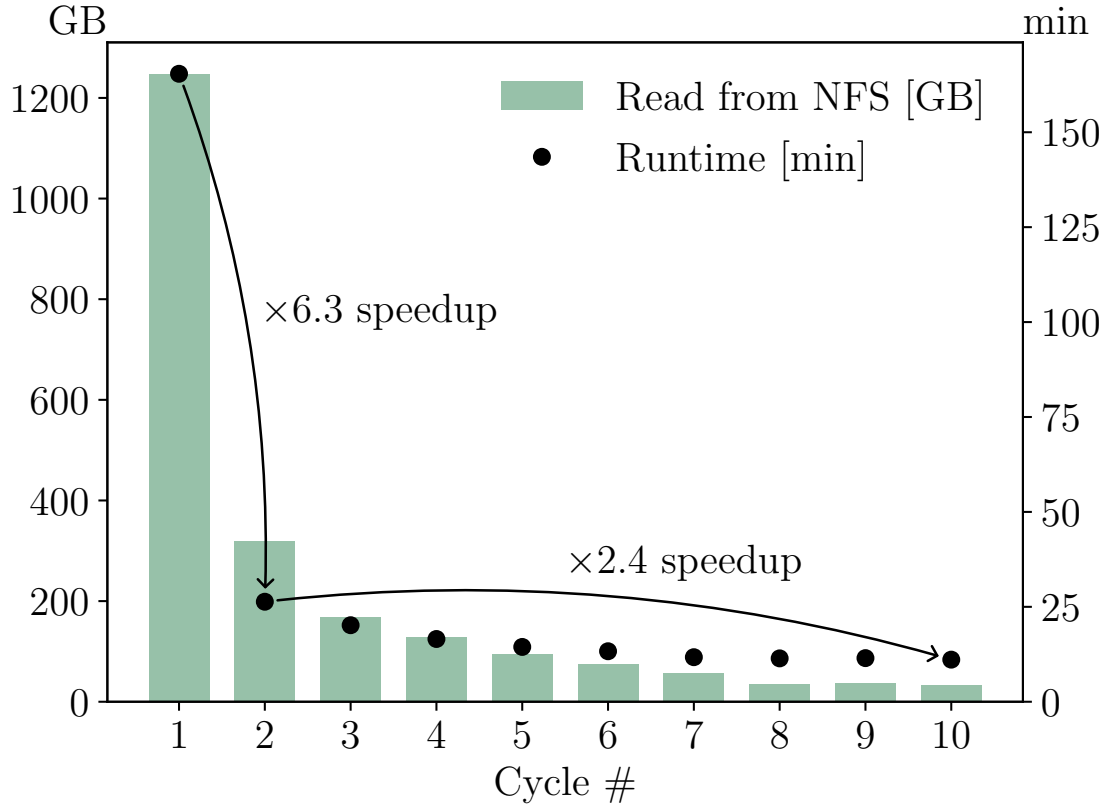
Figure 4.1: Performance benchmark results for ten consecutive cycles.

The key message is that the runtime decreases substantially for the first few cycles. In total, the improvement amounts to a factor of 14.9. The amount of data that is still read from the NFS of the VISPA computing cluster is vastly reduced as more data is read from the on-worker SSD caches. This effect converges for later cycles until almost all data is cached directly in the on-worker SSDs. The close overlay of runtime and the amount of data, which is still read from the NFS of the VISPA computing cluster, show a strong correlation between runtime reduction and caching. The cache usage gradually converges to a maximum since a work-stealing mechanism minimizes each cycle's runtime at the cost of slight degraded deterministic cache usage.

This analysis uses the above described cache-aware Dask cluster together with the resources of the VISPA computing cluster and has benefited substantially from the reduced runtime.

4.3 Deep Learning

The research field of deep learning is on the rise thanks to major successes in pattern recognition [146, 147], natural language processing [147–149], and self-driving cars [150]. The rapid progress in this field has been driven by the growing amount of data and the continuous

improvement of hardware accelerators, which led to the breakthrough of deep learning in the early 2010s.

In the presented analysis, deep learning is utilized to identify physics processes, which aims to isolate collision events involving Higgs boson pair production from other physics processes. This section focusses on the deep learning methods, which have been used in this analysis.

The structure of this section is as follows. First, the mathematical foundations of deep learning are introduced. Subsequently, the training process of deep neural networks is explained. Finally, the specifics of multi-class classification are discussed. The mathematical foundations are based on Ref. [151].

4.3.1 Mathematical Foundation

Neural networks connect input data to output data of arbitrary dimensions through a mapping. The most fundamental mapping is a linear model:

$$y(x) = W \cdot x + b, \tag{4.1}$$

where y denotes the output data, x the input data, W the weight matrix, and b the bias vector. The weight matrix and the bias vector often comprise a vast amount of adjustable parameters, which make neural networks known to be general function approximators. Figure 4.2 shows a visualization of such a linear mapping in the case of a three dimensional input vector and a two dimensional output vector. The arrows denote the six weights of the weight matrix and the two biases are depicted by the dark gray circles.
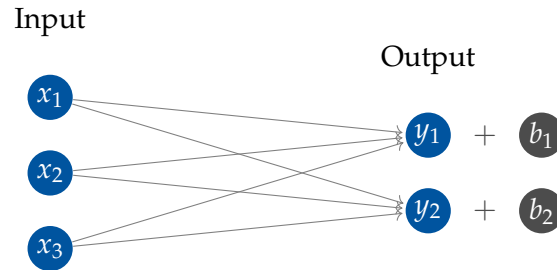
Figure 4.2: Sketch of linear mapping.

In general, processes in nature are non-linear requiring the extension of the linear mapping by at least one so-called "activation" function. This extension is defined as follows:

$$y(x) = \sigma \left(W \cdot x + b \right), \tag{4.2}$$

where σ denotes a non-linear function. Typical activation functions are depicted in Fig. 4.3.

This linearity-breaking component (Eq. 4.2) also allows to chain multiple mappings, whereas multiple nested linear mappings can always be rewritten as a single linear mapping. Such a chained architecture of multiple "layers" is shown in Fig. 4.4 defining a basic deep learning architecture, the so-called "fully-connected neural network". In general, deep learning architectures are referred to as "deep neural networks" (DNNs).

The activation functions, number of layers (depth), and number of nodes per layer (width) of a fully-connected neural network are part of the "hyperparameter" family. Such parameters

Common activation functions $\sigma(x)$
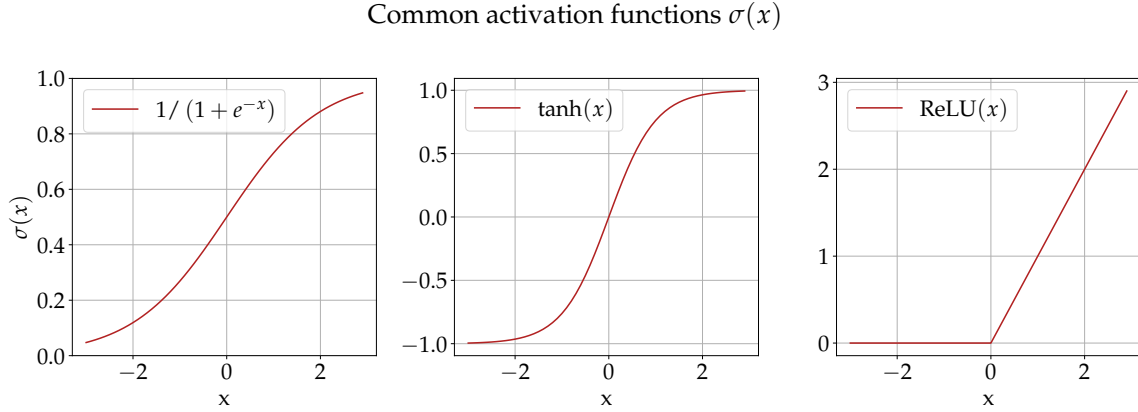
Figure 4.3: Common activation functions (left: "sigmoid", middle: "tanh", right: "ReLU").
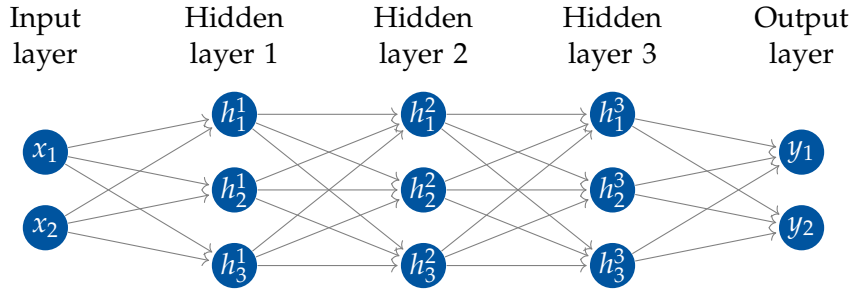
Figure 4.4: Sketch of a deep neural network.

are decided at the discretion of the user, and are usually domain-specific, i.e., they depend on the used dataset and learning task. The output layer for regression problems usually features a linear activation function in the last layer, whereas classification problems typically have a "softmax" (Eq. 4.3) activation function in the last layer:

$$\text{softmax}(x) = \frac{e^x}{\sum_{i=1}^{n} e^{x_i}}. \tag{4.3}$$

Hyperparameters are not learned during the training phase of a neural network; in contrast to the weights and biases, that are iteratively adjusted during the training phase through a gradient-based optimization procedure. This procedure is explained below.

4.3.2 Neural Network Training

A neural network is primarily characterized by its set of trainable parameters $\theta := \{W, b\}$ connecting the input vector to the output vector. Finding the most suitable values for these trainable parameters is of utmost importance for the performance of a neural network. At first, these parameters are initialized using a probability density function (pdf), usually a uniform or Gaussian distribution, enforcing the activations to have unit variance and zero mean prior training. This initial randomness in the model avoids the learning of the same parameters in each node. In "supervised" learning, which presupposes labeled data, the parameter adjustment procedure is based on quality measures between the neural network output prediction $y(x, \theta)$ and the truth $\hat{y}$. This measure is usually called "loss function" ($\mathcal{L}$);

common choices are the "mean squared error" (MSE) for regression tasks and the "cross-entropy" (CE) for classification tasks:

$$\mathcal{L}_{\text{MSE}} = \left(y(x,\theta) - \hat{y}\right)^2, \tag{4.4}$$

$$\mathcal{L}_{\text{CE}} = \sum_{j=1}^{k} -\hat{y}_j \log\left(y_j(x,\theta)\right), \tag{4.5}$$

where k denotes the number of classes for the $\mathcal{L}_{\text{CE}}$ loss function.

The optimization procedure for all trainable parameters relies on the "gradient descent" algorithm, which is based on the partial ("first order") derivatives of the loss function with respect to the trainable parameters. The gradient descent algorithm is defined as:

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta \mathcal{L}(\theta), \tag{4.6}$$

where α denotes the learning rate, and θ_t the trainable parameters at iteration t. The gradient descent algorithm is depicted in Fig. 4.5 for a linear regression.
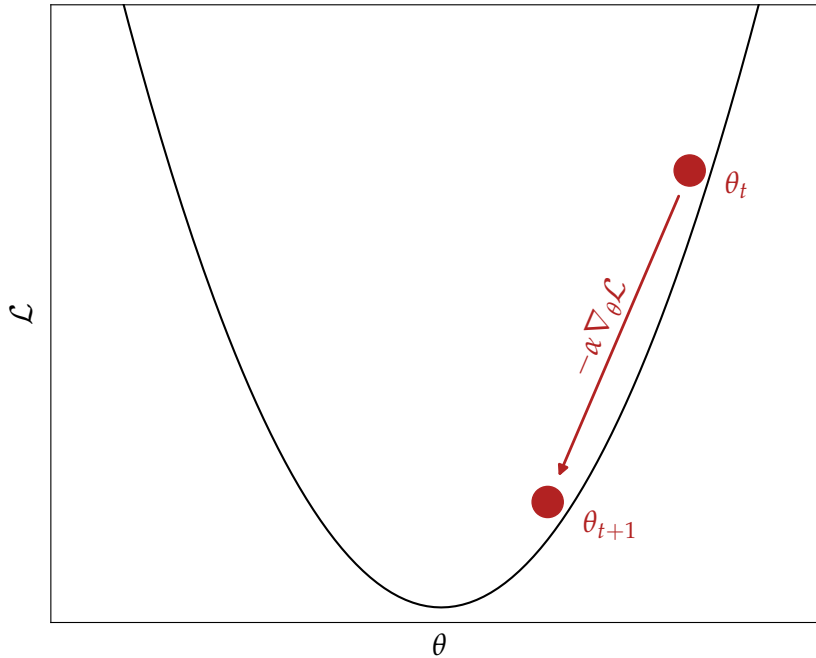

Figure 4.5: Sketch of the gradient descent algorithm for a linear regression.

The gradient descend algorithm updates the trainable parameters in the direction of the steepest descent of the loss function. This is done by iteratively adjusting the trainable parameters in the direction of the negative gradient of the loss function. The learning rate α

is a hyperparameter determining the step size of the algorithm. It is a rather sensitive hyperparameter, as too small learning rates might lead to a slow convergence or even prevent convergence to the minimum, whereas too large learning rates might overlook it and hence never reach it.

In practice, the gradient descent algorithm is usually applied to a small subset ("mini batch") of the input data. This modification of the algorithm is called "stochastic gradient descent" (SGD). The update of the parameters is calculated by the average of gradients of the mini batch:

$$\mathbb{E}\left[\nabla_\theta \mathcal{L}(\theta)\right] = \frac{1}{n}\sum_{i=1}^{n}\nabla_\theta \mathcal{L}_n(\theta), \tag{4.7}$$

where n denotes the mini batch size. The SGD algorithm has a few practical advantages: a fast gradient calculation on the mini batch in contrast to the computational burden for the whole dataset, and the fact that the mini batch is randomly sampled from the dataset helps the algorithm to overcome unwanted local minima.

Nowadays, optimization algorithms, such as ADAGRAD [152, 153], RMSPROP [154], and ADAM [155], incorporate in addition "adaptive learning rates" [156] and "momentum" [157] for faster and more stable convergence. These modifications improve the optimization procedure by leveraging information about gradients from previous iteration steps.

The gradients are calculated using backpropagation, a recursive procedure that calculates the gradients of the loss function with respect to the trainable parameters. Consider the loss function $\mathcal{L}$ of neural network with L layers, where the input layer is $l = 1$ and the output layer is $l = L$:

$$\mathcal{L} = \mathcal{L}\left(a_L\left(\theta_L, a_{L-1}\left(\theta_{L-1}\ldots a_1\left(\theta_1, x\right)\ldots\right)\right)\right) \tag{4.8}$$

$$\text{with}\quad a_l(\theta_l, x) = \sigma_l\underbrace{\left(W_l x + b_l\right)}_{z_l}\quad\text{and}\quad \theta_l = \{W_l, b_l\}, \tag{4.9}$$

where σ_l denotes the activation function of layer l, and θ_l the trainable parameters of layer l. The partial derivative can be calculated using the chain rule. Consider the following definition for the (last) layer l of a neural network:

$$z_l = W_l a_{l-1} + b, \tag{4.10}$$

$$a_l = \sigma_l(z_l), \tag{4.11}$$

$$\mathcal{L} = (a_l - y)^2. \tag{4.12}$$

The partial derivative of the loss function with respect to the trainable parameters of layer l is calculated as follows:

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial W_l} = \frac{\partial \mathcal{L}}{\partial a_l} \cdot \frac{\partial a_l}{\partial z_l} \cdot \frac{\partial z_l}{\partial W_l} = 2\left(a_l - y\right) \cdot \sigma'_l(z_l) \cdot a_{l-1} \tag{4.13}$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial b_l} = \frac{\partial \mathcal{L}}{\partial a_l} \cdot \frac{\partial a_l}{\partial z_l} \cdot \frac{\partial z_l}{\partial b_l} = 2\left(a_l - y\right) \cdot \sigma'_l(z_l) \cdot 1. \tag{4.14}$$

Since $\frac{\partial \mathcal{L}}{\partial a_l}$ (here: $\frac{\partial \mathcal{L}}{\partial a_l} = 2\left(a_l - y\right)$) and $\sigma'_l(z_l)$ can be computed very efficiently, the gradients can also be calculated efficiently. This algorithm can be applied to any layer l of the neural network, which allows to calculate a gradient through all layers backwards starting from the output layer. One update step for the trainable parameters comprises multiple steps. First, a full forward pass through the network is performed calculating the output a_L of the network, and the value of loss function $\mathcal{L}$. Then, the gradients are estimated by backpropagation. Here, the values a_l, a_{l-1}, and z_l are needed (see Eq. 4.14). In practice, they are cached after the first step (forward pass) and reused for the second step (backpropagation) enabling highly efficient computations. Finally, the trainable parameters are updated by the gradients according to Eq. 4.6.

One update step is performed on a previously described mini batch of the input data. The whole training procedure is repeated for a number of epochs, where an "epoch" denotes a full pass through the whole dataset. The training procedure is finished when the loss function converges to a minimum or when a user-defined maximum number of epochs is reached.

In order to avoid a statistical bias, the input data is usually split into three subsets: a training set, a validation set, and a test set. The training set is used to train the neural network, the validation set to evaluate the performance of the neural network during the training procedure and hyperparameter tests, and the test set for the final evaluation. A typical training procedure, including the training and validation set and the "generalization gap" between the neural networks performance on these sets, is shown in Fig. 4.6.

The validation set helps to detect "overfitting" of the neural network. Overfitting occurs, when the neural network learns statistical fluctuations of the training set. This becomes visible when the loss as a function of epochs improves on the training set, but worsens on the validation set as visible in Fig. 4.6 for $t > t_c$. A training is considered "successfully converged", when the training and validation loss have been minimized, are stagnating, and do not show signs of overfitting, i.e., at epoch t_c in Fig. 4.6. The test set is only used once, after the training procedure is finished.

4.3.3 Additional Techniques

This section describes additional techniques to improve, stabilize, and generalize the training and performance of neural networks. K-fold cross validation, regularization, and some solutions for the vanishing gradient problem are discussed. The section closes with a description of the Lorentz Boost Network (LBN) [158].

K-Fold Cross Validation

The classic dataset splitting into a training, validation, and test set comes with a pitfall: only parts of the input data are used for the training and the final application of the neural network. In addition, the validation set is typically chosen to be significantly smaller than the
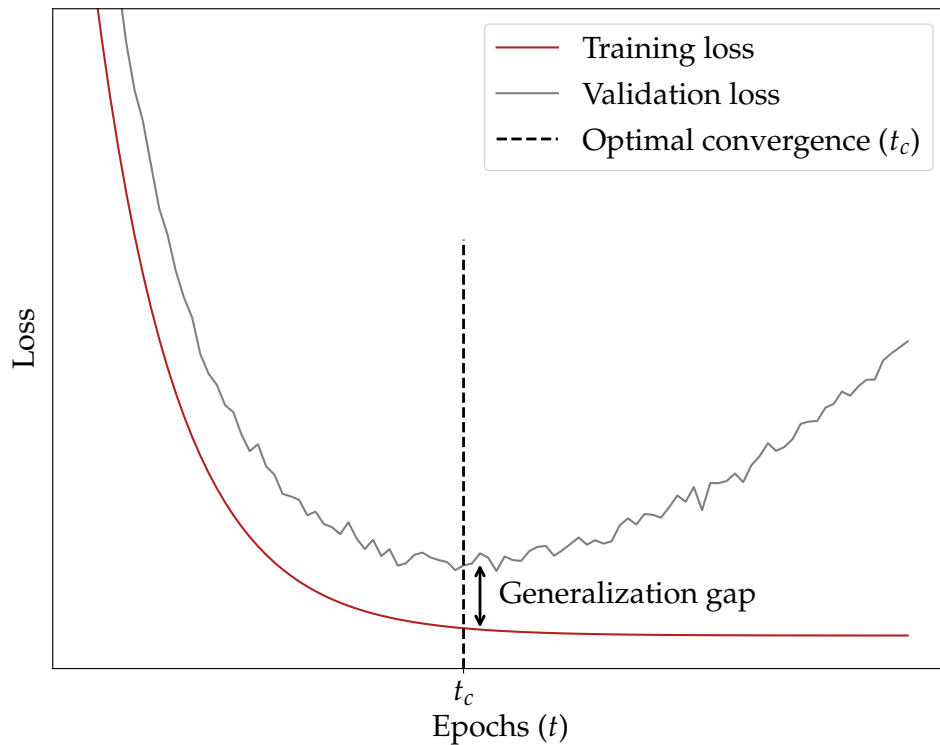
Figure 4.6: Sketch of the loss on the training and validation dataset as a function of epochs. Overfitting of the neural network is visible at $t > t_c$. The performance difference on the training and validation set is called "generalization gap".

training and test set in order to have the most data available for the training procedure and the final evaluation. This pitfall can be avoided by a splitting technique called "k-fold cross validation" described in Ref. [159]. The k-fold cross validation is shown in Fig. 4.7 for k=3.

The k-fold cross validation splits the data into k subsets, called "folds". The training procedure is repeated k times, where in each iteration a different fold is used as validation set and test set, whereas all remaining folds serve as the training set. This way the neural network can use the whole dataset for training, validation, and testing. The number of folds determines also the number of neural network trainings resulting in an averaged neural network performance among k different trainings, thus improving the generalization capacity.

Regularization

Regularization helps to avoid overfitting of neural networks. Several regularization techniques exist, of which the most common are parameter norm penalties, dropout, and early stopping. These regularization techniques are described in the following section.

Parameter norm penalties are penalty terms added to the loss function. These terms are primarily based on the weights of the network itself. Minimizing these prevents numerically large weights indicating overfitting. The two most common parameter norm penalties are the L1 [160] and L2 [161] norm penalties. Both extend the original loss function and are defined as:
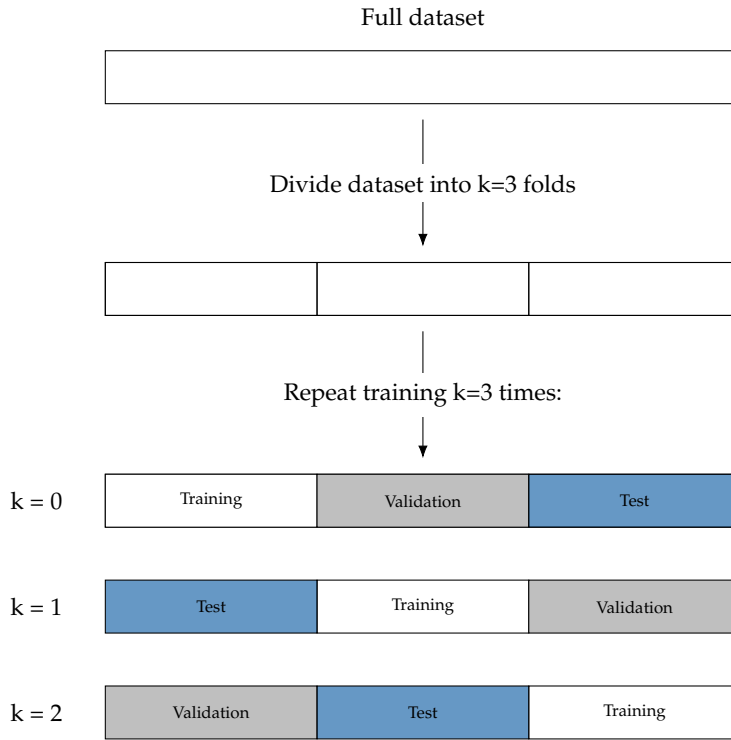
Full dataset

Figure 4.7: Sketch of k-fold cross validation with k=3. First the full dataset is split into three folds. Then, the training procedure is repeated three times, where in each iteration a different fold is used as training, validation, and test set. The final performance of the neural network is the average of the three iterations.

$$\text{L1 penalty:} \quad \mathcal{L}' = \mathcal{L} + \lambda \sum_{i=1}^{n} |w_i| \tag{4.15}$$

$$\text{L2 penalty:} \quad \mathcal{L}' = \mathcal{L} + \lambda \sum_{i=1}^{n} w_i^2, \tag{4.16}$$

where λ is a hyperparameter steering the strength of the penalty, and w_i are the weights of the neural network.

Dropout [162] is a technique, where during the training process random nodes are ignored (typically $10\% - 50\%$ of all nodes). Thus, in each pass the neural network "looks" slightly different resulting effectively in an ensemble training, which are less susceptible to overfitting. This mechanism is applied during training, for the validation and test dataset all nodes are used.

Early stopping is a very pragmatic regularization technique based on the fact that overfitting becomes visible when the validation loss increases again. In this case, the training process is stopped immediately and the neural network is used as it was trained in the previous epoch. This method prevents overfitting, but does not guarantee good training, since a good local minimum may not have been reached yet.

Vanishing Gradient Problem and its Solutions

The "vanishing gradient problem" might occur when training especially deep neural networks. The more layers (and thus parameters) a network comprises the more prone it is to receive vanishing small gradient updates, which in turn leads to a slow or even stagnating training. A few solutions to this problem are described in the following.

The choice of activation function can help to avoid vanishing (or even exploding) gradients. The hyperbolic tangent function has gradients between 0..1. Thus, chaining many layers with the hyperbolic tangent activation function leads to vanishing gradients, as these gradients are multiplied with each other according to the chain rule (see Eq. 4.14). In contrast, the ReLU activation is more robust with gradients of either 0 or 1.

Normalization of features within the neural network is another a common technique to avoid vanishing gradients. This technique is called "batch normalization" [163]. It applies a transformation to the output of a layer by subtracting the mean of that batch and dividing by its standard deviation. In this way, the output of the shift is centered around zero and has a standard deviation of one. This avoids numerical values in the extreme regions of typical activation functions (e.g. tanh or sigmoid), where the gradients are vanishingly small.

Another common approach to train very deep neural networks are residual neural networks [164] ("ResNets"). They consist of blocks of layers in which layer outputs are passed through residual shortcuts to later layers, essentially introducing shortcuts in the neural network. Such a shortcut, skipping two layers, is defined as follows:

$$z_l = \sigma \left(W_{l-1} \left[\sigma \left(W_{l-2} \cdot x + b_{l-2} \right) \right] + b_{l-1} + x \right), \tag{4.17}$$

where z_l defines the output of the l-th layer, σ the activation function, W and b the trainable parameters, and x the input to this block, which is additionally added to the output of the $l-1$-th block, creating a shortcut for x to bypass layers. These shortcuts can carry the gradient through many layers, without diminishing its numerical value excessively. Residual neural networks have become very popular after their highly successful application in computer vision challenges, such as winning the "ImageNet" [165] challenge in 2015.

Many more solutions, such as gradient clipping [166] and self-normalizing networks [167], exist to tackle this problem. However, these techniques were not used in the context of this thesis.

Lorentz Boost Network

Many successful deep neural network architectures are designed to exploit symmetries of the input data. A great example in pattern recognition is the translation invariance of convolutional neural networks [146]. In the field of particle physics, the Lorentz Boost Network (LBN) [158] architecture exploits the invariance of Lorentz transformations in the SM to autonomously learn high level features from the four momenta of input particles. The architecture of the LBN is shown in Fig. 4.8.

It creates global linear combinations through trainable weights between these particles. Two sets are defined, one for particle combinations and one for rest frames. Then, the LBN boosts pairwise combined particles into combined rest frames. Out of these boosted particle combinations new high level features are extracted, creating a so-called "physics embedding". The
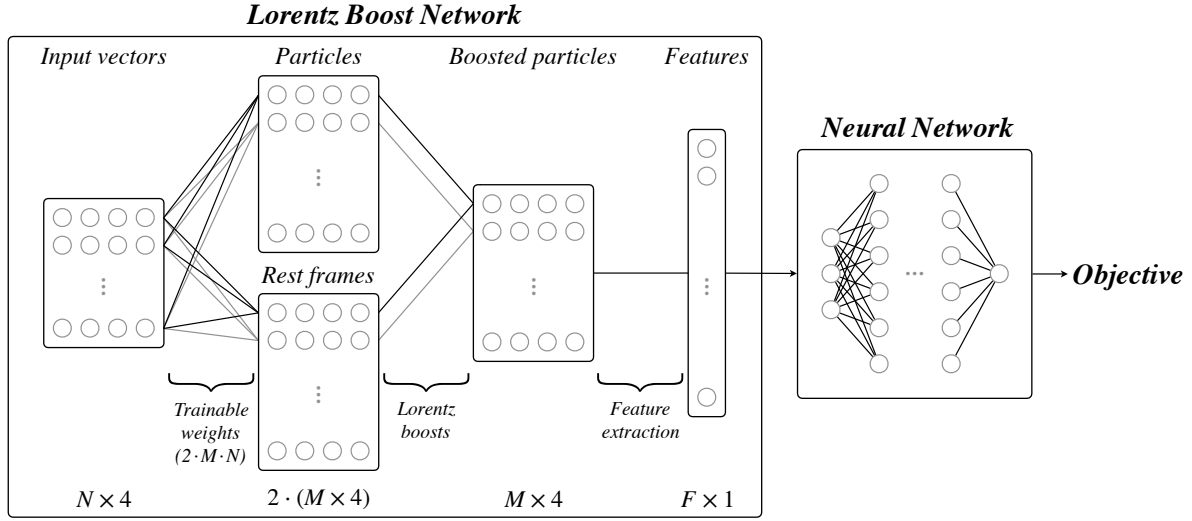
Figure 4.8: Sketch of the LBN architecture [158]. It comprises multiple layers starting from the input four vectors of final state particles of an event, then creating trainable particle and rest frame combinations, performing a lorentz boost between them, and finally extracting features from these boosted particles.

new high level features are the four momenta, p_T, p, and the invariant mass of each boosted particle combination, as well as $\cos \alpha := (\vec{p}_i \times \vec{p}_j) / (||\vec{p}_i|| \times ||\vec{p}_j||)$ of all pairs i, j of boosted particle combinations.

4.4 Statistical Inference

In order to measure the Higgs boson pair production, statistical methods are leveraged. Explicitly the signal strength parameter $\mu = \sigma/\sigma_\text{SM}$ and the coupling measurements are inferred from a maximum likelihood fit between an expected model based on simulation and recorded data. This fit includes systematic uncertainties that are subject to physics-inspired priors and correlations, and statistical uncertainties.

First, the construction of the likelihood function is described. This section comprises the description and treatment of different types of uncertainties, which are included in the likelihood function as nuisances parameters.

Then, the method to measure the signal strength parameter and the coupling measurements is presented. In addition, the concept of "upper limits" is introduced, which is commonly used in searches for new physics. The section of statistical methods finishes with the description of goodness-of-fit tests, that are used to quantify the agreement of the expected model and measured data.

The last section describes the toolkit used to perform the statistical inference.

4.4.1 Likelihood Construction

The likelihood observing the data count d given a model expectation λ for an n bin experiment is the product of Poisson distributions per bin and is defined as follows:

$$\mathcal{L}\left(\lambda|d\right) = \prod_{i}^{n} \underbrace{\frac{\lambda_i^{d_i}}{d_i!} e^{-\lambda_i}}_{\text{Poisson of bin } i} \quad . \tag{4.18}$$

The expected model λ can be further decomposed in a signal s and a background b expectation typically estimated from simulation, of which the signal is scaled by the free floating signal strength μ:

$$\lambda_i = \mu \cdot s_i + b_i. \tag{4.19}$$

In addition, the expectations s_i and b_i are subject to a set of statistical and systematic uncertainties, which may affect any combination of processes and bins with a certain strength. These are introduced as nuisance parameters θ, modeled by prior distributions π, and are inferred from physics-inspired measurements and predictions. The likelihood function is extended to:

$$\mathcal{L}\left(\mu, \theta|d\right) = \prod_{i}^{n} \frac{\left(\mu \cdot s_i\left(\theta\right) + b_i\left(\theta\right)\right)^{d_i}}{d_i!} e^{-\left(\mu \cdot s_i(\theta) + b_i(\theta)\right)} \cdot \prod_{j}^{p} \pi_j\left(\theta_j\right), \tag{4.20}$$

where θ is the set of all nuisance parameters, and p the total number of nuisance parameters.

For this thesis, three different types of nuisance parameters are of importance: rate-changing nuisances, shape-changing nuisances, and statistical uncertainties, discussed hereafter.

Rate-Changing Nuisances

Rate-changing nuisance parameters only affect the normalization of one or multiple processes. They can vary the yield of the affected processes, without alternating the shape of the corresponding histograms. These nuisance parameters are implemented with log-normal prior distribution with a width corresponding to the strength of the normalization uncertainty:

$$\pi\left(\theta\right) = \frac{1}{\sqrt{2\pi}\sigma\theta} \cdot \exp\left[-\frac{(\ln \theta)^2}{2\sigma^2}\right], \tag{4.21}$$

where σ denotes the width of the distribution. In contrast to the natural choice of a Gaussian prior, a log-normal probability density distribution avoids negative values for large widths. This prevents unphysical results, such as negative event counts.

Shape-Changing Nuisances

In contrast to rate-changing nuisance parameters, shape-changing nuisance parameters can additionally change the shape of histograms, i.e., they affect a set of bins with different strengths in potentially different directions simultaneously. They are derived by three reference points for each of these nuisance parameters: the nominal histogram, and the $\pm 1\sigma$ variations (often referred to as "up" and "down" variations) of the corresponding systematic uncertainty. The likelihood function (Eq. 4.20) demands a continuous functions for the

nuisance parameters, which is realized by a dedicated interpolation strategy [168] using the aforementioned reference points. Between the -1σ and $+1\sigma$ a quadratic interpolation is used, whereas a linear extrapolation is used outside of this range.

In rare occasions the interpolated function may lead to negative event counts in some bins. This is avoided by using a clipping strategy, that sets the event count of affected bins as:

$$\text{Bin yield} = 10^{-5} \cdot \min\left(1, \max\left[10^{-10}, \mathcal{N}\right]\right).$$ (4.22)

This strategy sets the bin yield to a value larger than zero, but smaller than 10^{-5}. In addition, it scales with the integral of the nominal histogram ($\mathcal{N}$), and thus, avoids an artificial injection of events. Such injection are in the order of 10^{-5} events expectation per bin (i.e., bin counts of signal processes).

Histogram variations with extremely large statistical fluctuations may become problematic for the fit and lead to unphysical results. This can happen if the estimation of the "up" and "down" variations stem from a smaller set of simulated events. In these cases, the variations are smoothed using the "353QH twice" algorithm [169]. The affected histograms are identified with the following χ^2 test:

$$\chi^2 = \frac{1}{N} \cdot \sum_i^N \left(\frac{v - n}{\sigma_v}\right)^2 < 1,$$ (4.23)

where N denotes the total number of bins, v (n) the bin content of the varied (nominal) histogram, and σ_v the statistical fluctuation of the varied bin. In case the χ^2 test statistic yields a numerical value smaller than one, both "up" and "down" variations are smoothed, even if only one of them triggered the χ^2 condition.

Statistical Uncertainties

Statistical uncertainties of data are included in the likelihood function as a Poisson error $\sigma_{\text{data, stat.}} = \sqrt{N}$. The picture is different for simulation as one simulated event does not always correspond to one expected event in a bin. With the help of event weights, the expected event yield and its statistical uncertainty in a bin reads:

$$N_{\text{exp.}} = \sum_i w_i,$$ (4.24)

$$\sigma_{\text{stat.}} = \sqrt{\sum_i w_i^2},$$ (4.25)

where $N_{\text{exp.}}$ denotes the expected event count from simulation, w_i the event weight of the i-th event, and $\sigma_{\text{stat.}}$ the statistical uncertainty. This introduces the statistical uncertainty as an additional nuisance parameter in the likelihood function for each process and bin, resulting in $N_{\text{bins}} \cdot N_{\text{processes}}$ nuisance parameters in total. In order to reduce the amount of nuisance parameters in this case, the "Barlow-Beeston"-lite [170] approach is used. It simplifies the statistical uncertainties within one bin to a single one with a Gaussian or Poisson

prior distribution in case a minimal amount of events is present in that bin, typically collapsing $N_{\text{bins}} \cdot N_{\text{processes}}$ nuisance parameters to at least N_{bins} nuisance parameters in total. In this analysis, the threshold for the minimal amount of events per bin is set to 10.

4.4.2 Statistical Methods

The starting point of the statistical inference is a binned summary statistic (histogram), one for recorded data and one for the expectation. The expectation is based on a model consisting of a background and a signal hypothesis typically estimated from simulation. The signal strength modifier μ scales the contribution of the signal hypothesis following Eq. 4.19. The compatibility between the "signal+background" ($\mu \neq 1$) and the recorded data is of major interest for the validity of the expectation model. In these searches exclusion limits are set based on the CL_s method [171, 172] commonly used among the CMS and ATLAS collaborations.

Exclusion limits

Exclusion limits are set with the help of the profiling likelihood ratio q_μ, which is the best discriminator between two hypotheses according to the "Neyman-Pearson lemma" [173]:

$$
q_\mu = -2 \cdot \ln \frac{\mathcal{L}\left(\mu, \hat{\hat{\theta}}|d\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\theta}|d\right)} \quad \text{with} \quad \hat{\mu} \leq \mu, \tag{4.26}
$$

where $\hat{\mu}$ and $\hat{\theta}$ maximize the likelihood function $\mathcal{L}$ in Eq. 4.20, and $\hat{\hat{\theta}}$ maximizes the likelihood function for a fixed μ. In order to obtain one-sided (upper) exclusion limits, μ has to be larger than $\hat{\mu}$. The profiling likelihood ratio q_μ is then used to calculate the CL_s values as a function of the signal strength modifier:

$$
\text{CL}_s = \frac{p_\mu}{1 - p_{\mu=0}} \quad \text{with} \quad p_\mu = \int_{q_\mu}^{\infty} p\left(q_\mu|\mu, \hat{\theta}\right) \, \mathrm{d}q, \tag{4.27}
$$

where p_μ is the p-value as a function of the signal strength modifier μ, and $p\left(q_\mu|\mu, \hat{\theta}\right)$ its underlying probability density function (pdf). This probability density function is obtained by sampling a sufficient number of data points (toys) from the full uncertainty model or by approximating this pdf with an asymptotic formulae [174]. The upper limit on the signal strength modifier is then calculated by solving Eq. 4.27 for μ with $\text{CL}_s = 0.05$.

High energy physics analyses typically introduce two kinds of upper limits: expected and observed upper limits. While the observed upper limits are calculated with real data, the expected upper limits are calculated by using a synthetic dataset [174] often referred to as "Asimov" dataset. By using the "Asimov" dataset, the sensitivity of an analysis can be determined without looking into the recorded data (blinded analysis), and thus avoids biases.

Profiled maximum likelihood fit

It is crucial to determine the best fit values $\hat{\mu}$ and $\hat{\theta}$ for the signal strength modifier and the nuisance parameters, respectively. The best fit values are found by maximizing the likelihood function, fulfilling the condition:

$$\hat{\mu}, \hat{\theta} = \arg\max_{\mu, \theta} \mathcal{L}\left(\mu, \theta | d\right). \tag{4.28}$$

In practice, the likelihood function is transformed into a negative log-likelihood function (NLL), where the best fit values are estimated by minimization instead. The negative log-likelihood function reads:

$$\mathrm{NLL} = -2\Delta \ln \mathcal{L}\left(\mu, \theta | d\right). \tag{4.29}$$

The NLL is scaled by a factor of 2 in order to transform the NLL distribution asymptotically into a χ^2 distribution. Besides, the NLL is corrected from an offset, which is denoted by Δ. With these transformations the NLL distribution can be used to extract the best fit value, and its one and two sigma uncertainty. The conditions for the measurement with uncertainty for the signal strength modifier with its uncertainties are:

$$\mathrm{NLL}(\hat{\mu}) = 0, \tag{4.30}$$
$$\mathrm{NLL}(\hat{\mu} \pm 1\sigma_{\hat{\mu}}) = 1, \tag{4.31}$$
$$\mathrm{NLL}(\hat{\mu} \pm 2\sigma_{\hat{\mu}}) = 4. \tag{4.32}$$

In reality, the simultaneous minimization with respect to μ and θ is often computationally expensive, especially in presence of a large number of nuisance parameters. Therefore, the concept of "profiling" is introduced, where for every value of μ the set of best fit values $\hat{\hat{\theta}}$ for the nuisance parameters is measured. This way, the NLL is not minimized with respect to all parameters, but only with respect to the nuisance parameters. In the resulting χ^2 distribution constructed from multiple values of μ, the best fit value and its uncertainties can be read off according to the conditions in Eq. 4.30. This procedure works likewise for each nuisance parameter. Two measures are introduced to validate the quality of the fit and the estimation of the corresponding nuisance parameter, called "pull" and "constrain":

$$\mathrm{pull} = \frac{\hat{\theta} - \theta_{\mathrm{pre\text{-}fit}}}{\Delta\theta_{\mathrm{pre\text{-}fit}}}, \tag{4.33}$$

$$\mathrm{constrain} = \frac{\Delta\theta_{\mathrm{post\text{-}fit}}}{\Delta\theta_{\mathrm{pre\text{-}fit}}}, \tag{4.34}$$

where $\theta_{\mathrm{pre\text{-}fit}}$ denotes the initial value of the nuisance parameter ("pre-fit"), $\hat{\theta}$ the best fit value ("post-fit"), and $\Delta\theta_{\mathrm{pre\text{-}/post\text{-}fit}}$ their uncertainties. The pull measures the deviation between the pre-fit estimation and the best fit value of the nuisance parameter in units of the pre-fit standard deviation. The pulls of all nuisance parameters are expected to follow a Gaussian distribution, which motivates most pulls to lie between -2σ and $+2\sigma$. Outliers can occur, but should be investigated to assure that their prior estimation is reasonable. The constrain is the relative difference between the uncertainty of the pre-fit estimation of the nuisance parameter, and the measured uncertainty of the best fit value. The pre-fit uncertainty has a width of -1 to $+1$. The post-fit uncertainty can be measured more precisely with the help of

dedicated control regions, leading to a smaller uncertainty on the final measurement. This effect is called "constrainment". Vice-versa, the post-fit uncertainty might also be larger than the pre-fit estimation, which hints that the pre-fit estimation was likely too optimistic. This is often referred to as a nuisance parameter "release".

Saturated goodness of fit test

A good agreement between recorded data and the expectation, especially in control regions, is essential to validate the quality of the expectation and the uncertainty model. The level of agreement can be quantified using "goodness of fit" (GOF) tests. A variety of GOF tests exist, as there is no optimal test statistic according to the "Neyman-Pearson lemma" due to the absence of an alternative hypothesis. In this analysis, a generalization of the χ^2 test [175] is used, that is applicable to histograms.

The underlying test statistic for this GOF test is based on the following likelihood ratio of a Gaussian distributed data model and a "saturated" model [176]:

$$q = \frac{\mathcal{L}}{\mathcal{L}_{\text{saturated}}}, \tag{4.35}$$

$$\text{with} \quad \mathcal{L} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(d_i - \mu_i)^2}{2\sigma_i^2}\right) \quad \text{and} \tag{4.36}$$

$$\mathcal{L}_{\text{saturated}} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}}, \tag{4.37}$$

where d_i mark the observation, μ_i the expected mean, and σ_i the standard deviation of the i-th bin with a total of N bins. The saturated model is obtained when setting the observation to the expected mean for each bin. The likelihood ratio is then transformed into a χ^2 distribution:

$$q = \prod_{i=1}^{N} \exp\left(-\frac{(d_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{4.38}$$

$$\rightarrow \chi^2 = -2\ln q = \sum_{i=1}^{N} \left(\frac{d_i - \mu_i}{\sigma_i}\right)^2. \tag{4.39}$$

In practice, the χ^2 distribution is obtained by drawing pseudo-experiments ("toys") from the post-fit expectation and its uncertainty model. For each toy the χ^2 function from Eq. 4.39 is evaluated by substituting d_i with the toy expectation, and using the post-fit expectation for μ_i. Likewise, one χ^2 value with recorded data can be calculated. Using the χ^2 distribution estimated with toys and the χ^2 value obtained from data, one can infer a p-value p, which quantifies the agreement between the expectation and the data. Typically, a threshold of at least $p \geq 0.05$ is used to quote a sufficiently good agreement. Very large p-values ($p \geq 0.95$) indicate a too conservative uncertainty model.

4.4.3 Higgs Pair Production Inference Tools

Higgs pair production analyses in the CMS experiment aim to measure the same coupling strength modifiers based on the techniques described in the previous section. Therefore, a common software package [18] with extensive documentation [19] called HH INFERENCE TOOLS has been developed for use with all Higgs pair production analyses. The HH INFERENCE TOOLS minimize the potential for individual errors, unify extraction and visualization techniques, and automatically ease the process of combining multiple analyses.

This is achieved by a mutual data structure of inputs ("datacards"). From there on, a pipeline of inference steps is applied with the LAW software package. The computational workload is carried out by the Higgs combine package [177] developed by the CMS collaboration. Defining standardized pipelines with LAW allows to scale these computations to distributed computing clusters and reuse as many intermediate computations as possible.

This toolbox contains pipelines for the statistical methods described in Sec. 4.4.2 and more: extracting one-sided confidence intervals (upper limits) from hypothesis tests, measuring couplings and nuisance parameters (likelihood profiles), measuring discovery strengths (significances), and many diagnostic measures for the quality of the fit performed. The latter include goodness-of-fit tests, differences between priors and posteriors ("pulls" and "constrains"), impacts of nuisance parameters on fit results, and correlations between parameters. In addition, the HH INFERENCE TOOLS provide a set of scripts for the manipulation of datacards, and the visualization of systematic uncertainties and event counts. Finally, two interactive browser-based viewers are provided to inspect potentially large datacards and correlation matrices.

Currently, the HH INFERENCE TOOLS are actively used throughout (almost) all Run 2 Higgs pair production analyses and their combination. They have already been successfully used to produce parts of the results presented in Ref. [69].

Chapter 5

Analysis Strategy

The presented analysis searches for HH production in the dileptonic $b\bar{b}W^+W^-$ decay channel. The considered production modes are gluon-gluon fusion (HH(GGF)) and vector boson fusion (HH(VBF)). The final state is composed of two leptons, two neutrinos, and two bottom quarks. Only leptonic final states, including e and μ, are considered, which results in three decay channels: $\mu\mu$, ee, and eμ. Both neutrinos cannot be measured by the CMS detector; instead, the missing transverse energy E_T^{miss} is used in this analysis.

This final state particle composition is the same as for the dileptonic $t\bar{t}$ + Jets process, which constitutes for the most dominant background process in this search. Also, single top production and Drell-Yan processes add large background contributions. The latter one and backgrounds, which arise from misidentified leptons, are estimated with data-driven methods. These estimations are described in detail in Secs. 6.5 and 6.6.

The challenge of this analysis is the small number of signal events, which vanish under an overwhelming amount of background events. In order to deal with this challenge, the analysis strategy is based on four steps, which are depicted in Fig. 5.1 and described in the subsequent paragraphs.

First, the general event selection is chosen to be as inclusive as possible, in order to keep as many signal events as possible. Second, the remaining events are analyzed and characterized by a deep learning based physics process classification [178]. Third, data-driven background estimations are used to improve the modelling of background processes. In addition, a binning optimization of the distribution of the sensitive variable is employed to enhance the sensitivity of the signal extraction. Finally, the measurement is performed with a binned maximum likelihood fit, as described in Sec. 4.4.2.

The data size of each consecutive step is significantly reduced, starting from approximately $\mathcal{O}(10\ \text{TB})$ in the first step to $\mathcal{O}(1\ \text{MB})$ in the last step. This is a common pattern for big data analyses, which requires dedicated software to deal with such large amounts of data. In this analysis, the reduction is facilitated using the techniques and tools as described previously in Secs. 4.1 and 4.2.

This chapter describes the challenges of the HH $\rightarrow b\bar{b}W^+W^-$ process in this analysis. Then, the physics process classification and the resulting event categorization is explained. Afterwards, the likelihood construction for Higgs pair production is explained. Finally, the analysis steps used to measure signal strength modifiers and coupling modifiers are summarized in order.
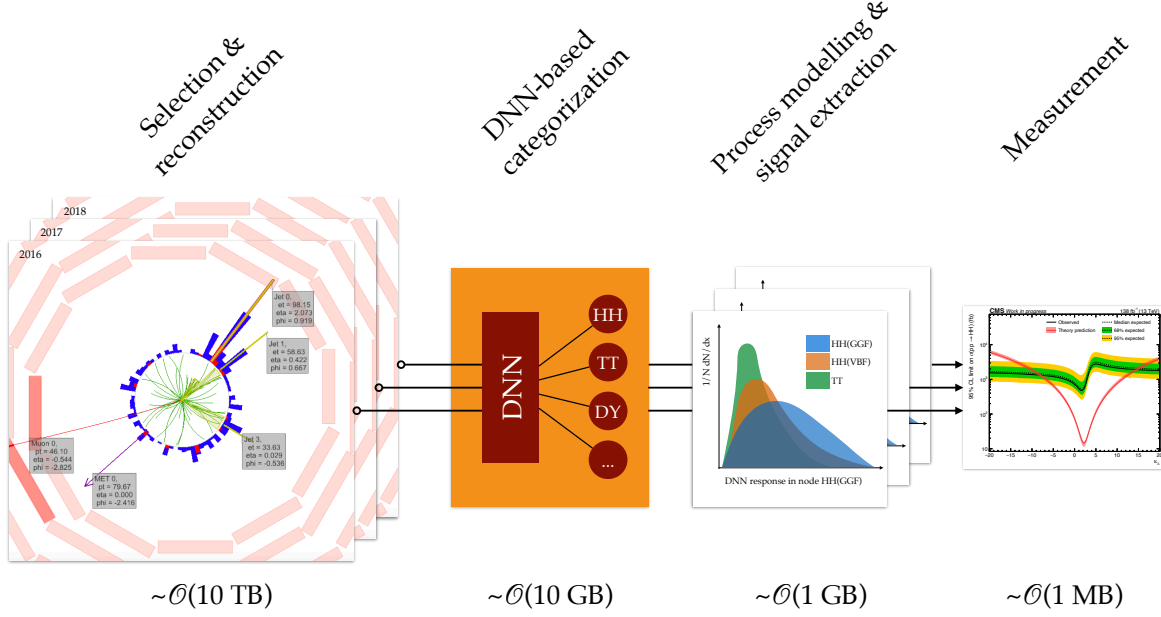
Figure 5.1: Overview of the (simplified) four-step strategy of this analysis.

5.1 Challenges of the $HH \rightarrow b\bar{b}W^+W^-$ Analysis

The production cross sections of the signal and background processes of this search are multiple magnitudes apart from each other (cf. Sec. 2.2.3). Figure 5.2 shows the latest cross section measurement results and the corresponding theory prediction for selected SM processes, including the inclusive HH process and the aforementioned background processes.

This leads to a large number of events originating from background processes, while only very few signal events are expected considering SM predictions. Table 5.1 shows the expected number of signal events for the different data-taking periods and the different production modes. In total, 4509 SM HH events are expected to be produced during the data-taking periods from 2016, 2017 and 2018 with a total amount of recorded data of $\mathcal{L}_{\text{Run 2}} = 137.60\,\text{fb}^{-1}$ [180–182]. These 4509 events are made of 4272 events coming from the HH(GGF) production mode ($\sigma_{\text{HH(GGF)}} = 31.05\,\text{fb}$ at NNLO precision [183]), and of 237 events coming from the HH(VBF) production mode ($\sigma_{\text{HH}(VBF)} = 1.726\,\text{fb}$ at N3LO precision [184]). This search primarily considers the dileptonic final state of the $HH \rightarrow b\bar{b}W^+W^-$ process, which corresponds to a $\mathcal{B} = 2.64\%$. Therefore, 112 signal events from the HH(GGF) production and 6 signal events from HH(VBF) production mode are expected to be recorded for the Run 2 data-taking period.

The expected amount of background events for the dileptonic $t\bar{t}$ + Jets process for the combined Run 2 data-taking period amounts to 114.450.176 expected events ($\mathcal{O}(10^8)$). The additional background processes contribute a smaller but still significant amount of events to this analysis. The great difference in expected events between signal and background processes has as a consequence that the modelling of the background processes needs to be extraordinarily accurate. Moreover, the event reconstruction is rather challenging to due the presence of two neutrinos and the broad width of the $H \rightarrow b\bar{b}$ mass resonance peak.

The physics processes are grouped together based on kinematic similarities, which result in

Figure 5.2: Latest cross section measurement results by the CMS collaboration and corresponding theory prediction for the SM processes, including the inclusive HH signal process and the main background processes of this analysis [179].
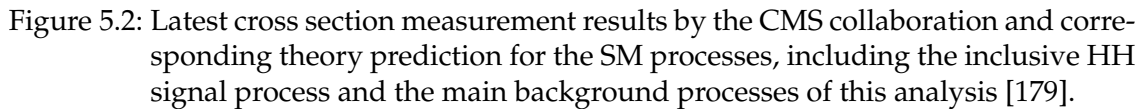
Table 5.1: Expected event numbers, considering SM cross sections, for the inclusive HH signal and the dileptonic final state of the signal process $HH \to b\bar{b}W^+W^-$ in the different data-taking periods of the CMS experiment. The values are calculated using $\mathcal{B}_{HH \to b\bar{b}W^+_\ell W^-_\ell} = 2.64\%$, cross sections for HH(GGF) (at NNLO precision) and HH(VBF) (at N3LO precision) from [183, 184] and the luminosities of the data-taking periods [180–182].

Year ($\mathcal{L}[\,\mathrm{fb}^{-1}]$)	Expected events
2016 (36.33)	1128.05
2017 (41.53)	1289.51
2018 (59.74)	1854.93
Run 2 (137.60)	4272.48

(a) $\sigma_{HH(GGF)} = 31.05\,\mathrm{fb}$.

Year ($\mathcal{L}[\,\mathrm{fb}^{-1}]$)	Expected events
2016 (36.33)	29.82
2017 (41.53)	34.09
2018 (59.74)	49.04
Run 2 (137.60)	112.95

(b) $\sigma_{HH(GGF)} \times \mathcal{B}_{HH \to b\bar{b}W^+_\ell W^-_\ell} = 0.82\,\mathrm{fb}$.

Year ($\mathcal{L}[\,\mathrm{fb}^{-1}]$)	Expected events
2016 (36.33)	62.71
2017 (41.53)	71.68
2018 (59.74)	103.11
Run 2 (137.60)	237.50

(c) $\sigma_{HH(VBF)} = 1.726\,\mathrm{fb}$.

Year ($\mathcal{L}[\,\mathrm{fb}^{-1}]$)	Expected events
2016 (36.33)	1.66
2017 (41.53)	1.90
2018 (59.74)	2.73
Run 2 (137.60)	6.28

(d) $\sigma_{HH(VBF)} \times \mathcal{B}_{HH \to b\bar{b}W^+_\ell W^-_\ell} = 0.046\,\mathrm{fb}$.

similar distributions of the signal extraction with deep neural networks (see Sec. 7.5).

A strategy to achieve a well-defined background modelling and signal extraction is based on the physics process classification, which is described in the following section.

5.2 Physics Process Classification

The physics process classification helps to simultaneously identify signal events, separate them from background events, and create regions in the phase-space mutually enriched by particular physics processes. The method is shown in Fig. 5.3.
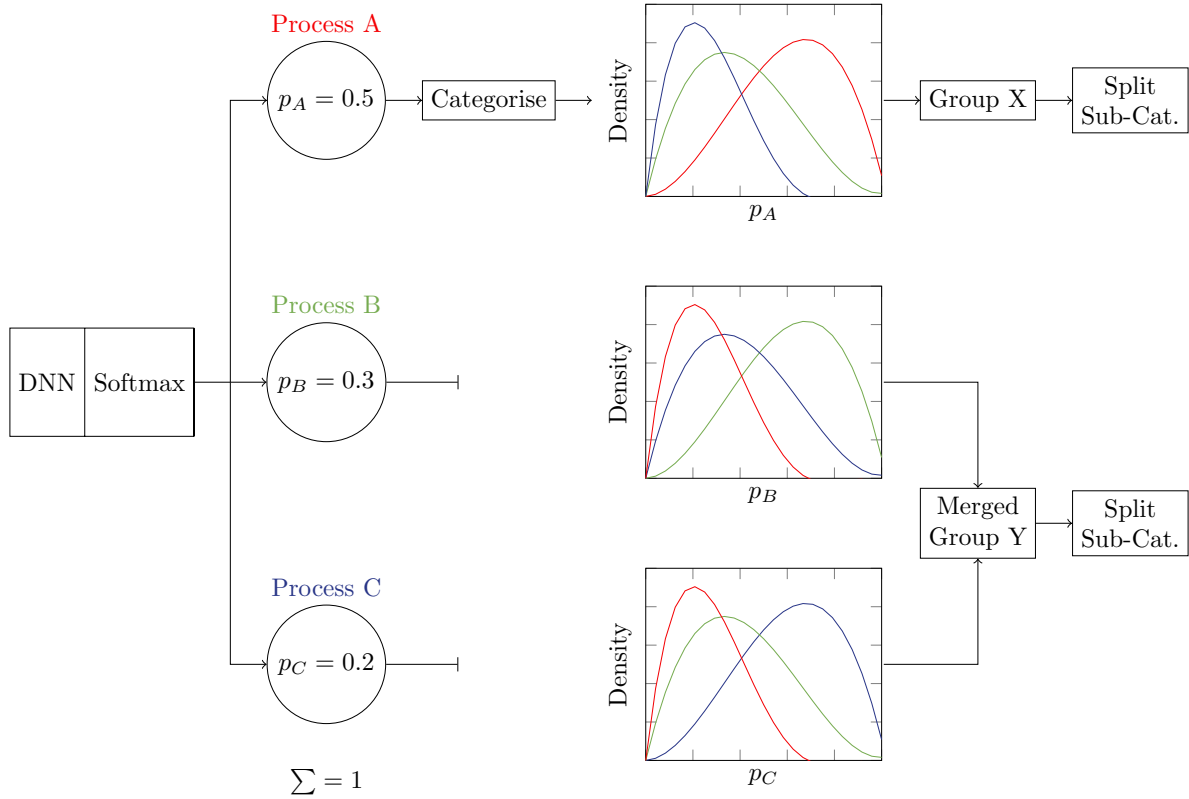
Figure 5.3: The physics process classification strategy.

The physics process classification analyzes each event as described in the following. Initially, an event passes through the DNN, which uses a softmax activation function in the last layer. The softmax activation function ensures that the DNN output is a normalized probability-like distribution, i.e., it predicts the probability for each event to belong to a certain physics process. The maximum probability is then used to assign an event to a physics process. This assignment is unambiguous and thus avoids double counting of events.

The resulting categories are then further split based on the number of b-quark jets and whether these b-quark jets are resolved or merged in case of a highly boosted event. This way, more sensitive signal regions can be defined, which are suited towards the $H \rightarrow b\bar{b}$ decay. In addition, categories with similar physics processes, i.e., resulting in similar DNN distributions, are combined in order to reduce the total number of categories insensitive to signal.

Finally, the DNN output prediction in these categories is used in a simultaneous fit in order to extract the HH signal and measure its most interesting coupling modifiers.

5.3 Likelihood Function for Higgs Pair Production

The analysis presented in this thesis considers two Higgs boson pair production modes (HH(GGF) and HH(VBF)), whose cross sections and kinematic distributions depend on the coupling modifiers as shown in Figs. 2.11, 2.14a and 2.14b. The signal hypothesis s contributes to the expectation λ together with the background hypothesis b as follows:

$$\lambda = \mu \cdot s + b, \tag{5.1}$$

$$\to \lambda = \mu \cdot \left[\underbrace{\mu_{\mathrm{HH(GGF)}} \cdot s_{\mathrm{HH(GGF)}}\left(\kappa_\lambda, \kappa_t\right)}_{\text{HH(GGF) signal}} + \underbrace{\mu_{\mathrm{HH(VBF)}} \cdot s_{\mathrm{HH(VBF)}}\left(\kappa_\lambda, \kappa_{2V}, \kappa_V\right)}_{\text{HH(VBF) signal}} \right] \tag{5.2}$$

$$+ \underbrace{b_{\mathrm{H}}\left(\kappa_\lambda, \kappa_t, \kappa_V\right)}_{\text{SM Higgs}} + b_{\text{other backgrounds}}. \tag{5.3}$$

Here, the signal expectation is expressed by the HH(VBF) and HH(VBF) components, which have individual signal strength modifiers $\mu_{\mathrm{HH(GGF)}}$ and $\mu_{\mathrm{HH(VBF)}}$. This allows to e.g. measure the signal strength of the HH(VBF) production mode by fixing the global modifier μ and the corresponding other signal strength modifier $\mu_{\mathrm{HH(GGF)}}$ to the SM expectation of one. In addition, the single Higgs boson processes, which are considered as backgrounds in this analysis, are also subject to the coupling modifiers κ_λ (through higher order corrections), κ_t, and κ_V; the scaling of their cross sections and branching fractions is considered in the statistical model. Finally, the parametrization of the expectation of the statistical model, shown in Eq. 5.3, is used in the likelihood function in Eq. 4.20.

The signal hypotheses $s_{\mathrm{HH(GGF)}}$ and $s_{\mathrm{HH(VBF)}}$ do not exist for every combination of coupling modifiers, as the computational effort to produce this large number of samples is too high. Instead, every signal hypothesis can be analytically inferred with a continuous morphing based on 3 (6) bases, $s_{\mathrm{HH(GGF)}}^{(i)}$ ($s_{\mathrm{HH(VBF)}}^{(i)}$), with specific coupling modifier settings for the HH(GGF) (HH(VBF)) production mode. The morphing strategy is defined as:

$$s_{\mathrm{HH(GGF)}}\left(\kappa_\lambda, \kappa_t\right) = \sum_{i=1}^{3} \alpha_i\left(\kappa_\lambda, \kappa_t\right) \cdot s_{\mathrm{HH(GGF)}}^{(i)}, \tag{5.4}$$

$$s_{\mathrm{HH(VBF)}}\left(\kappa_\lambda, \kappa_{2V}, \kappa_V\right) = \sum_{i=1}^{6} \alpha_i\left(\kappa_\lambda, \kappa_{2V}, \kappa_V\right) \cdot s_{\mathrm{HH(VBF)}}^{(i)}, \tag{5.5}$$

where the α_i denote the morphing fractions. These fractions can be calculated through the amplitude of the contributing Feynman diagrams (triangle loop: $\triangle$, box loop: $\square$) as:

$$\sigma_{\mathrm{HH(GGF)}}\left(\kappa_\lambda, \kappa_t\right) \propto |\mathcal{A}|^2 = \kappa_t^2 \kappa_\lambda^2 \left|\triangle\right|^2 + \kappa_t^4 \left|\square\right|^2 + \kappa_t^3 \kappa_\lambda \left|\triangle^*\square + \square^*\triangle\right|^2, \tag{5.6}$$

which can also be expressed as a vector product

$$\sigma_{\mathrm{HH(GGF)}}\left(\kappa_\lambda, \kappa_t\right) \propto \underbrace{\left(\kappa_t^2 \kappa_\lambda^2, \kappa_t^4, \kappa_t^3 \kappa_\lambda\right)^{\mathrm{T}}}_{\mathbf{c}^{\mathrm{T}}} \cdot \underbrace{\left(\left|\triangle\right|^2, \left|\square\right|^2, \left|\triangle^*\square + \square^*\triangle\right|^2\right)}_{\mathbf{v}}. \tag{5.7}$$

This equation can be extended with three sets of coupling modifiers ($\mathbf{c}_i$) and three known cross sections σ_i using a matrix vector product, and solved for the vector $\mathbf{v}$:

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{pmatrix} \cdot \mathbf{v}, \tag{5.8}$$

$$\sigma = \mathbf{C} \cdot \mathbf{v}, \tag{5.9}$$

$$\leftrightarrow \mathbf{v} = \mathbf{C}^{-1} \cdot \sigma. \tag{5.10}$$

Finally, this relation can be used to substitute the vector $\mathbf{v}$ in Eq. 5.7 to

$$\sigma_{\mathrm{HH}(GGF)}\left(\kappa_\lambda, \kappa_t\right) \propto \underbrace{\mathbf{c}^{\mathrm{T}} \cdot \mathbf{C}^{-1}}_{\alpha(\kappa_\lambda, \kappa_t)} \cdot \sigma. \tag{5.11}$$

Here, the morphing factors $\alpha\left(\kappa_\lambda, \kappa_t\right)$ can be identified. Both, $\mathbf{C}^{-1}$ and σ are known by the choice of bases (samples): their coupling modifiers are fixed and the cross sections can be calculated for the given set of coupling modifiers. This calculation does not only work for the overall cross section, but also for the morphing of the signal shapes $s_{\mathrm{HH}(GGF)}^{(i)}$ from Eq. 5.5. With this analytic morphing strategy, signal hypotheses for any combination of coupling modifiers can be calculated.

Similarly, a morphing strategy for the HH(VBF) production mode can be constructed. As the HH(VBF) production mode involves three coupling modifiers (see Fig. 2.11), more terms arise in the calculations, which requires at least six bases (samples) to be used.

Now, the expectation of the statistical model can be expressed as a function of different signal strength modifiers (μ, $\mu_{\mathrm{HH}(GGF)}$, and $\mu_{\mathrm{HH}(VBF)}$), and different coupling modifiers (κ_λ, κ_t, κ_{2V}, and κ_V). Therefore, the expectation can be evaluated for any set of modifiers. This allows to measure any of these modifiers individually by fixing all others to certain values or to measure arbitrary combinations of them simultaneously.

In this analysis, the first option is used to calculate upper limits on μ and $\mu_{\mathrm{HH}(VBF)}$ for standard model coupling modifier expectations, and as a function of κ_λ and κ_{2V}, respectively. In addition, measurements of the κ_λ and κ_{2V} coupling modifiers are performed through likelihood profiling.

5.4 Analysis Steps Towards Measurements

The challenges of this analysis has become clear in the previous sections. The small number of expected signal events almost vanish under the overwhelming amount of background events. A rather loose event selection is applied to keep as many signal events as possible. Isolating these signal events and controlling the major backgrounds is realized through a deep learning based physics process classification. Some backgrounds need a dedicated data-driven modelling to be well described. Finally, the resulting distributions are used to measure the signal strength modifiers and the coupling modifiers, which may affect the model expectation of the likelihood function as shown in Eq. 5.3.

A short step-by-step description of the analysis and the corresponding section in this thesis is given in the following:

1. Loose event selection for a high signal acceptance (Sec. 6.3).

2. Improved modelling of the Drell-Yan and Misid. leptons processes through data-driven estimations (Secs. 6.5 and 6.6).

3. Physics process classification for an optimized event categorization and most sensitive discriminant (chapter 7).

4. Measurements of the signal strength modifiers and the coupling modifiers (chapter 8).

Chapter 6

Datasets and Event Selection

The event selection for the search for the dileptonic $HH \rightarrow b\bar{b}W^+W^-$ process aims to keep a high acceptance for this signal process. Thus, a rather loose selection criteria are applied, leaving the separation of signal and background processes primarily to the physics process classification. The event selection and reconstruction steps are described in the internal analysis note of this analysis [185].

First, the used data and simulation samples are presented. Then, the event selection based on the identification and selection of final state particles and derived event-level quantities is described. Finally, corrections to simulation are discussed, which aim to mitigate inaccuracies in the simulation.

6.1 Data Samples

Data events used for this analysis are collected in pp collisions at 13 TeV and a bunch crossing period of 25 ns. Only data recording periods, so-called *runs*, are used where all detector components were fully operational. These runs are selected with the help of JSON certificates, which are provided by the CMS collaboration [186]. The selected runs correspond to a total integrated luminosity of 137.6 fb^{-1}, comprising 36.3 fb^{-1} in 2016, 41.5 fb^{-1} in 2017, and 59.7 fb^{-1} in 2018. The used JSON certificates are listed in Tab. 6.1.

Table 6.1: List of JSON certificates [186] used to select run periods in the 2016, 2017, and 2018 data-taking periods.

Year	JSON certificate
2016	`Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt`
2017	`Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt`
2018	`Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_` `PromptEraD_Collisions18_JSON.txt`

Datasets of these events are created with HLT triggers focussing on the presence of electrons and muons with minimal selection criteria. In addition, reconstruction steps are applied to all events. These steps are versioned by a reconstruction campaign used to identify all reconstruction steps in a certain version of the CMSSW software. In the following table the used datasets are listed including their availability in the three different data-taking periods.

Table 6.2: List of datasets for the different lepton channels. These datasets have been reconstructed with the `17Jul2018` / `31Mar2018` / [`17Sep2018, 22Jan2019,` `PromptReco`] reconstruction campaigns using the `9_4_X` / `9_4_X` / `10_2_X` version of the CMSSW software, respectively (2016 / 2017 / 2018). In 2018 the `EGamma` dataset is used instead of the `DoubleEG` and `EGamma` datasets, which is a combination of both. The dataset availability is given for the three years, in the form: 2016 / 2017 / 2018.

Channel	Dataset name	Year availability (2016/2017/2018)
e	SingleElectron	✓ / ✓ / ✗
ee	DoubleEG	✓ / ✓ / ✗
ee + e	EGamma	✗ / ✗ / ✓
μ	SingleMuon	✓ / ✓ / ✓
$\mu\mu$	DoubleMuon	✓ / ✓ / ✓
eμ	MuonEG	✓ / ✓ / ✓

6.2 Simulation

The signal events have been simulated using the POWHEG V2 [187–189] generator up to NLO precision of the hard interaction for the HH(GGF) process. The HH(VBF) process has been simulated with MADGRAPH_AMCATNLO [190] up to LO precision of the hard interaction. Datasets for both signal processes have been produced for different sets of coupling modifiers. This is needed for the morphing strategy as explained in Sec. 5.3.

The background events are simulated with POWHEG V2 and MADGRAPH_AMCATNLO with NLO precision of the hard interaction, except for the tH process, which is simulated using LO matrix elements. For the training of deep learning algorithms and the evaluation of the data-driven Drell-Yan process, samples are used, which are combined together using "inclusive" samples and samples binned in different jet multiplicities in order to increase the true number of generated events. The stitching procedure is documented in Ref. [191].

Parton showering and hadronization for all samples were simulated using PYTHIA 8 with the CUETP8M1, CUETP8M2, and CUETP8M2T4 [192, 193] tune in 2016 and the CP5 [194] tune in 2017 and 2018. The most often employed PDF sets to describe the proton structure are the NNPDF3.0 [195] and NNPDF3.1 [196] sets. Systematic variations of thereof are all calculated in terms of the NNPDF3.1 set only by reweighting all other PDF sets to this one to avoid a mixture of PDF sets. All samples are produced assuming a top quark mass of $m_\mathrm{t} = 172.5\,\mathrm{GeV}$ and Higgs boson mass of $m_\mathrm{H} = 125\,\mathrm{GeV}$. The cross sections of the Drell-Yan and the $t\bar{t}$ + Jets process are normalized to NNLO precision [53, 197, 198]. The GEANT4 [199] detector simulation toolkit is used for a full detector simulation for all simulated datasets.

Each simulated event has a (technical) generator weight $w_{\mathrm{gen},i}$ associated with it. This weight does not need to correspond to one. It is used by the event generator to correctly model distributions of physical observables of the events. The generator weight is necessary in a reweighting procedure to match N simulated events to data. This reweighting involves $w_{\mathrm{gen},i}$, the cross section of the process σ and the integrated luminosity $\mathcal{L}$, and is calculated as follows:

$$w_i = \frac{\sigma \mathcal{L}}{\sum_i^N w_{\text{gen},i}} \cdot w_{\text{gen},i} \,. \tag{6.1}$$

6.3 Event Selection

The event selection is applied to the events in data and simulation. These selection criteria are based on event-level quantities, such as the number of reconstructed leptons or jets. This section describes all selection criteria and the reconstruction of physics objects (leptons and jet) tailored towards the search of the $HH \rightarrow b\bar{b}W^+_\ell W^-_\ell$ process.

First, baseline selection criteria are applied to reject events with (large) unphysical missing transverse momentum or with an invalid primary vertex are described. Then, the reconstruction and selection of different types of leptons is explained. These lepton types and a selection of HLT triggers are used to define the three lepton channels ee, $\mu\mu$, and eμ. Finally, the jet reconstruction and selection is described, which concludes the event selection.

6.3.1 Missing Transverse Energy Filters and Primary Vertex Selection

Events with unphysical (typically too large) $E_{\text{T}}^{\text{miss}}$ are removed by so-called $E_{\text{T}}^{\text{miss}}$ filters. Unphysical $E_{\text{T}}^{\text{miss}}$ may originate from noise in detector parts (i.e., ECAL or HCAL), beam halo scraping, or from a poor muon track reconstruction step [200]. A list of used $E_{\text{T}}^{\text{miss}}$ filters is given in Tab. 6.3. The selection of $E_{\text{T}}^{\text{miss}}$ filters is based on recommendations of the CMS collaboration for the Run 2 data-taking period.

Table 6.3: $E_{\text{T}}^{\text{miss}}$ filters applied to events selected in data and to simulated events. A cross-mark (✗) indicates that the filter is not applied.

$E_{\text{T}}^{\text{miss}}$ filter name	Applied to data / simulation
Flag_goodVertices	✓ / ✓
Flag_globalSuperTightHalo2016Filter	✓ / ✓
Flag_HBHENoiseFilter	✓ / ✓
Flag_HBHENoiseIsoFilter	✓ / ✓
Flag_EcalDeadCellTriggerPrimitiveFilter	✓ / ✓
Flag_BadPFMuonFilter	✓ / ✓
Flag_eeBadScFilter	✓ / ✗

In addition, events are required to have a valid primary vertex. The primary vertex is required to have four tracks associated, and to be within 24 cm longitudinal and 2 cm transversal distance to the detector center.

6.3.2 Lepton Selection

In this analysis, only electrons and muons are considered as leptons in the final state. Two types of leptons are defined: *loose* and *medium*. The medium leptons together with a dedicated discriminant, from now on referred to as the "lepton MVA ID" [201], are later used to estimate the background, which involves misidentified leptons in all physics processes. To avoid potential biases in this background estimation, a modified definition of the transverse lepton momentum is used:

$$p_{\text{T,cone}}^{\ell} = \begin{cases} 0.9 \cdot (p_{\text{T}}^{\ell} + I^{\ell}) & \text{if } \ell \text{ fails the tight lepton MVA ID and } \Delta R\,(\ell, \text{jet}) > 0.4, \\ p_{\text{T}}^{\ell} & \text{otherwise.} \end{cases}$$

where p_{T}^{ℓ} denotes the transverse lepton momentum, I^{ℓ} the relative lepton isolation, and $\Delta R\,(\ell, \text{jet})$ the distance between the lepton and the closest jet. This quantity is equal to the original transverse momentum of leptons in the signal regions (the latter case), but differs for leptons in the control regions of the background estimation method (first case).

Both lepton flavors are well identified by criteria provided by the CMS collaboration. In addition, they are required to be isolated, which means that no (or minimal) deposited energy is found within a certain cone size around the lepton. This quantity is used to find leptons, that are outside of jets. The isolation I^{ℓ} is defined as follows:

$$I^{\ell} = \sum_{\text{charged}} p_{\text{T}} + \max\left(0, \sum_{\text{neutrals}} p_{\text{T}} - \rho\,\mathcal{A}\left(\frac{R}{0.3}\right)^2\right), \tag{6.2}$$

where ρ refers to the energy density of neutral particles reconstructed within the geometric acceptance of the tracking detectors [202, 203], $\mathcal{A}$ is the effective area, and R is the cone size, which varies for different transverse lepton momenta. Equation 6.2 is primarily based on the sum of transverse momenta of all charged and neutral particles within a cone of size R centered on the direction of flight of the lepton.

Detailed definitions of loose and medium leptons are described in the following paragraphs for muons and electrons.

Muon Selection

As a starting point, muons are selected and reconstructed as *global muons* as described in Sec. 3.2.8.2. The CMS collaboration provides a loose working point (cut based) identification for PF muons (reconstructed by PARTICLE-FLOW), which is used in this analysis. Muons are also required to pass several cuts based on distances d between the muon track and the interaction point. A full list of selection criteria for loose and medium muons is shown in Tab. 6.4.

The selection based on the b-tagging ID of the nearest jet is interpolated linearly from the medium working point for muons with $p_{\text{T,cone}} = 20\,\text{GeV}$ to the loose working point for muons with $p_{\text{T,cone}} = 45\,\text{GeV}$. The interpolation has proven to yield an improved performance for the estimation of misidentified muons [201].

Unless otherwise indicated, medium muons are used in this analysis.

Electron Selection

All electrons are identified using a multivariate algorithm, called FALL17NOISOV2 [204], which has been trained to separate electrons from jets. The loose working point of this algorithm is used, in order to retain the highest overall selection efficiency of electrons. In addition, electrons might stem from photons conversions. These electrons are removed by matching the electron track to a hit in each layer of the pixel subdetector. In total, only one missing hit in all layers is allowed. Similar to muons, selection criteria based on distances (*d*)

Table 6.4: List of selection criteria for loose and medium muons. The cut on the b-tagging ID of the nearest jet is interpolated between muons with low transverse momenta to high transverse momenta. The description of the b-tagging ID of jets can be found in Sec. 6.3.4.

Observable	Loose	Medium
p_T	$> 5\,\text{GeV}$	$> 10\,\text{GeV}$
$|\eta|$	< 2.4	< 2.4
$|d_{xy}|$	$< 0.05\,\text{cm}$	$< 0.05\,\text{cm}$
$|d_z|$	$< 0.1\,\text{cm}$	$< 0.1\,\text{cm}$
d/σ_d	< 8	< 8
I^μ	$< 0.4 \cdot p_T$	$< 0.4 \cdot p_T$
PF muon	$>$ loose working point	$>$ loose working point
b-tagging ID (nearest jet)	✗	$<$ interp. working point

between the electron track and the interaction point are imposed. A full list of selection criteria for loose and medium electrons is given in Tab. 6.5.

Table 6.5: List of selection criteria for loose and medium electrons. The cut on the width of the electron cluster ($\sigma_{i\eta i\eta}$) is distinguished between the barrel (< 0.011) and endcap (< 0.030) regions. The description of the b-tagging ID of jets can be found in Sec. 6.3.4.

Observable	Loose	Medium
p_T	$> 7\,\text{GeV}$	$> 10\,\text{GeV}$
$|\eta|$	< 2.5	< 2.5
$|d_{xy}|$	$< 0.05\,\text{cm}$	$< 0.05\,\text{cm}$
$|d_z|$	$< 0.1\,\text{cm}$	$< 0.1\,\text{cm}$
d/σ_d	< 8	< 8
I^e	$< 0.4 \cdot p_T$	$< 0.4 \cdot p_T$
$\sigma_{i\eta i\eta}$	✗	$< \{\,0.011\ /\ 0.030\,\}$
H/E	✗	< 0.10
1/E - 1/p	✗	> -0.04
Conversion rejection	✗	✓
Missing hits	≤ 1	$= 0$
FALL17NOISOV2	$>$ loose working point	$>$ loose working point
b-tagging ID (nearest jet)	✗	$<$ medium working point

Loose electrons are removed in case they overlap with loose muons within $\Delta R < 0.3$. Unless otherwise indicated, medium electrons are used in this analysis.

6.3.3 Decay Channel Selection

Based on the selection of a pair of opposite charged electrons and muons, three channels can be defined: ee, $\mu\mu$, eμ. Different compositions of HLT triggers are required to have fired for the three channels. The leptons are required to have a minimal transverse momentum of 25 GeV (15 GeV) for the leading (subleading) lepton. This cut ensures a phase-space, that

avoids poorly modeled HLT trigger efficiency regions (turn-on and below). An overview of the channel dependent HLT triggers is given in Tab. 6.6.

For the ee, $\mu\mu$ and eμ decay channels, the HLT trigger paths ee or e, $\mu\mu$ or μ and eμ or e or μ must be triggered respectively. Single lepton triggers (e, μ) are added to increase the signal acceptance. Datasets with recorded data are associated with specific HLT triggers that are applied only to them. The assignment is as follows:

$$\text{HLT trigger paths } (e) : \texttt{SingleElectron, EGamma}$$
$$\text{HLT trigger paths } (\mu) : \texttt{SingleMuon}$$
$$\text{HLT trigger paths } (ee) : \texttt{DoubleEG, EGamma}$$
$$\text{HLT trigger paths } (\mu\mu) : \texttt{DoubleMuon}$$
$$\text{HLT trigger paths } (e\mu) : \texttt{MuonEG}$$

In order to avoid selecting data events multiple times, a trigger logic is applied based on an priority order of the datasets, since an event can be triggered by more than one of the above mentioned triggers. The priority order is defined as follows (in decreasing order): `DoubleMuon, MuonEG, SingleMuon, SingleElectron / EGamma` (for 2016 and 2017 / 2018). The order is chosen to favor muons over electrons, as muons are typically better identified and reconstructed by the CMS detector. An example of this trigger logic is as follows: an event with two muons may be present in the `DoubleMuon` dataset and the `MuonEG` dataset. If the triggers of both datasets have fired, the `DoubleMuon` dataset gets prevalence over the `MuonEG` dataset. In this case, the event is selected for the $\mu\mu$ channel, and rejected for the eμ channel.

Low mass resonances that are usually poorly modeled in simulation, as well as the Z boson resonance, which is only part of background processes, are vetoed by requiring cuts on the invariant mass of two loose leptons in each event:

$$m_{\ell\ell} > 12\,\text{GeV} \quad \text{(low mass resonances veto)},$$
$$|m_{\ell\ell} - m_Z| > 10\,\text{GeV} \quad \text{(Z resonance veto)}.$$

6.3.4 Jet Selection

Jets are reconstructed and calibrated as explained in Sec. 3.2.8.4. Two jet sizes are used with a distance parameter of $\Delta R = 0.4$ and $\Delta R = 0.8$, respectively. These jets are referred to as small radius and large radius jets. Both jet types are requited to pass the loose working point identification in 2016, and the tight working point identification in 2017 and 2018 [205]. The identification is primarily based on cuts on the hadronic and electromagnetic energy fraction of the jet as well as the number of constituents within a jet. Small (large) radius jets are required to have a minimum transverse momentum of 25 GeV (200 GeV) and to be within $|\eta| \leq 2.4$. Large radius jets have to be composed of two small radius jets, which are required to have at least 20 GeV transverse momentum and $|\eta| \leq 2.4$. In addition, large

Table 6.6: List of HLT triggers used in the analysis. Checkmarks (✓) and crossmarks (✗) are added after each HLT path to indicate whether the path was enabled or disabled during the 2016, 2017, and 2018 data-taking periods.

Channel	HLT trigger paths	Year availability (2016/2017/2018)
e	`HLT_Ele25_eta2p1_WPTight_Gsf`	✓/✗/✗
	`HLT_Ele27_WPTight_Gsf`	✓/✗/✗
	`HLT_Ele27_eta2p1_WPLoose_Gsf`	✓/✗/✗
	`HLT_Ele32_WPTight_Gsf`	✗/✓/✓
	`HLT_Ele35_WPTight_Gsf`	✗/✓/✗
μ	`HLT_IsoMu22`	✓/✗/✗
	`HLT_IsoTkMu22`	✓/✗/✗
	`HLT_IsoMu22_eta2p1`	✓/✗/✗
	`HLT_IsoTkMu22_eta2p1`	✓/✗/✗
	`HLT_IsoMu24`	✓/✓/✓
	`HLT_IsoTkMu24`	✓/✗/✗
	`HLT_IsoMu27`	✗/✓/✓
ee	`HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ`	✓/✗/✗
	`HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL`	✗/✓/✓
$\mu\mu$	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL`	✓/✗/✗
	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ`	✓/✗/✗
	`HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL`	✓/✗/✗
	`HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ`	✓/✗/✗
	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8`	✗/✓/✗
	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8`	✗/✓/✓
eμ	`HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL`	✓/✗/✗
	`HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ`	✓/✓/✓
	`HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL`	✓/✗/✗
	`HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ`	✓/✗/✗
	`HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ`	✓/✓/✗
	`HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ`	✓/✓/✗
	`HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL`	✓/✓/✗
	`HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ`	✗/✓/✗

radius jets need to have a softdrop mass [206] between 30 GeV and 210 GeV, and their "N-subjettiness" [207] (τ_2/τ_1), which is a quality cut on the structure of the large radius jet to be smaller than 0.75 favoring a di-jet substructure. Small (large) radius jets are removed in case they overlap with medium leptons within $\Delta R < 0.4$ ($\Delta R < 0.8$). Jets arising from pileup interactions are removed using a dedicated pileup jet identification (PUJETID) as described in Sec. 3.2.8.4.

In order to enhance the sensitivity to the HH(VBF) production mode, additional jet selection criteria are applied. These jets are typically found in forward detector regions of large pseudorapidity, which is why they have an relaxed cut of $|\eta| \leq 4.7$. In addition, their transverse momentum threshold is raised to be at least 30 GeV. For a pseudorapidity range corresponding to the ECAL endcap region ($2.7 < |\eta| < 2.0$), the transverse momentum threshold threshold is further raised to a minimum of 60 GeV to reduce unwanted jets from detector noise. Jets for the HH(VBF) production mode are removed if they are found within $\Delta R < 0.8$ of the two leading small radius jets, or $\Delta R < 1.2$ of the leading large radius jet.

The DEEPJET algorithm, described in Sec. 3.2.8.6, is used to identify both b-quark jets, that are part of the final state of the signal process. These jets are required to pass the medium working point of the DEEPJET algorithm corresponding to a misidentification rate of 1% and an identification efficiency of approximately 80% (cf. Sec. 3.2.8.6). In the case of large radius jets, at least one small radius subjet is required to pass the medium working point with a minimum transverse momentum of 30 GeV.

Each event in the ee, $\mu\mu$ and eμ decay channels is required to have at least one small (large) radius jet, and at least one small (large) b-quark jet. Events selected by the number of small or large radius jets are referred to as "resolved" or "boosted", respectively.

6.4 Corrections to Simulation

Differences between simulation and data due to effects in the event generation, in the detector simulation, and in the reconstruction and identification algorithms, are corrected for in simulation. These corrections are typically in the form of weight factors, also called "scale factors", on an event by event basis. The scale factors of multiple corrections are multiplied all together, including the generator weight (see Eq. 6.1), into a single event weight. Each correction, and the corresponding correction factors, are described in the following.

Pileup Correction

The pileup distribution of the simulation does not match the measured one of data [208]. Scale factors are derived to correct the pileup distribution of the simulation. This is achieved by estimating the pileup per bunch crossing with the instantaneous luminosity of a single bunch, the beam orbit frequency, and the total pp inelastic cross section of 69.2 mb. The resulting pileup prediction is Poisson distributed for multiple bunch crossings. The scale factors are then calculated by dividing the normalized pileup distribution of data and simulation as a function of the number of pileup interactions.

Trigger Efficiency

The trigger efficiency between data and simulation may vary due to imperfect modelling of event kinematics and HLT triggers. Thus, scale factors are derived to correct the trigger efficiency in simulation [201, 209]. These factors are calculated with $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$

events with the `HLT_PFMET120_PFMHT120_IDTight` HLT trigger path. In order to mimic the lepton selection for this measurement, two opposite sign leptons with a minimum transverse momentum of 25 GeV (15 GeV) for the leading (subleading) lepton are required. The scale factors are calculated by dividing the trigger efficiency curves of data and simulation. It has been found that the trigger efficiency depends primarily on the $p_{T,\text{cone}}$ quantity of the leading lepton, which is why the scale factors are calculated as a function of $p_{T,\text{cone}}$.

Lepton Identification Efficiency

The identification, reconstruction, and isolation efficiencies of leptons (see Sec. 6.3.2) are different in data and simulation. Scale factors are measured in terms of the pseudorapidity and transverse momentum of both leptons. The measurement is performed in two steps. First, scale factors are calculated for loose leptons. Then, additional scale factors for medium leptons, which also pass the tight working point of the lepton MVA ID, are derived. Details on the measurement can be found in Ref. [209].

Jet Energy Corrections

The energy scale of jets is corrected in a factorized approach as described in Sec. 3.2.8.4. This correction is the only one which affects simulation and data. They correct differences coming from pileup effects, detector responses, and general differences between data and simulation, and are applied in strict order. In addition, the energy resolution in simulation needs to be smeared to match the resolution in data. The jet energy corrections correct the four momenta of all jets before any selection and are provided by the CMS collaboration [210].

b-quark Identification Efficiency

The performance of the DEEPJET algorithm to correctly identify b-quark jets differs in data and simulation. Scale factors are centrally provided by the CMS collaboration [211] to correct the simulation for this difference. The corrections depend on the jet flavor, its transverse momentum, its absolute pseudorapidity, and the DEEPJET output prediction score. In addition, the number of events before and after applying these scale factors should remain the same. This is ensured by calculating additional scale factors as a function of the number of jets by dividing the number of events before and after applying the scale factors. These additional scale factors are measured before applying any jet or lepton selection.

Pre-firing of the L1 ECAL Trigger

In the 2016 and 2017 data-taking periods, a gradual timing shift of the ECAL was not properly propagated to the L1 trigger [212]. The ECAL information was therefore associated to the previous bunch crossing, especially for energy deposits at $2.0 < |\eta| < 3.0$. Since the L1 trigger forbids two consecutive bunch crossings to fire, sometimes the ECAL information of an event comprises the information of two bunch crossings. This is called *pre-firing*. This effect is not present in simulation, and thus is corrected for in simulation. The correction is based on the product of probabilities of all physics objects in an event to not pre-fire. These probabilities are centrally provided by the CMS collaboration [213].

Modulation Correction of E_T^{miss}

In theory, there should be no ϕ dependency of E_T^{miss} due to the rotational symmetry around the beam axis. However, the reconstructed E_T^{miss} shows a sinusoidal ϕ dependency with a pe-

riod of 2π in data and simulation, which originates likely from detector inhomogeneities. A correction, provided by the CMS collaboration [214], as a function of the number of primary vertices is applied to $E_{\mathrm{T}}^{\mathrm{miss}}$ in both data and simulation.

t-quark Transverse Momentum Correction

It has been found that the transverse momenta spectrum of t-quarks in data is significantly softer, i.e., t-quarks have smaller transverse momenta than in simulation [215]. This effect is because of missing higher order calculations in the event generation. Scale factors based on the transverse momenta of t-quarks are derived for the $t\bar{t}$ + Jets datasets. These scale factors are parametrized by a function, which is obtained by fitting the ratio of the t-quark transverse momentum of the NNLO calculation [216] and the simulated events. The resulting function has been determined by the CMS collaboration to be:

$$\mathrm{SF}(p_{\mathrm{T}}^{\mathrm{t}}) = \exp\left(a + b \cdot p_{\mathrm{T}}^{\mathrm{t}} c \cdot (p_{\mathrm{T}}^{\mathrm{t}})^2 + \frac{d}{p_{\mathrm{T}}^{\mathrm{t}} + e}\right),$$

where $p_{\mathrm{T}}^{\mathrm{t}}$ denotes the transverse momentum of the (generator-level) t-quark. The constants (a, b, c, d, e) have been estimated with a fit. The event weight w for the $t\bar{t}$ + Jets process is calculated using the scale factor of the t-quark and $\bar{\mathrm{t}}$-quark:

$$w = \sqrt{\mathrm{SF}(p_{\mathrm{T}}^{\mathrm{t}}) \cdot \mathrm{SF}(p_{\mathrm{T}}^{\bar{\mathrm{t}}})}. \tag{6.3}$$

6.5 Estimation of Misidentified Leptons

Due to the high multiplicity of leptons and jets in the final state of this analysis, the background from misidentified leptons, called "Misid. leptons", plays a non-negligible role. Before the estimation procedure is explained, the concept of "lepton promptness" needs to be introduced: leptons are labelled as "prompt", if they origin from decays of H bosons, W bosons, Z bosons, or τ leptons, i.e., leptons of the signal and dominant background processes of this analysis; otherwise leptons are labelled as "non-prompt" (e.g. produced in hadronization process), and are usually misidentified jets. The background of remnant non-prompt leptons is estimated from data by the following procedure as these events are typically poorly modeled in simulation.

The estimation technique is based on the "Fake-Factor method" [201]. First, events are separated into the prompt (signal) region (SR) and a non-prompt ("fake") region (FR) based on the lepton MVA ID. Additional selection criteria on top of the medium leptons, as described in Sec. 6.3.2, are applied. These are listed for muons in Tab. 6.7, and for electrons in Tab. 6.8.

In addition, the truth information whether a lepton is prompt or not can be used in simulation to isolate the non-prompt part of the FR and the SR. The truth level matching of reconstructed leptons to H bosons, W bosons, Z bosons, or τ leptons yields generator lepton candidates, which require additionally $\Delta\mathrm{R}(\mathrm{gen}, \mathrm{reco}) < 0.3$ and $|p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{\mathrm{gen}}| < 0.5 \cdot p_{\mathrm{T}}^{\mathrm{gen}}$. Four regions can be defined with the generator truth information and the lepton MVA ID. These regions are shown in Fig. 6.1.

The background to be estimated is labelled as "non-prompt SR". It is estimated using the fake-factor (FF) to extrapolate events from the non-prompt FR into the non-prompt SR. The

Table 6.7: List of selection criteria for prompt and non-prompt muons. These muons are already selected as medium muons 6.3.2. The relative jet isolation (rel. I^{jet}) is defined as $p_\text{T}^{\text{jet}}/p_\text{T}^{\mu} - 1$ if $\Delta R(\text{jet}, \mu) < 0.4$, or as the PARTICLE-FLOW relative isolation for a cone with $\Delta R = 0.4$ otherwise.

Observable	non-prompt	prompt
lepton MVA ID	< 0.5	> 0.5
PF muon	$>$ loose working point	$>$ medium working point
b-tagging ID (nearest jet)	$<$ interp. working point	$<$ medium working point
rel. I^{jet}	< 0.8	✗

Table 6.8: List of selection criteria for prompt and non-prompt electrons. These electrons are already selected as medium electrons 6.3.2.

Observable	non-prompt	prompt
lepton MVA ID	< 0.3	> 0.3
FALL17NOISOV2	$>$ medium working point	$>$ loose working point
b-tagging ID (nearest jet)	$<$ tight working point	$<$ medium working point
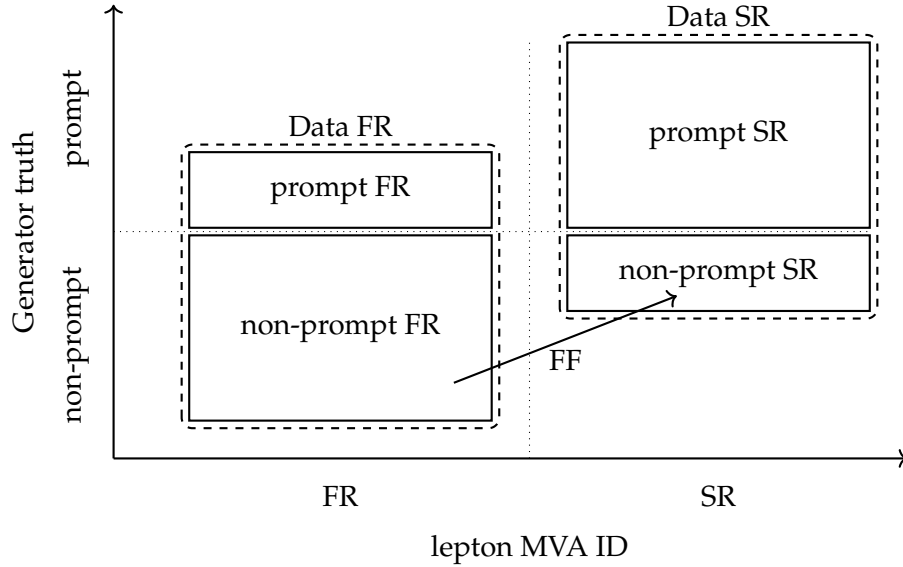
Figure 6.1: Sketch of the estimation procedure for the background of misidentified leptons. The "fake" region (FR) and signal region (SR) are defined by the lepton MVA ID, whereas the truth generator information is used to differentiate prompt and non-prompt lepton in simulation. The non-prompt background (non-prompt SR) can be estimated using the fake-factor (FF) and the non-prompt FR. The size of the regions schematically represent their event yields.

application procedure is as follows:

$$\boxed{\text{non-prompt SR}} = \text{FF} \cdot \boxed{\text{non-prompt FR}} \,, \tag{6.4}$$

$$\text{with } \boxed{\text{non-prompt FR}} \approx \boxed{\text{Data FR}} - \boxed{\text{prompt FR}} \,, \tag{6.5}$$

$$\rightarrow \boxed{\text{non-prompt SR}} = \text{FF} \cdot \left(\boxed{\text{Data FR}} - \boxed{\text{prompt FR}} \right) \,, \tag{6.6}$$

where the FF is defined using the probability f_i of a medium lepton to pass the prompt selection criteria for n leptons as follows:

$$\text{FF}(n) = (-1)^{n+1} \prod_{i=1}^{n} \frac{f_i}{1 - f_i} \,.$$

The probabilities f_i are dependent on the $p_{\text{T,cone}}$ and $|\eta|$ of medium prompt leptons. The measurement of these probabilities has been done in a QCD multijet enriched region, which is characterized by requiring exactly one prompt lepton and at least one jet (see Ref. [201]). More details on this procedure and its technicalities can be found in Ref. [209].

The intrinsic uncertainty of the estimation procedure has been investigated by a closure test performed in Ref. [209]. This uncertainty has been found to irrelevant for the presented analysis.

6.6 Estimation of the Drell-Yan Process

The event selection described in Sec. 6.3.3 and Sec. 6.3.4 comprises a Z boson mass veto and a minimum number of b-quark jets to be present in each event. These requirements reject the majority of Drell-Yan events, which leads to a degraded modelling due to low event statistics, i.e., large statistical fluctuations, and the remaining events to be present in the tails of the m_{ll} distribution. In addition, this particular phase-space, i.e., including multiple b-quark jets, of the Drell-Yan process is not well modeled in simulation by NLO matrix element calculations and parton shower simulation.

Thus, the Drell-Yan process is estimated with a data-driven "ABCD" method (see e.g. [217]). The ABCD method is characterized by two axes, which create four orthogonal regions. Here, the number of b-quark jets per event and the invariant mass of the two leptons are used as axes. The regions are shown in Fig. 6.2.

Each selection defines the Drell-Yan process in a certain phase-space region. The Drell-Yan process in regions A, B, and C is estimated by subtracting all processes except Drell-Yan from data in each region, respectively. The region of interest, or signal region (SR), is region D (Z veto, $\geq$ 1b). The Drell-Yan background is estimated by scaling Drell-Yan events from region C to region D with the help of a transfer factor (TF). The scaling is defined as follows:

$$\boxed{\text{D}} = \text{TF} \cdot \boxed{\text{C}} \,, \tag{6.7}$$

$$\text{with } \text{TF} = \boxed{\text{D}} / \boxed{\text{C}} = \boxed{\text{A}} / \boxed{\text{B}} \,, \tag{6.8}$$

$$\rightarrow \boxed{\text{D}} = \left(\boxed{\text{A}} / \boxed{\text{B}} \right) \cdot \boxed{\text{C}} \,. \tag{6.9}$$
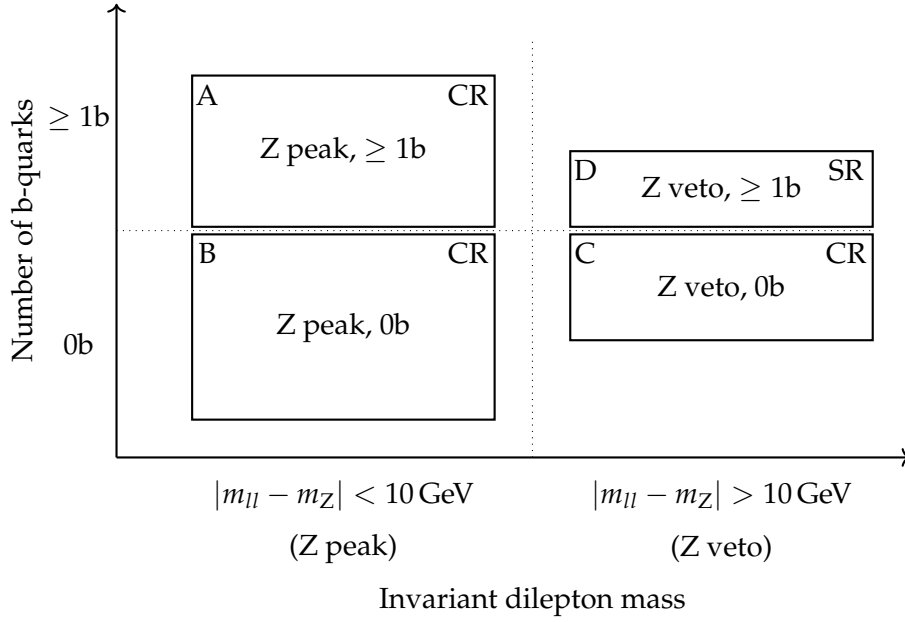
Figure 6.2: Sketch of the estimation procedure for the Drell-Yan background. The size of the regions represent schematically the number of events of the Drell-Yan process.

The transfer factor is parametrized as a function of the scalar sum of the transverse momenta of selected small radius jets (H_T) for resolved event topologies, and as a function of the soft-drop mass of the leading large radius jet for boosted jets.

The measurement of the transfer factor is done in an inclusive region, which incorporates all three lepton channels (ee, $\mu\mu$, eμ). This approach is valid, since the transfer factors of the ee and $\mu\mu$ channel are numerically close to each other. The Drell-Yan process of the eμ channel is highly suppressed as this particular lepton configuration can only be produced through leptonically decaying tau leptons. The benefit of measuring the transfer factors inclusively is a reduced systematic uncertainty of the transfer factors as the underlying event statistics is maximized.

Mismodelling effects of this method have been investigated with a dedicated closure test. The closure test is evaluated on the variable (DNN output prediction, see chapter 7), which is used in the final measurement. The closure test compares the Drell-Yan distribution using directly simulation ("DY (from MC)") and the Drell-Yan distribution estimated with the ABCD method but only using simulation ("DY (est., from MC)"). For reference and comparison the Drell-Yan distribution estimated with the ABCD method using data ("DY (est.)") is also shown. The results of the closure test are shown in Fig. 6.3 for the background control regions (see Sec. 7.5) in 2018. The closure test for the other years can be found in the appendix in Figs. 10.1 and 10.2.

The closure test reveals that the Drell-Yan background estimation introduces a slope, which needs to be corrected for. The correction is derived by dividing "DY (from MC)" by "DY (est., from MC)", and then applied to the Drell-Yan background estimation with data (DY (est.)). It has been found that the slope follows a linear function in the "Top+Other" regions, while in the "DY+VV(V)" region a quadratic function is needed. A fit of these function to the ratio of "DY (from MC)" and "DY (est., from MC)" is performed in each year and for each
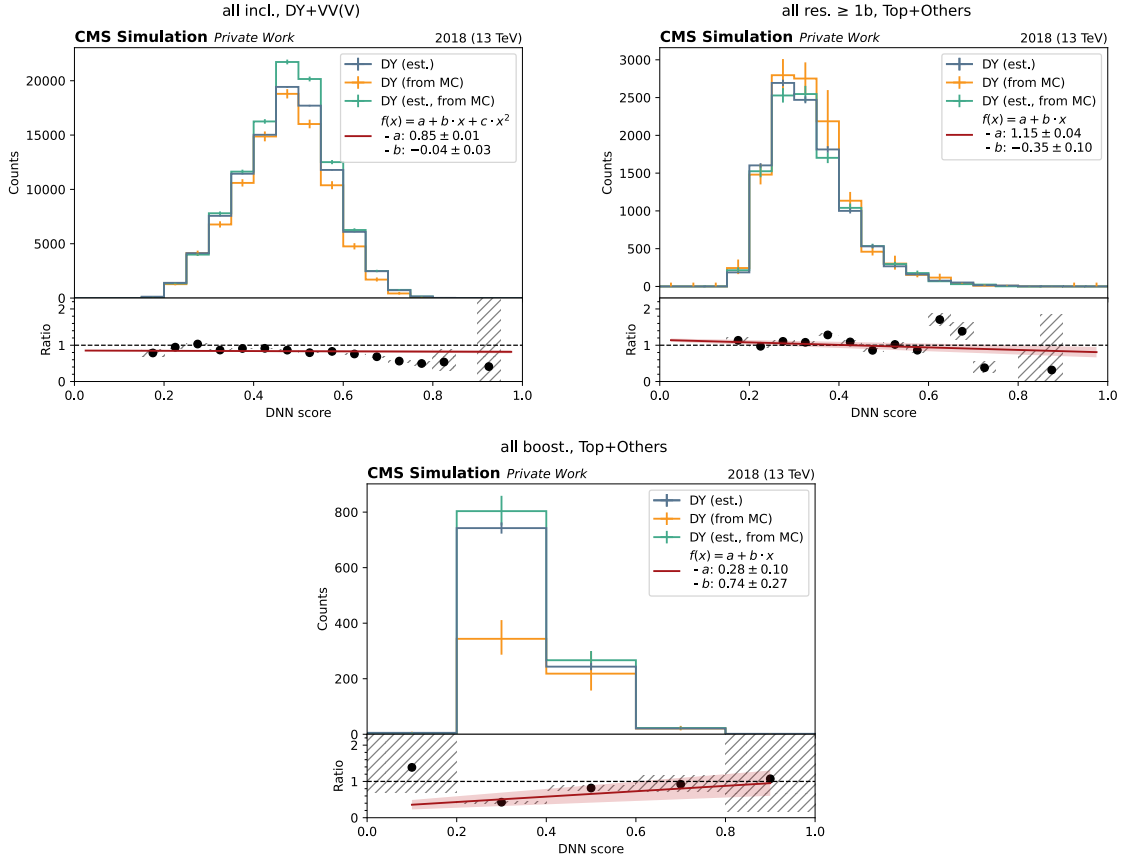
Figure 6.3: Closure test of the Drell-Yan background estimation for all background control regions in 2018. Shown are the Drell-Yan estimation itself (DY (est.)), the Drell-Yan estimation using simulation only (DY (est., from MC)), and the Drell-Yan distribution using simulation directly (DY (from MC)). The ratio between "DY (from MC)" and "DY (est., from MC)" is shown as well. From this ratio a correction is estimated by fitting a linear (quadratic) function to the "Top+Other" ("DY+VV(V)") region. The fit parameters are shown in the legend.

region. The resulting fit parameters can be found in the legend of Figs. 6.3, 10.1, and 10.2.

This correction comes with systematic uncertainties, which are based on the uncertainties of the fit parameters of the linear and quadratic functions. These parameters are correlated, and need to be disentangled in order to obtain decorrelated uncertainties of this method. This is done by diagonalizing the covariance matrix of the fit parameters:

$$f(x) = a + b \cdot x$$
$$\text{with} \quad \text{cov}(a, b) = \begin{pmatrix} \sigma_a^2 & \sigma_{a,b} \\ \sigma_{b,a} & \sigma_b^2 \end{pmatrix} .$$

The matrix $\text{cov}(a, b)$ is diagonalized with the eigenvalues λ_0, λ_1 and the associated eigenvectors $\vec{v}_0, \vec{v}_1$:

$$(\vec{v}_0, \vec{v}_1) \cdot \begin{pmatrix} \sigma_a^2 & \sigma_{a,b} \\ \sigma_{b,a} & \sigma_b^2 \end{pmatrix} \cdot \begin{pmatrix} \vec{v}_0 \\ \vec{v}_1 \end{pmatrix} = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

The decorrelated uncertainties, estimated through the systematic shifts of the Drell-Yan histogram ($h_{\text{DY,i}}^{\pm}$), are then given through this decomposition by:

$$h_{\text{DY,0}}^{\pm}(x) = \left(a \pm \sqrt{\lambda_0} \cdot \vec{v}_{0,0} \right) + \left(b \pm \sqrt{\lambda_0} \cdot \vec{v}_{1,0} \right) \cdot x,$$
$$h_{\text{DY,1}}^{\pm}(x) = \left(a \pm \sqrt{\lambda_1} \cdot \vec{v}_{0,1} \right) + \left(b \pm \sqrt{\lambda_1} \cdot \vec{v}_{1,1} \right) \cdot x.$$

This method works for quadratic functions too, where the matrix $\text{cov}(a, b, c)$ is a 3×3 matrix, which then ultimately results in three systematic shifts of the Drell-Yan histogram $h_{\text{DY,0..2}}^{\pm}$.

The signal categories (see Sec. 7.5) do not have enough statistics in simulation to perform a closure test. Therefore, no correction is applied, instead the ratio of "DY (from MC)" and "DY (est., from MC)" is calculated for each signal category from their total yields (instead of each bin) and used as a decorrelated normalization uncertainty in the final measurement. These ratios are listed in Tab. 6.9.

Table 6.9: Ratios of the histogram integrals between "DY (from MC)" and "DY (est., from MC)" for the signal regions.

Year	HH(GGF)			HH(VBF)		
	res. 1b	res. $\geq$ 2b	boost.	res. 1b	res. $\geq$ 2b	boost.
2016	0.77	1.16	0.18	0.44	0.75	0.21
2017	0.89	0.54	0.23	0.78	1.48	0.41
2018	0.94	0.24	0.15	0.82	0.82	0.42

The normalization uncertainty is calculated by taking the absolute difference of each value (r_i) in Tab. 6.9 from one ($\Delta_{\text{DY,sig}}^i = |r_i - 1|$), and use this difference as the width for a log-normal prior in each signal region and year respectively.

6.7 Agreement Between Data and Simulation After Event Selection

The full phase-space region of the presented analysis consists of the sum of lepton channels, and the different multiplicities and topologies of b-quark jets in an event. The resulting analysis regions are: resolved (res.) = 1b, resolved (res.) $\geq$ 2b, and boosted (boost.). The event yields after selection for these three regions are shown in Tab. 6.10.

This table includes the HH(GGF) and HH(VBF) signal processes, where $\mathcal{O}(1 - 10)$ events are retained after selection, most of them in the two resolved categories. The Drell-Yan and Misid. leptons background processes are estimated as described in the previous two sections. The Top and Other backgrounds are directly taken from simulation. S and B denote the sum of signal and background processes respectively. The data yields are also shown in the last row. The numbers in Tab. 6.10 include all statistical and systematic uncertainties, and

Table 6.10: Event yields after selection for the resolved (res.) = 1b, resolved (res.) $\geq$ 2b, and boosted (boost.) regions. The numbers include all statistical and systematic uncertainties.

Process(es)	res. = 1b	res. $\geq$ 2b	boost.
HH(GGF)	7 ± 1	7 ± 1	1 ± 1
HH(VBF)	≤ 1	≤ 1	≤ 1
Top	861715 ± 25211	831972 ± 33597	18915 ± 1544
Drell $-$ Yan $+$ Multiboson	223851 ± 8142	30977 ± 6652	8399 ± 1137
Misid. leptons	10104 ± 497	4917 ± 554	339 ± 29
Other	8173 ± 142	7562 ± 177	818 ± 20
S	8 ± 1	8 ± 1	1 ± 1
B	1103842 ± 26498	875429 ± 34255	28471 ± 1918
Data	1062827 ± 1031	867894 ± 932	23936 ± 155

are extracted before a fit was performed (*pre-fit*). At this level it can be seen that data and $S + B$ are already compatible within the 1σ uncertainties in the resolved $\geq$ 2b region. Minor deviations exist in the boosted and the resolved 1b categories, where the yields agree within $\leq 2\sigma$ of the pre-fit uncertainties.

In addition, pre-fit distributions of basic variables, such as the transverse momentum of the leading lepton, are shown in the most inclusive analysis phase-space in Fig. 6.4. Good agreement between data and the sum of background processes is observed in all distributions, with no significant deviation. These distributions have been produced with a simplified uncertainty model, where only rate-changing uncertainties are considered. The size of the uncertainty band is primarily driven by the rate-changing systematic uncertainties.

(a) Leading lepton p_{T}.

(b) Subleading lepton p_{T}.

(c) Leading jet p_{T}.

(d) Leading jet η.

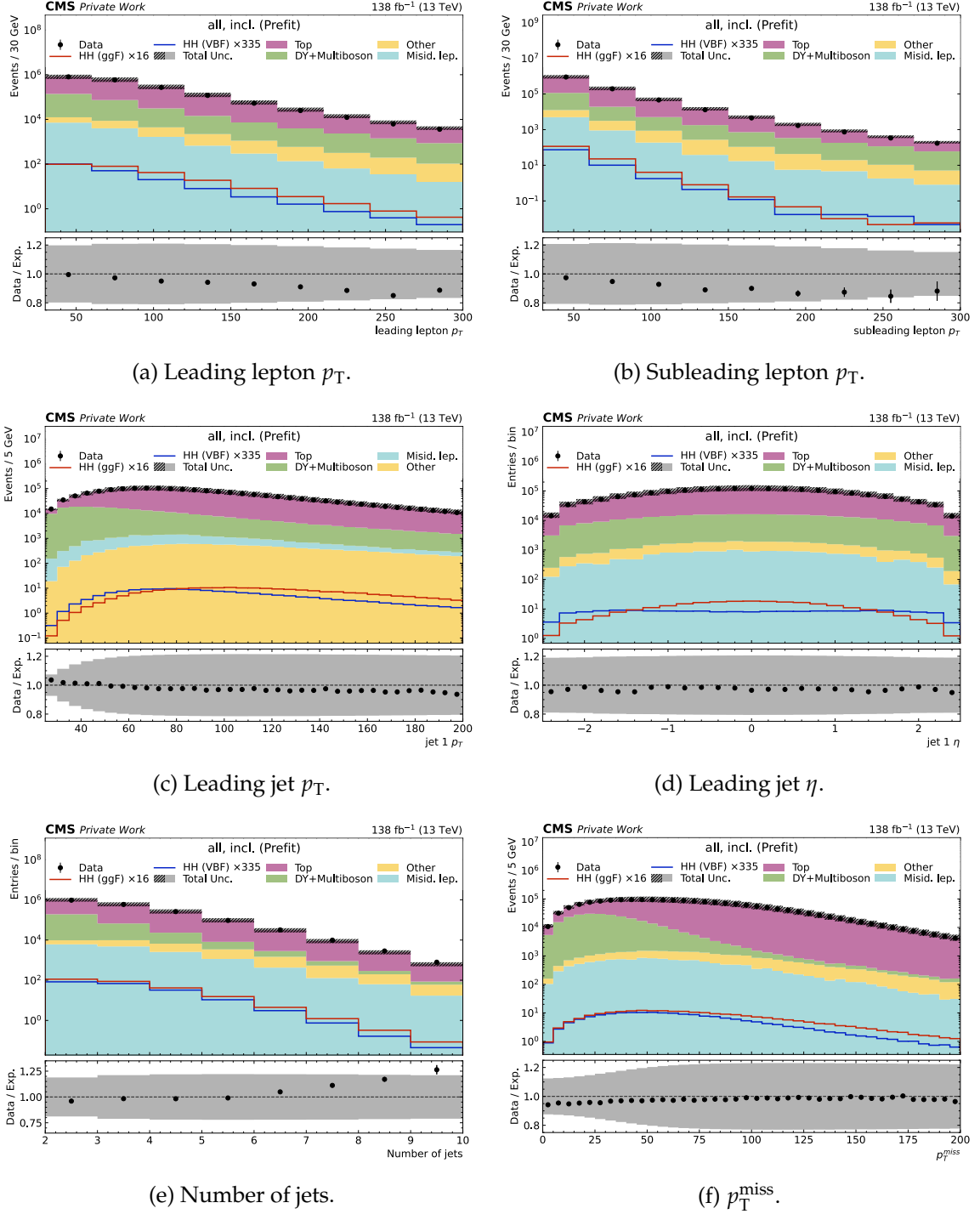(e) Number of jets.

(f) $p_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 6.4: Pre-fit distributions of basic variables in the most inclusive analysis regions after selection and corrections. A simplified uncertainty model is used, which neglects shape-changing uncertainties, but includes rate-changing uncertainties. The rate-changing uncertainties primarily determine the size of the uncertainty band.

Chapter 7

Signal Extraction and Event Categorization

This chapter explains the categorization of events, and in particular the extraction of signal events. Events are first categorized with respect to their jet topologies and b-quark jet multiplicities (cf. Sec. 6.3). In order to enhance the sensitivity, analysis regions are further divided into categories with mutually enriched physics processes. This is achieved with a two-stage deep neural network (DNN) trained as a physics process multi-class classification.

First, the neural network architecture is presented. Then, the input variables to the neural network and their reconstruction are described. The training strategy and the evaluation of the trained neural network is discussed subsequently. The output prediction of the DNN is histogrammed in each region and the respective binning is optimized with an algorithm described in the second last section. Finally, the resulting event yields and distributions are presented before a fit to data is performed.

7.1 Neural Network Architecture

The DNN used for the physics process classification has two components. The first component is the Lorentz Boost Network [158] (LBN) (see Sec. 4.3.3). The second component is a fully connected deep neural network with residual shortcuts (cf. ResNet 4.3.3). The complete architecture of the full network with its data flow from left to right is shown in Fig. 7.1.

Two sets of input features to the network exist: low level and high level features. Input to the LBN are only the low level features. It takes the four momenta of the final state particles, i.e., the neutrino via p_T^{miss}, one large radius jet, four small radius jets, and two leptons, as input. From these particles the LBN creates twelve output particles resulting in 229 output features (physics embeddings). The low level features, high level features, and the physics embeddings are then concatenated and fed into the ResNet architecture. The ResNet consists of three block with residual shortcuts. Each block comprises two layers of 229 nodes each except for the last layer which consists of nine output nodes corresponding to the aforementioned list of output classes. Batch normalization is applied after each layer of the ResNet. The used activation function is ReLU (see Fig. 4.3) except for the last layer, where the softmax activation function is used. The ResNet architecture then predicts based on all input information the probability of an event belonging to a certain physics process (output classes). A list of output classes is given as follows:

- Gluon fusion production mode of Higgs boson pair production (HH(GGF))
- Vector boson fusion production mode Higgs boson pair production (HH(VBF))
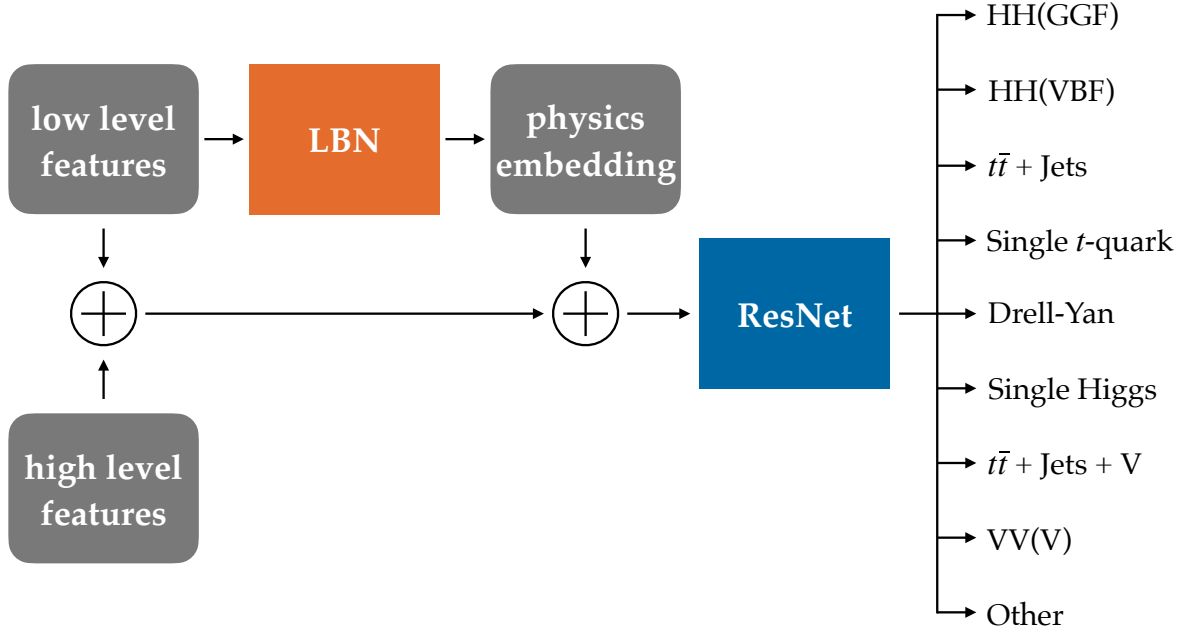- Top quark pair production ($t\bar{t}$ + Jets)

Figure 7.1: Sketch of the full network architecture including the LBN, the ResNet, and the data flow through them. Rounded gray boxes denote event information, colored boxes the LBN (orange) and the ResNet (blue) components, the black arrows the flow of event information, and the circled plus symbol the concatenation of all incoming event information.

- Single t-quark (ST) production
- Drell-Yan
- Single Higgs boson (H) production
- Vector boson associated top quark pair production ($t\bar{t}$ + Jets + V)
- Multiple vector boson production (Multiboson / VV(V))
- Other: all other, including W boson production with additional jets (W + Jets)

The architecture design exploits as much information of an event as possible by including basic information (low level), hand-reconstructed event information (high level), and autonomously reconstructed event information (physics embeddings). Maximizing input information results in a maximized prediction performance given the large enough network capacity of the ResNet.

7.2 Input Features and Event Reconstruction

The low and high level input features require the reconstruction of events. Events are considered as inputs to the neural network if they pass the event selection described in Sec. 6.3. In the following all low and high level event features are listed and described.

Low level features

Low level features correspond to event and particle information, which do not involve dedicated reconstruction algorithms. They include the four momenta of the p_T-leading two leptons, of the p_T-leading four small radius jets, of the p_T-leading large radius jet, and of the four

momentum representation of the missing transverse energy. These four momenta are the input variables to the LBN. In addition, the PDG-ID, the charge, and the $p_{\mathrm{T,cone}}$ of the leading two leptons are included. The softdrop mass and the subjettiness variables $\tau_i(i = 1..4)$ are also used as low level information.

High level features

High level features comprise event shape variables, features of intermediate resonances, and multi-particle information. Features of intermediate particles are calculated by summing the four momenta of a set of final state particles. Multiple intermediate particles are build with this strategy:

- $\ell\ell$: sum of the two p_{T}-leading leptons
- jj: sum of the two p_{T}-leading small radius jets
- $H_{b\bar{b}}$: sum of the two b-tagging probability leading small radius jets.
- $H_{W^+W^-}$: sum of $\ell\ell$ and $p_{\mathrm{T}}^{\mathrm{miss}}$
- HH: sum of $H_{b\bar{b}}$ and $H_{W^+W^-}$

From these intermediate particle hypotheses different features can be extracted:

- ΔR between the two p_{T}-leading leptons, the two b-tagging probability leading small radius jets, $\ell\ell$ and jj, and between $\ell\ell$ and $H_{b\bar{b}}$.
- minimum ΔR between the p_{T}-(sub)leading lepton and all small radius jets, the b-tagging probability (sub)leading small radius jet and both p_{T}-leading leptons, and between all small radius jets.
- $|\Delta\phi|$ between $p_{\mathrm{T}}^{\mathrm{miss}}$ and $\ell\ell$, and between $p_{\mathrm{T}}^{\mathrm{miss}}$ and $H_{b\bar{b}}$.
- minimum $|\Delta\phi|$ between all small radius jets.
- invariant mass m of $\ell\ell$, $H_{b\bar{b}}$ (b-quark jet energy regression applied, see Sec. 3.2.8.6), $H_{W^+W^-}$, and HH (b-quark jet energy regression applied, see Sec. 3.2.8.6).
- HH(VBF)-tag: existence of two small radius jets selected for the HH(VBF) production mode (see Sec. 6.3.4).
- boosted-tag: existence of a large radius jet with at least one b-quark subjet (see Sec. 6.3.4).
- the data-taking era (2016, 2017, or 2018).

In addition, two event shape variables are used:

$$H_{\mathrm{T}} = \sum_{j=1}^{\#\mathrm{jets}} p_{\mathrm{T}}^j \tag{7.1}$$

$$\text{and} \quad p_{\mathrm{T\,LD}}^{\mathrm{miss}} = 0.4 \cdot p_{\mathrm{T}}^{\mathrm{miss}} + 0.6 \cdot H_{\mathrm{T}}^{\mathrm{miss}}, \tag{7.2}$$

where $H_{\mathrm{T}}^{\mathrm{miss}}$ is the transverse momentum of the sum of four momenta of all loose leptons and small radius jets in an event. The $H_{\mathrm{T}}^{\mathrm{miss}}$ quantity is less sensitive to pileup effects than $p_{\mathrm{T}}^{\mathrm{miss}}$ as no hadrons with low momenta, which predominantly arise from pileup, are included in its calculation. The $p_{\mathrm{T\,LD}}^{\mathrm{miss}}$ variable is a linear combination of $p_{\mathrm{T}}^{\mathrm{miss}}$ and $H_{\mathrm{T}}^{\mathrm{miss}}$. A subset of high-level variables is shown in Fig. 7.2. Good agreement is observed between data and

simulation in all variables. The uncertainty model for the variables in Fig. 7.2 is simplified, neglecting shape-changing uncertainties.

(a) H_{T}.

(b) $p_{\mathrm{T\,LD}}^{\mathrm{miss}}$.

(c) Invariant mass m of HH.

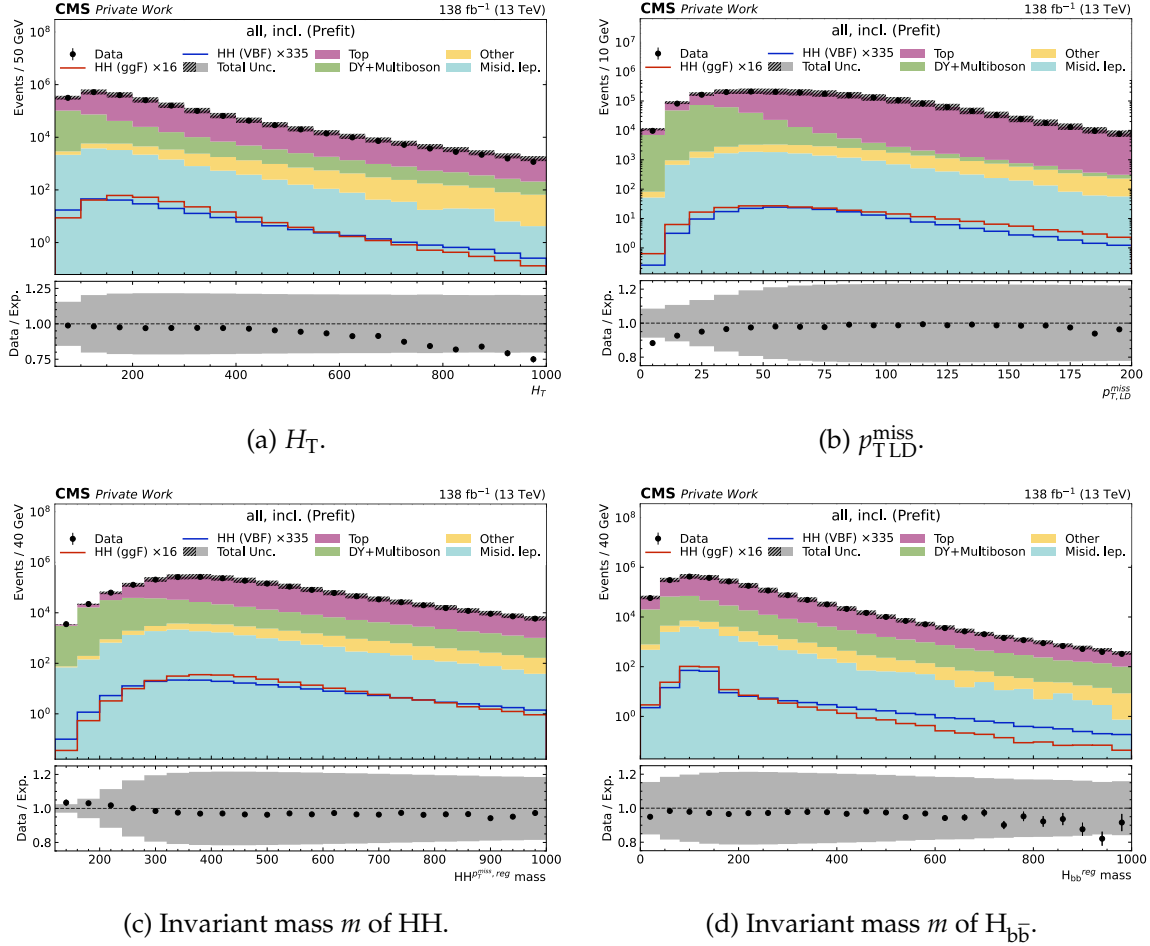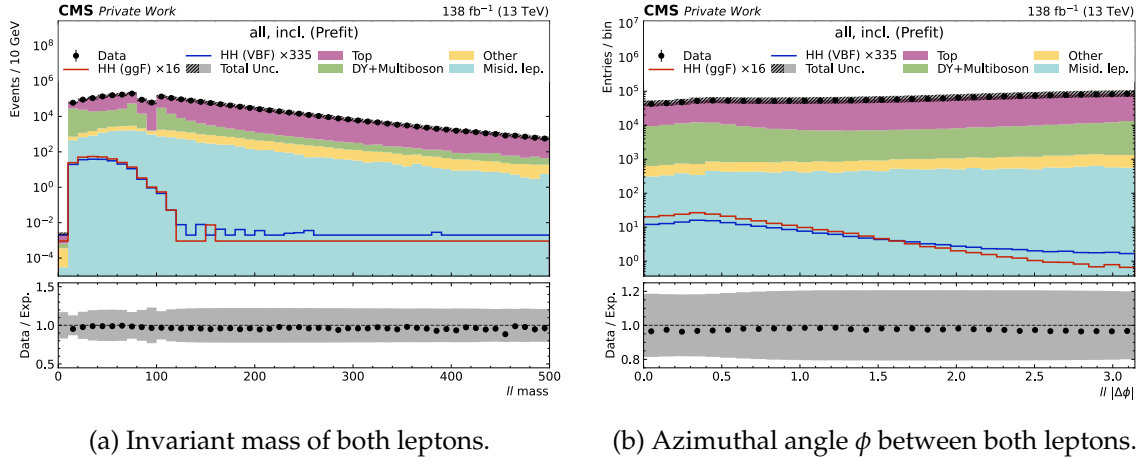(d) Invariant mass m of $\mathrm{H_{b\bar{b}}}$.

Figure 7.2: Distributions (pre-fit) of high-level variables in the most inclusive analysis regions after selection and corrections. A simplified uncertainty model is used, which neglects shape-changing uncertainties.

The dileptonic $\mathrm{H} \rightarrow \mathrm{W^+W^-}$ decay has the benefit of a narrow opening angle between the direction of flight of the two leptons as described in Sec. 2.2.3. The main background process ($t\bar{t}$ + Jets) does not have this feature as the W bosons originate from t-quark decays, which has a spin of $1/2$. The opening angle of the leptons are therefore significantly wider. Thus, the distribution of ΔR between the two p_{T}-leading leptons has strong discrimination power between the signal process and the $t\bar{t}$ + Jets background process; the same argumentation applies to the Drell-Yan process. In addition, other angles, such as the angle between the $p_{\mathrm{T}}^{\mathrm{miss}}$ and $\ell\ell$ system, also include the effect of the spin correlation. Moreover, the kinematic distributions of the two leptons, such as their transverse momentum and their invariant mass, also show a trend towards lower values. Figure 7.3 shows the invariant mass of the $\ell\ell$ system and the azimuthal angle between the two leptons in the most inclusive analysis regions.

The input variables are partly correlated with each other. The strength of their correlation is shown in the appendix in Fig. 10.3a for the low-level (kinematic) features of the input

(a) Invariant mass of both leptons.   (b) Azimuthal angle ϕ between both leptons.

Figure 7.3: Distributions (pre-fit) of high-level variables in the most inclusive analysis regions after selection and corrections, which benefit from the spin correlation effect in the fully leptonic $H \to W^+W^-$ decay. A simplified uncertainty model is used, which neglects shape-changing uncertainties.

particles, and in Fig. 10.3b for the high-level features.

7.3 Training Strategy

The neural network is trained with simulation only as described in Sec. 4.3.2, i.e., the Drell-Yan and Misid. leptons background process are not estimated as described in Sec. 6.5 and Sec. 6.6. Simulated events with negative event weights are removed as they effectively flip the sign of their contribution in the loss function by a flip of their true label. The training data is split in order to perform a 5-fold cross-validation, as described in Sec. 4.3.3, where the splits are 60% for training, 20% for validation and 20% for testing.

The events are weighted according to the corrections and scale factors described in Sec. 6.4. This removes potential biases when evaluating the neural network on real data and corrected simulation. In addition to these weights, events are weighted based on their physics process (w_p). It has been found empirically that the best weighting strategy is to weight both signal processes by a factor of $1/8$ and all background processes by a factor of 1.

The cross-entropy loss function, which is used here for the classification problem, is extended by the weights w_p as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_p^P w_p \sum_i^{N_p} \hat{y}_{i,p} \log\left(y_{i,p}\right), \tag{7.3}$$

where N denotes the total number of events in a batch, P the physics processes, N_p the number of events of a given physics process in a batch, $\hat{y}_{i,p}$ the true label of event i, and $y_{i,p}$ the predicted label of event i. In addition, the loss function is extended by an L2 penalty term (see Eq. 4.16) with a regularization factor of 10^{-8} to prevent overfitting. The weights of the LBN do not enter the L2 penalty term as they have a physical meaning and would be artificially suppressed.

The neural network is trained iteratively with the ADAM optimizer with a learning rate of 10^{-3}, which is reduced by a factor of 0.5 in case the validation loss does not improve for at least five epochs. A full epoch consists of 12000 batches with a batchsize of 256 events.

The hyperparameters of the training procedure and also of the neural network architecture (described in Sec. 7.1) are optimized with a Bayesian optimization algorithm [218]. After finding the optimal set of hyperparameters the process weights w_p have been further tuned by checking the effect of different combinations of process weights with respect to a measurement with Asimov data including all systematic uncertainties explained in Sec. 8.1. The chosen set of process weights is the one minimizing the expected upper limit on the total signal strength modifier ($\mu_{\text{HH(GGF)+HH(VBF)}}$) and on the HH(VBF) signal strength modifier ($\mu_{\text{HH(VBF)}}$).

7.4 Performance Evaluation and Introspection

Once the neural network is trained, it can be analyzed with respect to its performance and further understood through introspection. For this, three key metrics are discussed in this section: confusion matrices (performance), feature importances (introspection), and weights of the LBN (introspection).

Confusion matrices

The confusion matrix shows the true process label on the y-axis and the predicted process label on the x-axis. The entries of the matrix shows the fraction of events, which are predicted correctly on the diagonal, and the fraction of events, which are predicted wrongly on the off-diagonal elements. Two sets of matrices are produced, one with entries normed along the predicted label axis (Fig. 7.4a) and one with entries normed along the true label axis (Fig. 7.4b).

(a) Predicted labels are normed (columns). (b) Truth labels are normed (rows).
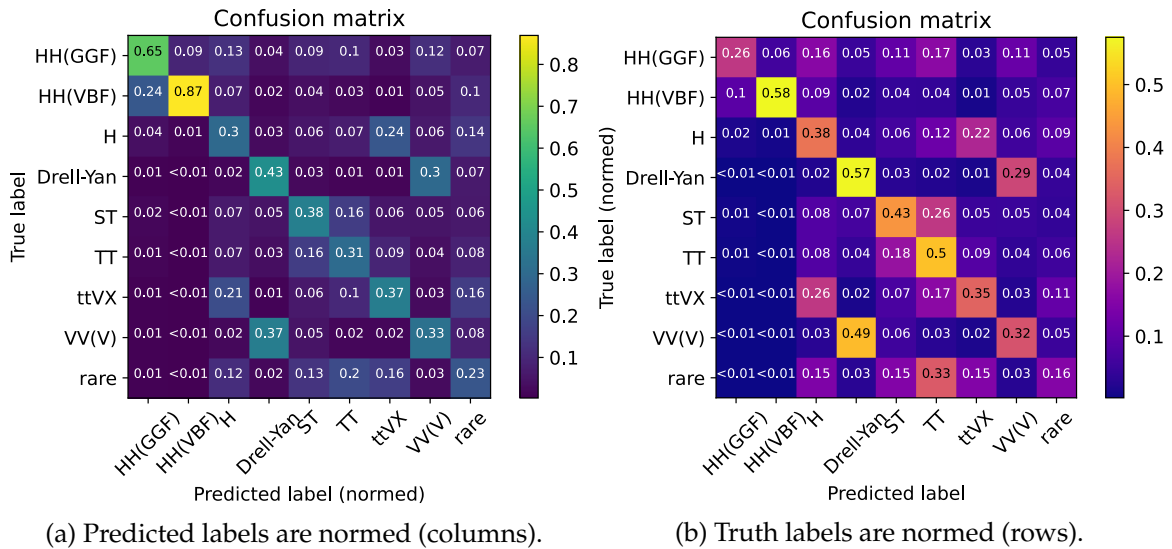
Figure 7.4: Confusion matrices for the nine output classes after training.

Most importantly it can be seen in both confusion matrices, that the main diagonal has the largest numerical values. Figure 7.4a shows that the fraction of HH(GGF) and HH(VBF) events in their corresponding output classes are 65% and 87%. However, Fig. 7.4b illustrates

that only 26% of HH(GGF) events are categorized into the HH(GGF) output class, similarly only 58% of HH(VBF) events into the HH(VBF) output class. This is due to the fact that the signal processes are weighted down by a factor of $1/8$. Although the down-weighting diminishes the number of signal events in their respective output classes, it has a clear benefit: the confusion of background processes in the HH(GGF) and HH(VBF) signal output classes are diminishingly small with approximately $\leq 1\%\ldots 2\%$. In addition, confusions between background can be identified in groups primarily due to their similar event topologies: single-top (ST) and $t\bar{t}$ + Jets (TT), Drell-Yan and Multiboson (VV(V)), single Higgs boson processes, ttVX ($t\bar{t}$ + Jets + V), and other rare processes.

Feature importances

The importance of all input features (see Sec. 7.2) can be estimated with the value of the loss function. By permuting, or "shuffling", the input position of a feature and re-evaluating the loss function, it can be compared to the original loss value. It is expected that the re-evaluated loss value is larger, because the relation between the input feature and the output class is essentially lost. The difference between the original and the re-evaluated value of the loss function is used as a measure of importance. The larger the difference, the more important is the feature. This method is known as *permutation importance* [219]. Figure 7.5 shows the feature importance using the *permutation importance* method for the high level features.

The feature importance of the high-level variables shows clear trends. The H_T event shape variable is ranked the highest in importance. The other event shape variable ($p_{\mathrm{T\,LD}}^{\mathrm{miss}}$) closely follows on rank three. Invariant mass variables of the $\ell\ell$, the $\mathrm{H}_{\mathrm{b\bar{b}}}$, and the HH system are the second highest group of features. Especially, the invariant $\ell\ell$ mass is a strong discriminant. Other high-level features, such as angular relations between leptons, jets, and combinations thereof benefit from the spin correlation effect, and thus rank also high. The boosted tag and VBF related features rank lower as they describe unique event topologies, which rather rarely occur.

The importances of the LBN inputs and the high-level variables cannot be shown together as the order of the LBN inputs have an intrinsic relation: particles are boosted into rest frames. The shuffling would break the Lorentz boosts of particles into rest frames resulting in arbitrary predictions of the neural network. The importances of these two sets of input variables can therefore not be directly compared, and are analyzed independently. The feature importance for the LBN input variables can be found in the appendix in Fig. 10.4. For the LBN input particles the p_T-leading small radius jet and the missing transverse energy have the largest importance for the classification. The least important features are the large radius jet features as only few if any large radius jets exist in an event, and the z-component of the missing transverse energy, which is always 0 by definition.

Weights of the Lorentz Boost Network

The weights of the linear combinations build by the LBN are shown in Fig. 7.6a for the particle combinations and in Fig. 7.6b for the rest frame combination. In both figures the weights are normalized along each column to the absolute value of 100.

These two dimensional weight matrices depict the input particles on the y-axis and the combined particles (or rest frames) on the x-axis. Particle combinations resulting in intermediate particle candidates of the signal process can be identified in Fig. 7.6a. LBN particle number 0
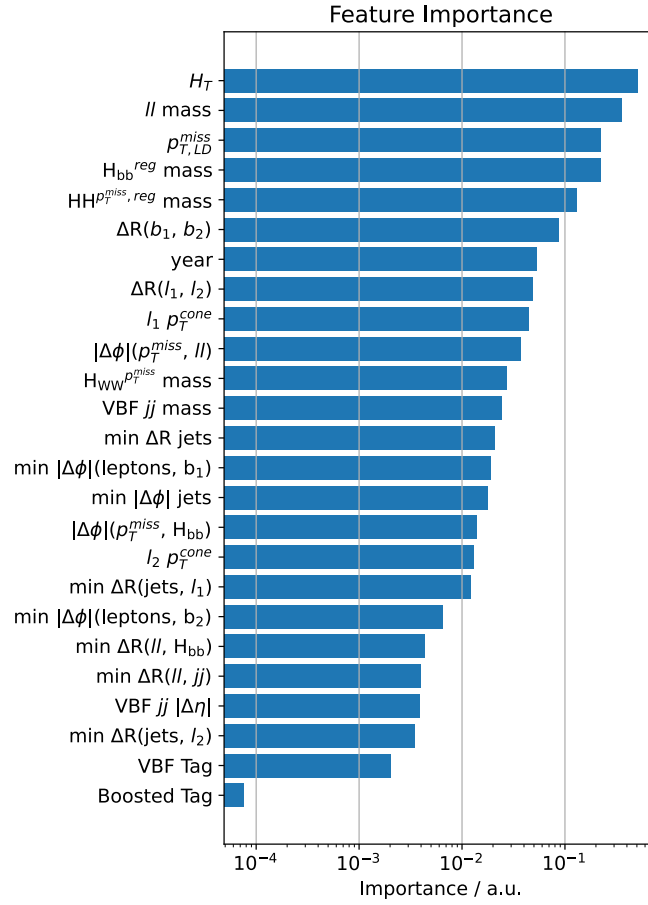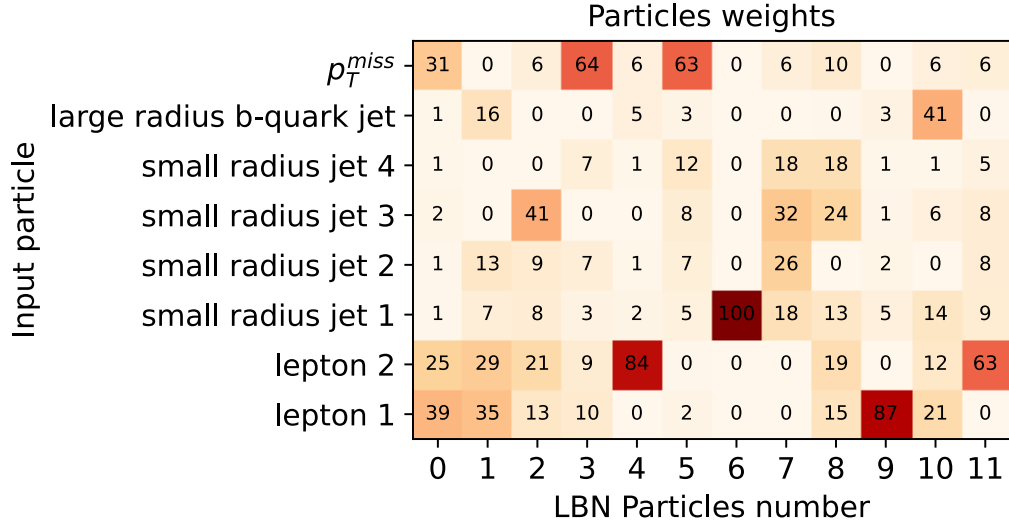
Figure 7.5: Feature importances for the high-level features based on the *permutation importance* method.
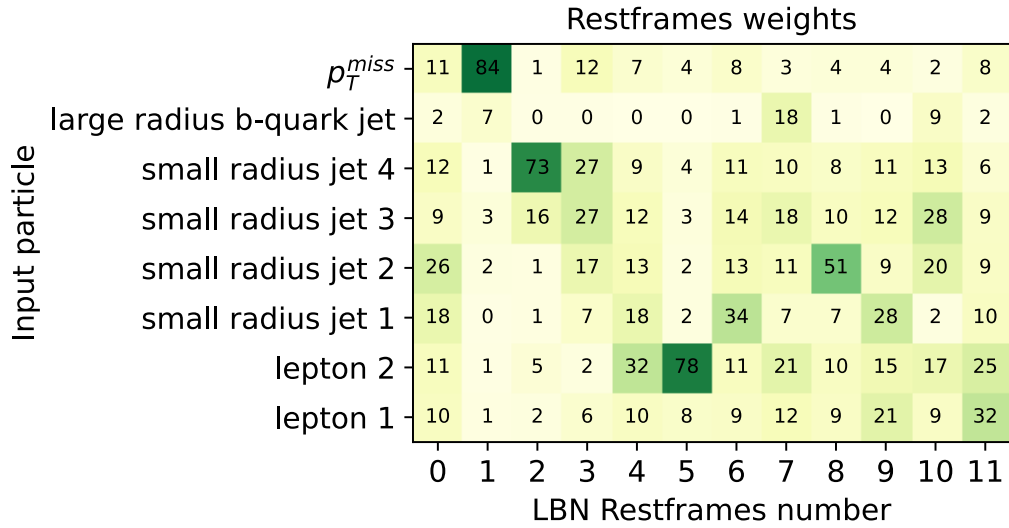
combines the four momenta of both leptons with p_T^{miss}, which resembles the leptonic branch of the H $\rightarrow$ W$^+$W$^-$ decay. Furthermore, LBN particle number 7 combines the four small radius jets (with a focus on jet 2 and jet 3). This combination resembles the H $\rightarrow$ b$\bar{\text{b}}$ candidate, where two additional jets may stem from the VBF production mode. Other combinations (LBN particle number: 2, 8, and 10), which involve leptons and jets, may resemble the main background $t\bar{t}$ + Jets. Finally, it can be seen that a major focus in the other LBN particles is on a single lepton or p_T^{miss}. The rest frame combinations (Fig. 7.6b) often involve all input particle with similar fractions (i.e., LBN rest frames 0, 4, 6, 7, 9, and 10). Special focus on the leptons and p_T^{miss} can be noticed in the LBN rest frames 1, 5, and 11.

7.5 Event Categorization

An event is categorized by selecting the maximum output of the probability-like DNN prediction, its jet topology and its b-quark jet multiplicity. Some of the predicted physics processes, listed in Sec. 7.1, are further grouped together based on their kinematic similarities, resulting in similar DNN output distribution shapes. In addition, these similarities can also be found in the confusion matrices (Fig. 7.4a and Fig. 7.4b) where similar processes often are confused with each other. This grouping significantly reduces the computational cost of the

Particles weights

Input particle	0	1	2	3	4	5	6	7	8	9	10	11
p_T^{miss}	31	0	6	64	6	63	0	6	10	0	6	6
large radius b-quark jet	1	16	0	0	5	3	0	0	0	3	41	0
small radius jet 4	1	0	0	7	1	12	0	18	18	1	1	5
small radius jet 3	2	0	41	0	0	8	0	32	24	1	6	8
small radius jet 2	1	13	9	7	1	7	0	26	0	2	0	8
small radius jet 1	1	7	8	3	2	5	100	18	13	5	14	9
lepton 2	25	29	21	9	84	0	0	0	19	0	12	63
lepton 1	39	35	13	10	0	2	0	0	15	87	21	0

LBN Particles number

(a) LBN weights (particle combinations).

Restframes weights

Input particle	0	1	2	3	4	5	6	7	8	9	10	11
p_T^{miss}	11	84	1	12	7	4	8	3	4	4	2	8
large radius b-quark jet	2	7	0	0	0	0	1	18	1	0	9	2
small radius jet 4	12	1	73	27	9	4	11	10	8	11	13	6
small radius jet 3	9	3	16	27	12	3	14	18	10	12	28	9
small radius jet 2	26	2	1	17	13	2	13	11	51	9	20	9
small radius jet 1	18	0	1	7	18	2	34	7	7	28	2	10
lepton 2	11	1	5	2	32	78	11	21	10	15	17	25
lepton 1	10	1	2	6	10	8	9	12	9	21	9	32

LBN Restframes number

(b) LBN weights (rest frame combinations).

Figure 7.6: Two dimensional matrices of LBN weights for particle and rest frame combi-
nations. The weights are normalized along each column to the absolute value
of 100.

analysis, while preserving its sensitivity. The list of (grouped) processes considered for the
event categorization is as follows:

- HH(GGF)
- HH(VBF)
- Top + Other: $t\bar{t}$ + Jets, ST, H, $t\bar{t}$ + Jets + V, Other
- Drell − Yan + Multiboson: Drell-Yan, Multiboson

The second categorization scheme is based on the jet topology and b-quark jet multiplicity (see Sec. 6.3.4). The resulting final categories are shown in Fig. 7.7.

Figure 7.7: The event categorization based on the physics process classification, and the multiplicity and topology of the b-quark jets.

The HH signal regions (SR) are defined by the HH(GGF) and HH(VBF) production modes. In addition, they are categorized by the number of b-quark jets and their topology (resolved and boosted) resulting in six signal regions. The categorization into two b-quark jets enhances the sensitivity to the signal process as one of the Higgs bosons decays to a pair of bottom quarks. Moreover, distinguishing between resolved and boosted event topologies particularly enhances the sensitivity to κ_{2V} close to the SM expectation $\kappa_{2V} \approx 1$ (resolved) and far away from the SM expectation $|\kappa_{2V}| \gg 1$ (boosted). The two control regions (CR) are defined primarily by the two most dominant background process groups, namely the $t\bar{t}$ + Jets and the Drell-Yan process. The Top + Other control region is further split into two sub categories, based on the b-quark jet topology.

7.6 Binning Optimization

Binning optimization is crucial for the final sensitivity, as the likelihood is defined as a product of all bins. Thus, the signal to background ratio needs to be maximized while keeping as few bins as possible to minimize computational costs without the loss of sensitivity as well as having a sufficient statistical description in each bin.

At first, the output prediction of the physics process multi classification is histogrammed with 400 equidistant bins between $0 \ldots 1$ per histogram. The number of initial bins (400) is arbitrary and is chosen to be large enough to have enough flexibility for combining bins

afterwards. The binning strategy for signal and background categories bases on a quantile binning, but differs in their details. A quantile binning is constructed by calculating the cumulative distribution of a reference histogram and then split it into equally sized portions. In other words, a histogram is rebinned such that in each bin the same fraction of events is found, essentially flattening the reference histogram.

The signal categories are rebinned iteratively using this procedure. The procedure starts by rebinning the 400 bins to 15 bins where the number of signal events is the same in each bin. If the total bin content of the last (most sensitive) bin is larger than 10 and has a statistical uncertainty smaller than 30% the binning is applied. In case both conditions are not met the procedure is repeated with 14 bins (one less as the iteration before). This is done iteratively until either the conditions are met, or only one bin is left. The quantiles are calculated with respect to the matching signal process of the respective category, i.e., the HH(GGF) process in the HH(GGF) categories. This works well for neural network based classification scores, as they are designed such, that signal is populated towards one and background towards zero.

The background categories are expected to have a low signal to background ratio and a high overall integral. Thus, a simple rebinning strategy is applied of five quantile bins. The quantiles are calculated with respect to the sum of all background processes instead of the signal histogram.

The number of bins within each category across all three eras are the same defined by the era with the least number of bins.

7.7 Neural Network Predictions

The neural network is not only used for classifying events into physics processes. In addition, the output score of the neural network is leveraged to build the most sensitive variable in each category. The maximum prediction score of an event decides in which output class the event will end up. The score itself is histogrammed in the respective output class. The resulting neural network output distributions are shown in Fig. 7.8.

No significant deviations between data and the expectation is observed in these distributions. In the signal categories (Fig. 7.8a and Fig. 7.8b) the signal to background ratio improves the closer the score is to one. The two Top + Other regions (Fig. 7.8c) are enriched with the $t\bar{t}$ + Jets and single t-quark production process. Likewise, the Drell − Yan + Multiboson region is enriched with the Drell − Yan + Multiboson background process (Fig. 7.8d), especially for the most right bins. These enriched and pure background regions allow to control the Top and Drell − Yan + Multiboson processes in the measurement. This helps to precisely determine their normalization and shape in the signal regions in a simultaneous fit of all categories.

(a) HH(GGF).

(b) HH(VBF).

(c) Top + Other.

(d) Drell − Yan + Multiboson.
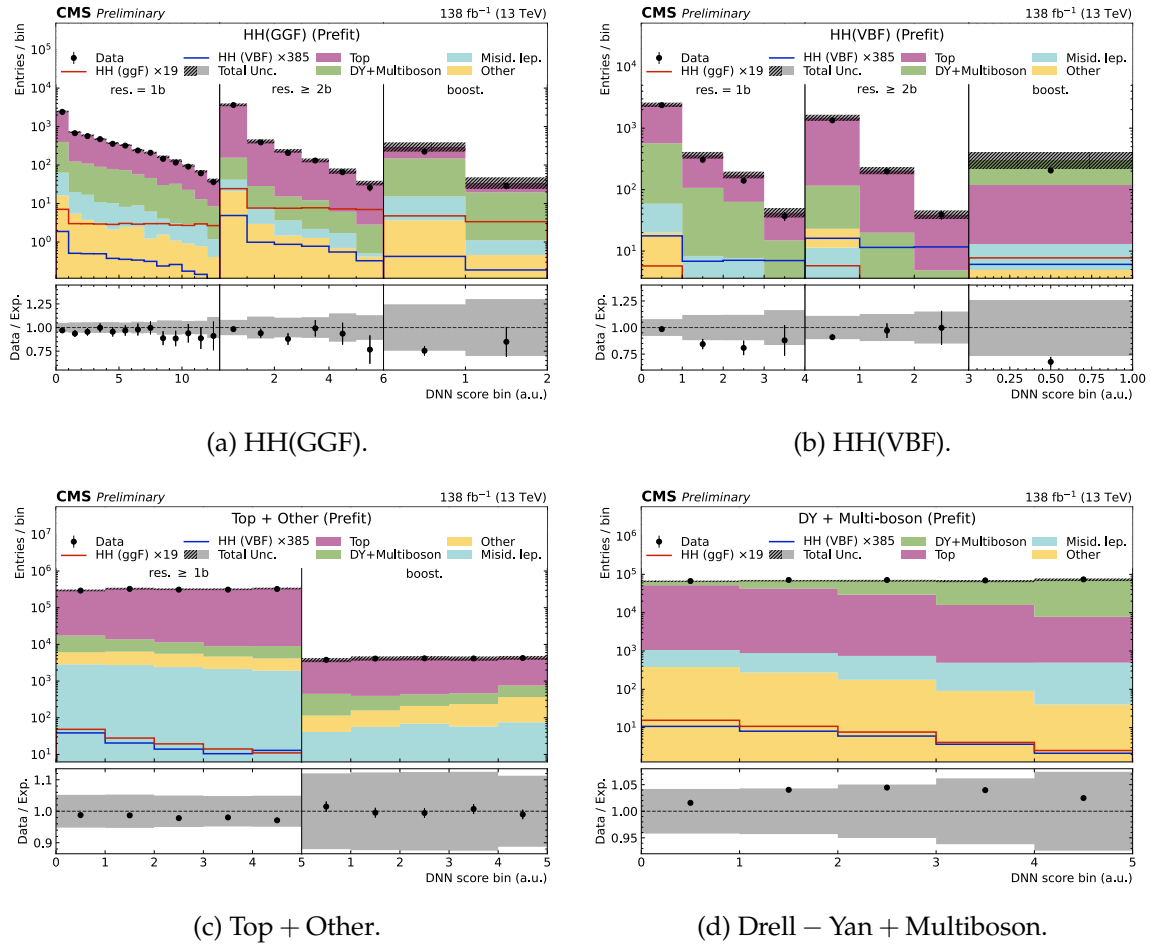
Figure 7.8: Neural network output scores for the six signal and three background categories (see Sec. 7.5). The distributions include the full uncertainty model and are shown before a fit was performed.

Chapter 8

Statistical Inference

This chapter presents the measurement procedure, the uncertainty model, and the results of the search for the HH $\to$ b$\bar{\text{b}}$W$^+_\ell$W$^-_\ell$ process. The measurement is based on a likelihood model, which is parametrized by parameters of interest, such as the signal strength or coupling modifier strengths. The likelihood model comprises the recorded data, the expectation (estimated from simulation and data-driven methods), and statistical and systematic uncertainties. The likelihood function (see Sec. 4.4.1 and Sec. 5.3) is as follows:

$$\mathcal{L}\left(\lambda, \theta | d\right) = \prod_i^n \frac{\lambda_i^{d_i}(\theta)}{d_i!} e^{-\lambda_i(\theta)} \cdot \prod_j \pi_j\left(\theta_j\right), \tag{8.1}$$

where n is the number of bins in the DNN output distribution, d_i the number of data events in bin i, λ_i the number of expected events in bin i, θ the set of nuisance parameters, and $\pi_j\left(\theta_j\right)$ the penalty functions for the nuisance parameters. The expectation $\lambda(\theta)$ describes the signal plus background hypotheses, which depends on a set of nuisance parameters θ. In addition, signal processes and single Higgs boson processes depend on different signal strengths μ and coupling modifiers κ:

$$\text{with} \quad \lambda(\theta) = \mu \cdot \left[\underbrace{\mu_{\text{HH(GGF)}} \cdot s_{\text{HH(GGF)}}\left(\kappa_\lambda, \kappa_t, \theta\right)}_{\text{HH(GGF) signal}} + \underbrace{\mu_{\text{HH(VBF)}} \cdot s_{\text{HH(VBF)}}\left(\kappa_\lambda, \kappa_{2V}, \kappa_V, \theta\right)}_{\text{HH(VBF) signal}} \right] \tag{8.2}$$

$$+ \underbrace{b_{\text{H}}\left(\kappa_\lambda, \kappa_t, \kappa_V, \theta\right)}_{\text{SM Higgs}} + b_{\text{other backgrounds}}. \tag{8.3}$$

By performing a fit of the free yet potentially constrained parameters to the observed data, which maximizes the likelihood function (Eq. 8.1), parameter of interests, such as the signal strengths μ and coupling modifiers κ, as well as all nuisance parameters can be measured.

The subsequent section first explains the different sources of uncertainties, that may modify the expectation λ through a set of nuisances parameters θ. These uncertainties are split into two different sources: experimental and theoretical uncertainties. Afterwards, the results of the fit with the full uncertainty model is discussed. Finally, post-fit distributions and their modelling quality, and impacts of the uncertainties are shown.

8.1 Systematic Uncertainties

This section describes the different sources of systematic uncertainties. These uncertainties may modify the model expectation λ of Eq. 8.1, and are included in the likelihood function as nuisance parameters θ. These have typically prior values from theory prediction of auxiliary measurements, and are constrained through penalty functions $\pi_j\left(\theta_j\right)$ (Eq. 8.1). Two types of systematic uncertainties are included which act as described: rate-changing nuisance parameters with a log-normal penalty term and shape-changing nuisance parameters. The latter may also affect the rate. In addition, dedicated (unconstrained) rate-changing parameters for some of the background processes are included which are not constrained through penalty functions. In total, 239 nuisance parameters are included in the likelihood function, excluding nuisance parameters accounting for the limited number of generated events in simulation.

All systematic uncertainties are described in the following. In addition, the correlation and naming scheme for each type of uncertainty is explained, and exemplary distributions of some of the leading systematic variations are shown.

8.1.1 Experimental Uncertainties

Experimental uncertainties primarily arise from measurement inaccuracies of the CMS detector, differences between simulation and recorded data, and uncertainties related to analysis methods. The following paragraphs describe all experimental uncertainties considered in this analysis in greater detail.

Integrated luminosity

The integrated luminosity is required to to scale the normalization of simulated background contributions to to the expectation in 2016, 2017, and 2018 as described in Eq. 6.1. Table 8.1 shows the individual relative uncertainties of the integrated luminosity. They are given for each year and for correlations between years.

Table 8.1: Correlation scheme of the integrated luminosity uncertainties.

Name	2016	2017	2018
Luminosity 2016	1.0%	✗	✗
Luminosity 2017	✗	2.0%	✗
Luminosity 2018	✗	✗	1.5%
Luminosity 2017 & 2018	✗	0.6%	0.2%
Luminosity Run 2	0.6%	0.9%	0.2%

These uncertainties act as rate-changing nuisance parameters, whose expectations are modeled with a log-normal distribution.

Pileup

The pileup distributions differs between simulation and recorded data [208]. The pileup distribution of simulation is therefore corrected as described in Sec. 6.4. This correction procedure involves the minimum bias cross section, which is varied by $\pm 4.6\%$ [220] and propagated through to varied event weights applied for the correction. This results in a shape-changing effect on the DNN output distributions. This uncertainty is correlated among all

years. It is named "Pileup" in the scope of this analysis. The up and down histogram variations are shown in Fig. 8.1.
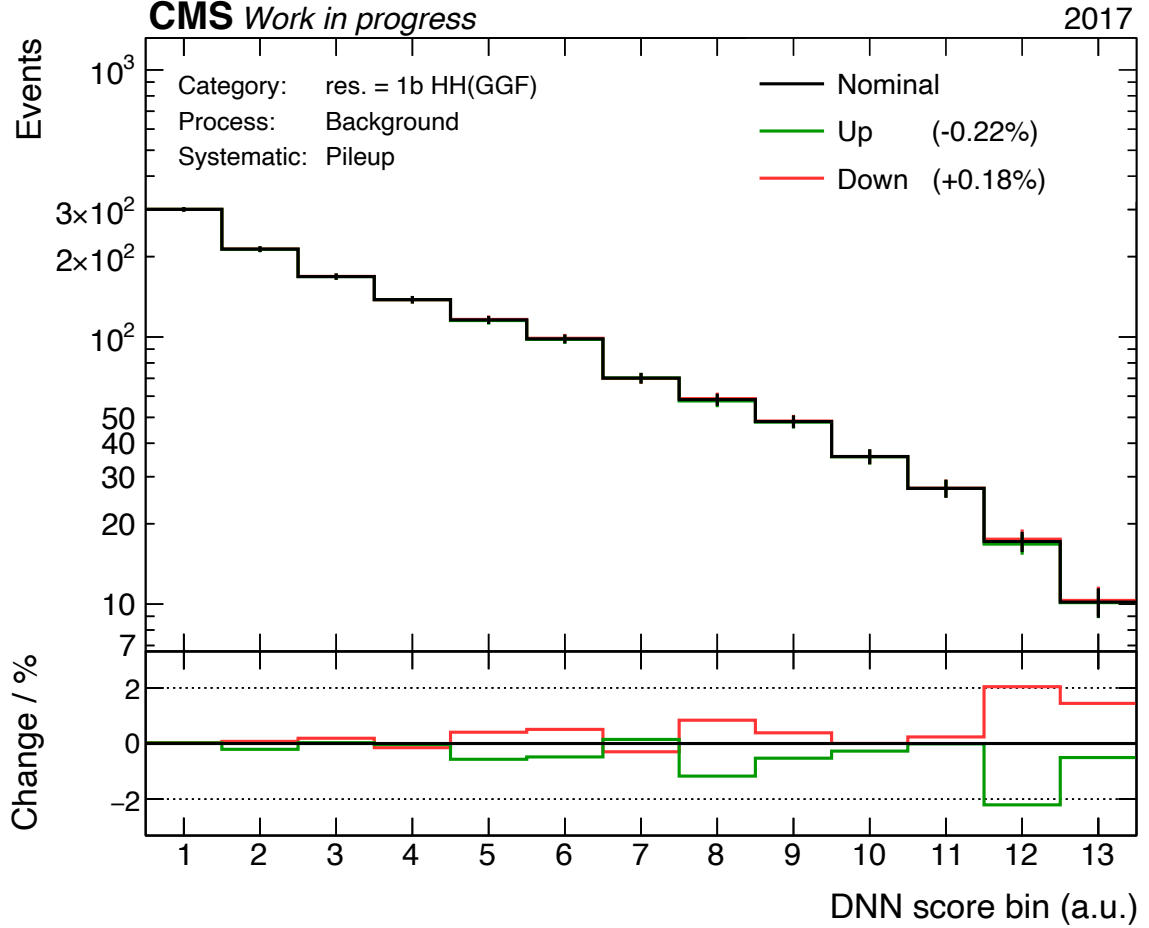
Figure 8.1: Systematic variations of the pileup uncertainty in the res. = 1b HH(GGF) category in 2017 are shown. The histograms correspond to the sum of all backgrounds. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

Trigger efficiencies

Corrections factors, depending on the $p_{T, cone}$ of leading lepton, are derived to correct differences between simulation and recorded data after applying triggers (described in Sec. 6.4). The measurement of these correction factors comes with systematic uncertainties and are provided by the ttH multilepton analysis [201, 209] of the CMS collaboration. These uncertainties have a shape-changing effect on the DNN output distribution. They are decorrelated across the different lepton channels (ee, eμ, $\mu\mu$) (see Sec. 6.3.3) and years. The naming scheme for these uncertainties is "Trigger (DL) {ee, eμ, $\mu\mu$} {*year*}", e.g., "Trigger (DL) ee 2016".

Lepton efficiencies

Correction factors for the lepton reconstruction and identification efficiencies are derived to remove differences between their performance on simulation and recorded data. The mea-

surement procedure of these correction factors is described in Sec. 6.4. Uncertainties of this procedure are provided by the ttH multilepton analysis [201, 209] of the CMS collaboration. Similar to the trigger efficiency uncertainties, the ones for the lepton efficiencies have also a shape-changing effect on the DNN output distribution, and are decorrelated across the different lepton channels ($\ell\ell$) (see Sec. 6.3.3) and years. These uncertainties follow the naming scheme "Lepton ({e, μ}) eff. ({loose,tight,ttH}) {*year*}", e.g., "Lepton (e) eff. (ttH) 2016".

Jet energy scales and resolution

The four momenta of jets are corrected in simulation and data as described in Sec. 3.2.8.4. This correction comprises an adjustment of the energy scale and its resolution. The uncertainties of the energy scale correction are factorized into eleven different sources [221] of which some are decorrelated per year, correlated among all years, or both. Table 8.2 summarizes the correlation scheme of the eleven different energy scale contributions. The uncertainties for the jet energy scale follow the naming of "JES ({*type*}) {*year*}", e.g., "JES (Abs) 2016", where the *year* is omitted for correlated uncertainties in Run 2.

Table 8.2: Correlation scheme of the jet energy scale uncertainties. Run 2 denotes the correlation across all three years.

Name	2016	2017	2018	Run 2
JES (Abs)	✓	✓	✓	✓
JES (BBEC1)	✓	✓	✓	✓
JES (EC2)	✓	✓	✓	✓
JES (FlavQCD)	✓	✗	✗	✗
JES (HF)	✓	✓	✓	✓
JES (RelBal)	✓	✗	✗	✗
JES (RelSample)	✗	✓	✓	✓

In addition, one uncertainty related to the resolution of the energy of jets is considered. This uncertainty is estimated by varying the respective scale factors provided by the CMS collaboration [222]. The uncertainty for the jet energy resolution is decorrelated between all three years and named "JER {*year*}" (e.g. "JER 2016"). Exemplary up and down histogram variations are shown for the JES (Abs) and the JER uncertainties in Fig. 8.2.

Unclustered energy

The calculation of p_T^{miss} involves clustered objects (i.e., reconstructed particles), but also unclustered energy deposits [200]. An uncertainty on the unclustered energy is propagated to the calculation of p_T^{miss} by varying these deposits within their energy resolution. This uncertainty has a shape-changing effect and is labeled by the naming scheme "Unclustered energy (MET) {*year*}" (e.g. "Unclustered energy (MET) 2016"). Exemplary up and down histogram variations are shown for the unclustered energy uncertainty in Fig. 8.3.

Jet pileup identification efficiencies

The jet pileup identification [101, 102] is used to remove jets arising from pileup collisions as described in Sec. 6.3.4. Uncertainties with respect to this identification are considered in this analysis. They are estimated by varying the efficiency and mistag rates provided by the CMS collaboration. These uncertainties vary the shape of the DNN output distributions,
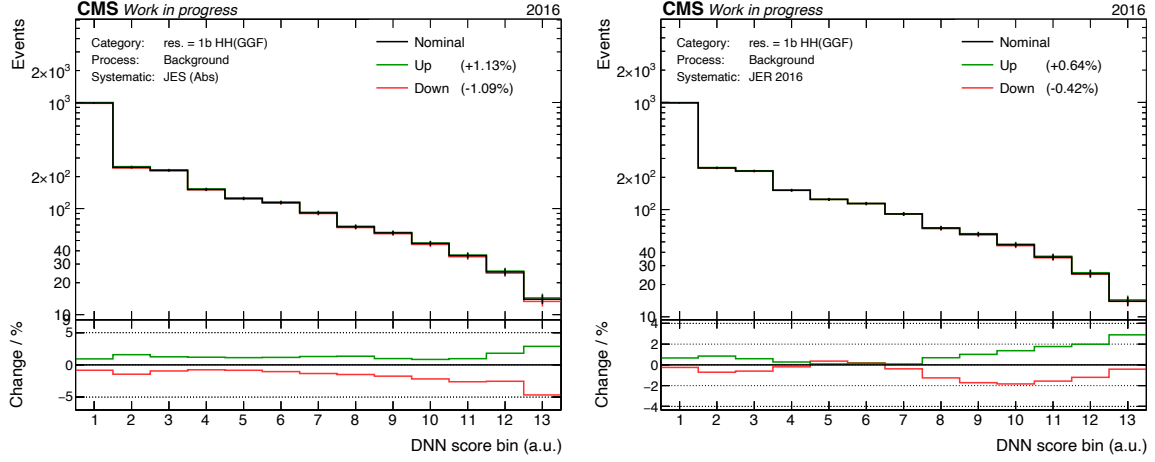
Figure 8.2: Systematic varations of the JES (Abs) (left) and JER (right) uncertainties in the res. = 1b HH(GGF) category in 2016 are shown. The histograms correspond to the sum of all backgrounds. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

are decorrelated across the three years, and are named according to the scheme "Pileup Jet ({ID,mistag}) {*year*}" (e.g. "Pileup Jet (ID) 2016").

b-quark identification efficiencies

b-quark jets are identified as described in Sec. 6.4. Uncertainties are estimated by varying the respective scale factors within their uncertainties. This results in a multitude of uncertainties, which can be roughly separated related to four different groups: heavy flavor, light flavor, charm flavor, and subjets of large radius jets. Uncertainties related to the contamination of heavy and light flavor jets are split into statistical uncertainties, which are decorrelated per year, and a systematic uncertainty each, which is correlated among all years. The uncertainties related to the charm flavor only consider a systematic effect correlated across all three years; no statistical uncertainties are considered related to the charm flavor. Finally, one uncertainty is included for the b-quark subjet identification of large radius jets decorrelated across the three years. All of these uncertainties introduce a shape-changing effect of the DNN output distributions. These uncertainties are prefixed with "B-tagging", (e.g. "B-tagging (heavy flav. stats 1) 2017"). The up and down variations of the "B-tagging (light flav.) Run 2" uncertainty are shown in Fig. 8.4.

t-quark transverse momentum correction

The p_T spectrum of t-quarks on generator level are modeled too hard in simulation, and thus these jets are corrected as explained in Sec. 6.4. An uncertainty of this correction is estimated by not applying the correction at all, and by applying it twice. It is correlated among all three years, and labeled "Top p_T reweighting". The up and down variations of the "Top p_T reweighting" uncertainty are shown in Fig. 8.5.

HEM issue

Due to power supply interruptions during the end of the 2018 data-taking period, two sectors, namely HEM15 and HEM16, of the HCAL endcap detector system were not operating.
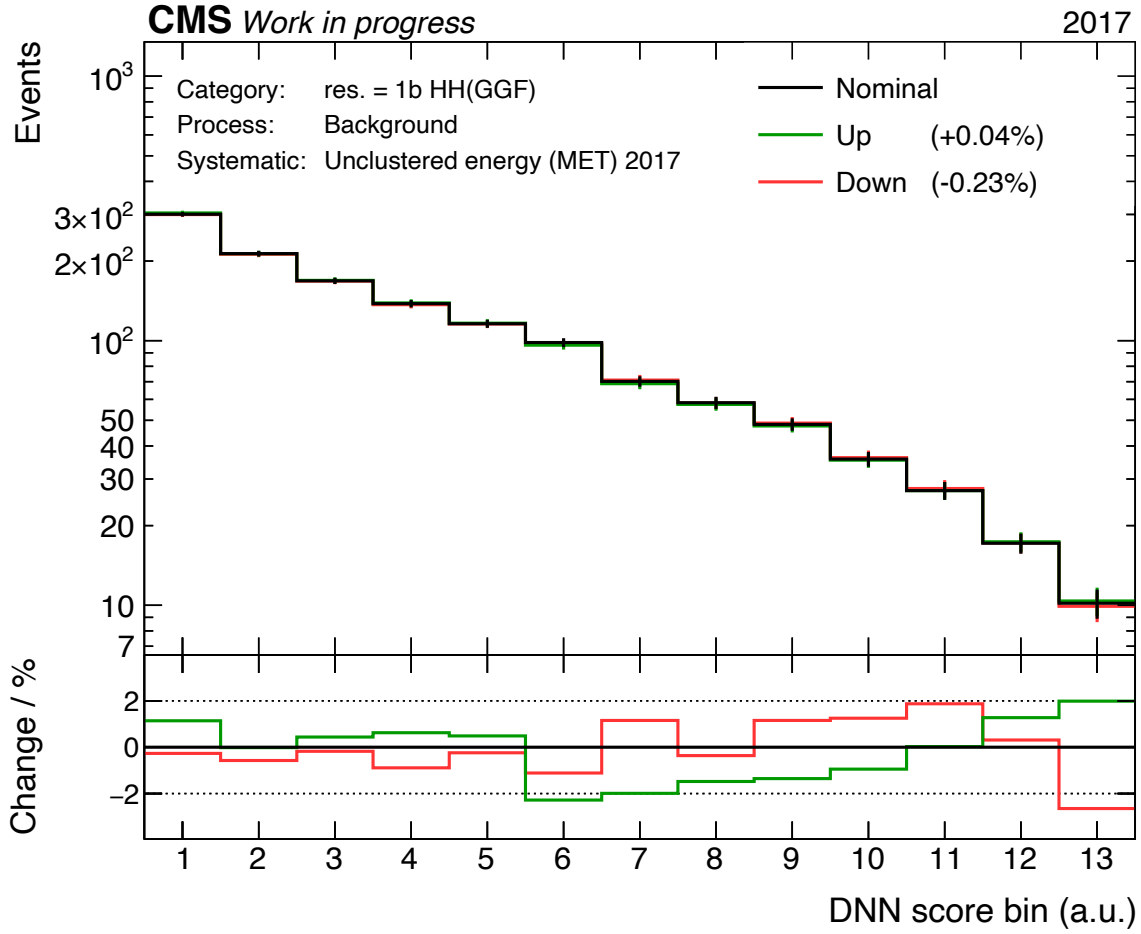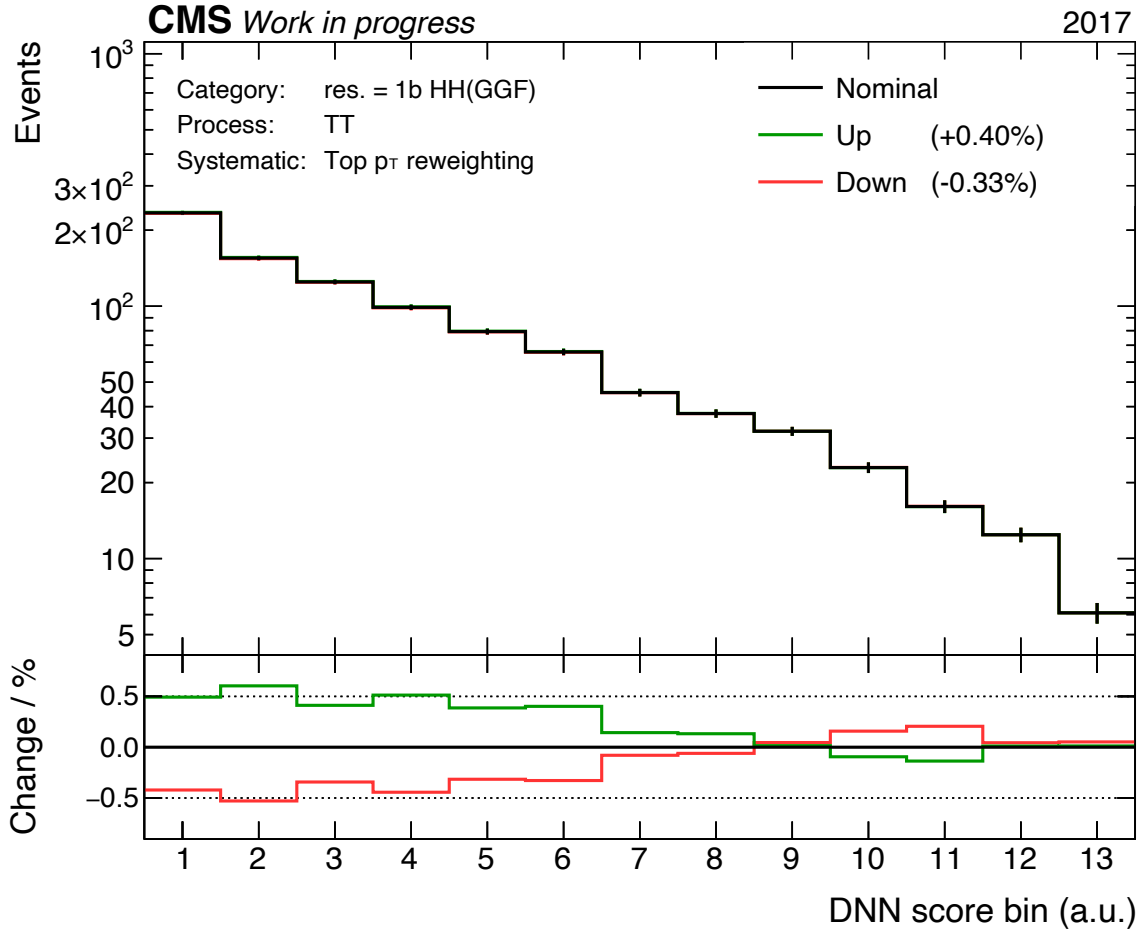
Figure 8.3: Systematic variations of the unclustered energy uncertainty in the res. = 1b
HH(GGF) category in 2017 are shown. The histograms correspond to the
sum of all backgrounds. Green denotes the up variation, and red the down
variation. The ratio of the up and down variation to the nominal histogram
is shown in the ratio plot.

The energy of jets in these regions need to be varied to compensate for this effect. Jets with
transverse momentum of at least 15 GeV in the region of $-1.57 < \phi < -0.85$ passing the
tight working point identification are affected. Their energy is scaled down by 20% in the
region of $-2.5 < \eta < -1.3$, and down by 35% in the region of $-3.0 < \eta < -2.5$. This energy
variation is propagated to event-level quantities, such as p_T^{miss}, which are affected by the jets.
This variation affects the DNN output distribution with a shape-changing effect. It is named
"JES (HEM, 2018)".

L1 ECAL trigger pre-firing

The pre-firing effect of the L1 ECAL trigger and its correction for simulation is described in
Sec. 6.4. An uncertainty for this is estimated by varying the pre-firing probabilities within
their own uncertainties. It has a shape-changing effect and only affects the data-taking pe-
riods of 2016 and 2017; thus it is decorrelated across these two years. It follows the naming
scheme of "L1 ECAL Prefiring ({*year*})", e.g., "L1 ECAL Prefiring (2017)".
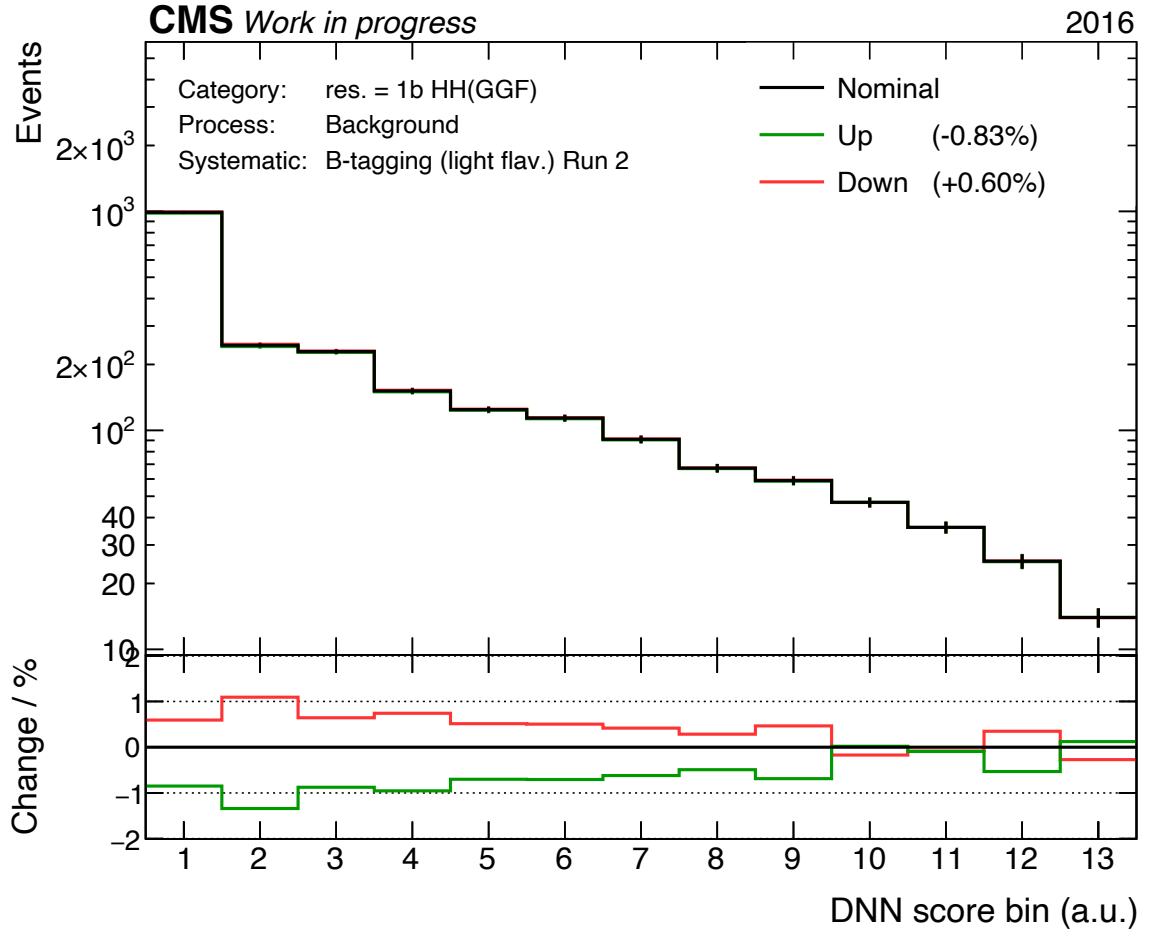
Figure 8.4: Systematic variations of the "B-tagging (light flav.) Run 2" uncertainty in the res. = 1b HH(GGF) category in 2016 are shown. The histograms correspond to the sum of all backgrounds. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

Estimation of misidentified leptons

Misidentified leptons are typically not well modeled in simulation. Thus, a data-driven method is used to estimate their contribution as explained in Sec. 6.5. It involves so-called fake-factors to extrapolate events from the non-prompt "fake" region (FR) to the non-prompt signal region (SR). The measurement of these fake-factors are subject to the statistical precision of their measurement. The fake-factors are provided by the ttH multilepton analysis [201, 209] of the CMS collaboration and are varied according to their uncertainties. The resulting uncertainties have a shape-changing effect on the estimation of the Misid. leptons background. It has been found that their effect depends on the lepton flavor and p_T, the detector region (barrel or endcap), and the year. Thus, each combination of these parameters contributes as an individual uncertainty. The uncertainties related to the Misid. leptons estimation follow the naming scheme of "Fakes est. (DL) $\{e, \mu\}$, $\{Barrel, Endcap\}$ $\{p_\mathrm{T}\ region\}$ $\{year\}$", e.g., "Fakes est. (DL) μ, Barrel $15 < \mathrm{pT} < 20$ 2018".

Figure 8.5: Systematic variations of the "Top p_T reweighting" uncertainty in the res. = 1b HH(GGF) category in 2017 are shown. The histograms correspond to the $t\bar{t}$ + Jets process. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

Estimation of the Drell-Yan process

The Drell-Yan process is estimated with a data-driven method as described in Sec. 6.6. Uncertainties related to the estimation procedure are twofold: one based on statistical uncertainties in the calculation of transfer factors, and one based on the described closure correction of the estimation. The former ones are decorrelated by the categorization, by the number and topology of b-quark jets in an event and by each year. They introduce a shape-changing effect and are named in the scheme "DY est., {*category*} {*year*}", e.g., "DY est., res. $\geq$ 2b 2018". The other type of uncertainties are related to the closure correction of the estimation method. These are split into the background enriched categories, where shape-changing variations of the fit parameters are estimated as explained in Sec. 6.6. They are labeled following the scheme "DY est. ncc ({1,2[,3]}), {*category*} {*year*}", e.g., "DY est. ncc (2), incl. DY + Multiboson 2017". As the number of events is not sufficient for these fits in the signal regions the total event counts are used to estimate a rate-changing uncertainty for the estimated Drell-Yan process in these regions. These uncertainties are calculated as explained in Sec. 6.6 from the values of Tab. 6.9. They are labeled in the scheme "DY est. ncc, {*category*} {*year*}", e.g.,

"DY est. ncc, boost. HH(GGF) 2017".

Limited number of generated events in simulation

Simulated datasets are generated with finite amount of events resulting in a bin-wise uncertainty for each physics process. Section 4.4.1 explains how statistical uncertainties for limited number of generated events in simulation are included in this analysis based on the "Barlow-Beeston"-lite [170] approach with a threshold of 10 events per bin.

8.1.2 Theoretical Uncertainties

Theoretical uncertainties are included in this analysis as they affect the cross section, branching fractions, or in general the simulation procedure of a physics process directly. The calculations of a cross section depend for example on the strong coupling constant α_s, which is measured with an uncertainty. Likewise, branching fractions of decays have uncertainties. Furthermore, some physical processes, such as additional QCD jets from parton showering, involve quantities with uncertainties. Finally, some physical phenomena, such as color flow, are different in event generators when applying different sets of parameters, or even when using different event generators.

Cross sections

The calculation of the cross section of a physics process depends on parameters, which are known to certain precision. The uncertainty of these parameters is propagated through this calculation to the cross section. These parameters are α_s, the QCD scale, and the parton distribution function (PDF). In addition, the value of the t-quark mass m_t used for the cross section calculation introduces an uncertainty for processes, which involve the t-quark. For the $t\bar{t}$ + Jets + V processes an additional uncertainty related to the electroweak correction (EWK) is added [66]. An overview of all cross section related uncertainties is given in Tab. 8.3 for the different physics processes.

The uncertainty of the HH(GGF) process that is affected by the precision of the QCD scale and m_t depends on κ_λ. In the case of $\kappa_\lambda = 1$ the uncertainty is given in Tab. 8.3. The uncertainty is calculated for different values of κ_λ following Eq. 2.78. The PDF uncertainties are correlated according to the initial state of the hard parton interaction, i.e., gg-, gq- or qq-induced processes, but separated between Higgs boson and non-Higgs boson processes (naming scheme: "PDF ({*process*})", e.g., "PDF (gluon-induced)"). The α_s uncertainties, labeled "Strong coupling constant (α_s)", are correlated across all non-signal processes, and decorrelated for HH(GGF) and HH(VBF). The QCD scale uncertainties are decorrelated across all physics processes, except for t-quark related processes, i.e., $t\bar{t}$ + Jets (+ V) and t-quark production, which are correlated (naming scheme: "QCD scale ({*process*})", e.g., "QCD scale (VV)"). All uncertainties are correlated among the three years. The m_t uncertainty is labeled "Top mass unc.", and the EWK related uncertainties "EWK corrections ({*ttW,ttZ*})", e.g., "EWK corrections (ttW)".

Higgs boson branching fractions

The branching fractions $\mathcal{B}$ of the different SM Higgs boson decays have uncertainties. Four different decay modes are considered, of which two are of the signal process H $\rightarrow$ b$\bar{\mathrm{b}}$ and H $\rightarrow$ W$^+$W$^-$. An overview of the branching fractions and their relative uncertainties $\sigma_{\mathcal{B},\mathrm{rel.}}$ are given in Tab. 8.4.

Table 8.3: Cross section uncertainties for the different physics processes. The uncertainties are split into different sources: PDF, α_s, QCD scale, m_t, and EWK. A crossmark denotes that the uncertainty is not considered for a given process. The last column gives the reference for the uncertainties.

Process	PDF	α_s	QCD scale	m_t	EWK	Ref.
HH(GGF) ($\kappa_\lambda = 1$)	±3.0%		-23% / +6%		✗	[223]
HH(VBF)	±2.1%		-0.04% / + 0.03%	✗	✗	[223]
H(GGF)	±1.9%	±2.6%	±3.9%	✗	✗	[224]
H(VBF)	±2.1%	±0.5%	-0.3% / +0.4%	✗	✗	[224]
ZH	±1.3%	±0.9%	-3.0% / +3.8%	✗	✗	[224]
WH	±1.7%	±0.9%	-0.7% / +0.5%	✗	✗	[224]
t$\bar{\text{t}}$H	±3.0%	±2.0%	-9.2% / +5.8%	✗	✗	[224]
tHq	±3.5%	±1.2%	-14.7% / +6.5%	✗	✗	[224]
tHW	±6.1%	±1.5%	-6.7% / +4.9%	✗	✗	[224]
$t\bar{t}$ + Jets	±4.2%		-3.5% / +2.4%	-2.7% / +2.8%	✗	[225]
$t\bar{t}$ + Jets + Z	±2.8%	±2.8%	-11.3% / 9.6%	✗	-0.2%	[226]
$t\bar{t}$ + Jets + W	±2.0%	±2.7%	-11.5% / + 12.9%	✗	-3.2%	[226]
t-quark (s-chan.)	±2.6%		-2.3% / +2.8%	✗	✗	[227]
t-quark (t-chan.)	±2.8%		-2.1% / +3.1%	✗	✗	[227]
t-quark (tW-chan.)	✗	✗	±2.5%	✗	✗	[227]
VV	±4.6%	✗	-2.2% / +2.5%	✗	✗	[226]
VVV	±0.1%			✗	✗	[226]

Table 8.4: Branching fractions with their relative uncertainties for different Higgs boson decays in the SM [228].

Decay	$\mathcal{B}$	$\sigma_{\mathcal{B},\text{rel.}}$
H $\rightarrow$ b$\bar{\text{b}}$	58.24%	-1.26% / +1.24%
H $\rightarrow$ W$^+$W$^-$	21.37%	-1.52% / +1.53%
H $\rightarrow$ $\tau\tau$	6.27%	-1.63% / +1.65%
H $\rightarrow$ ZZ	2.62%	-1.52% / +1.53%

The uncertainties $\sigma_{\mathcal{B},\text{rel.}}$ have a rate-changing effect on the signal processes and the SM single Higgs boson processes. Each uncertainty $\sigma_{\mathcal{B},\text{rel.}}$ is fully correlated among all categories and years. They are labeled "BR (*decay*)", e.g., "BR (H $\rightarrow$ b$\bar{\text{b}}$)", in this thesis.

Parton shower

Additional jets originating from the hard interaction in the generation of a physics process are modeled by the parton shower. They are separated into jets which are radiated off the initial state (ISR) and the final state (FSR). The parton shower depends on the value of α_s, which is varied up and down [229] for an uncertainty estimate of the parton shower. This results in a shape-changing effect of the DNN output distributions. These uncertainties are only applied to the signal processes and the $t\bar{t}$ + Jets process, and are decorrelated between

both. In addition, they are considered correlated among all three years. The uncertainties are called "Parton shower ({*ISR,FSR*}) {*HH,TT*}", e.g., "Parton shower (ISR) HH", in this thesis. Exemplary up and down histogram variations are shown for the parton shower FSR uncertainty in Fig. 8.6.
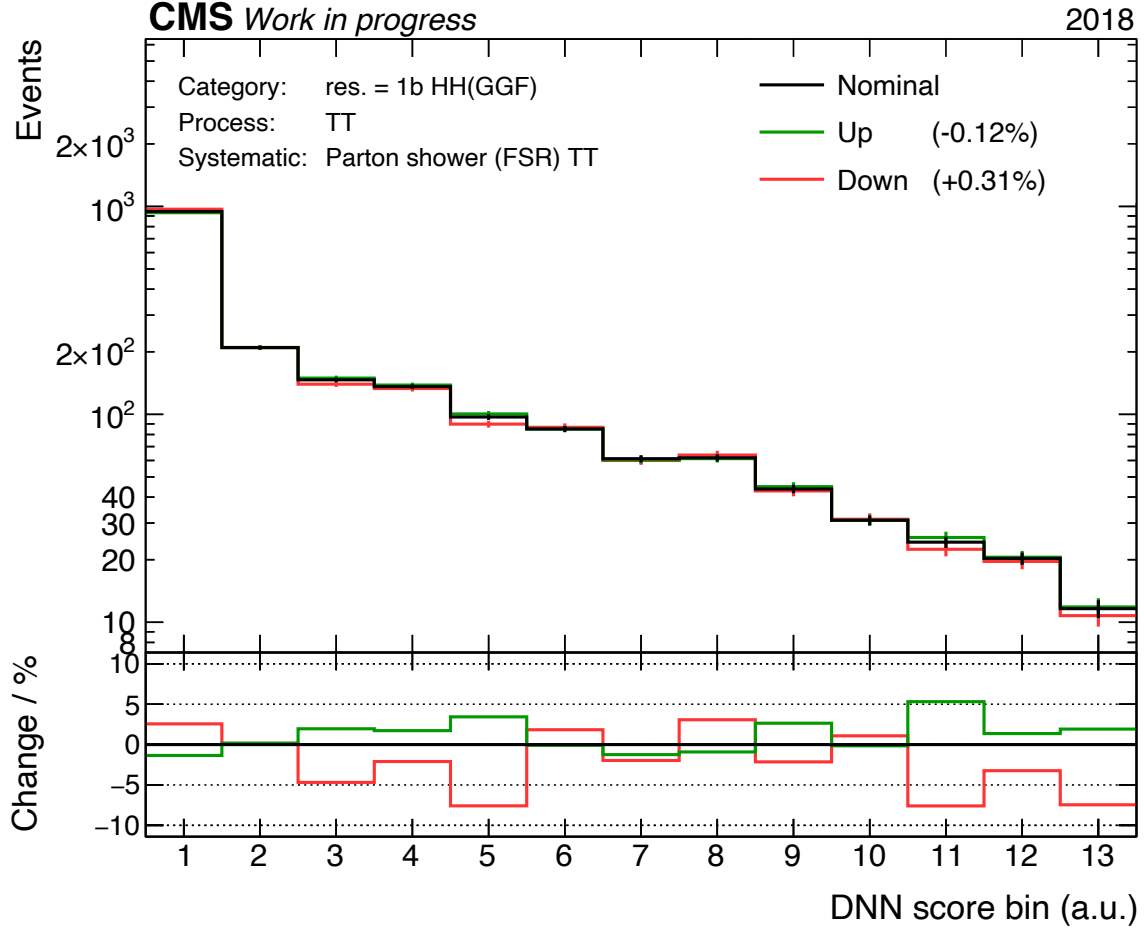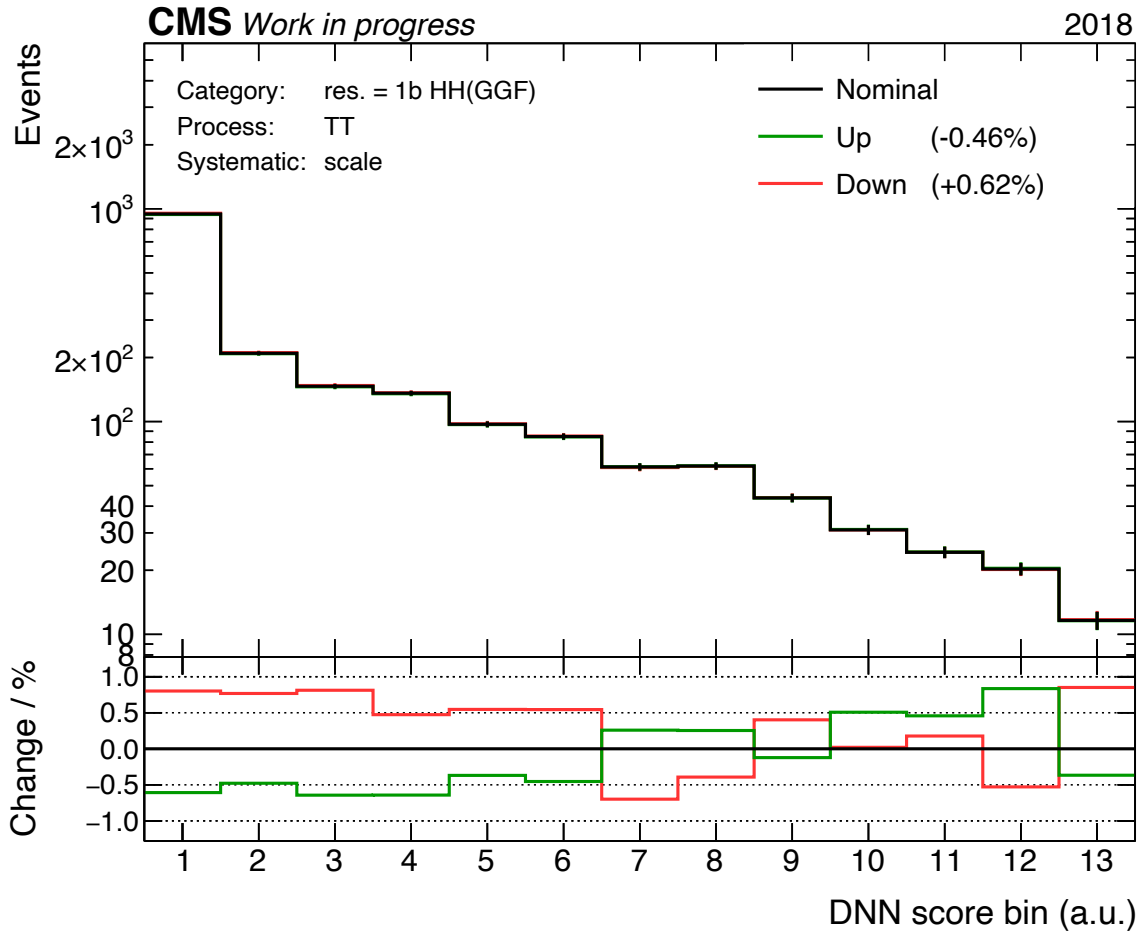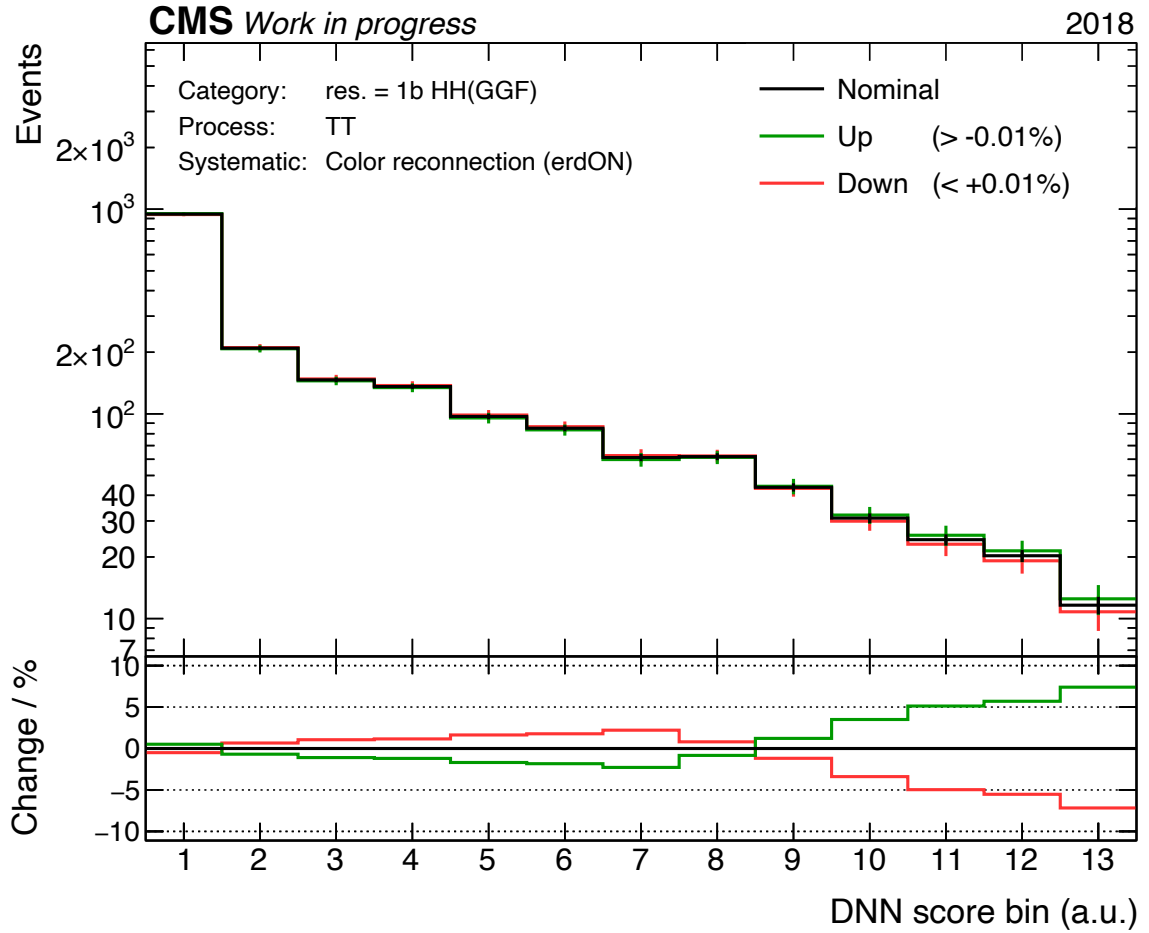

Figure 8.6: Systematic variations of the parton shower FSR uncertainty in the res. = 1b HH(GGF) category in 2018 are shown. The histograms correspond to the $t\bar{t}$ + Jets process. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

The parton shower for the ISR of the HH(VBF) process is of great importance for its precise modelling, as the VBF production is characterized by two jets with large pseudorapidity in the initial state. The generation of these initial state jets involves a global recoil scheme. An improved scheme, called "dipole recoil" [230], is used to generate more precise HH(VBF) events. Due to technical reasons, only some HH(VBF) samples are generated with this scheme. Thus, the difference of event counts between the global and dipole recoil scheme in the analysis phase-space regions, in particular the signal regions, has been investigated for the HH(VBF) process. It has been found, that the difference of event yields is approximately 10%. This difference is taken as a rate-changing uncertainty for the HH(VBF) process. It is correlated among all categories and years, and labeled "HH (VBF) dipole-recoil".

Renormalization and factorization scale

The renormalization μ_R and factorization scale μ_F are part of calculations for proton-proton collision cross sections as described in Eq. 2.75. Seperate variations of these scales are performed, by varying both scales by a factor of two up and down independently and simultaneously. One shape-changing uncertainty is estimated by calculating the envelope of all varied shape templates. The uncertainty is decorrelated between physics processes, but correlated among all three years. It follows the naming scheme of "Renormalization scale ({*process*})", e.g., "Renormalization scale (TT)". The up and down variations for the "Renormalization scale (TT)" are shown in Fig. 8.7.

Figure 8.7: Systematic variations of the "Renormalization scale (TT)" uncertainty in the res. = 1b HH(GGF) category in 2018 are shown. The histograms correspond to the $t\bar{t}$ + Jets process. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

Color reconnection

The color reconnection model [231] describes how multiple partons interact in a collision with each other, which are beyond the hard interaction. These multi-parton interactions may modify the color structure of the physics process until color-neutral hadrons are formed in the final state through a mechanism called "color reconnection". Three different models [231]

exist to describe this effect: "QCD-inspired", "Gluon move", and "ERDon". Dedicated simulation samples are generated for each model for the $t\bar{t}$ + Jets process. An uncertainty for each model is estimated by taking the shape difference of the DNN output distribution between the nominal sample and the sample with the modified color reconnection model. In order to obtain a two-sided uncertainty, the difference is symmetrized. These uncertainties only exist for the $t\bar{t}$ + Jets process and are correlated among all three years. They are named "Color reconnection (QCDbased)", "Color reconnection (GluonMove)", and "Color reconnection (erdON)", respectively. Figure 8.8 shows the up and down variations for the "Color reconnection (erdON)" uncertainty.

Figure 8.8: Systematic variations of the "Color reconnection (erdON)" uncertainty in the res. = 1b HH(GGF) category in 2018 are shown. The histograms correspond to the $t\bar{t}$ + Jets process. Green denotes the up variation, and red the down variation. The ratio of the up and down variation to the nominal histogram is shown in the ratio plot.

Underlying event

The underlying event describes additional processes in a collision, which is not directly related to the hard interaction. These additional processes may be e.g. the interaction of other partons in the pp collision. The underlying event is modeled by the event generators as described in Sec. 6.2. An uncertainty is estimated by varying the corresponding parameters of

the modelling in the event generators. This results in additional samples, which are used to estimate a shape-changing uncertainty by calculating the difference to the nominal sample. The uncertainty is only applied to the $t\bar{t}$ + Jets process and is correlated among all three years. It is labeled "Underlying event".

Matching between parton shower and matrix element

Additional jets in the initial and final state of a physics process can be generated by the parton shower as well as the matrix element in the event generator. A potential overlap of these emission arises which has to be accounted for. The POWHEG V2 event generator introduced a weighting function h_{damp}, which matches the parton shower and the matrix element, to avoid double counting. Dedicated samples are generated where the value of $h_{\text{damp}} = 1.379^{+0.926}_{-0.505}m_{\text{t}}$ [232] is varied up and down within its uncertainties. This shape-changing uncertainty is applied to the $t\bar{t}$ + Jets process and is correlated among all three years. It is named "ME - PS matching scale".

$t\bar{t}$ + Jets and single-top normalizations

The two largest backgrounds $t\bar{t}$ + Jets and single t-quark production need to be known as precisely as possible to obtain a good sensitivity. The uncertainties, which affect both processes and are listed above, may not be sufficient to control these backgrounds in the specific phase-space of this analysis. Thus, unconstrained scaling parameters are introduced for both processes. The scaling parameters are fully correlated among all categories and years and named "TT norm." and "ST norm.", respectively. Two control regions (Sec. 7.5) are created by the DNN to measure these to the highest precision. Additionally two rate-changing uncertainties are added for the boosted regions with an effect of 20% each. These are labeled "TT norm. boost. (DL, 20%)" and "ST norm. boost. (DL, 20%)", respectively.

8.1.3 Overview of Uncertainties

All uncertainties considered in the presented analysis are summarized in Tab. 8.5. They are split into experimental and theoretical uncertainties. For each uncertainty source the number of nuisance parameters N_p is given, their type is given, and a comment is added to briefly explain correlations. In total 239 nuisance parameters are considered, 195 from experimental sources, and 44 from theoretical sources. Uncertainties regarding the limited number of generated events in simulation are not included in the table.

8.2 Quality Assessment of the Likelihood Fit

A fit of the signal plus background (*S+B model*) expectation, including all aforementioned uncertainties, to recorded data is performed. The DNN output distributions of the *S+B model* may be varied by the fit within uncertainties in order to maximize the likelihood function of Eq. 8.1.

The quality of the fit has to be assessed and evaluated. For this, the DNN output distributions after the fit (*post-fit*) are shown. Furthermore, the event yields before and after the fit are compared. In addition, the quantitative measure of a goodness-of-fit test is used (see Sec. 4.4.2), to ensure good (post-fit) agreement between the *S+B model* and recorded data.

Figure 8.9 shows the post-fit DNN output distributions for the six signal and three background categories (see Sec. 7.5). No significant deviation between the *S+B model* and data is

Table 8.5: Summary of all systematic experimental and theoretical uncertainties. In addition, the number of nuisance parameters N_p is given, the type for each nuisance is given, and a comment is added to briefly explain correlations.

Source	N_p	Type	Comment
Experimental	**195**		
Luminosity	5	rate	1 per year, 1 for Run 2, 1 for 2017+2018
Pileup	1	shape	correlated among all years
Trigger	9	shape	1 per year and $\ell\ell$-channel
Lepton eff.	18	shape	1 per lepton type, year, working point
Jet energy scales	21	shape	correlation scheme in Tab. 8.2
Jet energy resolution	3	shape	1 per year
Unclustered energy	3	shape	1 per year
Jet pileup ID eff.	6	shape	1 per year and type (ID, mistag)
Small rad. b-quark ID eff.	16	shape	decorr. by flavor, year, stat. and syst.
Large rad. b-quark ID eff.	3	shape	1 per year
t-quark p_T correction	1	shape	correlated among all years
HEM issue	1	shape	only in 2018
L1 ECAL pre-firing	2	shape	only in 2016 and 2017
Misid. leptons est.	60	shape	1 per lepton type, year, η & p_T region
Drell-Yan est.	45		
Method	9	shape	1 per region (Sec. 6.7) and year
Closure (sig.)	18	rate	1 per signal category and year
Closure (bkg.)	18	shape	1 per background category and year
Theoretical	**44**		
Cross section	20	rate	see Tab. 8.3
Branching fraction	4	rate	1 per decay mode
Parton shower	4	shape	1 per ISR & FSR, HH & $t\bar{t}$ + Jets
Dipole recoil	1	rate	only for HH(VBF)
Renorm. & fact. scale	6	shape	1 per process
Color reconnection	3	shape	1 per model, only for $t\bar{t}$ + Jets
Underlying event	1	shape	only for $t\bar{t}$ + Jets
ME - PS matching scale	1	shape	only for $t\bar{t}$ + Jets
Background norm.	4		
all categories	2	-	1 per single-top & $t\bar{t}$ + Jets
boosted categories	2	rate	1 per single-top & $t\bar{t}$ + Jets
Total	**239**		

observed, i.e., they agree within $\leq 2\sigma$ uncertainty. Most bins are compatible directly within 1σ uncertainties. The ratio between the expectation and recorded data in the signal regions is in the order of $1 - 10\%$, whereas the agreement in the control regions is a magnitude smaller $(0.1 - 1\%)$ due to the high event yield in each bin.

In addition to the post-fit DNN output distributions, the event yields are shown in Tab. 8.6

(a) HH(GGF).

(b) HH(VBF).

(c) Top + Other.

(d) Drell − Yan + Multiboson.
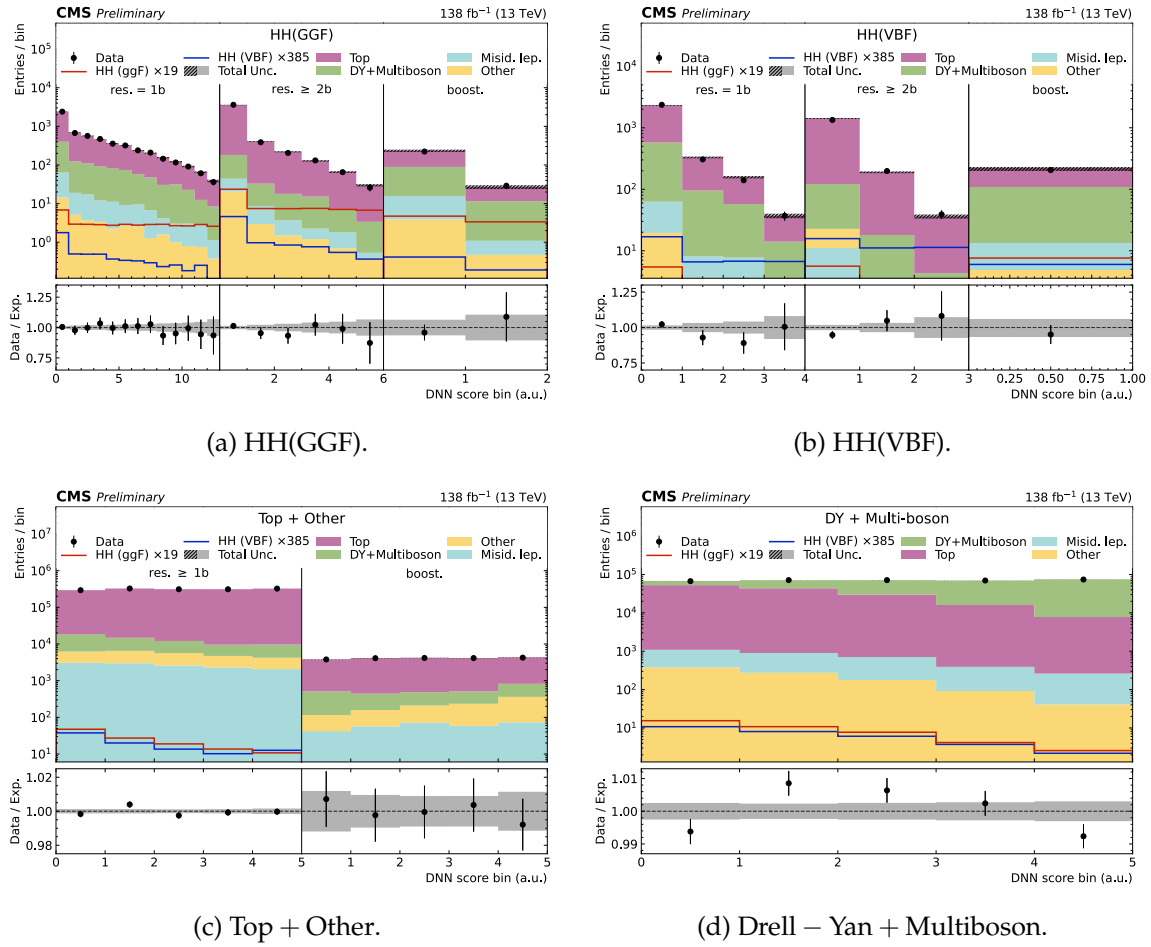
Figure 8.9: Neural network output scores for the six signal and three background categories (see Sec. 7.5). The distributions include the full uncertainty model and are shown *after* a fit was performed.

before and *after* the fit has been performed. The tables compare the post-fit and pre-fit event yields in the gray and white rows, respectively, for each physics process and category. The event yield for data is listed separately at the bottom of each table. The post-fit event yields of the *S+B model* agree within uncertainties with data. Additionally, one can see that the uncertainties are significantly smaller for the post-fit yields compared to the pre-fit yields. This is expected, as the fit may constrain the nuisance parameters (see Sec. 4.4.2).

The results of the saturated goodness of fit test are shown in Fig. 8.10. Reasonable agreement between the *S+B model* and recorded data is observed when the p-value exceeds a value of 5%. Three hundred toy experiments have been performed for each year and their combination to ensure a reasonable χ^2 distribution, which is used to calculate the p-value. It can be seen that in each individual year and for their combination (Run 2) the p-value is approximately between 10% and 50%.
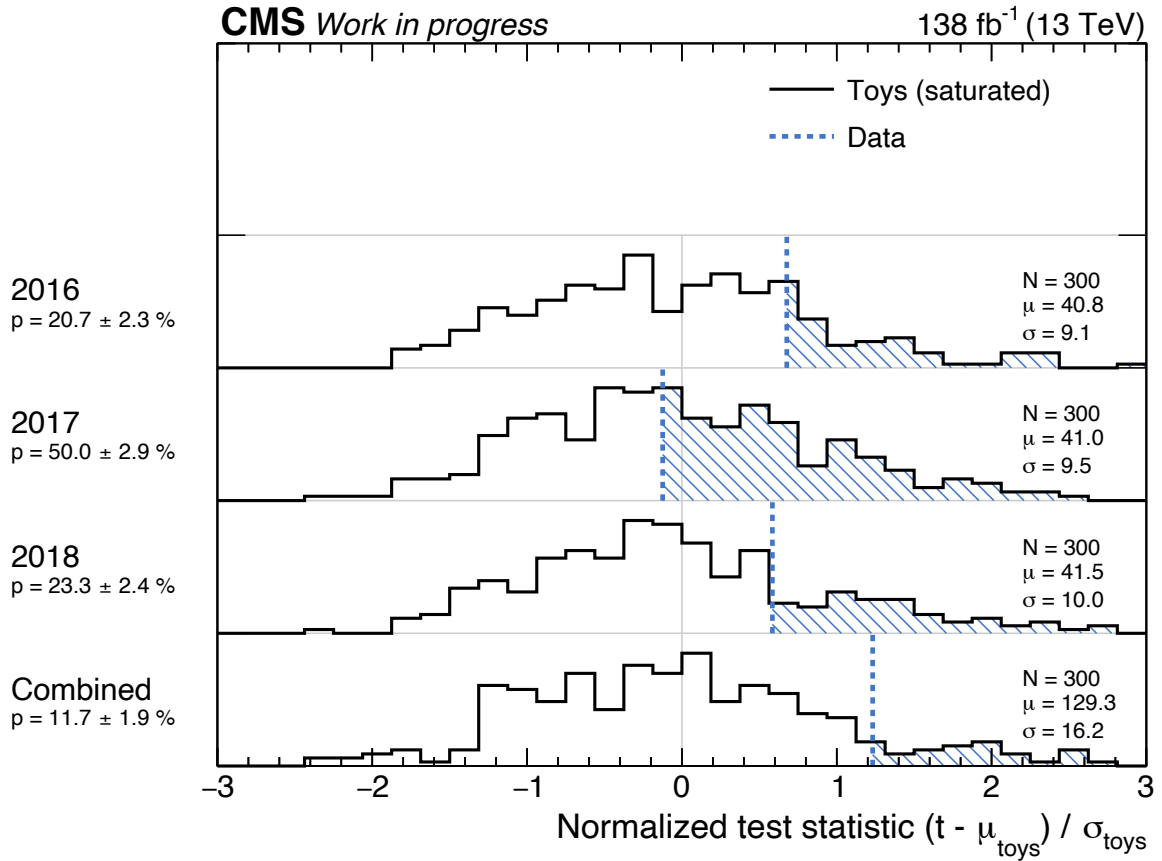
Figure 8.10: Saturated goodness of fit test for the individual years and their combination (Run 2). The p-value is quoted in percentage below the year label. For each year and their combination 300 toy experiments have been performed.

8.3 Expected and Observed Exclusion Limits

Upper exclusion limits at 95% confidence level are set on the global signal strength modifier μ affecting the HH(GGF) and HH(VBF) processes simultaneously, and on the signal strength modifier $\mu_{\text{HH(VBF)}}$ affecting only the HH(VBF) process. These limits are calculated as described in Sec. 4.4.2. The expected exclusion limits are shown with their 1σ and 2σ uncertain-

ties visualized by green and yellow color bands, respectively. Expected limits are calculated by replacing observed data with an Asimov dataset corresponding to the background-only hypothesis ($\mu = 0$).

The expected and observed exclusion limits for the SM scenario, where all coupling modifiers ($\kappa_\lambda, \kappa_{2V}, \kappa_t, \kappa_V$) are set to one, are shown in Fig. 8.11 sorted by the value of the expected exclusion limit.

(a) μ.
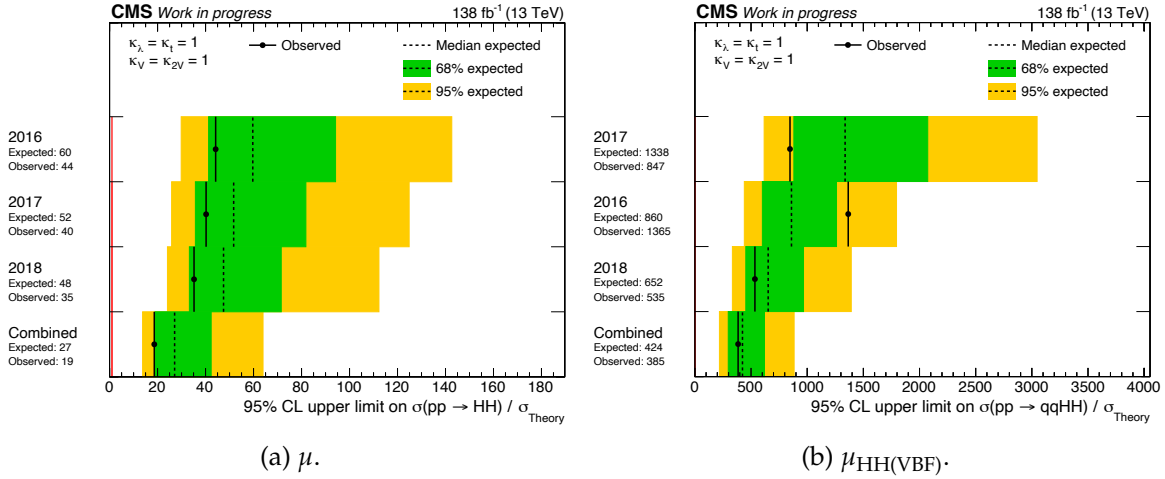
(b) $\mu_{\text{HH(VBF)}}$.

Figure 8.11: Upper exclusion limits at 95% confidence level for the SM scenario ($\kappa_X = 1$). Limits are set on the global signal strength modifier μ, and on the HH(VBF) modifier $\mu_{\text{HH(VBF)}}$. The one and two sigma uncertainties of the expected exclusion limit are shown by the green and yellow color bands.

The observed (expected) upper exclusion limit for μ considering the "Combined", i.e., Run 2, analysis is measured to be 19 (27). The observed (expected) upper exclusion limit for $\mu_{\text{HH(VBF)}}$ is measured to 385 (424). No significant deviation between the observed and expected exclusion limits on μ and $\mu_{\text{HH(VBF)}}$ is seen for any data-taking period.

In order to constrain the allowed values of the coupling modifiers κ_λ and κ_{2V} (see Fig. 2.11), exclusion limits are calculated as a function of those. The results are presented in Fig. 8.12.

Fig. 8.12a shows the sensitivity of the analysis on μ as a function of κ_λ. The strongest limit is set close to the SM scenario ($\kappa_\lambda = 1$) and the interference point ($\kappa_\lambda \approx 2.4$) at roughly 500 fb. Although the cross section is diminished in this region of κ_λ, the sensitivity is enhanced due to its unique differential m_{HH} distribution (see Fig. 2.14), which tends towards higher values of m_{HH}. This fact makes the signal more distinguishable from backgrounds. Vice versa, larger absolute values of κ_λ increase the total cross section of the HH signal, but the differential distribution of m_{HH} tend towards lower values. These two properties keep their balance, which leads to rather stable upper exclusion limits at around 1000 fb for larger absolute values of κ_λ. The value of κ_λ can be constrained with the observed (expected) limits towards $-8.64 \leq \kappa_\lambda \leq 15.21$ ($-11.43 \leq \kappa_\lambda \leq 17.83$).

Likewise, the upper exclusion limits on $\mu_{\text{HH(VBF)}}$ as a function of κ_{2V} are shown in Fig. 8.12b. The strongest limits are set for large absolute values of κ_{2V} to about 200 fb, while the weakest are at the SM scenario ($\kappa_{2V} = 1$) and close to the SM scenario at around 500 fb. This is due to the fact that the total cross section of HH(VBF) production can be factorized as follows: $\sigma_{\text{HH(VBF)}} \propto \kappa_{2V} - \kappa_V^2$ [233]. Here, κ_V is set to one (SM expectation), which leads to a small

(a) κ_λ.

(b) κ_{2V}.
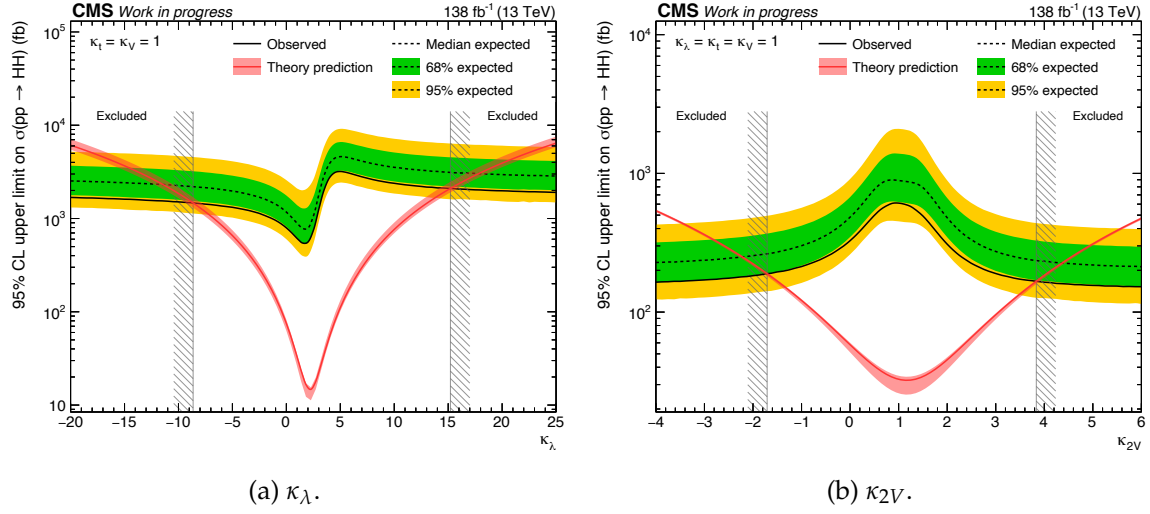
Figure 8.12: Upper exclusion limits at 95% confidence level as a function of κ_λ (left) and κ_{2V} (right). Limits are set on the global signal strength modifier μ (left), and on the HH(VBF) modifier $\mu_{\mathrm{HH(VBF)}}$ (right). The one and two sigma uncertainties of the expected exclusion limit are shown by the green and yellow color bands.

theory cross section for $\kappa_{2V} = \kappa_V = 1$. This fact decreases the analysis sensitivity significantly at the SM scenario. An observed (expected) constrain on κ_{2V} can be set towards $-1.71 \leq \kappa_{2V} \leq 3.84$ ($-2.22 \leq \kappa_{2V} \leq 4.37$).

8.4 Likelihood Scans

The likelihood is profiled, as explained in Sec.4.4.2, as a function of κ_λ and κ_{2V} in order to measure their best fit values with uncertainties. The results are shown in Fig. 8.13.

(a) κ_λ.

(b) κ_{2V}.
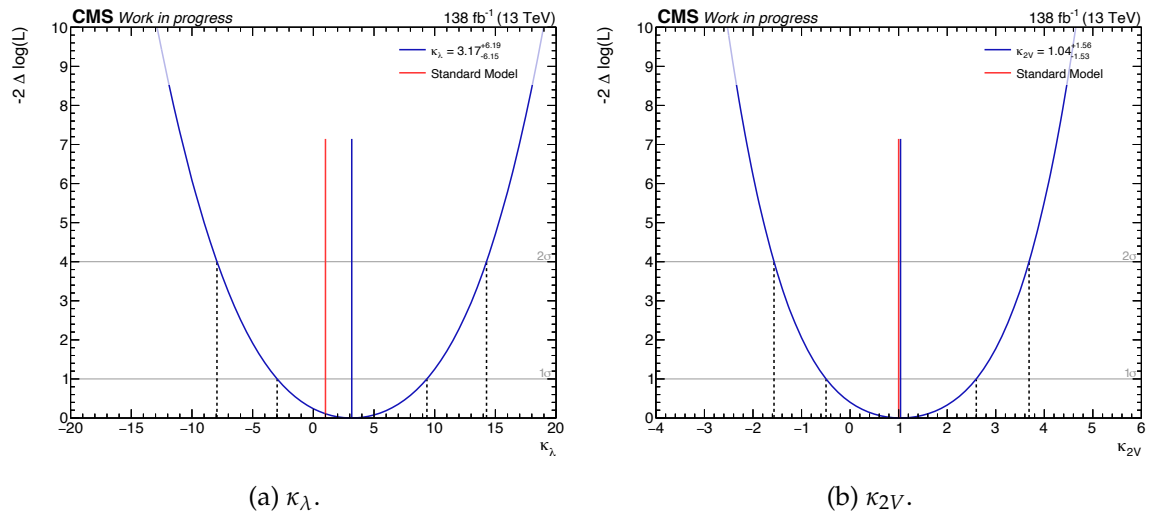
Figure 8.13: Likelihood profile of κ_λ (left) and κ_{2V} (right). The SM expectation is shown by the red line at one.

The best fit value is at the minimum of the likelihood profile. Its 1 (2) σ uncertainty can be

extracted, where the profile crosses $-2\Delta \log \mathcal{L} = 1\,(4)$. The measured best fit values with their 1σ uncertainties are as follows:

$$\hat{\kappa}_\lambda = 3.17^{+6.19}_{-6.15}, \tag{8.4}$$

$$\tag{8.5}$$

$$\hat{\kappa}_{VV} = 1.04^{+1.56}_{-1.53}. \tag{8.6}$$

Both measured values are well compatible with the SM expectation of one within their 1σ uncertainties.

8.5 Impact of Systematic Uncertainties

The impact of all systematic uncertainties on μ is evaluated per nuisance parameter by re-performing the likelihood fit for each uncertainty fixed to their post-fit value with uncertainty. In addition, pulls and constrains of all systematic uncertainties are calculated. The results are shown in Fig. 8.14 and Fig. 8.15 for all systematic uncertainties, except for uncertainties related to the limited number of generated events in simulation.

The systematic uncertainties are sorted by their impact, visualized by the $+1\sigma$ (red) and -1σ (blue) bands, on the global signal strength modifier μ. The black point (or value) denotes the pull, and the black line the constrain of each uncertainty, as defined in Eq. 4.33.

The leading uncertainties are mostly related to the largest backgrounds: $t\bar{t}$ + Jets and single t-quark production. The leading uncertainty is the "ST norm.". It controls the normalization of the single t-quark production by an unconstrained parameter. Dedicated control regions (res. Top + Other, and boost. Top + Other) have been produced with a DNN to control this normalization parameter. Uncertainties affecting the $t\bar{t}$ + Jets process typically rank high, such as "Renormalization scale (TT)", "Color reconnection (erdON)", and "Parton shower (FSR) TT". The theory uncertainty "QCD scale (ggHH + mtop)", which directly affects the signal process, ranks on the second place. Furthermore, uncertainties related to jets and p_T^{miss}, such as "JES (FlavQCD)" and "Unclustered energy (MET) 2017", are important, due to the high jet multiplicity and the two neutrinos in the final state. The last set of uncertainties among the leading 25 concern the closure of the data-driven Drell-Yan estimation in the signal regions.

The pulls of the uncertainties are within 2σ; most of them even within 1σ. The constraint of some nuisances can be observed, especially when these nuisances affect physics processes for which dedicated control regions exist in this analysis (e.g. "TT norm. boost. (DL, 20%)").

In order to estimate the total effect of all systematic uncertainties on the signal strength modifier μ, a dedicated fit has been performed without any systematic uncertainty, except for those related to the limit number of generated events in simulation. In addition, a second fit has been performed omitting also these uncertainties. The results are shown in Fig. 8.16.

Omitting the systematic uncertainties, except for those modelling the finite amount of generated events in simulation, results in an expected limit of $\mu = 20$ for the Run 2 data-taking period as shown in Fig. 8.16a. This shows that the systematic uncertainties do play a significant role, however this analysis is still significantly limited by the amount of recorded

data. In addition, uncertainties modelling the finite amount of generated events in simulation are omitted in Fig. 8.16b. The limits are improved only slightly compared to Fig. 8.16a, which shows that the modelling of physics processes with simulation is well described, and no more simulated events are needed to improve the sensitivity of this analysis.

Table 8.6: Event yields *after* (gray colored rows) and *before* (white colored rows) the fit has been performed. The yields are shown for the six signal and three background categories for each process individually. S and B denote the combination of all signal and all background processes, respectively. Finally, the event yield of recorded data is shown in the last row (blue colored).

Process	Category	HH(GGF) res. = 1b	res. ≥ 2b	boost.	HH(VBF) res. = 1b	res. ≥ 2b	boost.	Background categories res. Top + Other	boost. Top + Other	Drell-Yan + Multiboson
HH(GGF)		−14 ± 3	−21 ± 6	−3 ± 1	−3 ± 1	−3 ± 1	−3 ± 1	−42 ± 13	−3 ± 1	−14 ± 4
		2 ± 0	3 ± 0	≤ 1	≤ 1	≤ 1	≤ 1	6 ± 0	≤ 1	2 ± 0
HH(VBF)		≤ 1	≤ 1	≤ 1	−1 ± 0	−1 ± 0	≤ 1	−2 ± 1	≤ 1	−1 ± 0
		≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
Top		4598 ± 18	4160 ± 33	159 ± 5	2094 ± 21	1497 ± 19	107 ± 3	1502176 ± 426	17692 ± 34	146795 ± 124
		4841 ± 132	4341 ± 296	163 ± 22	2226 ± 174	1578 ± 158	108 ± 17	1536897 ± 36103	17953 ± 1097	144493 ± 3739
Drell-Yan + Multiboson		933 ± 15	197 ± 36	84 ± 14	662 ± 27	118 ± 18	95 ± 13	37407 ± 32	1683 ± 26	202172 ± 139
		924 ± 49	162 ± 67	151 ± 70	677 ± 91	114 ± 43	181 ± 77	33775 ± 1988	1414 ± 173	193097 ± 7732
Misid. leptons		122 ± 4	34 ± 2	12 ± 1	53 ± 4	12 ± 2	9 ± 1	13034 ± 20	303 ± 4	2380 ± 18
		117 ± 7	30 ± 7	12 ± 1	49 ± 7	13 ± 5	8 ± 1	12079 ± 517	300 ± 25	2713 ± 314
Other		40 ± 1	27 ± 0	4 ± 0	27 ± 0	14 ± 1	5 ± 0	14303 ± 8	778 ± 3	965 ± 1
		42 ± 3	28 ± 3	4 ± 0	27 ± 1	14 ± 2	5 ± 0	14683 ± 196	786 ± 17	964 ± 16
S		−14 ± 3	−21 ± 6	−3 ± 1	−4 ± 1	−4 ± 1	−3 ± 1	−43 ± 13	−3 ± 1	−15 ± 4
		2 ± 0	3 ± 0	≤ 1	1 ± 0	1 ± 0	≤ 1	7 ± 0	≤ 1	2 ± 0
B		5693 ± 24	4419 ± 49	259 ± 15	2837 ± 34	1640 ± 26	216 ± 13	1566921 ± 428	20455 ± 43	352313 ± 187
		5923 ± 141	4561 ± 304	330 ± 73	2980 ± 196	1719 ± 164	303 ± 79	1597435 ± 36162	20453 ± 1111	341268 ± 8594
Data		5690 ± 75	4432 ± 67	252 ± 16	2849 ± 53	1578 ± 40	205 ± 14	1566651 ± 1252	20452 ± 143	352548 ± 594
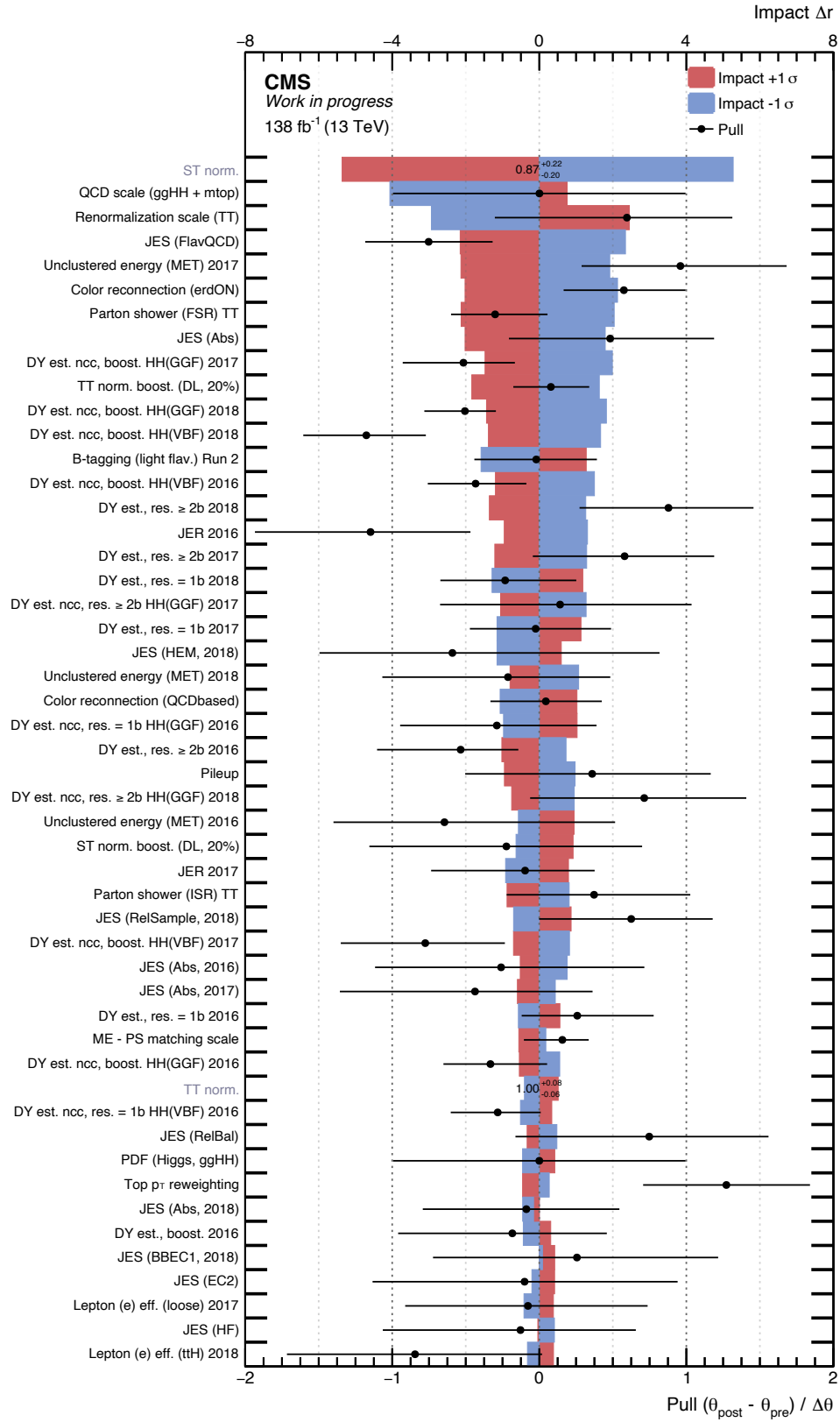
Figure 8.14: Impacts, pulls and constraints are shown for the leading 50 uncertainties sorted by their impact on μ for the Run 2 data-taking period.
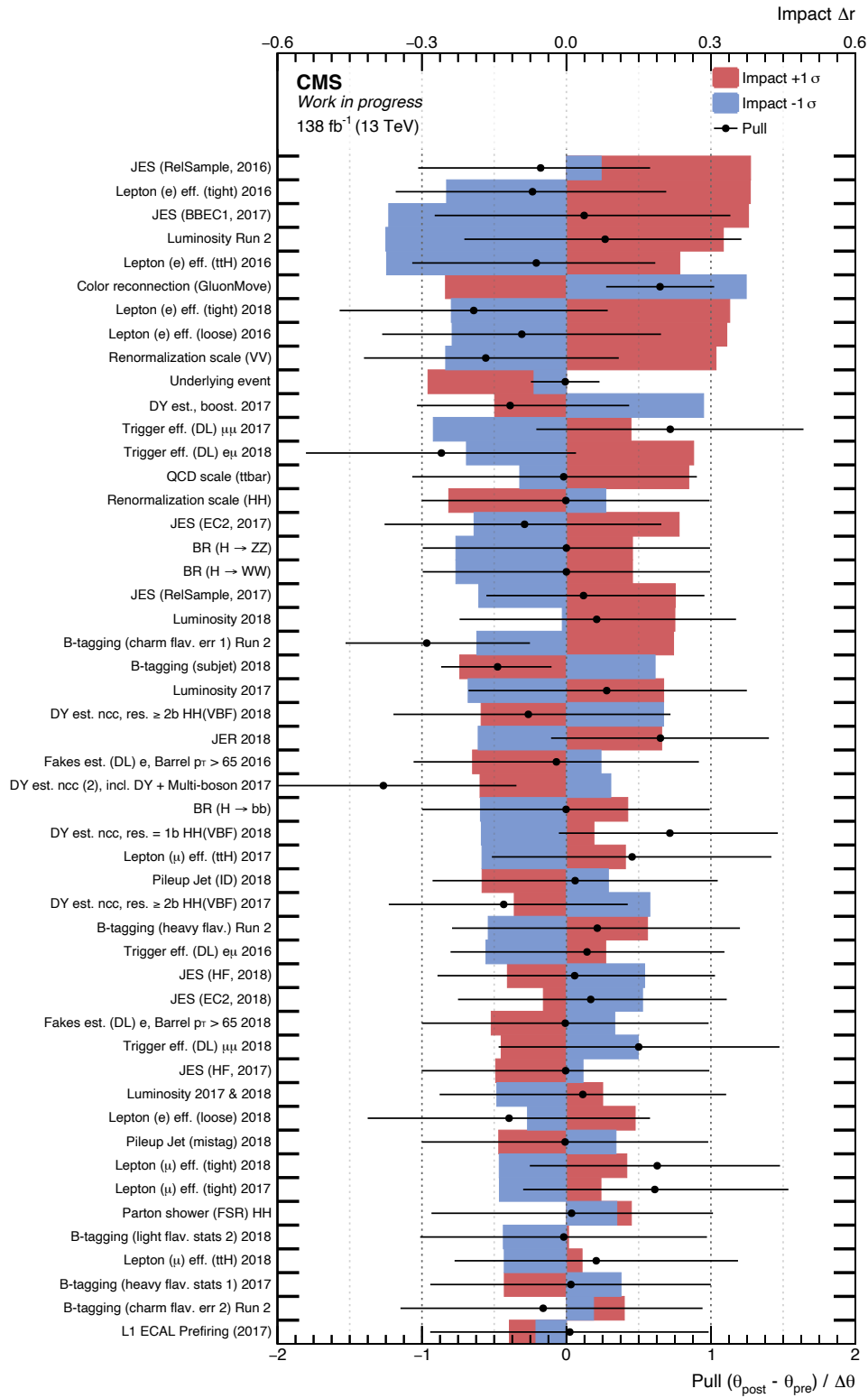
Figure 8.15: Impacts, pulls and constraints are shown for uncertainties (ranked 50 to 100) sorted by their impact on μ for the Run 2 data-taking period.
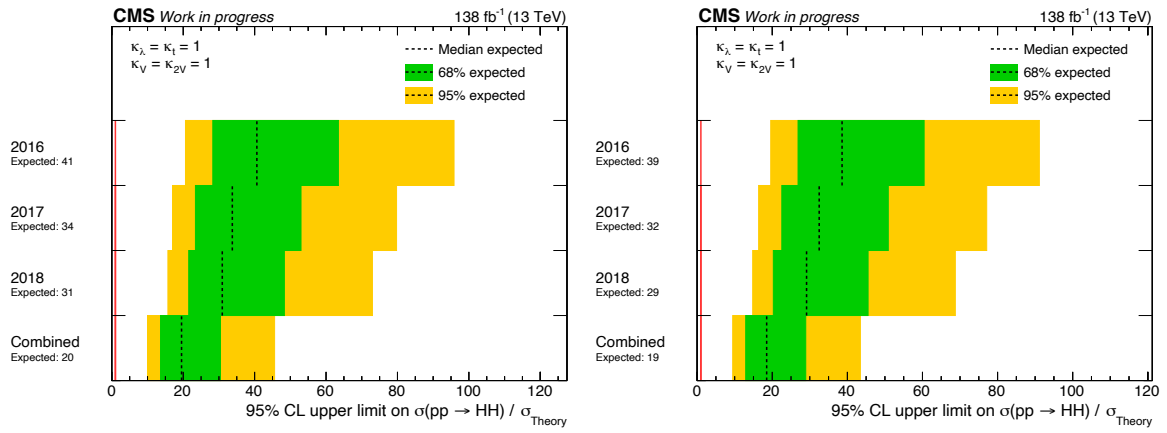
(a) Only the uncertainty for limited number of events in simulation is included.

(b) No systematic uncertainties are considered.

Figure 8.16: Upper exclusion limits at 95% confidence level for the SM scenario ($\kappa_X = 1$) for two sets of uncertainties: including only uncertainties related to the limited number of events in simulation (left), without any systematic uncertainties (right). Limits are set on the global signal strength modifier μ. The one and two sigma uncertainties of the expected exclusion limit are shown by the green and yellow color bands.

Chapter 9

Combination with the Semi-leptonic $HH \rightarrow b\bar{b}W^+W^-$ Analysis

Simultaneously to the presented analysis in this thesis, the semi-leptonic $HH \rightarrow b\bar{b}W^+W^-$ analysis has been performed by the CMS collaboration. Both analysis have been designed to be combined for a more sensitive measurement. The combination strategy and their results are presented in this chapter. The combination has been published in Ref. [10].

The semi-leptonic $HH \rightarrow b\bar{b}W^+W^-$ analysis follows a similar strategy as presented in chapter 5. Here, also a physics process classification is performed with the help of a DNN. Different signal regions and control regions are defined with this DNN and further event information, such as the number of b-quark jets and their topology. The major backgrounds of the semi-leptonic decay of the $HH \rightarrow b\bar{b}W^+W^-$ process are again the $t\bar{t}$ + Jets and single t-quark production. In addition, the W + Jets process plays a significant role, and marks one of the largest background contributions depending on the analysis phase-space. The W + Jets process is highly suppressed in the presented analysis in this thesis due to the presence of two leptons. Also the Misid. leptons background process has a significantly higher contribution as the higher jet multiplicity in the final state allow more possibilities for mis-reconstructed leptons from jets. The estimation of the Misid. leptons background process is the same as described in this thesis (see Sec. 6.5).

For the combined measurement of both analyses, the uncertainties have to be correlated properly. Systematic uncertainties, which affect both analyses are fully correlated. The only exception are the unconstrained parameters which affect the normalization of the major background processes. Uncertainties, which are unique to each analysis, such as trigger efficiencies, are fully decorrelated.

Finally, a fit including all regions from both analyses is performed. The results on the upper exclusion limits are shown in Fig. 9.1 for the SM scenario, and in Fig. 9.2 as a function of κ_λ and κ_{2V}. The limits are set on the same signal strength modifier μ and $\mu_{\text{HH(VBF)}}$ as described in the previous chapter.

The semi-leptonic $HH \rightarrow b\bar{b}W^+W^-$ analysis provides a similar sensitivity as the analysis presented in this thesis. Their combination results in an observed (expected) upper exclusion limit on μ of 14 (18), and on $\mu_{\text{HH(VBF)}}$ of 277 (301) at the SM scenario. The value of the observed limit on μ is roughly improved by a factor of 1.36 and on $\mu_{\text{HH(VBF)}}$ by approximately a factor of 1.39 compared to the analysis presented in this thesis.

More stringent constraints on the coupling modifier κ_λ and κ_{2V} are set when combining both analysis. The (observed) allowed range for κ_λ is $-7.22 \leq \kappa_\lambda \leq 13.81$, and for κ_{2V} the allowed
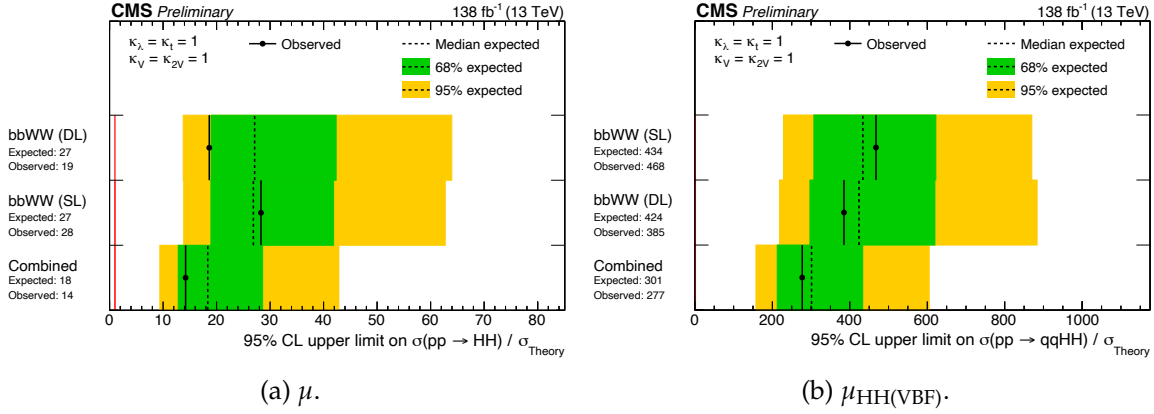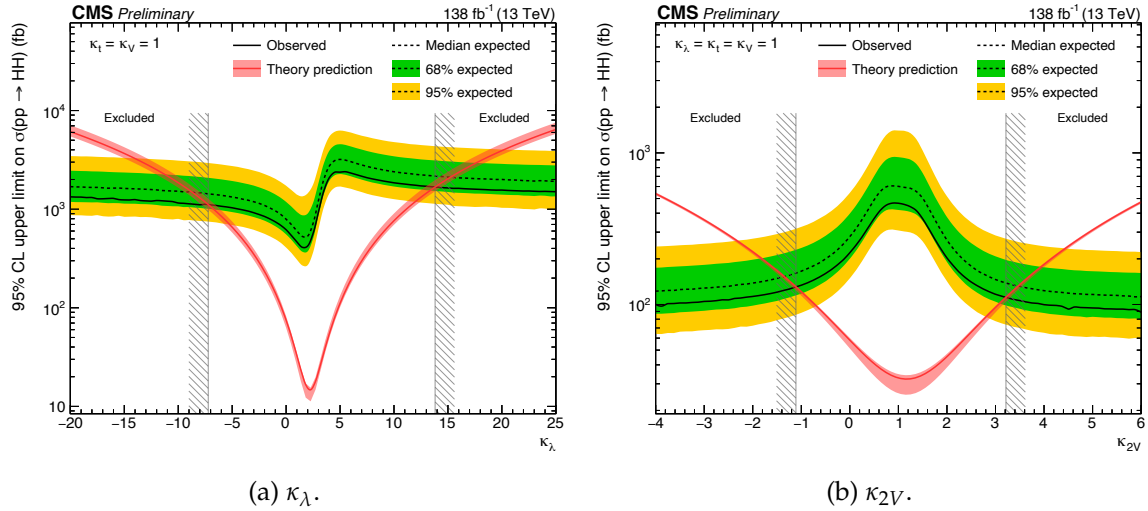
(a) μ.

(b) $\mu_{\text{HH(VBF)}}$.

Figure 9.1: Upper exclusion limits at 95% confidence level for the SM scenario ($\kappa_X = 1$) for the combination of the dileptonic and semi-leptonic HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ analysis [10]. Limits are set on the global signal strength modifier μ, and on the HH(VBF) modifier $\mu_{\text{HH(VBF)}}$. The one and two sigma uncertainties of the expected exclusion limit are shown by the green and yellow color bands.

(a) κ_λ.

(b) κ_{2V}.

Figure 9.2: Upper exclusion limits at 95% confidence level as a function of κ_λ (left) and κ_{2V} (right) for the combination of the dileptonic and semi-leptonic HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ analysis [10]. Limits are set on the global signal strength modifier μ (left), and on the HH(VBF) modifier $\mu_{\text{HH(VBF)}}$ (right). The one and two sigma uncertainties of the expected exclusion limit are shown by the green and yellow color bands.

range is $-1.11 \leq \kappa_{2V} \leq 3.22$. Especially the constrain on κ_{2V} benefits from the inclusion of the semi-leptonic HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ analysis.

In addition, two dimensional scans are performed in Ref. [10] as some of the coupling parameters are correlated. These two dimensional scans are shown in Fig. 9.3 for $\kappa_\lambda - \kappa_t$, $\kappa_\lambda - \kappa_{2V}$, and $\kappa_{2V} - \kappa_V$.

Upper exclusion limits are calculated in these two dimensional scans to measure the allowed values (white area) of the coupling parameters. In addition, the best fit value (black cross) is calculated by profiling the corresponding two dimensional likelihood function. In all scans,
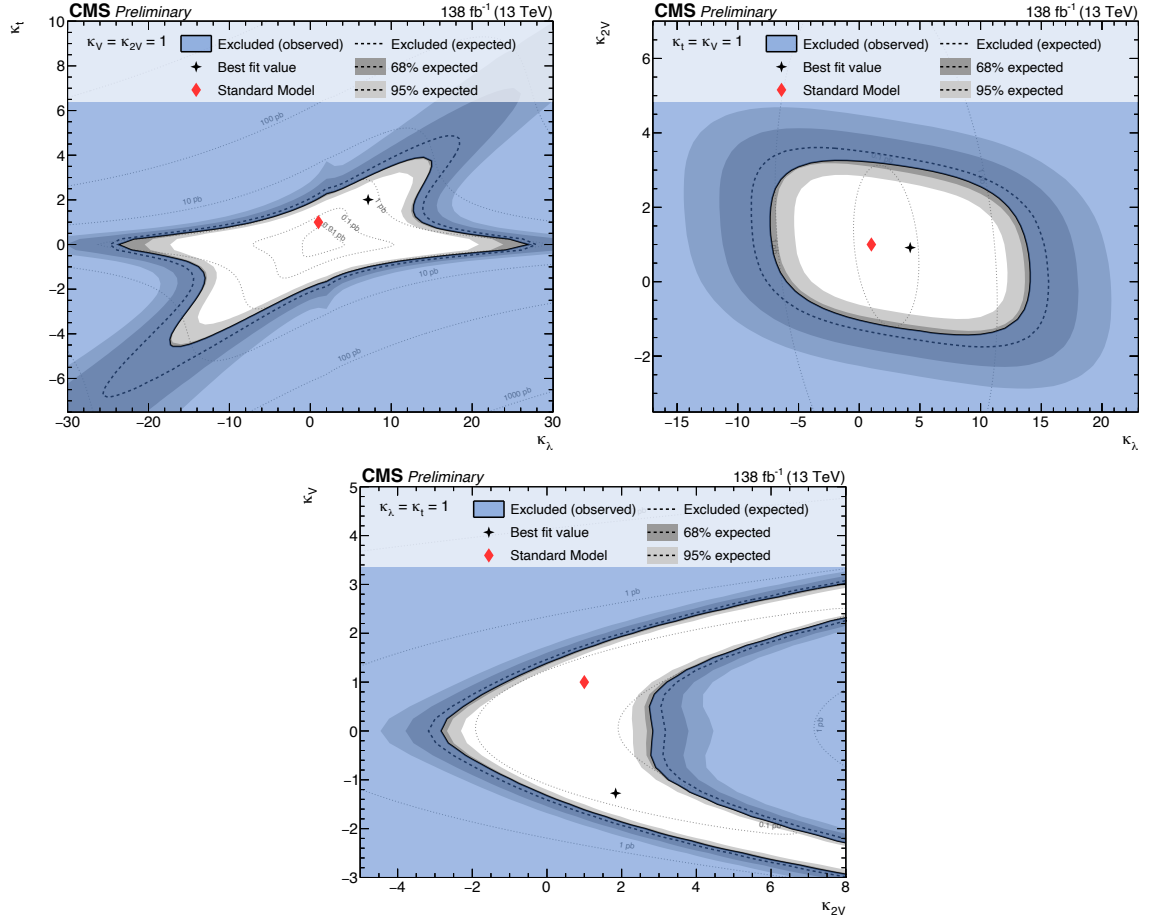
Figure 9.3: Two dimensional scans of different sets of coupling parameters for the combination of the dileptonic and semi-leptonic HH → b̄bW⁺W⁻ analysis [10]. Upper exclusion limits are set on μ to measure the allowed (white) and excluded (blue) regions. The best fit value, calculated with a two dimensional likelihood profile, is shown by the black cross. The SM expectation is shown by the red diamond.

the best fit value is within the allowed region.

Chapter 10

Conclusion

This thesis presents a search for Higgs boson pair production in the dileptonic decay mode of the HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ process. Data from proton-proton collisions produced at the LHC from 2016 to 2018 at a center-of-mass energy of $\sqrt{s} = 13\,$TeV has been analyzed. This data has been recorded by the CMS experiment and corresponds to an integrated luminosity of $138\,$fb^{-1}.

The presented analysis sets upper exclusion limits on the signal strength modifier of the inclusive and HH(VBF) production mode of Higgs boson pair production, and measures and constrains the Higgs boson self-coupling and the direct coupling between two Higgs bosons and two vector bosons. The analysis has been combined with the analysis of the semi-leptonic decay mode of the HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ process and published in Ref. [10].

The presented search for the dileptonic HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ process is particularly challenging due to the overwhelming amount of expected background events compared to the small number of expected signal events. Careful control over these backgrounds is crucial for reliable and precise measurements. A dedicated analysis strategy has been developed to tackle these challenges.

At first, a loose event selection is applied to maximize the acceptance for signal events. Events are selected if they contain a pair of opposite charged leptons, i.e., e$^+$e$^-$, $\mu^+\mu^-$, or e$^\pm\mu^\pm$. Additionally, more than two small radius b-quark jets or one large radius b-quark jet are required. This selection criteria characterizes resolved and boosted event topologies. Two backgrounds, namely the Drell-Yan process and the misidentified leptons process, are estimated using data-driven methods to improve their modelling. Then, a deep learning model is trained and used to create analysis regions mutually enriched by particular physics processes. Signal enriched categories provide the main sensitivity to the measurements while background enriched categories primarily constrain uncertainties related to the background modelling. Finally, all categories are fitted in a maximum likelihood simultaneously to measure parameters of interest.

No evidence for Higgs boson pair production has been found in the dileptonic HH $\rightarrow$ b$\bar{\text{b}}$W$^+$W$^-$ decay mode. Upper exclusion limits on the signal strength modifiers for the inclusive and HH(VBF) production modes are observed (expected) to be 19 (27) and 385 (424) at 95% CL$_s$, respectively. The Higgs boson self-coupling κ_λ has been measured to be $\kappa_\lambda = 3.17^{+6.19}_{-6.15}$, and could be constrained in the range of $-8.64 \leq \kappa_\lambda \leq 15.21$ ($-11.43 \leq \kappa_\lambda \leq 17.83$) based on observed (expected) upper exclusion limits at 95% CL$_s$. Likewise, the coupling between two Higgs bosons and two vector bosons κ_{2V} has been measured to be $\kappa_{2V} = 1.04^{+1.56}_{-1.53}$, and is constrained in the range of $-1.71 \leq \kappa_{2V} \leq 3.84$ ($-2.22 \leq \kappa_{2V} \leq 4.37$). A combina-

tion with the semi-leptonic $HH \rightarrow b\bar{b}W^+W^-$ analysis improved the sensitivity resulting in an upper observed (expected) exclusion limit of the inclusive signal strength modifier of 14 (18).

The sensitivity of the presented analysis is limited by the amount of recorded data. Combining multiple Higgs boson pair production decay modes is necessary to approach its observation and to probe SM expectations with significant precision. The presented analysis will contribute to the joined effort of the CMS collaboration aiming at such a combination.

Novel computing technologies have been developed during this thesis to handle the huge amounts of data and complex analysis workflows. The processing of events is vectorized, I/O latencies are reduced by caching data on SSDs [13], and complex pipelines of fit routines are automatized [18, 19]. These novel technologies significantly shorten turnaround times, lead to a more sustainable analysis, and improve its scientific results.

In summary, a search for Higgs boson pair production in the dileptonic $HH \rightarrow b\bar{b}W^+W^-$ decay mode has been presented. Novel analysis technologies have been successfully developed and applied to analyze sustainably and efficiently collision events, and to enhance the scientific results of the presented search. Future analyses may benefit from these technologies. The observation of Higgs boson pair production is a key measurement for the current and future LHC physics program to test the SM to its limits and search for new physics beyond the SM. The presented analysis contributes to this effort. The results presented in this thesis set the most stringent and precise measurements to date in the search for the $HH \rightarrow b\bar{b}W^+W^-$ process.

Editorial Tools

Different tools have been used for editorial help. These tools did not generate any content of this thesis. Their only purpose was to assist the author in technical tasks, especially with respect to the markup language LaTeX, which has been used to write this thesis. A list of tools is given in the following:

Visual Studio Code [234] has been used as code and text editor. The *LaTeX Workshop Extension* has been used to simplify the building process, synchronize between `.tex` and `.pdf` files, view the output `.pdf` on-the-fly, and leverage Visual Studio Code's *IntelliSense* for code completion.

GitHub Copilot [235] is a code generating AI tool using the Codex language model from OpenAI. This tool has been used as a "dynamic LaTeX macro generator", for e.g., LaTeX environments, such as `\begin{figure}...\end{figure}`. No content has been generated with this tool.

Grammarly [236] is an AI writing assistance tool, which helps to identify spelling errors and to improve the linguistic expression. This tool has been used primarily for the identification of spelling errors and misplacements of commas.

DeepL [237] is an AI powered language translation tool. Is has been used to improve the linguistic expression of this thesis and to assist the author translating the abstract into German.

Appendix

Drell-Yan Closure Test
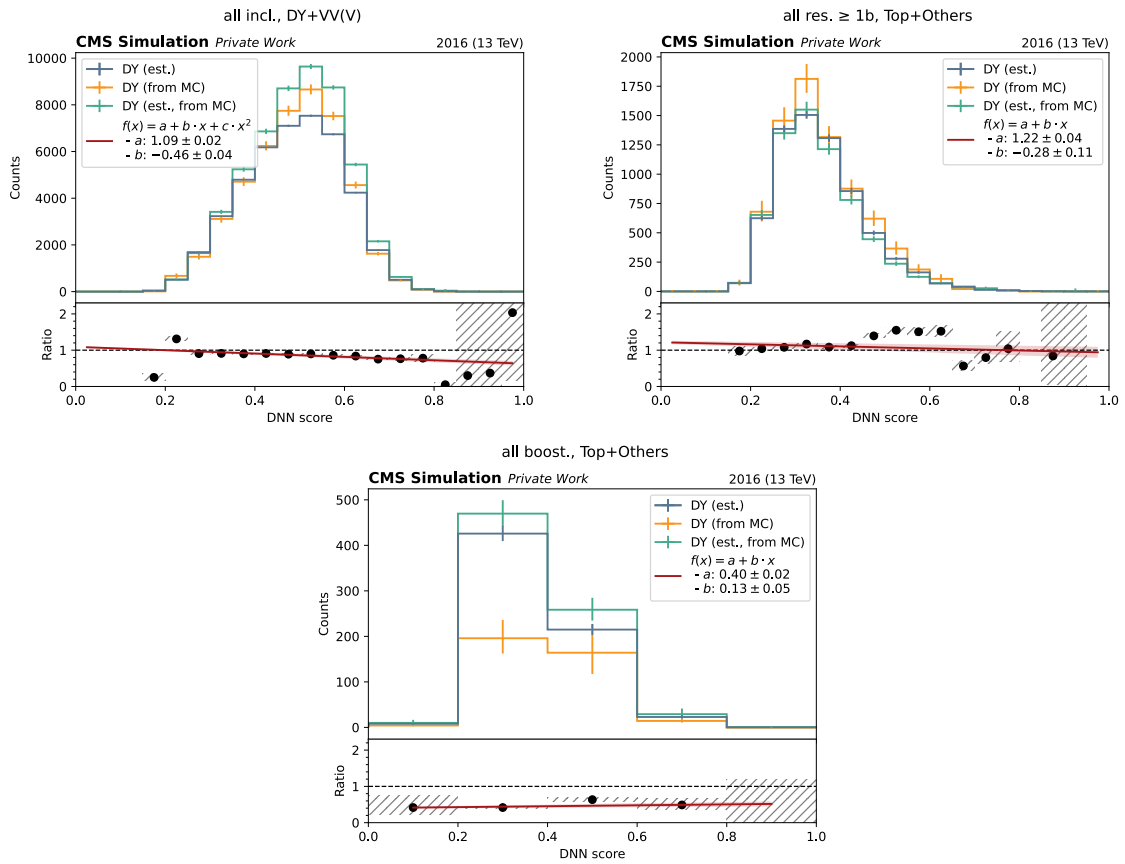
Figure 10.1: Closure test of the Drell-Yan background estimation for all background control regions in 2016. Shown are the Drell-Yan estimation itself (DY (est.)), the Drell-Yan estimation using simulation only (DY (est., from MC)), and the Drell-Yan distribution using simulation directly (DY (from MC)). The ratio between "DY (from MC)" and "DY (est., from MC)" is shown as well. From this ratio a correction is estimated by fitting a linear (quadratic) function to the "Top+Other" ("DY+VV(V)") region. The fit parameters are shown in the legend.
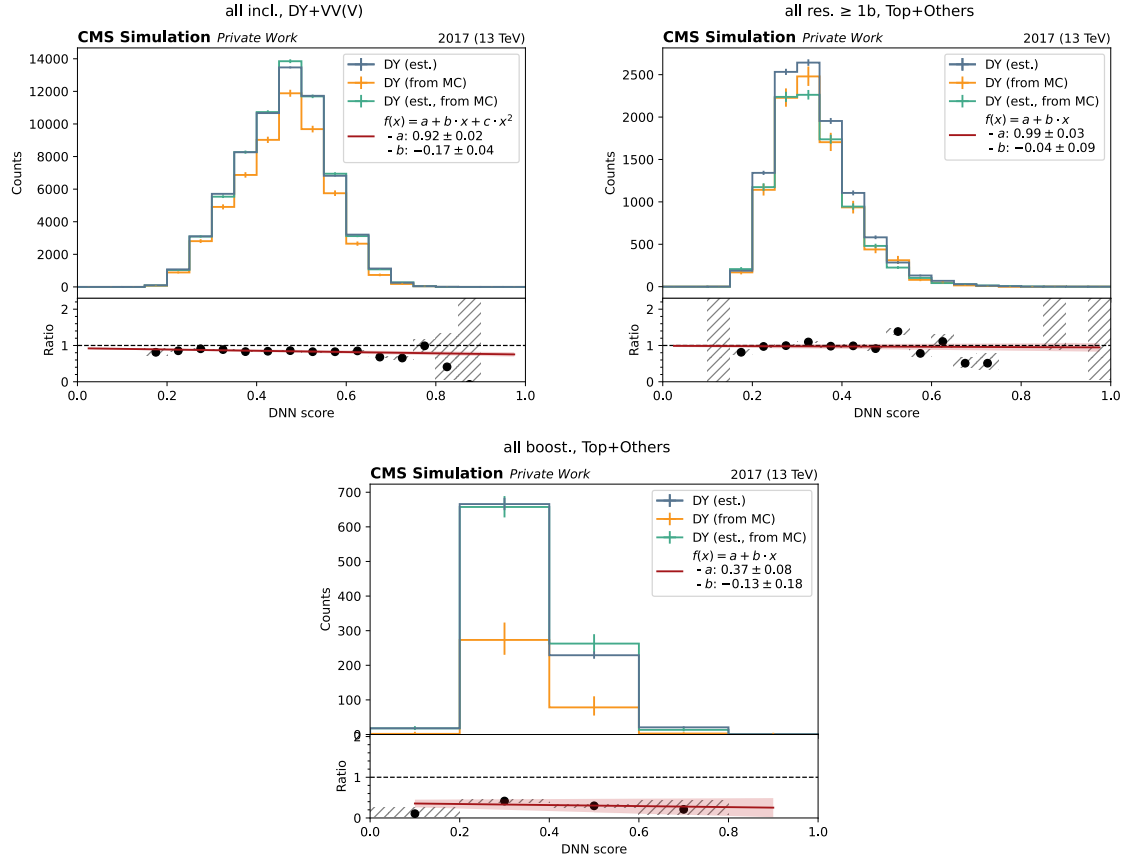
Figure 10.2: Closure test of the Drell-Yan background estimation for all background control regions in 2017. Shown are the Drell-Yan estimation itself (DY (est.)), the Drell-Yan estimation using simulation only (DY (est., from MC)), and the Drell-Yan distribution using simulation directly (DY (from MC)). The ratio between "DY (from MC)" and "DY (est., from MC)" is shown as well. From this ratio a correction is estimated by fitting a linear (quadratic) function to the "Top+Other" ("DY+VV(V)") region. The fit parameters are shown in the legend.

Correlation Matrices of DNN Input Features

(a) Correlations of the low-level features.

(b) Correlations of the high-level features.
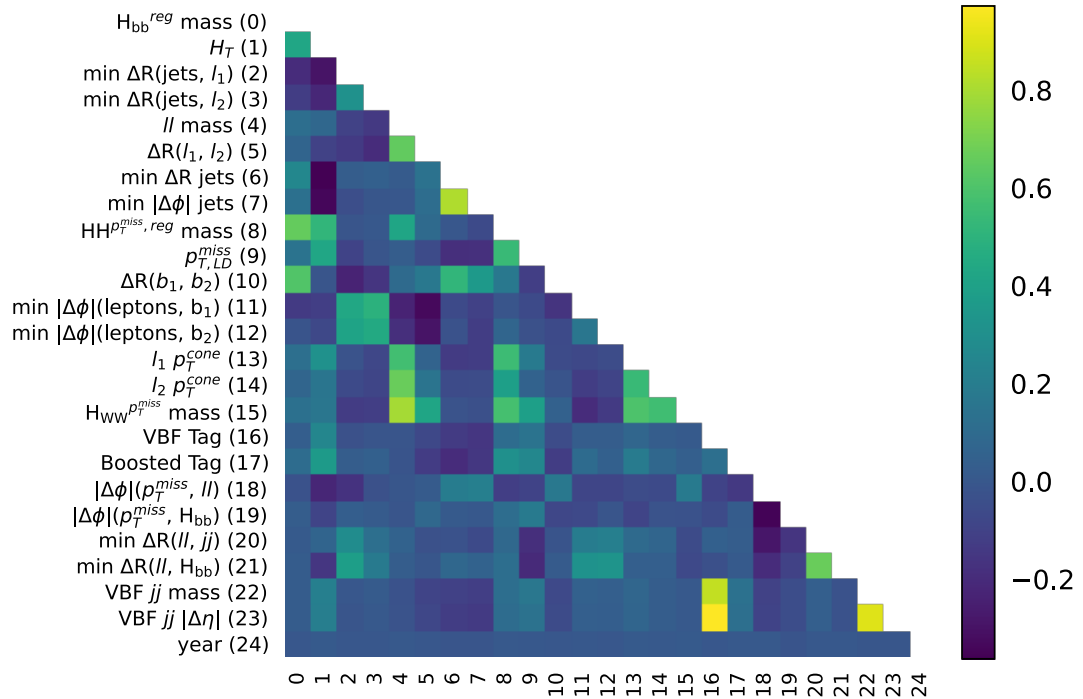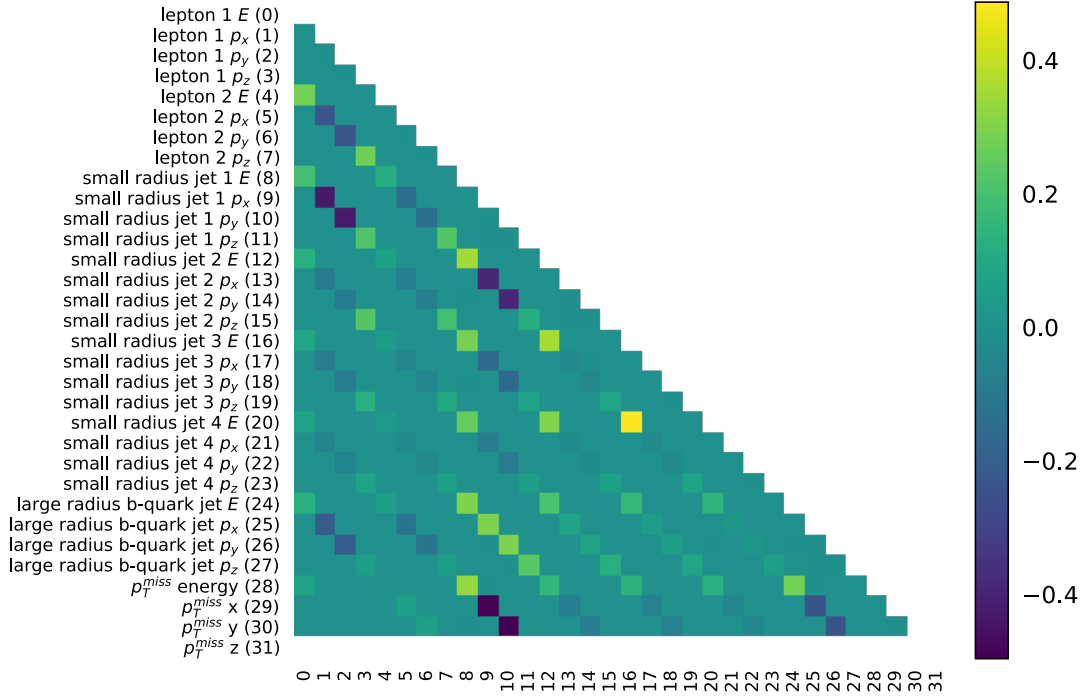
Figure 10.3: Feature correlations of the low-level and high-level features of the input particles. The numbering scheme of the particles denote the ranking with respect to the transverse momentum.

DNN Input Feature Importance

Feature Importance
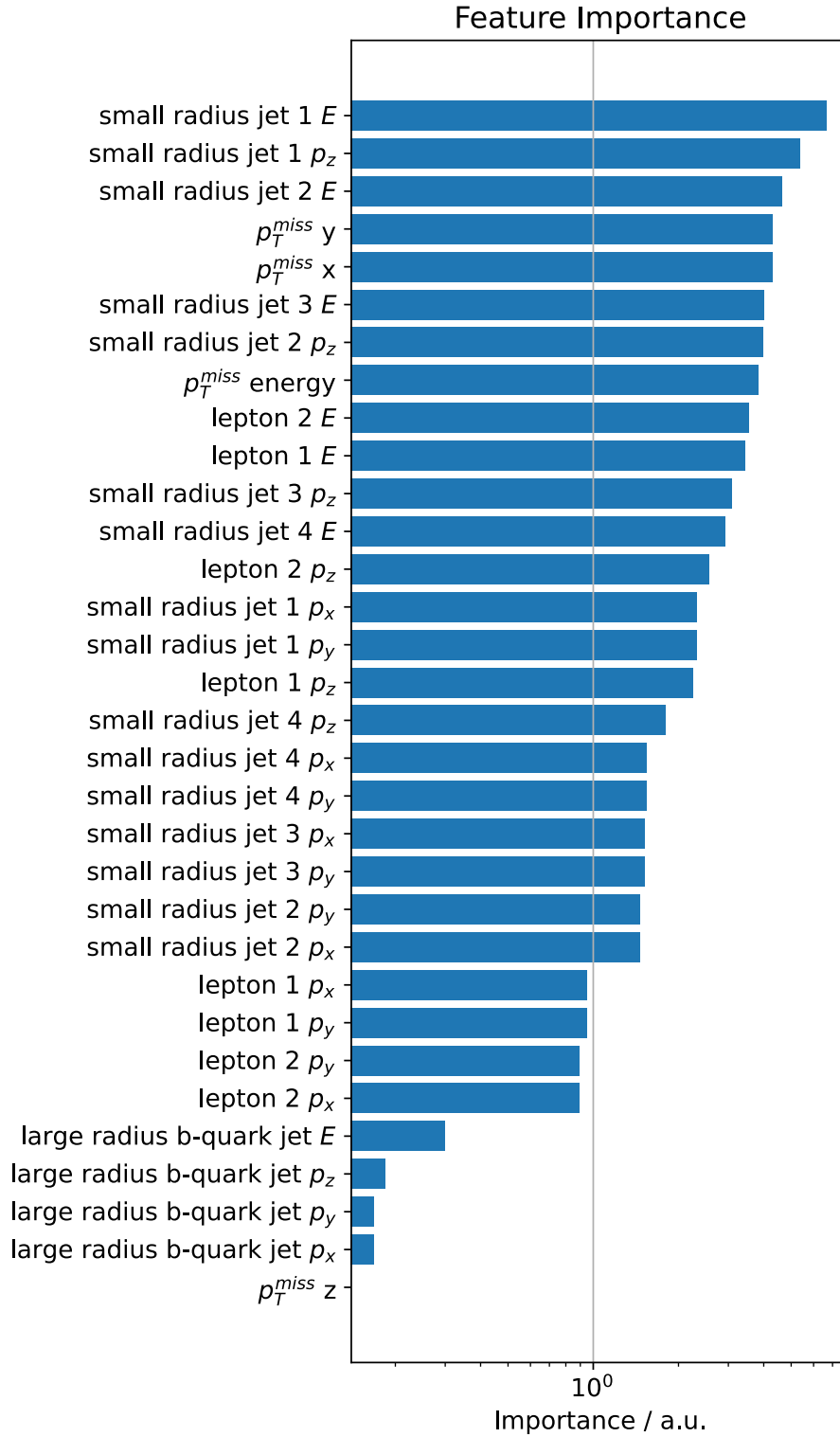
Figure 10.4: Feature importances for the LBN input variables based on the *permutation importance* method.

References

[1] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133, `doi:10.1016/0031-9163(64)91136-9`.

[2] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, `doi:10.1103/PhysRevLett.13.508`.

[3] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, `doi:10.1103/PhysRevLett.13.321`.

[4] Guralnik, Gerald S. and Hagen, Carl R. and Kibble, Tom W. B., "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, `doi:10.1103/PhysRevLett.13.585`.

[5] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, `doi:10.1103/PhysRevLett.19.1264`.

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

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

[8] D. Abercrombie et al., "Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum", *Physics of the Dark Universe* **27** (2020) 100371, `doi:https://doi.org/10.1016/j.dark.2019.100371`.

[9] P. J. E. Peebles and B. Ratra, "The cosmological constant and dark energy", *Reviews of Modern Physics* **75** (Apr, 2003) 559–606, `doi:10.1103/revmodphys.75.559`.

[10] CMS Collaboration, "Search for HH production in the bbWW decay mode", technical report, CERN, Geneva, 2023.

[11] H. P. Bretz et al., "A development environment for visual physics analysis", *Journal of Instrumentation* **7** (Aug, 2012) T08005–T08005, `doi:10.1088/1748-0221/7/08/t08005`.

[12] M. Erdmann et al., "Experiment Software and Projects on the Web with VISPA", *Journal of Physics: Conference Series* **898** (Oct, 2017) 072045, `doi:10.1088/1742-6596/898/7/072045`.

[13] N. Eich et al., "Fast Columnar Physics Analyses of Terabyte-Scale LHC Data on a Cache-Aware Dask Cluster", 2022. `doi:10.48550/ARXIV.2207.08598`.

[14] Dask Development Team, "Dask: Library for dynamic task scheduling", 2016, `https://dask.org`. `https://dask.org` (accessed on 24.08.2023).

[15] M. Rocklin, "Dask: Parallel Computation with Blocked algorithms and Task Scheduling", in *Proceedings of the 14th Python in Science Conference*, pp. 130 – 136. 2015.

[16] The Dask-Jobqueue Authors, "Dask-Jobqueue Source Code". `https://github.com/dask/dask-jobqueue` (accessed on 03.08.2022).

[17] The Dask-Jobqueue Authors, "Dask-Jobqueue Blog Entry". `https://blog.dask.org/2018/10/08/Dask-Jobqueue` (accessed on 03.08.2022).

[18] M. Rieger et al., "HH Inference Tools - Source Code". `https://gitlab.cern.ch/hh/tools/inference` (accessed on 11.01.2023).

[19] M. Rieger et al., "HH Inference Tools - Documentation". `https://cms-hh.web.cern.ch/cms-hh/tools/inference` (accessed on 11.01.2023).

[20] Super-Kamiokande Collaboration Collaboration, "Evidence for Oscillation of Atmospheric Neutrinos", *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567, `doi:10.1103/PhysRevLett.81.1562`.

[21] A. Bellerive et al., "The Sudbury Neutrino Observatory", *Nuclear Physics B* **908** (Jul, 2016) 30–51, `doi:10.1016/j.nuclphysb.2016.04.035`.

[22] B. Pontecorvo, "Mesonium and anti-mesonium", *Sov. Phys. JETP* **6** (1957) 429.

[23] B. Pontecorvo, "Neutrino Experiments and the Problem of Conservation of Leptonic Charge", *Zh. Eksp. Teor. Fiz.* **53** (1967) 1717–1725.

[24] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2020** (2020), no. 8, 083C01, `doi:10.1093/ptep/ptaa104` and 2021 update.

[25] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215, `doi:10.1016/S0031-9163(64)92001-3`.

[26] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 1", Jan, 1964. CERN-TH-412.

[27] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 2", Feb, 1964. CERN-TH-412, NP-14146, PRINT-64-170.

[28] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields", *Nuclear Physics B* **44** (1972), no. 1, 189 – 213, `doi:https://doi.org/10.1016/0550-3213(72)90279-9`.

[29] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (1961), no. 4, 579 – 588, `doi:10.1016/0029-5582(61)90469-2`.

[30] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367–377, `doi:10.1142/9789812795915_0034`.

[31] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD electroweak, heavy flavour groups, "Precision electroweak measurements on the Z

resonance", May, 2006. `doi:10.1016/j.physrep.2005.12.006`.

[32] T. Plehn and M. Rauch, "Quartic Higgs coupling at hadron colliders", *Physical Review D* **72** (Sep, 2005) `doi:10.1103/physrevd.72.053008`.

[33] D. B. Chitwood et al., "Improved Measurement of the Positive-Muon Lifetime and Determination of the Fermi Constant", *Physical Review Letters* **99** (Jul, 2007) `doi:10.1103/physrevlett.99.032001`.

[34] T. Markkanen, A. Rajantie, and S. Stopyra, "Cosmological Aspects of Higgs Vacuum Metastability", *Frontiers in Astronomy and Space Sciences* **5** (Dec, 2018) `doi:10.3389/fspas.2018.00040`.

[35] W. Buchmüller and D. Wyler, "Effective lagrangian analysis of new interactions and flavor conservation", *Nuclear Physics B* **268** (1986), no. 3, 621–653, `doi:https://doi.org/10.1016/0550-3213(86)90262-2`.

[36] B. Grzadkowski, M. Iskrzyński, M. Misiak, and J. Rosiek, "Dimension-six terms in the Standard Model Lagrangian", *Journal of High Energy Physics* **2010** (Oct, 2010) `doi:10.1007/jhep10(2010)085`.

[37] A. Noble and M. Perelstein, "Higgs self-coupling as a probe of the electroweak phase transition", *Physical Review D* **78** (Sep, 2008) `doi:10.1103/physrevd.78.063518`.

[38] J. R. Espinosa and D. Racco and A. Riotto, "Cosmological Signature of the Standard Model Higgs Vacuum Instability: Primordial Black Holes as Dark Matter", *Physical Review Letters* **120** (Mar, 2018) `doi:10.1103/physrevlett.120.121301`.

[39] A. Carvalho, "Gravity particles from Warped Extra Dimensions, predictions for LHC", 2014. `doi:10.48550/ARXIV.1404.0102`.

[40] R. W. Assmann, "LEP Operation and Performance with Electron-Positron Collisions at 209 GeV", `https://cds.cern.ch/record/567226`.

[41] R. D. Ball et al., "Parton distributions from high-precision collider data", *The European Physical Journal C* **77** (Oct, 2017) `doi:10.1140/epjc/s10052-017-5199-5`.

[42] H1,ZEUS Collaboration, "Combined measurement and QCD analysis of the inclusive e ± p scattering cross sections at HERA", *Journal of High Energy Physics* **2010** (Jan, 2010) `doi:10.1007/jhep01(2010)109`.

[43] HERA Collaboration, "Combined measurement and QCD analysis of the inclusive e ± p scattering cross sections at HERA", *Journal of High Energy Physics* **2010** (Jan, 2010) `doi:10.1007/jhep01(2010)109`.

[44] J. C. Collins, D. E. Soper, and G. Sterman, "Factorization of hard processes in QCD", pp. 1–91. Adv. Ser. Direct. High Energy Phys. 5, 1989. `arXiv:https://inspirehep.net/literature/25808`. `doi:10.1142/9789814503266\_0001`.

[45] J. de Blas et al., "Higgs Boson studies at future particle colliders", *Journal of High Energy Physics* **2020** (Jan, 2020) 139, `doi:10.1007/JHEP01(2020)139`.

[46] C. T. Potter et al., "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group", 2013.

```
doi:10.5170/CERN-2013-004.
```

[47] L. H. C. S. W. Group et al., "LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle", 2012.

[48] R. Frederix et al., "Higgs pair production at the LHC with NLO and parton-shower effects", *Physics Letters B* **732** (May, 2014) 142–149,
```
doi:10.1016/j.physletb.2014.03.026.
```

[49] The LHC Higgs Cross section group (HH subgroup). `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGHH?redirectedfrom=LHCPhysics.LHCHXSWGHH#Current_recommendations_for_HH_c` (accessed on 29.07.2022).

[50] B. D. Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, "Higgs boson potential at colliders: Status and perspectives", *Reviews in Physics* **5** (Nov, 2020) 100045,
```
doi:10.1016/j.revip.2020.100045.
```

[51] G. Degrassi, P. P. Giardino, F. Maltoni, and D. Pagani, "Probing the Higgs self coupling via single Higgs production at the LHC", *JHEP* **12** (2016) 080,
```
doi:10.1007/JHEP12(2016)080, arXiv:1607.04251.
```

[52] F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, "Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC", *Eur. Phys. J. C* **77** (2017), no. 12, 887, ```doi:10.1140/epjc/s10052-017-5410-8,
arXiv:1709.08649.```

[53] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) 2930,
```
doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
```

[54] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", 2011.

[55] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections", *The European Physical Journal C* **64** (Oct, 2009) 653–680,
```
doi:10.1140/epjc/s10052-009-1164-2.
```

[56] J. Gao et al., "CT10 next-to-next-to-leading order global analysis of QCD", *Physical Review D* **89** (Feb, 2014) ```doi:10.1103/physrevd.89.033009.```

[57] Ball, Richard D. and Bertone, Valerio and Carrazza, Stefano and Deans, Christopher S. and Del Debbio, Luigi and Forte, Stefano and Guffanti, Alberto and Hartland, Nathan P. and Latorre, José I. and Rojo, Juan and et al., "Parton distributions with LHC data", *Nuclear Physics B* **867** (Feb, 2013) 244–289,
```
doi:10.1016/j.nuclphysb.2012.10.003.
```

[58] M. Aliev et al., "HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR", *Computer Physics Communications* **182** (Apr, 2011) 1034–1046,
```
doi:10.1016/j.cpc.2010.12.040.
```

[59] P. Kant et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Computer Physics Communications* **191** (Jun, 2015) 74–89,
```
doi:10.1016/j.cpc.2015.02.001.
```

[60] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", 2011.

[61] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC", *The European Physical Journal C* **63** (Jul, 2009) 189–285, `doi:10.1140/epjc/s10052-009-1072-5`.

[62] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections", *The European Physical Journal C* **64** (Oct, 2009) 653–680, `doi:10.1140/epjc/s10052-009-1164-2`.

[63] H.-L. Lai et al., "New parton distributions for collider physics", *Physical Review D* **82** (Oct, 2010) `doi:10.1103/physrevd.82.074024`.

[64] R. D. Ball et al., "Parton distributions with LHC data", *Nuclear Physics B* **867** (Feb, 2013) 244–289, `doi:10.1016/j.nuclphysb.2012.10.003`.

[65] CMS Collaboration, "Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing (CMS - TWiki)". `https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns?rev=153#Summary_table_of_samples_produce` (accessed on 12.07.2023).

[66] CERN, "CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector", 2017. `doi:10.23731/CYRM-2017-002`.

[67] CMS Collaboration, "Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2018** (Jan, 2018) 54, `doi:10.1007/JHEP01(2018)054`.

[68] ATLAS Collaboration, "Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physics Letters B* **801** (Feb, 2020) 135145, `doi:10.1016/j.physletb.2019.135145`.

[69] CMS Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery", *Nature* **607** (Jul, 2022) 60–68, `doi:10.1038/s41586-022-04892-x`.

[70] ALICE Collaboration et al., "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (Aug, 2008) S08002, `doi:10.1088/1748-0221/3/08/S08002`.

[71] ATLAS Collaboration et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (Aug, 2008) S08003, `doi:10.1088/1748-0221/3/08/S08003`.

[72] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, `doi:10.1088/1748-0221/3/08/S08004`.

[73] LHCb Collaboration et al., "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (Aug, 2008) S08005, `doi:10.1088/1748-0221/3/08/S08005`.

[74] C. Lefevre, "LHC: the guide (English version). Guide du LHC (version anglaise)", Feb, 2009. `http://cds.cern.ch/record/1165534`.

[75] TOTEM Collaboration, "Total cross-section, elastic scattering and diffraction

dissociation at the Large Hadron Collider at CERN: TOTEM Technical Design Report". Technical design report. TOTEM. CERN, Geneva, 2004.

[76] CMS Collaboration, "CMS Public Luminosity Results". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults?rev=168` (accessed on 03.02.2022).

[77] T. Sakuma, "SketchUpCMS Gallery". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS#5_Detector_figures` (accessed on 28.02.2022).

[78] G. L. Bayatian et al., "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical design report. CMS. CERN, Geneva, 2006.

[79] G. L. Bayatian et al., "CMS Physics: Technical Design Report Volume 2: Physics Performance", volume 34. CERN, 2007.

[80] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade". CERN, Sep, 2012.

[81] G. L. Bayatyan et al., "CMS computing: Technical Design Report". Technical design report. CMS. CERN, Geneva, 2005.

[82] CMS Collaboration, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *Journal of Instrumentation* **5** (Mar, 2010) T03012–T03012, `doi:10.1088/1748-0221/5/03/t03012`.

[83] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s}$=7 TeV at the LHC", *Journal of Instrumentation* **8** (Nov, 2013) P11002–P11002, `doi:10.1088/1748-0221/8/11/p11002`.

[84] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV", *Journal of Instrumentation* **13** (Jun, 2018) P06015–P06015, `doi:10.1088/1748-0221/13/06/p06015`.

[85] CMS Collaboration, "The CMS trigger system", *Journal of Instrumentation* **12** (Jan, 2017) P01020–P01020, `doi:10.1088/1748-0221/12/01/p01020`.

[86] CMS Collaboration, S. Cittolin, A. Rácz, and P. Sphicas, "CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project". Technical design report. CMS. CERN, Geneva, 2002.

[87] CMS Collaboration, "The CMS trigger in Run 2", technical report, CERN, Geneva, 2018. `doi:10.22323/1.314.0523`.

[88] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (Jun, 2017) P10003. 82 p, `doi:10.1088/1748-0221/12/10/P10003`, `arXiv:1706.04965`.

[89] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (Oct, 2014) P10009–P10009, `doi:10.1088/1748-0221/9/10/p10009`.

[90] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (Oct, 2014) P10009,

doi:10.1088/1748-0221/9/10/P10009.

[91] P. Billoir, "Progressive track recognition with a Kalman-like fitting procedure", *Computer Physics Communications* **57** (1989), no. 1, 390–394, doi:https://doi.org/10.1016/0010-4655(89)90249-X.

[92] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the Kalman filtering method", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **294** (1990), no. 1, 219–228, doi:https://doi.org/10.1016/0168-9002(90)91835-Y.

[93] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **395** (1997), no. 2, 169–184, doi:https://doi.org/10.1016/S0168-9002(97)00705-5.

[94] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987), no. 2, 444–450, doi:https://doi.org/10.1016/0168-9002(87)90887-4.

[95] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, "Electron Reconstruction in the CMS Electromagnetic Calorimeter", https://cds.cern.ch/record/687345?ln=en.

[96] Frühwirth, R. and Speer, T., "A Gaussian-sum filter for vertex reconstruction", *Nucl. Instrum. Meth. A* **534** (2004) 217–221, doi:10.1016/j.nima.2004.07.090.

[97] CMS Collaboration, "Performance of electron and photon reconstruction in Run 2 with the CMS experiment", https://cds.cern.ch/record/2725004.

[98] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm", *Journal of High Energy Physics* (Apr, 2008) 063–063, doi:10.1088/1126-6708/2008/04/063.

[99] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *The European Physical Journal C* **72** (Mar, 2012) doi:10.1140/epjc/s10052-012-1896-2.

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

[101] CMS Collaboration, "Pileup Jet Identification", technical report, CERN, Geneva, 2013. https://cds.cern.ch/record/1581583.

[102] CMS Collaboration, "Performance of the pile up jet identification in CMS for Run 2",.

[103] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (Nov, 2011) P11002–P11002, doi:10.1088/1748-0221/6/11/p11002.

[104] CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018", https://cds.cern.ch/record/2715872.

[105] UA1 Collaboration, "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540 GeV$", *Phys. Lett. B* **122** (Jan, 1983) 103–116. 31 p, doi:10.5170/CERN-1983-004.123.

[106] A. Astbury et al., "A 4π solid angle detector for the SPS used as a proton-antiproton collider at a centre of mass energy of 540 GeV", technical report, CERN, Geneva, Jan, 1978. `https://cds.cern.ch/record/319371`.

[107] E. Bols et al., "Jet Flavour Classification Using DeepJet", *JINST* **15** (2020), no. 12, P12012, `doi:10.1088/1748-0221/15/12/P12012`, `arXiv:2008.10519`.

[108] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", *Proceedings of the IEEE* **86** (Dec, 1998) 2278 – 2324, `doi:10.1109/5.726791`.

[109] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory", *Neural computation* **9** (Dec, 1997) 1735–80, `doi:10.1162/neco.1997.9.8.1735`.

[110] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector", `https://cds.cern.ch/record/2646773`.

[111] CMS Collaboration, "Heavy flavor identification at CMS with deep neural networks", `https://cds.cern.ch/record/2255736`.

[112] CMS Collaboration, "b Tag & Vertexing Physics Object Group (CMS - TWiki)". `https://twiki.cern.ch/twiki/bin/view/CMS/BtagPOG?rev=291` (accessed on 06.07.2023).

[113] A. Sirunyan, A. Tumasyan, W. Adam, and F. Ambrogi, "Evidence for the Higgs boson decay to a bottom quark–antiquark pair", *Physics Letters B* **780** (May, 2018) 501–532, `doi:10.1016/j.physletb.2018.02.050`.

[114] CMS Collaboration, "NANOAOD: a new compact event data format in CMS", *EPJ Web Conf.* **245** (2020) 06002, `doi:10.1051/epjconf/202024506002`.

[115] Y. Collet and M. Kucherawy, "Zstandard Compression and the 'application/zstd' Media Type". RFC 8878, Feb, 2021. `doi:10.17487/RFC8878`.

[116] D. Spiga et al., "CRAB: the CMS distributed analysis tool development and design", *Nucl. Phys. B, Proc. Suppl.* **177-178** (2008) 267–268, `doi:10.1016/j.nuclphysbps.2007.11.124`.

[117] M. Erdmann et al., "A field study of data analysis exercises in a bachelor physics course using the internet platform VISPA", *European Journal of Physics* **35** (Apr, 2014) 035018, `doi:10.1088/0143-0807/35/3/035018`.

[118] M. Erdmann et al., "The VISPA internet-platform in deep learning applications", *Journal of Physics: Conference Series* **1085** (Sep, 2018) 042044, `doi:10.1088/1742-6596/1085/4/042044`.

[119] Beer, Max et al., "Knowledge sharing on deep learning in physics research using VISPA", *EPJ Web Conf.* **245** (2020) 05040, `doi:10.1051/epjconf/202024505040`.

[120] The Jupyter Project, "JupyterLab computational environment". `https://github.com/jupyterlab/jupyterlab` (accessed on 25.03.2022).

[121] Erdmann, Martin et al., "Evolution of the VISPA-project", *EPJ Web Conf.* **214** (2019) 05021, `doi:10.1051/epjconf/201921405021`.

[122] E. Rodrigues et al., "The Scikit HEP Project – overview and prospects", *EPJ Web Conf.* **245** (2020) 06028, `doi:10.1051/epjconf/202024506028`, `arXiv:2007.03577`.

[123] C. R. Harris et al., "Array programming with NumPy", *Nature* **585** (Sep, 2020) 357–362, `doi:10.1038/s41586-020-2649-2`.

[124] "Anaconda Software Distribution", 2020, `https://docs.anaconda.com/`.

[125] M. Rieger, M. Erdmann, B. Fischer, and R. Fischer, "Design and Execution of make-like, distributed Analyses based on Spotify's Pipelining Package Luigi", 2017.

[126] M. Rieger, "Design Pattern for Analysis Automation on Distributed Resources using Luigi Analysis Workflows", *EPJ Web of Conferences* **245** (Jan, 2020) 05025, `doi:10.1051/epjconf/202024505025`.

[127] The Luigi Authors, "Luigi (Spotify)". `https://github.com/spotify/luigi` (accessed on 03.08.2022).

[128] J. Pivarski et al., "scikit-hep/uproot", Jul, 2022, `https://doi.org/10.5281/zenodo.6885371`.

[129] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth. A* **389** (1997) 81–86, `doi:10.1016/S0168-9002(97)00048-X`.

[130] The pandas development team, "pandas-dev/pandas: Pandas", Feb, 2020. `doi:10.5281/zenodo.3509134`.

[131] Wes McKinney, "Data Structures for Statistical Computing in Python", in *Proceedings of the 9th Python in Science Conference*, Stéfan van der Walt and Jarrod Millman, eds., pp. 56 – 61. 2010. `doi:10.25080/Majora-92bf1922-00a`.

[132] R. Okuta et al., "CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations", in *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*. 2017. `http://learningsys.org/nips17/assets/papers/paper_16.pdf`.

[133] J. Pivarski et al., "Awkward Array", Oct, 2018. `doi:10.5281/zenodo.4341376`.

[134] L. Gray et al., "CoffeaTeam/coffea". `doi:10.5281/zenodo.5762406`.

[135] HTCondor Team, "HTCondor". `doi:10.5281/zenodo.5750673`.

[136] The Boost::Histogram Authors, "Boost::Histogram". `https://github.com/boostorg/histogram` (accessed on 03.08.2022).

[137] H. Schreiner et al., "scikit-hep/boost-histogram", Feb, 2022. `doi:10.5281/zenodo.6107728`.

[138] J. D. Hunter, "Matplotlib: A 2D graphics environment", *Computing in Science & Engineering* **9** (2007), no. 3, 90–95, `doi:10.1109/MCSE.2007.55`.

[139] A. Novak et al., "scikit-hep/mplhep", Jul, 2022. `doi:10.5281/zenodo.6807166`.

[140] TensorFlow Developers, "TensorFlow", May, 2022. `doi:10.5281/zenodo.6574269`.

[141] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous

Systems", 2015. Software available from `https://www.tensorflow.org/`.

[142] The TensorFlow Serving Authors, "TensorFlow Serving Source Code". `https://github.com/tensorflow/serving` (accessed on 03.08.2022).

[143] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters", in *OSDI'04: Sixth Symposium on Operating System Design and Implementation*, pp. 137–150. San Francisco, CA, 2004.

[144] L. Kernel, "General Filesystem Caching", 2022.

[145] J. W. Shay Gueron, Simon Johnson, "SHA-512/256". `https://eprint.iacr.org/2010/548.pdf` (accessed on 19.07.2023).

[146] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks", 2015.

[147] A. Vaswani et al., "Attention Is All You Need", 2017.

[148] J. Hochreiter, "Untersuchungen zu dynamischen neuronalen Netzen". `http://www.bioinf.jku.at/publications/older/3804.pdf`.

[149] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation* **9** (Nov, 1997) 1735–1780, `doi:10.1162/neco.1997.9.8.1735`.

[150] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies", *IEEE Access* **8** (2020) 58443–58469, `doi:10.1109/access.2020.2983149`.

[151] M. Erdmann, J. Glombitza, G. Kasieczka, and U. Klemradt, "Deep Learning for Physics Research". World Scientific, 2021.

[152] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization", *Journal of Machine Learning Research* **12** (2011), no. 61, 2121–2159.

[153] H. B. McMahan and M. Streeter, "Adaptive Bound Optimization for Online Convex Optimization", 2010.

[154] K. S. G. Hinton, Nitish Srivastava, "Neural Networks for Machine Learning, Lecture 6a, Overview of mini-batch gradient descent". `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`.

[155] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", 2017.

[156] S. Ruder, "An overview of gradient descent optimization algorithms", 2017.

[157] N. Qian, "On the momentum term in gradient descent learning algorithms", *Neural Networks* **12** (1999), no. 1, 145–151, `doi:https://doi.org/10.1016/S0893-6080(98)00116-6`.

[158] M. Erdmann, E. Geiser, Y. Rath, and M. Rieger, "Lorentz Boost Networks: autonomous physics-inspired feature engineering", *Journal of Instrumentation* **14** (Jun, 2019) P06006–P06006, `doi:10.1088/1748-0221/14/06/p06006`.

[159] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Springer series in statistics. Springer, 2009.

[160] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal*

Statistical Society. Series B (Methodological) **58** (1996), no. 1, 267–288.

[161] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 Regularization for Learning Kernels", 2012.

[162] N. Srivastavaet al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research* **15** (2014), no. 56, 1929–1958.

[163] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", 2015.

[164] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. 2016. `doi:10.1109/CVPR.2016.90`.

[165] J. Deng et al., "ImageNet: A large-scale hierarchical image database", in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. 2009. `doi:10.1109/CVPR.2009.5206848`.

[166] T. Mikolov, "Statistical Language Models based on Neural Networks". `https://www.fit.vutbr.cz/~imikolov/rnnlm/thesis.pdf` (Sec. 3.2.2, accessed on 11.07.2023).

[167] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks", 2017.

[168] M. Baak, S. Gadatsch, R. Harrington, and W. Verkerke, "Interpolation between multi-dimensional histograms using a new non-linear moment morphing method", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **771** (2015) 39–48, `doi:https://doi.org/10.1016/j.nima.2014.10.033`.

[169] CERN, "Proceedings of the 1974 CERN School of Computing: Godøysund, Norway 11 - 24 Aug 1974. 3rd CERN School of Computing". CERN, Geneva, (1974). `doi:10.5170/CERN-1974-023`.

[170] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Computer Physics Communications* **77** (1993), no. 2, 219–228, `doi:https://doi.org/10.1016/0010-4655(93)90005-W`.

[171] A. L. Read, "Presentation of search results: the CLs technique", *Journal of Physics G: Nuclear and Particle Physics* **28** (Sep, 2002) 2693, `doi:10.1088/0954-3899/28/10/313`.

[172] T. Junk, "Confidence level computation for combining searches with small statistics", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **434** (Sep, 1999) 435–443, `doi:10.1016/s0168-9002(99)00498-2`.

[173] N. Jerzy and P. E. Sharpe, "On the problem of the most efficient tests of statistical hypotheses", *Philosophical Transactions of the Royal Society of London* (Feb, 1933).

[174] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C* **71** (Feb, 2011) `doi:10.1140/epjc/s10052-011-1554-0`.

[175] R. D. Cousins, "Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms", 2013. `https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf` (accessed on 25.01.2023).

[176] J. Lindsey and P. Lindsey, "Parametric Statistical Inference". Oxford science publications. Clarendon Press, 1996.

[177] "CMS Higgs Combination toolkit". `https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit` (accessed on 25.01.2023).

[178] CMS Collaboration, "Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Journal of High Energy Physics* **2019** (Mar, 2019) `doi:10.1007/jhep03(2019)026`.

[179] "PhysicsResultsCombined < CMSPublic < TWiki". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined#CMS_Cross_Section_measurements` (accessed on 18.11.2021).

[180] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021), no. 9, 800, `doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927`.

[181] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", CMS-PAS-LUM-17-004.

[182] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", CMS-PAS-LUM-18-002.

[183] M. Grazzini et al., "Higgs boson pair production at NNLO with top quark mass effects", *Journal of High Energy Physics* **2018** (May, 2018) 59, `doi:10.1007/JHEP05(2018)059` arXiv: 1803.02463.

[184] Dreyer, Frédéric A. and Karlberg, Alexander, "Vector-Boson Fusion Higgs Pair Production at NLO", *Physical Review D* **98** (Dec, 2018) 114016, `doi:10.1103/PhysRevD.98.114016` arXiv: 1811.07906.

[185] CMS Collaboration, "Search for Higgs boson pair production in the decay channel to two bottom quarks and two W bosons using the full Run 2 dataset". CMS internal note: AN-2020/119.

[186] CMS Collaboration, "How to work with files for Good Luminosity Sections in JSON format (CMS - TWiki)". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGoodLumiSectionsJSONFile?rev=30` (accessed on 20.07.2023).

[187] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *Journal of High Energy Physics* **2004** (Dec, 2004) 040, `doi:10.1088/1126-6708/2004/11/040`.

[188] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *Journal of High Energy Physics* **2007** (Nov, 2007) 070, `doi:10.1088/1126-6708/2007/11/070`.

[189] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *Journal of High Energy Physics* **2010** (Jun, 2010) `doi:10.1007/jhep06(2010)043`.

[190] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *Journal of High Energy Physics* **2014** (Jul, 2014) `doi:10.1007/jhep07(2014)079`.

[191] K. Ehatäht and C. Veelken, "Stitching Monte Carlo samples", *The European Physical Journal C* **82** (May, 2022) `doi:10.1140/epjc/s10052-022-10407-9`.

[192] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *The European Physical Journal C* **76** (Mar, 2016) `doi:10.1140/epjc/s10052-016-3988-x`.

[193] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *The European Physical Journal C* **74** (Aug, 2014) `doi:10.1140/epjc/s10052-014-3024-y`.

[194] CMS Collaboration, "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements", *The European Physical Journal C* **80** (Jan, 2020) `doi:10.1140/epjc/s10052-019-7499-4`.

[195] R. D. Ball et al., "Parton distributions for the LHC run II", *Journal of High Energy Physics* **2015** (Apr, 2015) `doi:10.1007/jhep04(2015)040`.

[196] The NNPDF Collaboration et al., "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO", 2011.

[197] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$", *Phys. Rev. D* **74** (Dec, 2006) 114017, `doi:10.1103/PhysRevD.74.114017`.

[198] I. Neutelings, "Drell-Yan + jets cross section at NNLO". `https://indico.cern.ch/event/746829/contributions/3138541/attachments/1717905/2772129/Drell-Yan_jets_crosssection.pdf` (accessed on 13.07.2023).

[199] 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), no. 3, 250–303, `doi:https://doi.org/10.1016/S0168-9002(03)01368-8`.

[200] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at 13 TeV using the CMS detector", *Journal of Instrumentation* **14** (Jul, 2019) P07004–P07004, `doi:10.1088/1748-0221/14/07/p07004`.

[201] CMS Collaboration, "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s}$=13TeV", `doi:10.1140/epjc/s10052-021-09014-x`.

[202] M. Cacciari, G. P. Salam, and G. Soyez, "The catchment area of jets", *Journal of High Energy Physics* **2008** (Apr, 2008) 005–005, `doi:10.1088/1126-6708/2008/04/005`.

[203] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Physics Letters B* **659** (Jan, 2008) 119–126, `doi:10.1016/j.physletb.2007.09.077`.

[204] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *Journal of Instrumentation* **16** (May, 2021) P05014, `doi:10.1088/1748-0221/16/05/p05014`.

[205] CMS Collaboration, "Jet algorithms performance in 13 TeV data", technical report, CERN, Geneva, 2017.

[206] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft drop", *Journal of High Energy Physics* **2014** (May, 2014) `doi:10.1007/jhep05(2014)146`.

[207] J. Thaler and K. V. Tilburg, "Identifying boosted objects with N-subjettiness", *Journal of High Energy Physics* **2011** (Mar, 2011) `doi:10.1007/jhep03(2011)015`.

[208] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *Journal of Instrumentation* **15** (Sep, 2020) P09018–P09018, `doi:10.1088/1748-0221/15/09/p09018`.

[209] CMS Collaboration, "Study of Higgs boson production in association with top quarks final states with electrons, muons, and taus using the full Run 2 dataset". CMS internal note: AN-2019/111.

[210] CMS Collaboration, "Recommended Jet Energy Corrections and Uncertainties For Data and MC (CMS - TWiki)". `https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC?rev=202` (accessed on 18.07.2023).

[211] CMS Collaboration, "Identification of heavy-flavor jets with the CMS detector in pp collisions at 13 TeV", *Journal of Instrumentation* **13** (May, 2018) P05011–P05011, `doi:10.1088/1748-0221/13/05/p05011`.

[212] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Journal of Instrumentation* **15** (Oct, 2020) P10017–P10017, `doi:10.1088/1748-0221/15/10/p10017`.

[213] CMS Collaboration, "Reweighting recipe to emulate Level 1 ECAL and Muon prefiring (CMS - TWiki)". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe?rev=3` (accessed on 13.07.2023).

[214] CMS Collaboration, "MET Analysis (CMS - TWiki)". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis#xy_Shift_Correction?rev=200` (accessed on 13.07.2023).

[215] CMS Collaboration, "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons", *Journal of High Energy Physics* **2019** (Feb, 2019) `doi:10.1007/jhep02(2019)149`.

[216] M. Czakon et al., "Top-pair production at the LHC through NNLO QCD and NLO EW", *Journal of High Energy Physics* **2017** (Oct, 2017) `doi:10.1007/jhep10(2017)186`.

[217] CDF Collaboration, "Measurement of $\sigma B(W \to e\nu)$ and $\sigma B(Z^0 \to e^+e^-)$ in $\overline{p}p$

collisions at $\sqrt{s} = 1800$ GeV", *Phys. Rev. D* **44** (Jul, 1991) 29–52,
doi:10.1103/PhysRevD.44.29.

[218] R. Garnett, "Bayesian Optimization". https://bayesoptbook.com/ (accessed on 20.06.2023).

[219] L. Breiman, "Random Forests", *Machine Learning* **45** (Oct, 2001) 5–32,
doi:10.1023/A:1010933404324.

[220] ATLAS Collaboration, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC", *Phys. Rev. Lett.* **117** (Oct, 2016) 182002, doi:10.1103/PhysRevLett.117.182002.

[221] CMS Collaboration, "Sources of jet energy scale uncertainties (CMS - TWiki)".
https://twiki.cern.ch/twiki/bin/viewauth/CMS/
JECUncertaintySources#Jet_energy_scale_uncertainty_sou (accessed on 26.06.2023).

[222] CMS Collaboration, "Jet energy resolution scale factors and uncertainty (CMS - TWiki)". https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution (accessed on 27.06.2023).

[223] CMS Collaboration, "LHC Higgs Cross Section HH Sub-group (CMS - TWiki)".
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGHH?rev=74 (accessed on 28.06.2023).

[224] CMS Collaboration, "SM Higgs production cross sections at $\sqrt{s} = 13$ TeV (CMS - TWiki)". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/
CERNYellowReportPageAt13TeV?rev=23 (accessed on 28.06.2023).

[225] CMS Collaboration, "NNLO+NNLL top-quark-pair cross sections (CMS - TWiki)".
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO?rev=17
(accessed on 28.06.2023).

[226] CMS Collaboration, "Standard Model Cross Sections for CMS at 13 TeV (CMS - TWiki)". https://twiki.cern.ch/twiki/bin/viewauth/CMS/
StandardModelCrossSectionsat13TeV?rev=27 (accessed on 28.06.2023).

[227] CMS Collaboration, "NLO single-top channel cross sections (CMS - TWiki)".
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/
SingleTopRefXsec?rev=34 (accessed on 28.06.2023).

[228] CMS Collaboration, "Higgs branching fractions (CMS - TWiki)". https://twiki.
cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR?rev=22
(accessed on 28.06.2023).

[229] CMS Collaboration, "TOP Systematic Uncertainties (Run2) (CMS - TWiki)". https:
//twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics?rev=198
(accessed on 28.06.2023).

[230] B. Jäger et al., "Parton-shower effects in Higgs production via vector-boson fusion",
The European Physical Journal C **80** (Aug, 2020) 756,
doi:10.1140/epjc/s10052-020-8326-7.

[231] CMS Collaboration, "A new set of CMS tunes for novel colour reconnection models in PYTHIA8 based on underlying-event data", technical report, CERN, Geneva, 2021.

[232] A. M. Sirunyan et al., "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements", *The European Physical Journal C* **80** (Jan, 2020) 4, `doi:10.1140/epjc/s10052-019-7499-4`.

[233] F. Bishara, R. Contino, and J. Rojo, "Higgs pair production in vector-boson fusion at the LHC and beyond", *The European Physical Journal C* **77** (Jul, 2017) 481, `doi:10.1140/epjc/s10052-017-5037-9`.

[234] Microsoft, "Visual Studio Code". `https://code.visualstudio.com` (accessed on 16.07.2023).

[235] GitHub and OpenAI, "GitHub Copilot". `https://github.com/features/copilot/` (accessed on 16.07.2023).

[236] Grammarly Inc., "Grammarly". `https://www.grammarly.com` (accessed on 16.07.2023).

[237] DeepL SE, "DeepL". `https://www.deepl.com/translator` (accessed on 16.07.2023).

List of Publications

Main Publications

CMS Collaboration, *Search for HH production in the bbWW decay mode* [10], Physics Analysis Summary (CMS-PAS-HIG-21-005).

Niclas Eich, Martin Erdmann, Peter Fackeldey, Benjamin Fischer, Dennis Noll and Yannik Rath, *Fast Columnar Physics Analyses of Terabyte-Scale LHC Data on a Cache-Aware Dask Cluster* [13], DOI: `10.1007/s41781-023-00095-9`, Computing and Software for Big Science **7**, 3 (2023).

Lisa Benato, Erik Buhmann, Martin Erdmann, Peter Fackeldey, Jonas Glombitza, Nikolai Hartmann, Gregor Kasieczka, William Korcari, Thomas Kuhr, Jan Steinheimer, Horst Stöcker, Tilman Plehn and Kai Zhou, *Shared Data and Algorithms for Deep Learning in Fundamental Physics*, DOI: `10.1007/s41781-022-00082-6`, Computing and Software for Big Science **6**, 9 (2022).

Conference Proceedings

Max Beer, Niclas Eich, Martin Erdmann, Peter Fackeldey, Benjamin Fischer, Katharina Hafner, Dennis Daniel Nick Noll, Yannik Alexander Rath, Marcel Rieger, Alexander Temme, Max Vieweg and Martin Urban, *Knowledge sharing on deep learning in physics research using VISPA*, 24th International Conference on Computing in High Energy and Nuclear Physics, DOI: `10.1051/epjconf/202024505040`, EPJ Web of Conferences **245**, 05040 (2020).

Dennis Noll, Martin Erdmann, Peter Fackeldey and Benjamin Fischer, *Vectorised Neutrino Reconstruction by Computing Graphs*, 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, DOI: `10.1088/1742-6596/2438/1/012133`, Journal of Physics: Conference Series **2438** 012133 (2023).

Niclas Steve Eich, Martin Erdmann, Svenja Diekmann, Manfred Peter Fackeldey, Benjamin Fischer, Dennis Noll and Yannik Alexander Rath, *Going fast on a small-size computing cluster*, 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, DOI: `10.1088/1742-6596/2438/1/012042`, Journal of Physics: Conference Series **2438** 012042 (2023).

Other

Original developer and maintainer of the *HH Inference Tools* software package [18, 19]. The package has been used by the CMS collaboration in the following Higgs pair production publications:

- CMS Collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, Nature volume 607, pages 60-68 (2022).

- CMS Collaboration, *Search for HH production in the bbWW decay mode*, Physics Analysis Summary (CMS-PAS-HIG-21-005).

- CMS Collaboration, *Search for Higgs boson pairs decaying to WWWW, WWττ, and ττττ in proton-proton collisions at $\sqrt{s} = 13\,TeV$*, Submitted to the Journal of High Energy Physics (JHEP).

- CMS Collaboration, *Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13\,TeV$*, Phys. Lett. B 842 (2023) 137531

- CMS Collaboration, *Search for Higgs boson pair production with one associated vector boson in proton-proton collisions at $\sqrt{s} = 13\,TeV$*, Physics Analysis Summary (CMS-PAS-HIG-22-006)

Acknowledgements (Danksagungen)

Du hast mich durch die gesamte Zeit begleitet, mir immer zugehört und mich bedingungslos unterstützt. In jeder Situation warst du für mich da und hast die gesamte Zeit zu einem unvergesslichen Erlebnis gemacht. Ich danke dir für deine Liebe und Unterstützung.

Danke!

Eidesstaatliche Erklärung

Manfred Peter Fackeldey erklärt hiermit, dass diese Dissertation und die darin dargelegten Inhalte die eigenen sind und selbstständig, als Ergebnis der eigenen originären Forschung, generiert wurden.

Hiermit erkläre ich an Eides statt

1. Diese Arbeit wurde vollständig oder größtenteils in der Phase als Doktorand dieser Fakultät und Universität angefertigt;

2. Sofern irgendein Bestandteil dieser Dissertation zuvor für einen akademischen Abschluss oder eine andere Qualifikation an dieser oder einer anderen Institution verwendet wurde, wurde dies klar angezeigt;

3. Wenn immer andere eigene- oder Veröffentlichungen Dritter herangezogen wurden, wurden diese klar benannt;

4. Wenn aus anderen eigenen- oder Veröffentlichungen Dritter zitiert wurde, wurde stets die Quelle hierfür angegeben. Diese Dissertation ist vollständig meine eigene Arbeit, mit der Ausnahme solcher Zitate;

5. Alle wesentlichen Quellen von Unterstützung wurden benannt;

6. Wenn immer ein Teil dieser Dissertation auf der Zusammenarbeit mit anderen basiert, wurde von mir klar gekennzeichnet, was von anderen und was von mir selbst erarbeitet wurde;

7. Kein Teil dieser Arbeit wurde vor deren Einreichung veröffentlicht.

Aachen, den

_____

Manfred Peter Fackeldey