

# **Advancing methods for probing $\mathcal{CP}$ violation in Higgs couplings**

Von der Fakultät für Mathematik und Physik  
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades  
Doktor der Naturwissenschaften  
Dr. rer. nat.

genehmigte Dissertation von

**M.Sc. Marco Menen**

2025

Referentin: Prof. Dr. Elina Fuchs  
Korreferent: Prof. Dr. Tilman Plehn  
Korreferent: Priv.-Doz. Dr. Philip Bechtle

Tag der Promotion: 29.09.2025

Charge-parity ( $\mathcal{CP}$ ) violation in the couplings of the Higgs boson is a promising candidate for physics beyond the SM and could play a key role in explaining the observed asymmetry of matter and antimatter in the Universe. While probing  $\mathcal{CP}$  violation at the Large Hadron Collider (LHC) is a complex task, the continuously growing amounts of data provide new opportunities. Therefore, in this thesis, we propose techniques to optimally exploit the growing dataset and find potential improvements in sensitivity and interpretability compared to current approaches. In particular, we examine analysis strategies targeting  $\mathcal{CP}$  violation in the Higgs couplings to top quarks, gluons, and weak vector bosons. The gluon fusion channel in association with two jets (ggF2j) is investigated using machine learning techniques. We find that a dedicated signal region for this channel outperforms phase space regions with vector boson fusion (VBF)-like kinematics. This could potentially improve current experimental constraints on the  $\mathcal{CP}$  structure of an effective Higgs-gluon coupling and be competitive with other channels for constraining the top-Yukawa coupling. For a more direct probe, we discuss an extension of the Simplified Template Cross Sections (STXS) framework for top-associated Higgs production ( $t\bar{t}H$ ). We propose to extend the current binning in the Higgs transverse momentum by a second  $\mathcal{CP}$ -sensitive dimension. Three observables are identified as optimal candidates for this extension. Finally, we use symbolic regression (SR) to build analytical observables for the VBF and  $t\bar{t}H$  processes, which offer maximal interpretability. This is especially important for a fundamental symmetry such as  $\mathcal{CP}$ , which can only be unambiguously identified by  $\mathcal{CP}$ -odd observables. Using information at the detector level, we learn equations for  $\mathcal{CP}$ -odd as well as  $\mathcal{CP}$ -sensitive, but  $\mathcal{CP}$ -even, observables. They can be directly analyzed and used in an analysis, while outperforming other methods. We compare the sensitivity of our various techniques to traditional methods in the respective Higgs production and decay channels. Our results are relevant for studies of the  $\mathcal{CP}$  character of the Higgs boson using the full Run 2 and Run 3 data, but also for the high luminosity phase of the LHC.

*Keywords:* Higgs boson,  $\mathcal{CP}$ -violation, Machine Learning

Die Verletzung der Ladungskonjugations-Parität ( $\mathcal{CP}$ ) in den Kopplungen des Higgs-Bosons ist ein vielversprechender Kandidat für Physik jenseits des Standardmodells und könnte eine Schlüsselrolle bei der Erklärung der beobachteten Asymmetrie von Materie und Antimaterie im Universum spielen. Die Untersuchung der  $\mathcal{CP}$ -Verletzung am Large Hadron Collider (LHC) ist eine komplexe Aufgabe, doch die stetig wachsende Menge an Daten bietet gleichzeitig neue Möglichkeiten. Daher schlagen wir in dieser Arbeit Möglichkeiten vor, um den wachsenden Datensatz optimal zu nutzen und potenzielle Verbesserungen in Bezug auf Sensitivität und Interpretierbarkeit im Vergleich zu bisherigen Ansätzen zu finden. Insbesondere untersuchen wir Techniken für Analysen, die auf die Verletzung der  $\mathcal{CP}$ -Symmetrie in den Higgs-Kopplungen an Top-Quarks, Gluonen und schwache Vektorbosonen abzielen. Die Produktion eines Higgs durch Fusion zweier Gluonen mit zwei zusätzlichen Jets (ggF2j) wird mit Hilfe von Techniken des maschinellen Lernens untersucht. Wir stellen fest, dass ein speziell auf diesen Kanal zugeschnittener Phasenraumbereich bessere Ergebnisse liefert als Phasenraumbereiche für die Higgs-Produktion via Fusion zweier schwacher Vektorbosonen (VBF). Dies könnte möglicherweise die aktuellen experimentellen Beschränkungen der  $\mathcal{CP}$ -Struktur einer effektiven Higgs-Gluon-Kopplung verbessern und mit anderen Kanälen zur Beschränkung der Top-Yukawa-Kopplung konkurrieren. Um Letztere direkter zu testen, untersuchen wir eine Erweiterung der STXS-Struktur (Simplified Template Cross Sections) für die gleichzeitige Produktion eines Higgs-Bosons und zwei Top-Quarks ( $t\bar{t}H$ ). Wir schlagen vor, die derzeitige Einteilung in transverse Impulse des Higgs in eine zweite,  $\mathcal{CP}$ -sensitive Dimension zu erweitern. Drei Observablen werden als optimale Kandidaten für diese Erweiterung identifiziert. Schließlich verwenden wir symbolische Regression (SR), um analytische Observablen mit maximaler Interpretierbarkeit für die VBF- und  $t\bar{t}H$ -Prozesse zu konstruieren. Dies ist besonders wichtig für eine fundamentale Symmetrie wie  $\mathcal{CP}$ , die nur durch  $\mathcal{CP}$ -ungerade Observablen eindeutig identifiziert werden kann. Anhand von kinematischen Informationen auf Detektorebene lernen wir Gleichungen für  $\mathcal{CP}$ -ungerade sowie  $\mathcal{CP}$ -sensitive,  $\mathcal{CP}$ -gerade Observablen. Diese können einfach untersucht und unmittelbar in einer Analyse verwendet werden, wobei sie andere Methoden übertreffen. Wir vergleichen die Sensitivität unserer verschiedenen Techniken mit traditionellen Methoden in den jeweiligen Higgs-Produktions- und Zerfallskanälen. Unsere Ergebnisse sind relevant für Studien der  $\mathcal{CP}$ -Struktur des Higgs-Bosons bei Verwendung der vollständigen Daten aus Run 2 und Run 3, aber auch für die zukünftige "High-Luminosity"-Phase des LHC.

*Schlagwörter:* Higgs Boson,  $\mathcal{CP}$ -Verletzung, Maschinelles Lernen

---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The theory of particle physics</b>	<b>3</b>
2.1	The Standard Model of particle physics	3
2.1.1	The electroweak interaction	5
2.1.2	The strong interaction	10
2.1.3	The Higgs mechanism	11
2.1.4	CP violation in the Standard Model	14
2.2	Physics beyond the Standard Model	16
2.2.1	Shortcomings of the Standard Model	16
2.2.2	Parameterizing BSM physics	18
<b>3</b>	<b>Probing the Higgs sector</b>	<b>23</b>
3.1	The Large Hadron Collider	23
3.1.1	The ATLAS and CMS experiments	24
3.2	Event generation for phenomenology	25
3.3	Higgs physics at the LHC	29
3.3.1	Production and decay channels	29
3.3.2	Current experimental constraints	32
3.3.3	The Simplified Template Cross Sections framework	36
<b>4</b>	<b>Statistical learning in high energy physics</b>	<b>38</b>
4.1	Supervised learning	38
4.2	Boosted decision trees	40
4.2.1	XGBoost	42
4.3	Neural networks	43
4.4	Interpretability of machine learning models	46
4.4.1	Shapley values	46
4.4.2	Symbolic regression	47
4.4.3	Symbolic regression with PySR	47
4.4.4	Symbolic regression with SymbolNet	49
<b>5</b>	<b>Classifying the <math>\mathcal{CP}</math> properties of the <math>ggH</math> coupling in <math>H + 2j</math> production</b>	<b>54</b>
5.1	The topology and generation of $H + 2j$ events	55

5.2	Analysis strategy	58
5.2.1	Defining the signal regions	59
5.2.2	Classification of the ggF2j $\mathcal{CP}$ states	62
5.3	Sensitivity to the Higgs-gluon coupling	64
5.3.1	Constraints in the ggF2j-SR	64
5.3.2	Constraints in the VBF-SR	66
5.3.3	Combination	68
5.3.4	Relative importance of the observables	68
5.4	Impact of $\mathcal{CP}$ violation in the $HVV$ coupling	72
5.5	Evaluation in terms of the top-Yukawa coupling	73
5.6	Conclusions	75
<b>6</b>	<b><math>\mathcal{CP}</math>-sensitive simplified template cross-sections for <math>t\bar{t}H</math></b>	<b>76</b>
6.1	$\mathcal{CP}$ -sensitive observables in $t\bar{t}H$ events	77
6.1.1	The generation of $t\bar{t}H$ events	77
6.1.2	Observable definitions	78
6.1.3	Parton-level distributions	80
6.2	Evaluation of the observable performance	82
6.2.1	Detector effects and selection efficiency	83
6.2.2	Evaluation of the sensitivity	85
6.3	Sensitivity to the top-Yukawa coupling	87
6.3.1	Significance evaluation	87
6.3.2	Binning optimization	89
6.3.3	Comparison to a multivariate analysis	90
6.3.4	Limits on the $\mathcal{CP}$ state of the top-Yukawa coupling	94
6.4	Extending the STXS framework	96
6.4.1	Impact of the background distributions	96
6.4.2	Final $t\bar{t}H$ STXS extension proposal	99
6.5	Conclusions	100
<b>7</b>	<b>Symbolic regression for Higgs <math>\mathcal{CP}</math> analyses</b>	<b>102</b>
7.1	Optimal $\mathcal{CP}$ -odd observables	103
7.2	$\mathcal{CP}$ -odd observables in VBF	104
7.2.1	Event generation and training	104
7.2.2	Learned equations and sensitivities	106
7.3	$\mathcal{CP}$ -odd observables in $t\bar{t}H$	110
7.3.1	Event generation and training	110
7.3.2	Learned equations and sensitivities	111
7.4	Reconstruction of the Collins-Soper angle for $t\bar{t}H$	113
7.4.1	Event generation and training	113
7.4.2	Predicted equations for the Collins-Soper angle	116
7.4.3	Performance and sensitivity of the equations	118
7.5	Conclusions	120
<b>8</b>	<b>Conclusions</b>	<b>122</b>

<b>A Symmetry groups</b>	<b>124</b>
<b>B Details of the event generation</b>	<b>127</b>
<b>C Likelihood tests</b>	<b>130</b>
<b>D Observable distributions</b>	<b>133</b>
D.1 $ggF2j$ distributions	133
D.2 $t\bar{t}H$ distributions	136
<b>E Training uncertainty</b>	<b>139</b>
<b>F Significance tables</b>	<b>141</b>
<b>G Equations from PySR and SymbolNet</b>	<b>146</b>
<b>Bibliography</b>	<b>151</b>
<b>List of Figures</b>	<b>176</b>
<b>List of Tables</b>	<b>179</b>

# CHAPTER 1

---

## Introduction

---

Particle physics aims to provide an understanding of the most fundamental building blocks of the Universe. Over many decades of research, a description of these building blocks and their interactions has been developed, which is known today as the Standard Model of particle physics. Ultimately, it led to the discovery of a particle compatible with the predictions of the SM for the Higgs boson within the experimental and theoretical uncertainties [1, 2]. While this discovery happened in 2012, the Higgs boson was already predicted 50 years earlier [3–6]. Now, the SM is considered the most precisely tested theory of nature. The experimental measurements of the last decades agree well with the SM predictions, with only sparse exceptions (e.g. [7–10]).

Despite its great success, the SM is not a full theory of nature and has only been tested directly up to energy scales of a few TeV. Consequently, it is unable to provide answers to many fundamental questions. One of these is the overabundance of matter over antimatter in the visible Universe, without which life would not be possible. While this so-called BAU can be measured precisely from the cosmic microwave background and the abundance of light elements [11, 12], its origin remains unknown. The necessary Sakharov conditions for producing it demand a first-order phase transition, baryon number violation, and sizable amounts of  $C$  and  $CP$  violation [13], which can be realized by baryogenesis or leptogenesis. However, the SM cannot explain the BAU as it fails to fulfill these conditions sufficiently. Especially, the amount of  $CP$  violation in the SM is too low by several orders of magnitude [14, 15]. Therefore, new sources of  $CP$  violation beyond the SM are needed.

The Higgs sector is an interesting candidate for looking for beyond the SM (BSM) physics for several reasons. As the latest discovered particle of the SM, only a few of its properties and interactions with other particles are determined with high precision. The partially large uncertainties leave room for effects that have so far avoided detection. In electroweak baryogenesis [16, 17], the Higgs sector is directly related to the generation of the matter-antimatter asymmetry via the electroweak phase transition. A BSM Higgs sector can therefore help fulfill the Sakharov conditions by providing a first-order phase transition or additional sources of  $CP$  violation (see e.g. [18–29]).

In the SM, all Higgs boson interactions are predicted to be  $CP$ -even. However, higher-dimensional operators in effective theories are not constrained to be  $CP$ -conserving (see e.g. [30]). Furthermore, many extended Higgs models or UV completions naturally introduce  $CP$  violation in the Higgs boson interactions (see e.g. [31, 32]). Therefore, while a pure  $CP$ -odd state of the Higgs boson has been ruled

out experimentally [33–35], it could be a mixed  $\mathcal{CP}$  state. In this case,  $\mathcal{CP}$  violation can appear at tree level in the Yukawa couplings. This makes the top-Yukawa coupling an ideal candidate for probes of  $\mathcal{CP}$  violation in the Higgs sector, as it has the largest Yukawa coupling and is easily accessible at the LHC. Closely related is the Higgs coupling to gluons, which is dominated by a top quark loop in the SM and can therefore be used to place indirect constraints. Finally, while the  $\mathcal{CP}$  state of the Higgs couplings to the massive vector bosons is suppressed in many BSM models, it can be probed very precisely experimentally [36–39].

In many cases, the most stringent constraints on the amount of  $\mathcal{CP}$  violation present in the Higgs boson couplings are imposed by the current bounds on the electron EDM [40] and the neutron EDM [41]. Especially, the electron EDM constrains the  $\mathcal{CP}$ -violating part of the top-Yukawa coupling to  $\mathcal{O}(10^{-3})$  [25, 42, 43]. However, these constraints rely on SM-like couplings of the Higgs boson to the first-generation fermions, which themselves are only very weakly constrained [44–55]. In case of an already small deviation, the constraints from the electron EDM may be lifted [27]. The LHC therefore offers the only way of unambiguously searching for  $\mathcal{CP}$  violation in individual Higgs couplings.

Performing  $\mathcal{CP}$  analyses at the LHC is a highly complex task, as many effects can wash out the sensitivity of the experiment. These include background events with a similar signature to that of the signal process, parton showering and hadronization, limited detector resolution, or uncertainties in the theory prediction. Especially, polarization information is often unavailable, either due to the hadronization of strongly interacting particles or due to the insufficient statistics and detector resolution in the case of photons [56–58]. It is therefore key to design the analyses optimally in terms of signal extraction, reconstruction of the parton-level objects, and measuring the quantities that are most sensitive to the  $\mathcal{CP}$  state of the coupling under examination.

Several methods have been developed over the last years to aid with these challenges. The Simplified Template Cross Section (STXS) framework provides a way to combine channels and results across experiments to isolate BSM effects while keeping the theoretical uncertainties small [59, 60]. Furthermore, many techniques employing machine learning have been developed in recent years, tailored to the challenges met in LHC analyses. They offer higher sensitivity towards BSM effects as well as improved efficiency when handling large datasets, but also rare processes with few signal events. This thesis determines how methods of probing  $\mathcal{CP}$ -violating Higgs couplings can be advanced and used to improve experimental analyses at the LHC by comparing them to techniques that have been used in the past. For this, the gluon fusion process in association with two jets (ggF2j), top-associated Higgs production ( $t\bar{t}H$ ), and vector boson fusion (VBF) are examined. A special focus is put on achieving optimal sensitivities while keeping the results highly interpretable.

The thesis begins with a brief introduction to the SM and discusses some of its shortcomings, as well as its extensions, with a focus on the Higgs sector in Chapter 2. An overview of experimental techniques and recent results is provided in Chapter 3. Next, Chapter 4 introduces methods of statistical learning, especially boosted decision trees and neural networks. The following chapters discuss the results. Chapter 5 probes the constraints on the effective Higgs gluon coupling from ggF2j production using multivariate techniques and interprets it in terms of the top-Yukawa coupling. Then, Chapter 6 investigates a  $\mathcal{CP}$ -sensitive extension of the STXS framework for the  $t\bar{t}H$  process. Analytical expressions for  $\mathcal{CP}$  sensitive and optimal observables are obtained and analyzed in Chapter 7 by employing two methods of symbolic regression. Finally, Chapter 8 concludes the findings of this thesis and provides an outlook for future studies.

# CHAPTER 2

---

## The theory of particle physics

---

Particle physics aims to describe the fundamental interactions of the Universe on the smallest and largest scales. It provides mechanisms that possibly describe the dynamics of the very early universe as well as its evolution to the present day. The ultimate goal of particle physics is to find nothing less than a "theory of everything" – an accurate description of all fundamental forces valid at any energy scale.

The exact time and place of birth of particle physics is subject to debate. One could argue that it began with the discovery of the first fundamental particle, namely the electron, by J. J. Thompson in 1897 [61]. In the years following this breakthrough, great progress was made in the description of the photon as a point-like particle and its interaction with electrons, the two most notable works being the discovery of quantum mechanics [62] and the explanation of the photoelectric effect [63]. Another two decades later, Paul Dirac published a dynamic and relativistic description of *quantum electrodynamics* (QED) [64], which is nowadays known as *quantum field theory* (QFT).

This chapter is structured as follows. First, a detailed introduction to the *Standard Model of particle physics* (SM) is given in Section 2.1. This includes a discussion of the *electroweak* (EW) theory, *quantum chromodynamics* (QCD), a description of the Higgs mechanism and the Higgs sector in general, and a discussion of  $\mathcal{CP}$  violation in the SM. Afterwards, physics *beyond the SM* (BSM) is introduced in Section 2.2 by discussing shortcomings of the SM as well as possible extensions that can provide explanations for said shortcomings.

### 2.1 The Standard Model of particle physics

The SM encapsulates our current knowledge of particle physics and is backed by a huge amount of experimental confirmation. It is a QFT that successfully describes three of the four known fundamental forces – the electromagnetic, weak, and strong forces – with only gravity missing, which is instead described by general relativity. Historically, the SM emerged as a combination of findings, the first being the Yang-Mills theory [65], which describes particle interactions as a gauge theory with a non-abelian Lie algebra. This led to the foundation for both the EW unification as well as QCD. The former was completed by several works describing spontaneous symmetry breaking in 1964 [4–6], the unified description of electromagnetism and the weak force in 1967 [66–68], and finally the proof that such a theory is renormalizable in 1971 [69]. QCD, on the other hand, owes its success to the invention

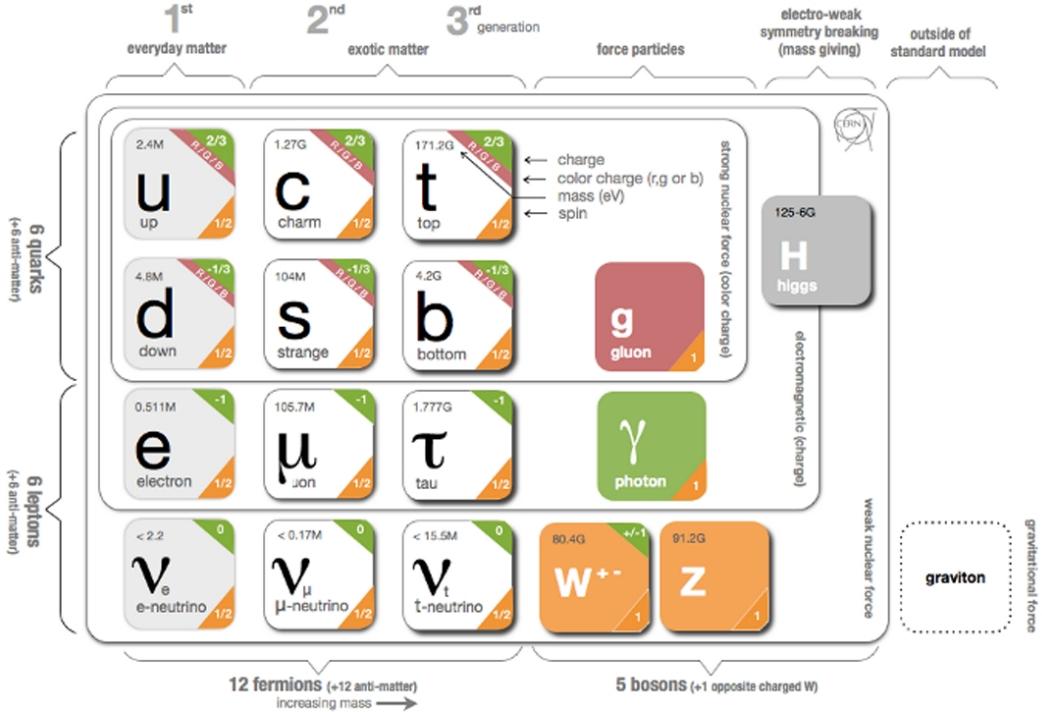


Figure 2.1: Particle content of the SM. Figure taken from [75].

of the first quark model in 1961 [70, 71] and the discovery of asymptotic freedom in 1973 [72, 73]. The SM combines both theories and is based on a  $SU(3) \times SU(2) \times U(1)$  gauge symmetry group (see Appendix A for a detailed discussion).<sup>1</sup>

The particle content of the SM is shown in Fig. 2.1. It can be split into two groups of fundamental particles: Fermions, which are spin-1/2 particles and the building blocks of matter and antimatter, and bosons, which are integer-spin particles propagating the fundamental interactions between fermions (with the Higgs boson as a special case). The fermions can be further split up into leptons and quarks. Both come in three generations, and the leptons and quarks of a specific generation are placed in doublets for the electroweak interaction, as detailed below. The fermions differ in their mass, which increases with generation, and particles of each generation are generally allowed to mix.

Neutrinos are leptons that are not electrically charged and therefore only participate in the weak interaction, mediated by the  $W^\pm$  and  $Z$  bosons. They have very small but non-zero masses, which is a direct consequence of the observed neutrino oscillations. The origin of the neutrino masses is still unclear at this time. If the neutrinos are Dirac particles, their mass could originate from the known SM Higgs mechanism. If instead they are Majorana particles, a BSM source is needed. The remaining leptons are the electron, muon, and  $\tau$ , all of which are electrically charged and therefore interact both weakly and electromagnetically. Their masses are several magnitudes higher than those of the neutrinos, and the reason for this discrepancy is also currently unclear.

Just like leptons, the quarks can be grouped into two categories, up- and down-type quarks, that exist

<sup>1</sup> Technically, due to the particle content of the SM and the respective quantum numbers of the particles, the  $SU(3) \times SU(2) \times U(1)/Z_6$  subgroup is sufficient. This can be important, for example, in grand unified theories [74].

in three generations. Quarks carry both color and electric charge and are the only fermions that can interact strongly via the exchange of gluons, as well as weakly and electromagnetically. While the up and down quarks are the building blocks of the nucleons, the top quark holds a special role when it comes to physics involving the Higgs boson. With a mass of about 172 GeV, it is not only the heaviest quark but also the overall heaviest fundamental particle known.

The behavior and interactions of all particles in the SM are described by Lagrangian densities  $\mathcal{L}(\phi, \partial_\mu \phi)$  (also just called Lagrangians) with the corresponding action  $S = \int d^4x \mathcal{L}$ . From the principle of least action  $\delta S = 0$ , one can derive the Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial \phi} = \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \quad (2.1)$$

which result in the equations of motion for the fields  $\phi$  in the given theory.

Lagrangians need to respect the symmetries of the respective theory by being invariant under transformations of the symmetry group (a short introduction is provided in Appendix A). The symmetries in the SM can be grouped into global and local ones. Global symmetries that affect spacetime are defined via the invariance under Poincaré transformations, which is required by special relativity. The Poincaré transformations consist of translations and Lorentz transformations and lead to the conservation of energy, linear momentum, and angular momentum via the Noether theorem. Lorentz invariance is ensured by building a Lagrangian only from Lorentz scalars, while translational invariance is trivial for the fields of the SM. Global internal symmetries of the SM are, e.g., the baryon number and lepton number, which are said to be accidental symmetries.

Local internal symmetries in the SM are dubbed gauge symmetries and describe the invariance of the Lagrangian under transformations of the fields. This ensures that the theory is independent of the point of spacetime. In the SM, gauge invariance is achieved by promoting global to local symmetries. Starting from a Lagrangian describing free fermions, the need for gauge invariance leads to the introduction of the gauge bosons, as shown below.

Finally, it is possible to spontaneously break internal symmetries. In contrast to explicit symmetry breaking, where terms of the Lagrangian violate the invariance, *spontaneous symmetry breaking* (SSB) happens when the potential has a symmetry but the vacuum of the theory does not correspond to its origin. Although the vacuum will have a degeneracy with the same symmetry as the potential, choosing a specific vacuum (gauge) will break the symmetry. This concept is necessary for the completion of the SM via the Higgs mechanism and will be discussed further in Section 2.1.3.

### 2.1.1 The electroweak interaction

The electroweak theory unifies electromagnetism and the weak interaction at energy scales  $E \gtrsim 80$  GeV corresponding to the mass of the  $W$  boson. At energies below the mass of the heavy gauge bosons, quantum electrodynamics (QED) appears as a standalone force resulting from the interaction of electrically charged particles with photons. To derive the QED Lagrangian, only two ingredients are needed: the Lagrangian for a free Dirac particle  $\Psi$

$$\mathcal{L}_{\text{Dirac}} = \bar{\Psi} \left( i\gamma^\mu \partial_\mu - m \right) \Psi \quad (2.2)$$

and the internal  $U(1)$  symmetry (see Appendix A). While Eq. (2.2) is invariant under a global  $U(1)$  transformation of  $\Psi(x) \rightarrow e^{iQa}\Psi(x)$ , where  $Q$  and  $a$  are constants, it is not invariant under  $\Psi(x) \rightarrow e^{iQa(x)}\Psi(x)$  where  $a(x)$  is some spacetime dependent function. To ensure gauge invariance, the derivative  $\partial_\mu$  is replaced by the covariant derivative  $\partial_\mu \rightarrow D_\mu = \partial_\mu + ieQA_\mu$  with the newly introduced vector boson  $A_\mu$ . The latter transforms as  $A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu a(x)$ . Substituting the covariant derivative into Eq. (2.2) leads to an interaction term of the form

$$\mathcal{L}_{\text{int}} = -eQ\bar{\Psi}\gamma^\mu A_\mu \Psi \quad (2.3)$$

where  $e$  specifies the coupling strength of the interaction and is an intrinsic property of  $A_\mu$ , while  $Q$  is a quantum number of  $\Psi$  and identified as the electric charge of a particle. Finally, the properties of the newly introduced gauge boson have to be described. While a mass term  $m_A^2 A_\mu A^\mu$  breaks the gauge invariance, a dynamic term of the form

$$\mathcal{L}_A = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.4)$$

with

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (2.5)$$

can be added without violating the imposed gauge invariance. Such a term is needed, as otherwise the Euler-Lagrange equations of  $A_\mu$  would require  $eQ = 0$  and therefore no interaction would be possible. The full QED Lagrangian then reads

$$\mathcal{L}_{\text{QED}} = \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{int}} + \mathcal{L}_A = \bar{\Psi} \left( i\gamma^\mu D_\mu - m \right) \Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} . \quad (2.6)$$

With the interaction term for the gauge boson, the Euler-Lagrange equation for  $A_\mu$  now results in

$$\partial_\mu F^{\mu\nu} = eQ\bar{\Psi}\gamma^\nu\Psi \quad (2.7)$$

which is the four-vector representation of the Maxwell equations. It is important to note, however, that the quantities appearing in Eq. (2.6) are not necessarily the physical objects that can be measured in an experiment. This is especially important for the parameter  $e$ , which is identified with the electron charge. Higher order corrections to the interaction between a fermion  $\Psi$  and a photon  $A_\mu$  lead to the creation of virtual electron-positron pairs. This is referred to as vacuum polarization and will screen the bare electron charge, with the strength of the screening being dependent on the energy scale. The scale dependence of the electron charge can be calculated from the modification of the photon propagator after renormalization. It follows the renormalization group equation

$$\beta\alpha_{\text{em}} = q^2 \frac{\partial\alpha_{\text{em}}}{\partial q^2} \quad (2.8)$$

with the beta function ( $\alpha_{\text{em}}$ ) and the electromagnetic coupling constant  $\alpha_{\text{em}} = e_{\text{phys}}^2/4\pi$ . At leading order, the resulting energy dependence is

$$e_{\text{phys}}^2(q^2) = \frac{e_{\text{phys}}^2(\mu^2)}{1 + \beta_0 \frac{e_{\text{phys}}^2(\mu^2)}{4\pi} \ln \frac{q^2}{\mu^2}}, \quad \alpha_{\text{em}}(q^2) = \frac{\alpha_{\text{em}}(\mu^2)}{1 + \beta_0 \alpha_{\text{em}}(\mu^2) \ln \frac{q^2}{\mu^2}} \quad (2.9)$$

with  $\beta_0 = -\frac{1}{3\pi}$ . Evaluating Eq. (2.9) at a specific energy scale yields a physical charge. For example, the value of  $\alpha_{\text{em}}(\mu^2 \lesssim m_e^2) = 1/137$  is well known and can therefore be used to predict the renormalized charge at any energy. The scale dependence of the QED and other couplings is referred to as *running of the coupling*.

The weak interaction appears as a four-point interaction at energy scales below the mass of the  $W$  boson. It was first fully predicted in 1934 [76] to describe radioactive  $\beta$ -decays via the introduction of the neutrino. The original form of the interaction between neutron, proton, electron, and neutrino was predicted to have a pure vector-like structure in correspondence with the interaction in QED. However, with the inclusion of parity violation as measured in 1956 [77], the only possibility for the structure of the weak interaction was a combination of two terms: One that transformed as a vector and the other as an axial-vector (V-A) under parity transformations [78–80]. Therefore, an interaction of the form

$$\mathcal{L}_{\text{int}} = \frac{G_F}{\sqrt{2}} [\bar{p}\gamma_\mu(1 - g_A\gamma_5)n] [\bar{e}\gamma_\mu(1 - \gamma_5)v] \quad (2.10)$$

was obtained with the Fermi constant  $G_F$  and a constant  $g_A \approx 1$ .

Nowadays, it is known that the weak interaction is not a four-point interaction but occurs via the exchange of gauge bosons. It follows an internal  $SU(2)$  gauge symmetry, and particles that participate in the weak interaction are therefore placed in  $SU(2)$  doublets. The  $1 - \gamma_5$  vector-axial-vector structure of the weak interaction corresponds to the projection operator for left-handedness, and consequently, only left-handed particles participate in it. A combined transformation of charge-conjugation and parity ( $C\mathcal{P}$ ) preserves this structure so that right-handed antiparticles also interact weakly. Right-handed (left-handed) (anti)particles do not participate in the weak interaction and are placed in  $SU(2)$  singlets. With the particle content of the SM, there are a total of six  $SU(2)$  doublets and nine  $SU(2)$  singlets. They read

<u>Particle content</u>	<u>Charged under</u>
$\Psi_L = \left\{ \begin{bmatrix} \nu_e \\ e \end{bmatrix}_L, \begin{bmatrix} \nu_\mu \\ \mu \end{bmatrix}_L, \begin{bmatrix} \nu_\tau \\ \tau \end{bmatrix}_L, \begin{bmatrix} u \\ d \end{bmatrix}_L, \begin{bmatrix} c \\ s \end{bmatrix}_L, \begin{bmatrix} t \\ b \end{bmatrix}_L \right\}$	$SU(2)_L \times U(1)_Y$
$\Psi_R = \{e_R, \mu_R, \tau_R, u_R, d_R, c_R, s_R, t_R, b_R\}$	$U(1)_Y$

where the subscripts  $L$  and  $R$  label the respective chirality, and the symmetry groups are needed for the electroweak unification.

The weak interaction is accompanied by a new quantum number, *weak isospin*  $T_3$ , which serves a role

similar to the electric charge in QED.  $SU(2)$  doublets consist of two particles where the particle with  $T_3 = +(-)1/2$  is placed in the upper (lower) part of the doublet.  $SU(2)$  singlets have  $T_3 = 0$ . The  $SU(2)$  gauge group of the weak interaction gives rise to three generators, which in this case form a weak isospin triplet. This means there are three particles mediating this interaction: Two that are charged ( $W^1$  and  $W^2$  with  $T_3 = \pm 1$ , not the physical  $W$  bosons) and give rise to *charged currents* and one that is not charged ( $W^3$  with  $T_3 = 0$ ) and gives rise to *neutral currents*. The exchange of these bosons is only allowed for particles with  $T_3 \neq 0$ .

The Lagrangian of the weak interaction is derived similarly to that of QED. Starting from the Lagrangian for free fermions, the global gauge symmetry is promoted to a local one by replacing the derivative with the covariant derivative

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig J^i W_\mu^i \quad (2.11)$$

where  $g$  is the weak coupling constant and the generators  $J^i$  follow Eq. (A.5). The transformation behavior of the three newly introduced bosons is fixed by requiring gauge invariance

$$W_\mu^i \rightarrow W_\mu^{i'} = W_\mu^i - \partial_\mu \alpha^i(x) - g \epsilon^{ijk} \alpha^j(x) W_\mu^k \quad . \quad (2.12)$$

Analogous to QED, a kinematic term for the bosons is needed for a non-zero interaction, which reads

$$\mathcal{L}_W = -\frac{1}{4} W^{i,\mu\nu} W_{\mu\nu}^i \quad (2.13)$$

with

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g \epsilon^{ijk} W_\mu^j W_\nu^k \quad . \quad (2.14)$$

The additional term in Eq. (2.14) with respect to Eq. (2.5) leads to three- and four-point interactions of the  $W$  bosons.

At and above the energy scale of the physical  $W$  and  $Z$  boson masses, the electromagnetic and weak interactions appear as one singular force. Their combined description is therefore dubbed the *electroweak (EW) unification*. Such a unified theory is needed to explain the observed difference between the  $Z$  and  $W$  boson masses. Within the electroweak interaction, the  $U(1)$  gauge symmetry of QED is modified to conserve *hypercharge*

$$Y = 2(Q - T_3) \quad (2.15)$$

where the relation between the three conserved quantities is required to restore the photons' properties at low energies. As in QED, from the new  $U(1)_Y$  symmetry emerges one gauge boson  $B_\mu$  with properties similar to those of the photon. Only the coupling to fermions is modified

$$\begin{aligned} U(1)_{\text{EM}} &\longrightarrow U(1)_Y \\ eQ &\longrightarrow g' \frac{Y}{2} \end{aligned}$$

with the new coupling strength  $g'$ . Left- and right-handed fermions have different, but non-zero couplings to  $B_\mu$  as can be seen from Eq. (2.15).

Writing out the covariant derivative of the unified EW theory explicitly yields

$$D_\mu = \partial_\mu + i \frac{g}{2} \begin{pmatrix} W^3 & W^1 - iW^2 \\ W^1 + iW^2 & -W^3 \end{pmatrix}_\mu + ig' \frac{Y}{2} \begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix}_\mu. \quad (2.16)$$

The off-diagonal elements can be defined as

$$W^\pm = \frac{1}{\sqrt{2}} (W^1 \mp iW^2) \quad (2.17)$$

where  $W^\pm$  are now the physical  $W$  bosons. The exchange of a  $W$  boson corresponds to a rotation in the weak isospin space and transforms the upper particle of an  $SU(2)$  doublet into the lower one or vice versa. The diagonal elements must be manipulated so that one of the bosons couples to fermions like the QED photon. This can be achieved by a rotation

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} \quad (2.18)$$

with the *weak mixing angle*  $\theta_W$  which therefore has to fulfill

$$eQ = \sin \theta_W g T_3 + \cos \theta_W g' Y. \quad (2.19)$$

This subsequently also fixes the coupling strength between the  $Z$  boson and the fermions

$$g_Z = \cos \theta_W g T_3 - \sin \theta_W g' Y \quad (2.20)$$

which allows a coupling to right-handed fermions in contrast to the  $W$  boson. In total, the EW Lagrangian with the unbroken  $SU(2) \times U(1)$  symmetry reads

$$\mathcal{L}_{\text{EW}} = \bar{\Psi} i \gamma^\mu \partial_\mu \Psi - \bar{\Psi}_L \gamma^\mu \frac{g}{\sqrt{2}} \begin{pmatrix} 0 & W^+ \\ W^- & 0 \end{pmatrix}_\mu \Psi_L - \bar{\Psi} \gamma^\mu (eQ A_\mu + g_Z Z_\mu) \Psi \quad (2.21)$$

where  $\Psi = \Psi_L + \Psi_R$ .

There are several problems with this theory regarding the masses of both fermions and bosons. Eq. (2.21) contains no term of the form  $m \bar{\Psi} \Psi$ . This is because the different transformation behaviors of left- and right-handed particles explicitly break the gauge invariance of the electroweak theory if any of the fermions are massive. Furthermore, giving mass to the bosons by just adding a mass term to the Lagrangian breaks gauge symmetry. Both are in contrast to experimental observations, and unlike in QCD, there is no confinement mechanism to explain the short range of the interaction. To resolve these problems, the different symmetries must be (spontaneously) broken to a common symmetry, which is the idea of the Higgs mechanism discussed in Section 2.1.3. The Higgs mechanism also prevents the cross section of the 4-point interaction from diverging at high energies for longitudinally polarized gauge bosons [81].

### 2.1.2 The strong interaction

The last fundamental force in the SM is the strong interaction, which is described by QCD. Historically, while it was clear that a strong interaction must exist to explain the stability of atomic nuclei, it was first attributed to the exchange of pions between protons and neutrons [82]. At the beginning of the 1960s, the first models describing quarks as constituents of nucleons were proposed [70, 71]. In the same decade, evidence for the existence of a substructure of protons was found, proving the existence of quarks [83]. Furthermore, the existence of particles like the  $\Delta^{++}$  baryon with quark-substructure  $uuu$  gave hints towards the color charge of quarks to avoid violating Pauli's exclusion principle.

Similar to QED, the strong interaction emerges when the QCD lagrangian is required to be invariant under  $SU(3)$ , where the participating fermions live in a 3-dimensional color space. Starting again from the Lagrangian for a free Dirac spinor, the invariance under  $SU(3)$  leads to a covariant derivative

$$D_\mu = \partial_\mu + ig_s T^a G_\mu^a \quad (2.22)$$

with the Gell-Mann matrices  $\lambda^a = 2 T^a$  and the strong coupling constant  $g_s$ . There are eight generators  $G_\mu^a$  that are identified with the gluons. These gauge bosons interact with quarks by changing their color and therefore act as rotations in color space. The kinetic (Yang-Mills) term for the gluons is

$$\mathcal{L}_G = -\frac{1}{4} G^{a,\mu\nu} G_{\mu\nu}^a \quad (2.23)$$

with

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c . \quad (2.24)$$

The structure constants  $f^{abc}$  are fully determined by the  $su(3)$  algebra.

Eq. (2.23) gives rise to cubic and quartic couplings of gluons. While self-interactions also exist in the EW theory, the gauge bosons involved in these interactions acquire a mass via the Higgs mechanism, leading to the short-range nature of the weak interaction. There is, however, no evidence for gluons having a non-zero mass. Instead, the reason for the limited range of the strong interaction comes from the running of  $g_s$ . In contrast to QED, where virtual electron-positron pairs screen the bare electron charge, the self-interaction of gluons dominates the contribution of virtual quark-antiquark pairs and leads to an anti-screening of the color charge. As in Eq. (2.9), the running of the coupling is described at leading order via

$$g_s^2(q^2) = \frac{g_s^2(\mu^2)}{1 + \beta_0 \frac{g_s^2(\mu^2)}{4\pi} \ln \frac{q^2}{\mu^2}} , \quad \alpha_s(q^2) = \frac{\alpha_s(\mu^2)}{1 + \beta_0 \alpha_s(\mu^2) \ln \frac{q^2}{\mu^2}} \quad (2.25)$$

but with  $\beta_0 = \frac{21}{12\pi}$  for the six known quark flavors. The positive sign of  $\beta_0$  makes  $\alpha_s$  increase with lower  $q^2$  which results in a Landau pole near the so-called QCD scale  $\mu = \Lambda_{QCD} \sim 200$  MeV.<sup>2</sup> This is known as *infrared slavery*. The QCD scale corresponds to the radius of a proton when converted to a distance. For distances larger than that, the creation of a quark-antiquark pair is energetically favored over a long-range flux tube. Individual gluons and quarks can therefore not be resolved beyond this distance, which is known as *color confinement*.

<sup>2</sup> The exact value of  $\Lambda_{QCD}$  depends on the renormalization scheme.

### 2.1.3 The Higgs mechanism

While the EW theory manages to unify two fundamental forces of nature, it fails to explain the origin of the (different) masses of its interacting fermions and bosons. This problem can be fixed by spontaneously breaking the EW symmetry from which a new physical particle emerges, which is known as the Brout-Englert-Higgs (BEH) mechanism [3–6]. The new particle is therefore required to couple to fermions and weak bosons to generate mass terms, while leaving the QED boson massless, and acquire a non-zero *vacuum expectation value* (vev) that breaks the symmetry. To fulfill all three criteria, the new particle must transform non-trivially under  $SU(2) \times U(1)$ , have hypercharge  $Y = 1$ , and must be a scalar not to break Lorentz invariance. Therefore, the EW Lagrangian is extended by

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger (D^\mu \Phi) - \underbrace{\left( \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \right)}_{V(\Phi)} \quad (2.26)$$

with the covariant derivative  $D_\mu$  of the EW interaction (see Eq. (2.16)), and the complex scalar doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (2.27)$$

consisting of real scalar fields  $\phi_{1,\dots,4}$ . The Higgs potential  $V(\Phi)$  contains all terms allowed by gauge invariance and renormalizability and introduces two new parameters  $\mu^2$  and  $\lambda$ . For vacuum stability,  $\lambda > 0$  is required while  $\mu^2 < 0$  is needed to obtain a non-zero vev. This form of the potential is known as the "Mexican hat potential" and is depicted in Fig. 2.2. In this case the symmetric state at  $\Phi^\dagger \Phi = 0$  is unstable and instead

$$\Phi^\dagger \Phi = v^2 = \frac{-\mu^2}{\lambda} \quad (2.28)$$

minimizes the potential with the vev  $v$ .

From a gauge symmetry point of view, there is a freedom in choosing the  $\phi_i$  from Eq. (2.27) so that Eq. (2.28) is fulfilled. However, to identify the physical degrees of freedom, the *unitary gauge* is the most helpful.<sup>3</sup> As a first step, the ground state is fixed to

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (2.29)$$

with  $\phi_1 = \phi_2 = \phi_4 = 0$ , from which it is easy to see that the QED charge is preserved, as only the  $Q = 0$  component of the Higgs doublet acquires a non-zero vev. The three (for this particular ground state) zero-valued fields correspond to massless Nambu-Goldstone bosons [84–86]. These bosons are unphysical since they vanish in the unitary gauge, which is a consequence of the locality of the broken gauge symmetry. However, the degrees of freedom  $\phi_i$  become related to the degrees of freedom of physical bosons via the gauge invariance of the potential. More precisely, these are longitudinal degrees of freedom for the  $W^\pm$  and  $Z$  bosons arising due to their non-zero mass.

<sup>3</sup> Other gauges (e.g.  $R_\xi$ ) introduce additional Goldstone interactions and propagators.

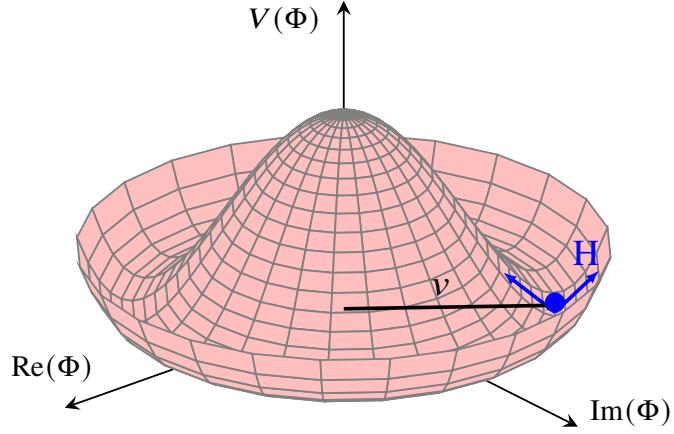


Figure 2.2: Schematic of the so-called Mexican hat potential after EWSB. Both the vev as the ground state, as well as the Higgs boson as its perturbation, are shown.

In contrast to a movement along the degenerate minima leading to unphysical Nambu-Goldstone bosons, a movement transverse to them corresponds to an excitation. This can be parameterized by a spacetime-dependent perturbation of the vacuum state

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (2.30)$$

where  $H(x)$  is an operator that can create (and annihilate) a physical scalar boson, the Higgs boson  $H$ .

### Higgs dynamics

Inserting Eq. (2.30) into the Higgs potential  $V(\Phi)$  of Eq. (2.26) yields

$$V(\Phi) = \frac{1}{4} \mu^2 v^2 - \mu^2 H^2 + \lambda v H^3 + \frac{1}{4} \lambda H^4 \quad (2.31)$$

with the terms linear in  $H$  canceling out. The constant term in Eq. (2.31) does not affect the dynamics of the Higgs field and therefore cannot be observed. The term quadratic in  $H$  is responsible for the mass of the Higgs boson

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2} \quad (2.32)$$

which is generated by the vev. Due to the two free parameters in the Higgs potential,  $m_H$  is not predicted by the BEH mechanism and has to be measured. The terms cubic and quartic in  $H$  fix the strength of the Higgs self-interactions (3-point and 4-point, respectively).

### Higgs-gauge interactions

The masses of the gauge bosons and their interactions with the Higgs are generated by the first part of Eq. (2.26). Using Eq. (2.16), this becomes

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \frac{1}{2} (\partial_\mu H) (\partial^\mu H) + \frac{g^2}{4} (v + H)^2 W_\mu^- W^{+\mu} + \frac{1}{8} (v + H)^2 (-g W_\mu^3 + g' B_\mu) (-g W^{3,\mu} + g' B^\mu) \quad (2.33)$$

in the unitary gauge, where the definitions for  $W^\pm$  have already been inserted. As already mentioned in Section 2.1.1, the  $W_\mu^3$  and  $B_\mu$  can be rotated to the mass eigenstates  $Z_\mu$  and  $A_\mu$ . This fixes the weak mixing angle to

$$\cos \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}, \quad \sin \theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \quad . \quad (2.34)$$

Therefore, all that is left is

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \frac{1}{2} (\partial_\mu H) (\partial^\mu H) + \frac{1}{4} (v + H)^2 \left( g^2 W_\mu^- W^{+\mu} + (g^2 + g'^2) Z_\mu Z^\mu \right) \quad (2.35)$$

where the first term describes the Higgs field dynamics and the other terms describe mass terms for the  $W^\pm$  and  $Z$  bosons, as well as their interactions with the Higgs boson. Importantly, there is no term of the form  $A_\mu A^\mu$ , which means that the theory successfully predicts a massless boson that couples to fermions with the coupling  $eQ$ . It also means that there is no tree-level coupling between the Higgs boson and the photon.

Eq. (2.35) predicts the masses of the heavy gauge bosons to be

$$m_W = \frac{1}{2} g v, \quad m_Z = \frac{1}{2} \sqrt{g^2 + g'^2} v, \quad \frac{m_W}{m_Z} = \cos \theta_W \quad . \quad (2.36)$$

It also predicts three-point ( $HWW$ ,  $HZZ$ ) and four-point ( $HHWW$ ,  $HHZZ$ ) interactions. The former read

$$\mathcal{L}_{HVV, \text{int}} = H \left( 2 \frac{m_W^2}{v} W_\mu^- W^{+\mu} + \frac{m_Z^2}{v} Z_\mu Z^\mu \right) \quad (2.37)$$

where the factor 2 difference comes from the different degrees of freedom. As Eq. (2.37) shows, the  $HVV$  couplings are predicted to be proportional to the square of the massive gauge bosons. Additionally, the vev is related to the Fermi constant via

$$\frac{1}{v^2} = \frac{g^2}{4m_W^2} = \sqrt{2} G_F \Rightarrow v \approx 246 \text{ GeV} \quad (2.38)$$

which is also dubbed the EW scale.

### Higgs-fermion interactions

In Eq. (2.21), a mass term for fermions was not allowed by gauge symmetry. However, this changes with the introduction of the Higgs doublet. All terms

$$\mathcal{L}_{\text{Yuk}} = \sum_{i,j=1}^3 \left( C_{ij}^u \overline{Q}_{i,L} \Phi_C u_{j,R} + C_{ij}^d \overline{Q}_{i,L} \Phi d_{j,R} + C_{ij}^l \overline{L}_{i,L} \Phi l_{j,R} + h.c. \right) \quad (2.39)$$

are singlets under  $SU(2)$  and therefore respect the gauge symmetry. This is the so-called Yukawa Lagrangian. The sum indicates the three fermion generations,  $Q$  and  $L$  are the quark and lepton doublets, and the  $C^f$  are complex matrices. To create masses for the up-type quarks, the charge conjugate

$$\Phi_C = -i\sigma_2 \Phi^* \quad (2.40)$$

is used, which has the same transformation behavior as  $\Phi$ . Inserting Eq. (2.30) into Eq. (2.39) leads to

$$\mathcal{L}_{\text{Yuk}} = \frac{1}{\sqrt{2}} (v + H) \sum_{i,j=1}^3 \left( C_{ij}^u \overline{u}'_{i,L} u'_{j,R} + C_{ij}^d \overline{d}'_{i,L} d'_{j,R} + C_{ij}^l \overline{l}'_{i,L} l'_{j,R} + h.c. \right) \quad (2.41)$$

where the prime indicates that the quarks and leptons are in their weak eigenstate. To obtain mass terms for the quarks and leptons, the matrices  $C_{ij}$  have to be diagonalized. This is achieved via transformations

$$V_{f_L}^\dagger C^f V_{f_R} = \text{diag}(y_{f_1}, y_{f_2}, y_{f_3}), \quad (2.42)$$

where  $y_f$  are the Yukawa couplings and  $y_{f_i}$  refers to the Yukawa coupling of the  $i$ th fermion generation. This results in mass terms of the form

$$m_f = y_f \frac{v}{\sqrt{2}} \quad (2.43)$$

and interaction terms

$$\mathcal{L}_{\text{Yuk,int}} = - \sum_f \frac{m_f}{v} H \bar{f} f \quad (2.44)$$

between the Higgs boson and fermions. In contrast to the  $HVV$  couplings, the Yukawa couplings are predicted to be directly proportional to the fermion masses. Mass terms for the neutrinos are not included in the SM due to the absence of right-handed neutrinos.

#### 2.1.4 CP violation in the Standard Model

The matrices in Eq. (2.42), which relate the weak eigenstate and the mass eigenstate of quarks and leptons, lead to an interesting phenomenon. Inserting the transformation behavior for the up-type and down-type quarks

$$q_i = (V_q)_{ij} q'_j \quad (2.45)$$

into the electroweak Lagrangian in Eq. (2.21) yields

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} \overline{u}_{i,L} \gamma^\mu \left( V_{u_L} V_{d_L}^\dagger \right)_{ij} \begin{pmatrix} 0 & W^+ \\ W^- & 0 \end{pmatrix}_\mu d_{j,L} \quad (2.46)$$

in the mass basis. The matrix product

$$V_{\text{CKM}} = V_{u_L} V_{d_L}^\dagger \quad (2.47)$$

forms the so-called Cabibbo-Kobayashi-Maskawa (CKM) matrix [87, 88], which provides the transition probabilities of an up-type quark to a down-type quark (and vice versa) in a charged-current interaction. The CKM matrix is complex, and of the initial 18 free parameters, only four survive due to the unitarity constraint and absorption of free phases into field redefinitions (up to one global phase). Three of the four remaining degrees of freedom are mixing angles between the quark generations, while the last one is a phase  $\delta_{\text{CKM}}$ , which is responsible for  $\mathcal{CP}$ -violating effects. The diagonal elements of the CKM matrix are close to one, which means that mixing between the three generations of quarks is strongly suppressed.

The CKM matrix is currently the only known source of  $\mathcal{CP}$  violation within the SM. It was first observed experimentally from the mixing of neutral kaons in 1964 [89]. A neutral kaon  $K^0$  and its antiparticle  $\bar{K}^0$  can mix via box diagrams in which two  $W$  bosons are exchanged. The imaginary part of the CKM matrix leads to a difference in the mixing rates  $\Lambda(K^0 \rightarrow \bar{K}^0) - \Lambda(\bar{K}^0 \rightarrow K^0) \propto 2 \text{Im } \mathcal{M}$  where  $\mathcal{M}$  is the matrix element of the process. This was experimentally confirmed by observation of the  $\mathcal{CP}$ -violating decay of a (physical) long-lived kaon  $K_L$  into two pions. The current value for the  $\mathcal{CP}$ -violating phase in the CKM matrix has been measured to be  $\delta_{\text{CKM}} = 65.5^\circ$  [12].

Additional  $\mathcal{CP}$  violation is expected in the lepton sector from the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [90, 91], which plays the same role as the CKM matrix for leptons and neutrinos. It is needed to explain the measured oscillations of neutrinos [92]. The PMNS matrix has 4 or 6 free parameters, depending on whether the neutrinos are Dirac or Majorana fermions, of which one is again a  $\mathcal{CP}$ -violating phase  $\delta_{\text{PMNS}}$ . In contrast to the quark sector, the mixing between generations in the lepton sector is not suppressed. Furthermore, current measurements of the  $\mathcal{CP}$ -violating phase average to  $\delta_{\text{PMNS}} \approx 210^\circ$  in the case of a normal-ordering of the mass eigenstates, which suggests a near-maximal  $\mathcal{CP}$  violation in processes involving neutrino mixing [12].

So far in this thesis, only the EW sector has been considered as a source of  $\mathcal{CP}$  violation. However, the SM allows for a  $\mathcal{CP}$ -violating term in the strong sector of the form

$$\mathcal{L}_\theta = \frac{g^2}{32\pi^2} \theta_{\text{QCD}} G^{a,\mu\nu} \tilde{G}_{\mu\nu}^a \quad (2.48)$$

with the dual field strength tensor  $\tilde{G}_{\mu\nu}^a = 1/2 \epsilon_{\mu\nu\rho\sigma} G^{a,\rho\sigma}$ . The SM does not provide a reason for a strong suppression of  $\theta_{\text{QCD}}$ , but measurements of the neutron EDM constrain  $\theta_{\text{QCD}} < 10^{-10}$  [41]. This is known as the strong  $\mathcal{CP}$  problem.

Since every EDM of a fundamental particle requires  $\mathcal{CP}$  violation, other EDMs also provide interesting tests of the SM. EDMs in the SM are strongly suppressed, as the only  $\mathcal{CP}$  violation comes from the CKM matrix. A non-zero EDM measurement that exceeds the SM prediction is therefore a clear sign of additional sources of  $\mathcal{CP}$  violation. For example, the EDM of the electron (eEDM) arises in the SM

at the four-loop level. The best current experimental constraint is  $|d_e| < 4.1 \cdot 10^{-30} e \text{ cm}$  [40], while estimates of the SM eEDM yield values several orders of magnitude below this [93, 94]. However,  $\mathcal{CP}$  violation from the PMNS matrix could already increase the eEDM prediction by a lot [95]. It is also sensitive to potential new sources of  $\mathcal{CP}$  violation in the Higgs sector (see e.g. [27]), which will be discussed further below. Consequently, EDM measurements provide a complementary method of probing  $\mathcal{CP}$  violation in addition to collider experiments. The latter are discussed in detail in Chapter 3.

## 2.2 Physics beyond the Standard Model

The Standard Model of particle physics is regarded as one of the most successful theories of all time. It is a fundamental theory based on causality, relativity, symmetries, and quantum field theory. Furthermore, it is renormalizable up to the Planck scale, has correctly predicted several new particles, and is unitary. Still, it fails to explain a significant number of phenomena in the observable universe. Efforts to find theories beyond the SM (BSM) that can correctly predict experimental deviations from the SM and incorporate missing aspects, such as gravity, are therefore consistently increasing. Since the shortcomings of the SM and possible explanations via BSM theories are a vast topic, this thesis will focus on a subset of shortcomings that BSM extensions of the Higgs sector may explain.

### 2.2.1 Shortcomings of the Standard Model

The many shortcomings of the SM can be split into experimental observations that it cannot explain and theoretical arguments. They include the hierarchy problem, the apparent meta-stability of the Higgs potential, the strong  $\mathcal{CP}$  problem as mentioned in Section 2.1.4, and simplicity arguments since the SM contains a total of 26 free parameters, of which 20 parameters are in the flavor sector. Among the most striking shortcomings are the inability of the SM to explain the origin of dark matter and dark energy, its missing description of gravity, and its failure to account for the observed asymmetry of matter and antimatter in the Universe.

#### The hierarchy problem

In the SM, the parameter  $\mu^2$  appearing in the Higgs potential (see Eq. (2.31)) is directly responsible for the Higgs boson mass. It is also responsible for the nature of EWSB, as only  $\mu^2 < 0$  leads to spontaneous symmetry breaking in the first place. Therefore, states of NP that appear at a higher mass scale  $M$  and couple to the Higgs boson introduce corrections to the Higgs mass

$$m_H^2 = (m_H^2) \Big|_{\text{bare}} - \Delta m_H^2 (M_{\text{NP}}^2) \quad . \quad (2.49)$$

If the SM is considered to describe physics accurately to the Planck scale  $M_P \sim 10^{18} \text{ GeV}$ , where quantum gravity effects are expected, the corrections to the Higgs boson mass are of order  $\Delta m_H^2 \sim 10^{36} \text{ GeV}$ . The fact that the SM does not explain the apparent fine-tuning of the Higgs boson mass to the EW scale, despite these large corrections, is referred to as the hierarchy problem. In other words, the SM does not explain the origin of its own symmetry breaking. Several BSM scenarios can deal with these issues, such as supersymmetry or composite Higgs models, where  $m_H^2$  is protected by new symmetries, or dynamical solutions via the relaxion. A review of the hierarchy problem can be found in [96].

## Dark Matter

There are a large number of experimental observations that hint at the existence of some form of dark matter, and it is estimated that dark matter makes up about 23% of the total matter in the observable universe. The first experimental hint was the motion of galaxies in the Coma cluster [97]. A comparison of the gravitational potential obtained via the virial theorem and a mass estimation via luminosity measurements revealed that most of the mass could not be seen. Furthermore, the rotation curves of galaxies show a constant velocity of the stars with respect to their distance from the galaxy nucleus, also hinting towards additional matter being present [98]. Finally, the power spectrum of the cosmic microwave background (CMB) shows peaks which can be well explained by dark matter [99]. While the SM neutrinos are potential candidates for hot dark matter, there is no SM particle with properties needed for the  $\Lambda$ CDM (cold dark matter) model, which is currently the preferred model of dark matter and dark energy. Therefore, extensions of the SM often include at least one particle that can act as a  $\Lambda$ CDM candidate. Examples are supersymmetry, in which the lightest supersymmetric particle is a dark matter candidate, or models that introduce any axion or axion-like particle (originally designed to solve the strong  $C\mathcal{P}$  problem). Reviews of dark matter with a focus on cosmology or the LHC are [100, 101].

## Gravity

The SM cannot be a complete description of nature as it only describes three of the four known fundamental forces. While special relativity is embedded in QFT, an accurate description of gravity is only provided by Einstein's general relativity (GR) [102]. As it is not a quantized theory, the SM and GR are currently not compatible. Consequently, a theory is needed that can unify both theories at energies around the Planck scale, which is the goal of quantum gravity. Theories that unify the four fundamental interactions while quantizing gravity are dubbed a "theory of everything," with string theory being a potential candidate. Reviews on this topic include [103, 104].

## Baryon asymmetry of the Universe

The baryon asymmetry of the Universe (BAU), or matter-antimatter asymmetry, can be measured from big bang nucleosynthesis (BBN) and the CMB. Both measurements yield approximately [11, 12]

$$\eta = \frac{n_B - n_{\bar{B}}}{s} \approx 8.7 \cdot 10^{-11} \quad (2.50)$$

where  $n_B$  ( $n_{\bar{B}}$ ) is the (anti-)baryon number density and  $s$  is the entropy density. To dynamically produce a baryon asymmetry during the early Universe, the three so-called Sakharov conditions have to be fulfilled [13]. They consist of

- violation of baryon number conservation,
- departure from thermal equilibrium,
- and  $C$ -violation and  $C\mathcal{P}$ -violation.

However, the SM is unable to generate enough baryon asymmetry to account for the observed value of the BAU. The baryon number is called an accidental symmetry of the SM, as no symmetry group imposes its conservation. Instead, the SM simply does not contain processes that violate the baryon number, except for hypothetical sphaleron processes, which have not been observed so far [105, 106]. Furthermore, the departure from thermal equilibrium, which corresponds to a strong first-order phase transition of the EW

vacuum for baryogenesis, is not possible for the observed Higgs mass of  $m_H = 125$  GeV [107, 108]. Finally, the amount of  $\mathcal{CP}$  violation present in the SM does not suffice to account for the observed BAU, even if a strong first-order phase transition is assumed [14, 15].<sup>4</sup>

One of the most popular proposed solutions to these problems is *electroweak baryogenesis* (EWBG) [16, 17]. In models of EWBG, the BAU is generated during the EW phase transition, which is required to be first-order. During the SSB, so-called bubbles arise in the Universe that contain the broken EW symmetry. These bubbles start to expand rapidly, leading to a departure from thermal equilibrium.  $\mathcal{CP}$ -violating processes at the bubble wall are responsible for the creation of a chiral asymmetry in the symmetric phase. This chiral asymmetry can be converted to baryon number violation via sphaleron processes, which describe transitions between EW vacua and violate  $B + L$ . Sphaleron processes are efficient in the symmetric phase but become suppressed in the broken one, so that the violated baryon number becomes frozen when it is caught by the expanding bubble. The amount of baryon asymmetry produced this way depends on the bubble parameters, as well as the sources of  $\mathcal{CP}$  violation [110–113]. A review of EWBG can be found in [114].

### Tensions with experimental results

In recent years, there have been some experimental measurements in high-energy physics that yielded observations in tension with the SM predictions. An example of this is a slim excess at 95 GeV in  $\gamma\gamma$  final state events, which was observed across experiments [115, 116]. Furthermore, there have been long-standing discrepancies between measurements and predictions of the lepton flavor universality anomalies and the  $g - 2$  of the muon. However, both of them have been weakened by recent results [117, 118]. Several measurements of the  $W$  boson mass from the CDF collaboration are in tension with the SM prediction [9], whereas the most recent measurements by ATLAS and CMS agree with the SM [119, 120]. Overall, no clear indication of BSM physics has been found in high-energy physics experiments as of today.

### 2.2.2 Parameterizing BSM physics

The term "BSM" is very general and contains a vast landscape of theories. They can be roughly classified into *effective field theories* (EFTs) [121] and *models*. The former are model-independent as they do not require new states of physics, and are further discussed below. On the other hand, models introduce concrete new degrees of freedom. They can be split into simple extensions of the SM, like singlet extensions or the 2HDM, and UV-complete models. While simple extensions of the SM are often designed to fix certain shortcomings of the SM and are usually again only valid up to some energy scale  $\Lambda$ , UV completions aim to provide full descriptions of nature that are valid at all energy scales.

#### Effective field theories

Since there is currently no evidence that favors a concrete BSM model, it is often helpful to work with an EFT approach. Such approaches are used when there are two theories, one being more underlying but unknown, which are separated by some energy scale. In this case, the known theory accurately describes

---

<sup>4</sup> There are models for which the amount of  $\mathcal{CP}$  violation in the SM is enough to generate the BAU, but they require other types of tailored NP, see e.g. [109].

physics at low energies. Deviations due to the underlying theory are suppressed by the energy scale at which new states appear, and where the EFT breaks down. A prominent example is the Fermi theory of the beta decay, which describes the weak interaction as a 4-point interaction; an assumption that is valid for  $E < m_W$ . The electroweak interaction of the SM takes on the role of the more underlying theory.

Nowadays, the same approach is used but with the SM as the low-energy theory. It is assumed that states of NP appear at  $E = \Lambda_{\text{NP}} > 1$  TeV, which are decoupled from the SM according to the Appelquist-Carazzone theorem [122]. They can therefore only be detected indirectly via modifications of SM parameters that are parameterized within an EFT. Two EFTs are widely used, which are the Standard Model Effective Field Theory (SMEFT) [123] and the Higgs Effective Field Theory (HEFT) [124, 125]. The former includes the same  $SU(2)$  Higgs doublet as the SM and keeps its  $SU(3) \times SU(2) \times U(1)$  gauge structure intact, whereas this assumption is relaxed in the HEFT, and operators can possess a different gauge structure. In single Higgs production processes, differences between the two might only be visible in the  $HVV$  couplings [126]. Therefore, and for simplicity, only the SMEFT is considered in this work.

In the SMEFT, the SM Lagrangian is extended by operators of mass-dimension  $D > 4$  which are suppressed by corresponding powers of the NP scale  $\Lambda$ . The total SMEFT Lagrangian is therefore an infinite sum

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{D>4} \frac{1}{\Lambda^{D-4}} C_i^{(D)} O_i^{(D)} \quad (2.51)$$

where the  $C_i$  are called Wilson coefficients that parameterize the strength of the interactions induced by the operators  $O_i$ . At  $D = 5$ , there is only one operator that keeps the gauge structure of the SM intact. It is called the Weinberg operator and generates a Majorana mass term for neutrinos, as well as lepton number-violating processes. The current upper limits on neutrino masses suggest that NP states responsible for them must exist at very high energies, so this operator is expected to be strongly suppressed [127]. The operators at  $D = 7$  are lepton number-violating as well and are even further suppressed.

The operators relevant for this thesis appear at  $D = 6$ . In total, 59 non-redundant operators conserve baryon and lepton number at this dimension, which form the so-called *Warsaw basis* [30, 123]. The operators can be grouped by their field contents. There are 25 four-fermion operators, which generate an effective four-point interaction involving mixtures of left- and right-handed fermions. Next, there are 15 purely bosonic operators that can modify the self-couplings of the strong, weak, and Higgs bosons as well as their interactions. Operators involving a dual field strength tensor  $\tilde{W}_{\mu\nu}$  or  $\tilde{G}_{\mu\nu}^a$  give rise to  $\mathcal{CP}$ -violating interactions. Finally, there are 19 operators that mix fermionic and bosonic fields and can modify their interactions, such as the Yukawa couplings. If the Wilson coefficient of such an operator is complex,  $O_i - O_i^\dagger \propto 2 \text{Im } C_i$  will result in a  $\mathcal{CP}$ -violating part of the modified interaction. However, as the SMEFT operators are defined in the weak eigenstate, a complex Wilson coefficient does not directly correspond to  $\mathcal{CP}$  violation measured in an experiment. The complete list of the  $D = 6$  operators in the Warsaw basis can be found in [30].

### The Higgs Characterization Model

The SMEFT is a very general and powerful approach for modeling potential effects of BSM physics. However, for analyses targeting the Higgs sector, it is sufficient to use a more specific effective framework to capture deviations from the SM. The so-called *Higgs Characterization model* (HCM) was designed

in 2013 to assess the nature and interactions of the Higgs boson discovered in 2012 [128]. The HCM includes different spin hypotheses for the Higgs boson. However, experiments have confirmed the zero-spin nature of the Higgs boson, and other values of the spin are therefore discarded (see the discussion in Section 3.3.2). Furthermore, the HCM allows for a  $\mathcal{P}$ -even and a  $\mathcal{P}$ -odd state of the Higgs boson. These two states can mix, resulting in  $\mathcal{CP}$  violation in the interactions of the Higgs boson. In contrast to the SMEFT, the HCM is defined in the mass basis and therefore,  $\mathcal{CP}$  violation in the HCM corresponds to the one observable in experiments.

In the SMEFT, the couplings between the Higgs boson and the SM fermions can be modified by dimension-6 operators of the form  $\Psi^2 \Phi^3$  in the Warsaw basis. The corresponding Lagrangian

$$\mathcal{L}_{\text{Yuk}}^{D=6} = \frac{C_{\Psi\Phi}}{\Lambda^2} (\Phi^\dagger \Phi) (\bar{\Psi}_L \Psi_R \Phi) \quad (2.52)$$

is added to the SM, where the value of the (complex) Wilson coefficient  $C_{\Psi\Phi}$  determines the strength of the Yukawa coupling modification. It modifies the existing three-point interactions, but also creates new four- and five-point interactions involving two or three Higgs bosons. However, these are experimentally less relevant compared to the three-point interaction and are not included in the HCM.

Instead, the modified Yukawa coupling is written as [128]

$$\mathcal{L}_{\text{Yuk}}^{\text{mod}} = - \sum_f \frac{y_f}{\sqrt{2}} \bar{f} \left( c_f + i\gamma_5 \tilde{c}_f \right) f H \quad (2.53)$$

in the HCM, where  $c_f$  is the  $\mathcal{CP}$ -even part of the coupling and  $\tilde{c}_f$  the  $\mathcal{CP}$ -odd part. This Lagrangian includes the SM Yukawa-couplings and corresponds to Eq. (2.44) when choosing  $c_f = 1$  and  $\tilde{c}_f = 0$ . Other parameterizations are possible, such as

$$\mathcal{L}_{\text{Yuk}}^{\text{mod}} = - \sum_f \frac{y_f}{\sqrt{2}} \bar{f} g_f \left( \cos \alpha_f + i\gamma_5 \sin \alpha_f \right) f H \quad (2.54)$$

where

$$g_f = \sqrt{c_f^2 + \tilde{c}_f^2}, \quad \alpha_f = \arctan \tilde{c}_f / c_f \quad (2.55)$$

are the rate modifier  $g_f$  and the  $\mathcal{CP}$ -mixing angle  $\alpha_f$ . CMS adopts the usage of an effective fractional cross section [129]

$$f_{\mathcal{CP}}^{Hff} = \frac{|\tilde{c}_f|^2}{|c_f|^2 + |\tilde{c}_f|^2} \text{sgn} \left( \frac{\tilde{c}_f}{c_f} \right) \quad (2.56)$$

which connects to the  $\mathcal{CP}$ -mixing angle via  $|f_{\mathcal{CP}}^{Hff}| = \sin^2 \alpha_f$ .

As mentioned above, a great advantage of using the coupling modifiers  $c_f$  and  $\tilde{c}_f$  is that they are defined in the mass basis. Consequently, they are related to the Wilson coefficients of the SMEFT via a rotation to the weak eigenstate. Performing the rotation, the relation between the coefficients in Eq. (2.52) and Eq. (2.53) is [25, 26]

$$c_f = \frac{1 + 4 \text{Re} \tilde{C}_{\Psi\Phi} + 3(\text{Re} \tilde{C}_{\Psi\Phi}^2 + \text{Im} \tilde{C}_{\Psi\Phi}^2)}{(1 + \text{Re} \tilde{C}_{\Psi\Phi})^2 + \text{Im} \tilde{C}_{\Psi\Phi}^2}, \quad \tilde{c}_f = \frac{2 \text{Im} \tilde{C}_{\Psi\Phi}}{(1 + \text{Re} \tilde{C}_{\Psi\Phi})^2 + \text{Im} \tilde{C}_{\Psi\Phi}^2} \quad (2.57)$$

with

$$\tilde{C}_{\Psi\Phi} = \frac{v^2}{2\Lambda^2} \frac{C_{\Psi\Phi}}{y_f} . \quad (2.58)$$

The modification of the Yukawa Lagrangian leads to a modification of the Feynman rules and consequently to signal strengths and kinematics that depend on the amount of  $\mathcal{CP}$ -mixing. Generally, the matrix element of a process can be split into

$$|\mathcal{M}_{\text{tot}}|^2 = c_i^2 |\mathcal{M}_{\mathcal{CP}-\text{even}}|^2 + 2c_i\tilde{c}_i \text{Re}(\mathcal{M}_{\mathcal{CP}-\text{even}}\mathcal{M}_{\mathcal{CP}-\text{odd}}^*) + \tilde{c}_i^2 |\mathcal{M}_{\mathcal{CP}-\text{odd}}|^2 \quad (2.59)$$

where the parts proportional to  $c_i^2$  and  $\tilde{c}_i^2$  are  $\mathcal{CP}$ -even and affect the rate, while the middle term is  $\mathcal{CP}$ -odd and gives the interference between the two contributions.

The HCM also allows for modifications of the Higgs-gauge boson couplings. In the SMEFT, they are generated by operators of the form  $X^2\Phi^2$ . More precisely, there are six operators

$$O_{\Phi X} = \Phi^\dagger \Phi X_{\mu\nu} X^{\mu\nu}, \quad \tilde{O}_{\Phi \tilde{X}} = \Phi^\dagger \Phi \tilde{X}_{\mu\nu} X^{\mu\nu} \quad (2.60)$$

where  $X_{\mu\nu} = W_{\mu\nu}^i, B_{\mu\nu}, G_{\mu\nu}^a$ . Additionally, two more operators allow mixing of the  $W$  and  $B$  field strength tensors. All of these operators can generate new four-point interactions between two Higgs and two gauge bosons, but just like in the case of the Yukawa couplings, these are not included in the HCM. Instead, only the three-point interactions are considered. The Higgs-gauge operators involving a dual field strength tensor are  $\mathcal{CP}$ -odd by definition.

The Lagrangians in the HCM describing the Higgs couplings to massless gauge bosons take on the form [128]

$$\mathcal{L}_{H\gamma\gamma}^{\text{mod}} = -\frac{1}{4v} \left( \frac{47\alpha_{\text{em}}}{18\pi} c_\gamma F_{\mu\nu} F^{\mu\nu} + \frac{4\alpha_{\text{em}}}{3\pi} \tilde{c}_\gamma F_{\mu\nu} \tilde{F}^{\mu\nu} \right) H \quad (2.61)$$

and

$$\mathcal{L}_{ggH}^{\text{mod}} = -\frac{1}{4v} \left( -\frac{\alpha_s}{3\pi} c_g G_{\mu\nu}^a G^{\mu\nu,a} + \frac{\alpha_s}{2\pi} \tilde{c}_g G_{\mu\nu}^a \tilde{G}^{\mu\nu,a} \right) H \quad (2.62)$$

for the Higgs-photon and Higgs-gluon coupling, respectively. The prefactors of the couplings are chosen so that they resemble the loop contributions of the SM when the top quark is integrated out. Therefore, the parameters  $c_g$  and  $\tilde{c}_g$  fully parameterize the loop in the gluon fusion production channel, and the SM is recovered for  $c_g = 1$  and  $\tilde{c}_g = 0$ .

The modified Higgs couplings to the heavy vector bosons are further split into parts [128]

$$\begin{aligned} \mathcal{L}_{HVV}^{\text{mod}} = & \frac{1}{v} \left( 2c_{\text{SM}} m_W^2 W_\mu^+ W^{-,\mu} + c_{\text{SM}} m_Z^2 Z_\mu Z^\mu \right) H \\ & - \frac{1}{4\Lambda} \left( c_Z Z_{\mu\nu} Z^{\mu\nu} + \tilde{c}_Z Z_{\mu\nu} \tilde{Z}^{\mu\nu} \right) H \\ & - \frac{1}{2\Lambda} \left( c_W W_{\mu\nu}^+ W^{-,\mu\nu} + \tilde{c}_W W_{\mu\nu}^+ \tilde{W}^{-,\mu\nu} \right) H \end{aligned} \quad (2.63)$$

where the first line describes modifications of the tree-level SM coupling and the other lines modify loop-induced couplings. A coupling between a  $\mathcal{CP}$ -odd Higgs boson and heavy gauge bosons is not allowed on tree level since such a coupling of mass-dimension 4 would introduce a hard breaking of the  $\mathcal{CP}$ -symmetry of the Lagrangian. The lowest mass-dimension at which such an interaction can be

generated is 6 and corresponds to the second term in the second and third lines of Eq. (2.63).<sup>5</sup>

On the amplitude level, the coupling between the Higgs boson and any gauge bosons  $\mathcal{V} = W, Z, \gamma, g$  can be written generally as [130]

$$A(H\mathcal{V}_1\mathcal{V}_2) \sim \left[ a_1^{\mathcal{V}\mathcal{V}} + \frac{\kappa_1^{\mathcal{V}\mathcal{V}} q_1^2 + \kappa_2^{\mathcal{V}\mathcal{V}} q_2^2}{\Lambda^2} \right] m_{\mathcal{V}_1}^2 \epsilon_{\mathcal{V}_1}^* \epsilon_{\mathcal{V}_2}^* + a_2^{\mathcal{V}\mathcal{V}} f_{\mu\nu}^{(1)} f^{(2)\mu\nu} + a_3^{\mathcal{V}\mathcal{V}} f_{\mu\nu}^{(1)} \tilde{f}^{(2)\mu\nu} \quad (2.64)$$

with the field strength tensor  $f^{(i)\mu\nu} = \epsilon_{\mathcal{V}_1}^\mu q_1^\nu - \epsilon_{\mathcal{V}_1}^\nu q_1^\mu$  and its dual  $\tilde{f}^{(i)\mu\nu} = 1/2 \epsilon^{\mu\nu\rho\sigma} f_{\rho\sigma}^i$ . The  $a_3^{\mathcal{V}\mathcal{V}}$  coupling parameterizes  $\mathcal{CP}$  violation. The first part of Eq. (2.64) describes tree-level couplings and so  $a_1^{\gamma\gamma} = a_1^{gg} = \kappa_1^{\gamma\gamma} = \kappa_1^{gg} = \kappa_2^{\gamma\gamma} = \kappa_2^{gg} = 0$ . For the gluon fusion loop, in contrast to Eq. (2.62), this corresponds to an EFT where the top quark is not integrated out. The parameters of the two equations can be related via

$$c_g = 1 - \frac{6\pi}{\alpha_s} a_2^{gg}, \quad \tilde{c}_g = \frac{4\pi}{\alpha_s} a_3^{gg} \quad (2.65)$$

assuming SM-like contributions to the gluon fusion loop.

Similar to the  $Hff$  case,  $\mathcal{CP}$  violation in the  $H\mathcal{V}\mathcal{V}$  coupling allows to separate the matrix element into three terms. The Higgs couplings to gluons and photons can be parameterized in the same way as in Eq. (2.59), because the SM coupling is loop-suppressed and therefore comparable to a dimension-6 effective interaction. The SM  $H\mathcal{V}\mathcal{V}$  coupling exists at tree level, but  $\mathcal{CP}$ -violating interactions are suppressed. Therefore, the matrix element is parameterized as

$$|\mathcal{M}_{\text{tot}}|^2 = |\mathcal{M}_{\mathcal{CP}-\text{even}}|^2 + 2d \operatorname{Re}(\mathcal{M}_{\mathcal{CP}-\text{even}} \mathcal{M}_{\mathcal{CP}-\text{odd}}^*) + d^2 |\mathcal{M}_{\mathcal{CP}-\text{odd}}|^2 \quad (2.66)$$

where  $d$  is, e.g., a  $\mathcal{CP}$ -violating Wilson coefficient, and the  $d^2$  term is strongly suppressed.

<sup>5</sup> The modified loop-induced  $H\mathcal{V}\mathcal{V}$  couplings in the HCM look like dimension-5 terms because  $v$  was used as a reference scale instead of  $\Lambda$ . The correct prefactors in correspondence with the SMEFT operators would be  $v/\Lambda^2$ .

# CHAPTER 3

---

## Probing the Higgs sector

---

Data taking at the Large Electron-Positron Collider (LEP) took place until the year 2000 and allowed for narrowing down the potential mass for the SM Higgs boson to a mass window between 114 GeV and 158 GeV [131, 132]. One of the main tasks of its successor, the Large Hadron Collider (LHC), was to find the Higgs boson and determine its mass and other properties. After the first launch of the LHC in late 2008, it took less than four years for the two multi-purpose experiments ATLAS and CMS to discover a particle that fits the SM predictions [1, 2]. Since then, numerous measurements have been conducted to determine the properties of the discovered Higgs boson with high precision, such as its mass, spin, width, and parity.

In this chapter, an overview of current Higgs boson studies and their most recent results will be given. Section 3.1 briefly discusses the setup of the LHC and its two biggest experiments, ATLAS and CMS. This is followed by a discussion of simulating events via Monte Carlo generators, as well as the subsequent parton showering and detector simulation in Section 3.2. Finally, an overview of Higgs physics at the LHC will be given in Section 3.3, and the most relevant experimental results for this thesis are discussed.

### 3.1 The Large Hadron Collider

The LHC, located at CERN in Geneva and its surroundings, is the largest and most powerful particle accelerator in the world. It is a circular proton-proton collider with a total length of 26.7 km that was designed to reach center-of-mass (COM) energies of up to 14 TeV [133]. The protons collide at four different interaction points at which events can be detected for the ATLAS, CMS, LHCb, and ALICE experiments. Before the protons enter the LHC, they are pre-accelerated in the so-called injector chain. It consists of a linear accelerator and three smaller synchrotrons which boost the protons to an energy of 450 GeV. The protons are not accelerated as a continuous beam but in bunches with a bunch spacing of 25 ns.

The LHC was designed to first test the validity of the SM via a discovery of the Higgs boson and subsequently look for clear signals of potential NP. Since its operation began in 2008, it has collected data in multiple runs, which are summarized in Table 3.1. Currently, Run 3 is operating at a COM energy of  $\sqrt{s} = 13.6$  TeV. This is the energy that the protons receive, while the actual collisions happen

	Years	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [ $\text{fb}^{-1}$ ]
Run 1	2010 – 2012	7 & 8	5 & 20
Run 2	2015 – 2018	13	139
Run 3	2022 – 2026	13.6	$\sim 300$

Table 3.1: Overview of the LHC runs up to now, including the COM energy and the (expected) integrated luminosity obtained by ATLAS and CMS.

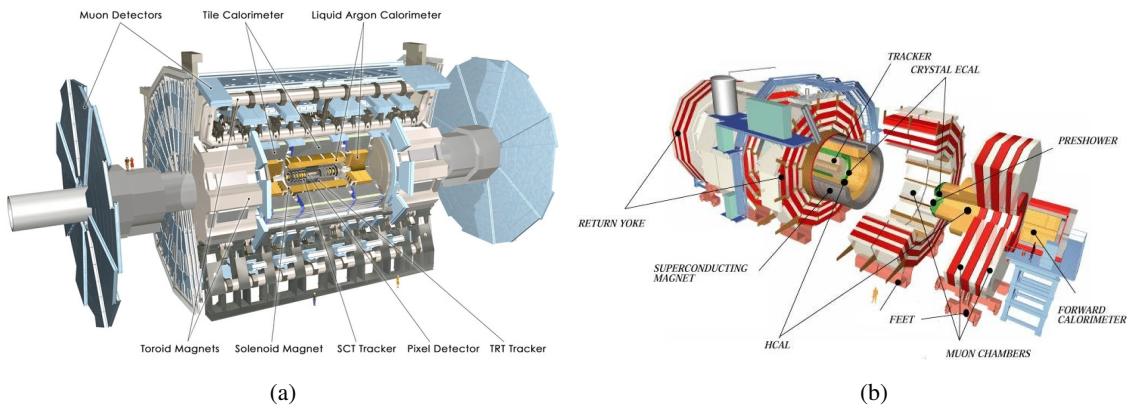


Figure 3.1: Schematics of the (a) ATLAS and (b) CMS detector. Figures taken from [136, 137].

between partons with only a fraction of the total energy, justifying the need for a multi-TeV collider to perform Higgs measurements. However, due to the low cross sections of the Higgs production channels compared to other SM processes, a high luminosity is needed as well. This is achieved by a high number of colliding protons and a strong focusing of the beams via quadrupole magnets.

### 3.1.1 The ATLAS and CMS experiments

The two biggest experiments at the LHC are the ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) experiments. They are also the two detectors capable of collecting the highest luminosities at the LHC [134, 135]. Both detectors are used for very similar analyses in the context of SM precision tests and BSM searches. Therefore, they offer the unique possibility to perform measurements independently and cross-check each other's results. Their structure was intentionally designed to be very similar for this purpose. A schematic overview of the ATLAS and CMS detectors, highlighting their most important systems, is shown in Fig. 3.1.

The kinematics of events detected in the experiments are commonly expressed in the *laboratory frame* (lab frame). The coordinate system of the lab frame is defined such that the  $z$ -axis aligns with the beam axis. As the detectors are of cylindrical shape, instead of  $p_x, p_y, p_z$  a more commonly used

parameterization consists of

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad \phi = \cos^{-1} p_x / p_T, \quad \eta = \sinh^{-1} p_z / p_T \quad (3.1)$$

with the *transverse momentum*  $p_T$ , the *azimuthal angle*  $\phi$  and the *pseudorapidity*  $\eta$ . Additionally, the angular distance between two particles can be defined as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \quad (3.2)$$

Both detectors must be able to identify particles, provide the best possible resolution, and have fast read-out times. To achieve this, they were built in an "onion-like" structure. The part of both detectors closest to the beam line is called the Inner Detector. It can precisely track the trajectory of particles and therefore measure the momenta and electric charge of incoming objects. This allows the identification of electrons. In both experiments, the Inner Detector covers a pseudorapidity region of  $|\eta| \leq 2.5$  and is subjected to a strong magnetic field of 2 T in ATLAS (4 T in CMS).

Covering the Inner Detector are the calorimeters, namely the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL). The ECAL surrounds the Inner Detector and is designed to fully stop electrons and photons. Their deposited energy can be measured because they produce electromagnetic showers within the ECAL volume. The HCAL is the outer part of the calorimeter system. It stops hadrons and measures their energy by producing hadronic showers. The calorimeters in ATLAS (CMS) both cover a pseudorapidity region of  $|\eta| \leq 3.2$  (3.0) with additional forward calorimeters covering  $|\eta| \leq 4.9$  (5.2).

Finally, the outermost part of the detector is the muon spectrometer. Compared to electrons, muons cannot be stopped in the ECAL due to their higher mass and lower ionization losses. Although the muon spectrometer is unable to fully stop the muons, a precise momentum reconstruction is possible via the ionization of a gas.

Other particles cannot be detected (directly) at the ATLAS and CMS detectors. Heavy particles, such as the top quark or massive bosons, decay before entering the detector volume. Furthermore, free quarks and gluons hadronize prior to detection. In both cases, the distribution of the measurable final-state particles contains information about the parton-level event. Finally, the interaction cross section of neutrinos with matter is very small, and they escape the detector without any detection possibility.<sup>1</sup> Information about neutrinos is mostly inferred from the *missing transverse energy*  $E_T^{\text{miss}}$  of the total event.

## 3.2 Event generation for phenomenology

### Fixed order calculations

In order to perform phenomenological studies for high-energy physics, the possibility to create events matching a specific underlying physics assumption is of great importance. Such a task is fulfilled by

---

<sup>1</sup> During the shutdown between Run 2 and Run 3, the two neutrino detectors FASER and SND@LHC were constructed at the LHC which are able to detect neutrinos coming from the interaction points [138–141]. However, it is not possible to match detected neutrinos to ATLAS or CMS events.

Monte Carlo event generators. The central idea of these generators is to fully automate the computation of any process in particle physics at a fixed order, independent of the types, number, and interactions of the particles. There are multiple event generators available, with one of the most versatile ones being `MadGraph5_aMC@NLO` [142–144].

In a perturbative theory, the non-trivial interactions that happen at the lowest order of the respective coupling constants produce events at *leading order* (LO). `MadGraph` can generate physical LO events at the parton-level by calculating the (squared) matrix elements  $|\mathcal{M}|^2$  from the generated Feynman diagrams and integrating the phase space via Monte Carlo sampling. The returned events are typically *unweighted* by rejecting events with a probability according to their weight. The resulting distributions are then reflective of their underlying theory. BSM theories can be incorporated into `MadGraph` with files in the Universal FeynRules Output (UFO) format. UFO files for the SMEFT and the HCM are available online [145–147] and are used for the event generation in this thesis (see the respective chapters and Appendix B).

A great advantage of event generators is the possibility to isolate the effects of BSM couplings at a specific interaction order. For example, `MadGraph` allows the usage of a restriction file, in which all Wilson coefficients in the SMEFT are set to zero, except for a  $\mathcal{CP}$ -conserving and a  $\mathcal{CP}$ -violating one of interest for a chosen process. In this case, the resulting squared matrix element can be split up into six terms describing the SM, interference, and squared contributions. Specifying the NP interaction order during the event generation will result in the isolation of the chosen contribution, as long as events are generated at LO.

`MadGraph` also allows for generating events at *next-to-leading order* (NLO). Real corrections to the LO diagrams are tree-level diagrams and can be calculated straightforwardly, while virtual corrections are handled by `MadLoop5`. The finite parts and divergencies of the amplitude are computed in the 't Hooft-Veltman scheme [148], and tensor integrals are evaluated via integral reduction. Divergencies are regularized dimensionally, while the renormalization scheme can be specified in the NLO model used.

The final step in the generation of parton-level events is the decay of unstable particles. For example, in top-associated Higgs production at the LHC,  $pp \rightarrow t\bar{t}H$ , none of the final state particles will be found in the detector due to their short lifetimes. However, generating the full (non-resonant) process, including the decay to detectable particles, makes the calculation very inefficient, especially at NLO. Instead, `Madspin` [149, 150] is used as a compromise between performance and information about spin correlations and off-shell effects. For this, the *narrow-width approximation* (NWA) [151, 152]

$$\left| \frac{1}{q^2 - m^2 - im\Gamma} \right|^2 \sim \frac{\pi}{m\Gamma} \delta(q^2 - m^2) \quad (3.3)$$

is used, which means that the width of the unstable particles is assumed to be close to zero. Consequently, the production and decay of the full process factorize, and the decay can be treated as an independent process.

### Parton showering and hadronization

To model the evolution of the parton-level objects to detectable particles, several effects have to be taken into account, which are visualized in Fig. 3.2. The first ones are initial-state radiation (ISR) and final-state radiation (FSR). ISR considers that photons, gluons, and other particles may be emitted from

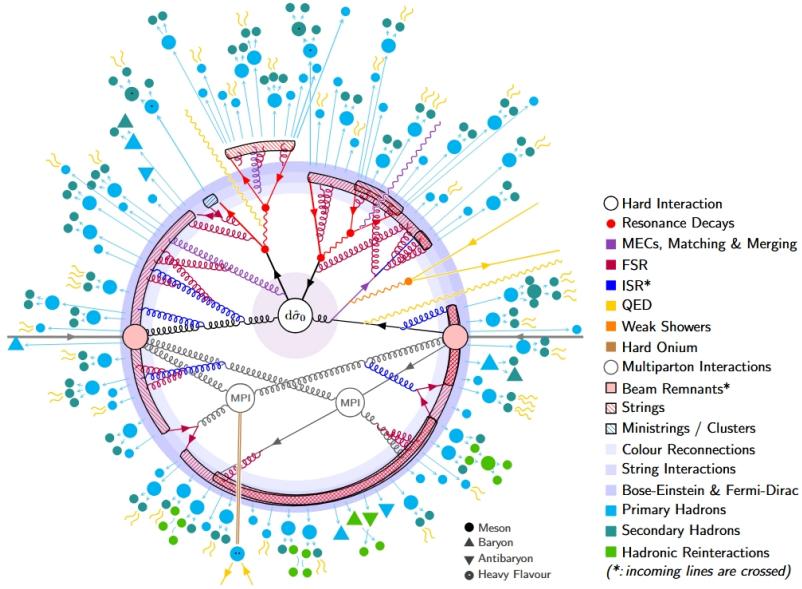


Figure 3.2: Simplified event display from **Pythia8** for  $t\bar{t}$  production including all relevant effects of the parton showering and hadronization. The beams enter from the left and right, and the hard interaction  $d\sigma_0$  is found in the center of the figure. Measurable particles propagate outwards to the detector. Figure taken from [153].

the initial-state partons before the hard interaction. FSR describes the same effect for particles emerging from the hard scattering process and any subsequent decays. The combination of both effects is called a *parton shower* and is modeled with **Pythia8** in this work [153]. It accurately accounts for the emission of soft and collinear gluons via resummation of the logarithmic divergences.

To ensure that the physical process is correctly modeled when going from the hard scattering process to the parton shower, it is necessary to perform *matching* and *merging* between them. Additional hard jets created at the parton level, either from a high jet multiplicity at LO, real corrections from NLO, or both, could also be present in the parton shower. This would lead to a double counting of the respective jets. **MadGraph5** automatically subtracts the possible double counting for an NLO event generation, making a matching to the parton shower trivial but also necessary to produce physical samples. LO events with several jet multiplicities need to be explicitly handled, e.g., by introducing a jet matching scale to smoothly separate the hard process from the parton shower.

In addition to the radiation effects from the hard scattering, multiple parton interactions (MPIs) describing secondary scattering processes between other partons of the beams are also taken into account by **Pythia8**. Once the energy of the free particles charged under QCD reaches  $\Lambda_{\text{QCD}}$ , they become confined in so-called strings, which are color-singlets and break apart into hadrons shortly after. This is known as the *hadronization*. These (primary) hadrons can further decay if they are unstable, and both primary and secondary hadrons can interact with other nearby particles. Ultimately, a set of color-neutral, stable particles and their momentum is obtained.

## Detector simulation

The final step of generating events that resemble measurements at the LHC is a simulation of the detector response. The input for the detector simulation consists of the hard scattering process after showering and hadronization. The most precise and versatile simulation tool for detector effects is `Geant4` [154]. However, it is not suited for phenomenological studies with a large number of events due to its complexity and slow computation time. Instead, `Delphes3` is used, which provides a framework for fast detector simulation specifically tailored towards hadron collisions at the LHC [155].

The processing of events in `Delphes3` closely follows the structure of the ATLAS and CMS detectors described in Section 3.1.1. In a first step, the inner detector is modeled by the propagation of hadrons and leptons in a magnetic field. Charged particles can be reconstructed as a track with limited energy and momentum resolution, as well as a smeared normalization of the transverse momentum. Afterwards, particles that are within the angular coverage of the calorimeters deposit their energy in cells of either the ECAL or the HCAL. Electrons and photons deposit their entire energy in the ECAL. Hadrons deposit their energy in the HCAL if they are stable or in both calorimeters if they decay within the detector volume. The hits in the calorimeters are classified by their energy and pseudorapidity with limited resolution. Multiple particles that deposit their energy in the same calorimeter cell are referred to as a tower.

After hits, tracks, and towers have been measured, the corresponding particles have to be identified. This is performed by the particle-flow algorithm, which is also used for particle reconstruction in CMS and is implemented in `Delphes3` as a simplified version [156]. It combines all available information from an event to improve the identification and resolution of particles and their momenta. The following particles can be identified:

- Particles that have no track and deposit their entire energy in the ECAL are identified as photons.
- If a track matches the energy deposit in the ECAL, the particle is instead identified as an electron.
- Muons are reconstructed if they have a track and are detected in the muon system.

Other fundamental particles cannot be identified based on the track and tower information alone due to their short lifetime (in the case of heavy particles) and hadronization. Instead, they have to be identified from a collection of particles via an algorithm that matches a group of detectable particles to an initial object. Objects that are reconstructed this way are dubbed *jets*. There are several algorithms designed for reconstructing jets, which are implemented in `Delphes` via the `FastJet` package [157]. One of the most used jet clustering algorithms is the so-called  $\text{anti-}k_t$  algorithm [158]. It computes the distances between two particles  $i$  and  $j$  and the beam  $B$  via

$$d_{ij} = \min \left( p_{T,i}^{-2}, p_{T,j}^{-2} \right) \Delta R_{ij}^2 / R^2, \quad d_{iB} = p_{T,i}^{-2} \quad (3.4)$$

and checks for the overall minimum value of all  $d$ . If  $d_{\min} \equiv d_{ij}$ , the particles  $i$  and  $j$  are merged. If instead  $d_{\min} \equiv d_{iB}$ , particle  $i$  is identified as a jet. The value of  $R$  in Eq. (3.4) can be chosen freely, with a common value of  $R = 0.4$  for light jets. The advantage of the  $\text{anti-}k_t$  algorithm is that it results in a mostly circular jet cone in the  $(\phi, \eta)$  plane.

Some parton-level objects can be identified from the properties of the reconstructed jet. A standard procedure, but also an active area of research for current experimental analyses, is the identification of jets originating from a  $b$ -quark or  $\tau$ -lepton, dubbed  $b$ -tagging or  $\tau$ -tagging (see e.g. [159–162]). Both

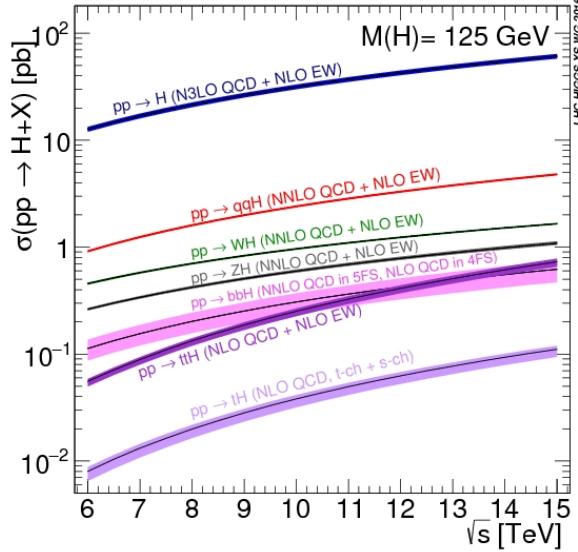


Figure 3.3: Main channels for the production of a single SM Higgs boson at the LHC as a function of  $\sqrt{s}$ . Figure taken from [169].

are available in `Delphes` via the "BTag" and "TauTag" attributes with tunable efficiency. Recently, much progress has been made in the identification of jets originating from a  $c$ -quark and in the differentiation between light quark jets and gluon jets [163, 164], although such options are not implemented in `Delphes` by default. Jets from more massive particles, such as the  $W$ ,  $Z$ , and Higgs boson, as well as the top quark, are generally reconstructed in several steps using more complicated techniques (see e.g. [165, 166]).

### 3.3 Higgs physics at the LHC

So far, most measurements of the 125 GeV Higgs boson are in excellent agreement with the predictions for a SM-like Higgs boson. More than 10 years after its discovery, many of the properties of the Higgs boson, such as its spin, the coupling strength to heavy particles, or its main production and decay channels at the LHC, have been measured with great precision [167, 168]. A collection of the most recent results can be found in [12]. However, many open questions remain, and there are large efforts to find hints of NP. These include numerous possibilities such as deviations from the SM prediction in rare decays, flavor-violating Higgs decays, or  $\mathcal{CP}$ -violating Higgs interactions. Especially, the light-Yukawa couplings are very weakly constrained [44–55].

#### 3.3.1 Production and decay channels

##### Higgs production channels

The SM Higgs boson can be produced in several different channels at the LHC. The production cross section of each channel depends on the COM energy  $\sqrt{s}$  and the availability of the initial states. Generally,

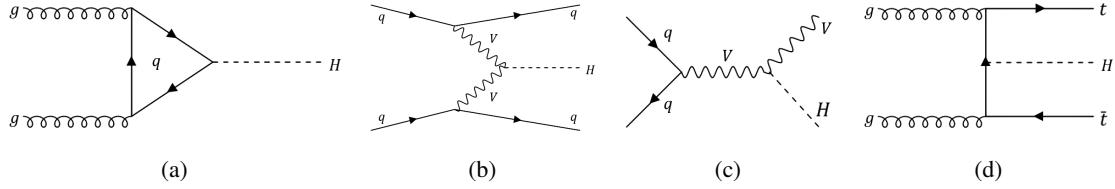


Figure 3.4: Feynman diagrams for the most relevant Higgs production mechanisms, namely (a) gluon fusion, (b) vector boson fusion, (c) Higgsstrahlung, and (d) top-associated Higgs production.

for the energies of the LHC runs between 7 and 14 TeV, the production of a single Higgs boson becomes increasingly more likely the higher  $\sqrt{s}$ , as can be seen in Fig. 3.3. The relative importance of the production channels stays approximately constant at these energies, with the exception of top-associated Higgs production.

By far the most relevant production channel is the so-called *gluon fusion* (ggF). With a cross section of  $\sigma_{\text{ggF}} \approx 52.2 \text{ pb}$  [12], it is responsible for about 88% of all Higgs bosons produced at the COM energy  $\sqrt{s} = 13.6 \text{ TeV}$  during Run 3 of the LHC. Two gluons fuse into a Higgs boson via a quark loop, as seen in Fig. 3.4(a), which is dominated by the top quark. The contribution of other quarks is suppressed due to the appearance of the respective Yukawa coupling in the diagrams.<sup>2</sup> Despite this process being loop-induced, it owes its relatively high cross section to the high abundance of gluons in the protons at medium Bjorken scale  $x$  [170], which is obtained from the *parton distribution functions* (PDFs) [171]. ggF stays the most relevant production mode, even in the exclusive case that two additional jets are radiated at the parton level [172].

The second most likely production channel of a single Higgs boson is *vector boson fusion* (VBF) with a cross section of  $\sigma_{\text{VBF}} \approx 4.1 \text{ pb}$  at  $\sqrt{s} = 13.6 \text{ TeV}$  [12]. In this process, two quarks from the proton beams radiate off a  $W$ - or  $Z$  boson (denoted jointly as  $V$  boson) which fuse into a Higgs boson, as shown in Fig. 3.4(b).<sup>3</sup> The  $V$  bosons are produced off-shell and, due to the high energies of the LHC, they are mostly longitudinally polarized. Since the initial quarks only give a fraction of their energy to the gauge bosons, VBF has a distinct topology of two very forward jets with high momenta and therefore large angular separation.

Higgs bosons can also be produced by being scattered off a weak vector boson. This is the underlying process of  $VH$  production, or *Higgsstrahlung*, and the corresponding Feynman diagram is depicted in Fig. 3.4(c). It is the main Higgs production mode at a lepton collider, but ranks third at the LHC with a cross section of  $\sigma_{\text{VH}} \approx 2.41 \text{ pb}$  [12]. For the LO s-channel diagram, the Higgs boson and  $V$  boson are produced back-to-back in  $p_T$ .  $ZH$  production receives a significant contribution at the loop level from a gluon-fusion-like diagram with a  $gg$  initial state.

Furthermore, Higgs bosons at the LHC can be produced in association with a pair of heavy quarks. The most prominent example for this is  $t\bar{t}H$  with  $\sigma_{t\bar{t}H} \approx 0.57 \text{ pb}$  [12]. Fig. 3.4(d) shows how two gluons turn into a  $t\bar{t}$  pair via a t-channel top quark, which emits a Higgs boson. The  $t\bar{t}$  pair can also be produced at LO via an s-channel gluon. Despite the two gluons in the initial state, the cross section for  $t\bar{t}H$  is small compared to ggF. This is due to the high energy required to produce the final state and the gluon

<sup>2</sup> Furthermore, there is also destructive interference between the top quark and the other quark loops.

<sup>3</sup> The term VBF can be misleading since gluons and photons are also vector bosons. Therefore, VBF is also sometimes called WBF (weak boson fusion).

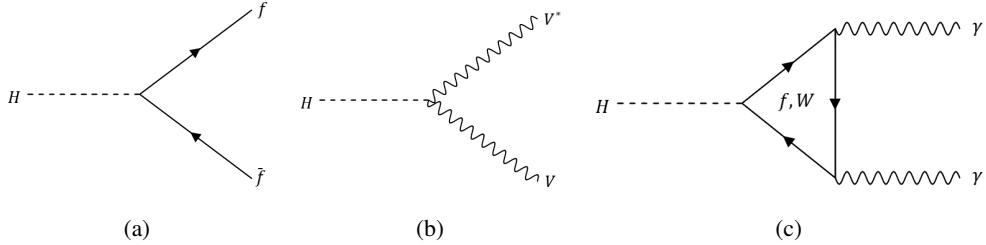


Figure 3.5: Feynman diagrams for the most relevant Higgs decay mechanisms, namely (a) the Higgs to fermion decay, (b) the Higgs decay into weak bosons, and (c) the Higgs decay into photons via a fermion or  $W$  boson loop.

luminosity decreasing heavily for high Bjorken scale  $x$ . Higgs production associated with a bottom quark pair has a similar cross section  $\sigma_{b\bar{b}H} \sim \sigma_{t\bar{t}H}$  [173]. The small bottom-Yukawa coupling compared to the top quark is balanced by the more easily available final state. However, VBF and  $VH$  are direct backgrounds of this process and make a measurement at the LHC challenging compared to  $t\bar{t}H$ .

Finally, it is also possible to generate a Higgs boson in association with a single top quark. The  $tH$  channel can be further categorized into channels with a light quark, a bottom quark, or a  $W$  boson in the final state. The emission of the Higgs boson is possible from top and bottom quark lines, as well as  $W$  boson lines. This leads to negative interference between the diagrams, reducing the total cross section of the process to about  $\sigma_{tH} \approx 0.09$  pb [12]. However, the large interference effects allow probing the sign of the top-Yukawa coupling in the context of Eq. (2.53) (see e.g. [174]).<sup>4</sup>

Apart from the processes for single Higgs production mentioned above, Higgs bosons can also be produced in pairs. This is important for measuring the self-coupling of the Higgs boson. While such events are rare, they are already being investigated with good precision [175, 176].

### Higgs decay channels

Unlike the Higgs production channels, its decay channels are not influenced by  $\sqrt{s}$  and the proton PDFs. Instead, the decay channels of the Higgs boson are completely defined via the masses and the couplings of the particles involved in the decay process. The SM predicts that the total decay width of the Higgs boson is

$$\Gamma_H \approx 4 \cdot 10^{-3} \text{ GeV} . \quad (3.5)$$

The fact that  $\Gamma_H \ll m_H$  allows usage of the NWA as defined in Eq. (3.3), for which the production and decay of the Higgs boson factorize.

The most likely decay channel is  $H \rightarrow b\bar{b}$  with a branching ratio of  $BR(H \rightarrow b\bar{b}) = 0.582$ , shown in Fig. 3.5(a). This is due to the  $b$  quark being the particle with the highest mass for which  $2m_X < m_H$  and the fermionic decay width scaling with  $m_f^2$ . Although over half of the produced Higgs bosons decay into a pair of  $b$  quarks, this channel usually does not provide the highest sensitivity in analyses at the LHC and consequently does not belong to the discovery channels of the Higgs boson. The large QCD background at  $pp$  colliders makes this final state difficult to extract. Nonetheless, it remains an important channel for many analyses and allows probing the bottom-Yukawa coupling. Other fermionic decay channels, such as  $H \rightarrow \tau^+\tau^-$ ,  $H \rightarrow c\bar{c}$ , or  $H \rightarrow \mu^+\mu^-$  are suppressed compared to the  $b\bar{b}$  decay and

<sup>4</sup> A similar interference pattern can be found in the sub-leading  $gg \rightarrow ZH$  process, but it is suppressed.

become increasingly difficult to detect. The leptonic decays have the advantage of emitting electrons or muons in the final state, which are significantly easier to tag than quark jets.

The second highest decay width of the Higgs boson can be found in the  $H \rightarrow WW^*$  decay with  $BR(H \rightarrow WW^*) = 0.214$ . Since  $2m_W > m_H$ , this decay channel is kinematically forbidden, and one of the  $W$  bosons must be produced off-shell, which is denoted by  $W^*$ . The large coupling of the Higgs boson to the  $W$  boson compensates for the off-shell suppression. Both  $W$  bosons are produced favorably with small momenta. The case of the  $H \rightarrow ZZ^*$  decay is similar, except that the even higher mass of the  $Z$  boson suppresses the branching ratio to  $BR(H \rightarrow ZZ^*) = 2.62 \cdot 10^{-2}$ . Despite this, the  $H \rightarrow ZZ^* \rightarrow 4l$  decay was among the two channels used to discover the Higgs [1, 2]. The four leptons in the final state make the decay very clean, and it is therefore dubbed the "golden channel". The weak boson decay is pictured in Fig. 3.5(b).

Finally, the Higgs boson can decay into massless particles via loop-induced processes. The most relevant channel of this form is the  $H \rightarrow \gamma\gamma$  decay, shown in Fig. 3.5(c), which is mediated by a fermion or  $W$  boson loop. As in the gluon fusion production, the amplitude of the fermionic loop scales with  $m_f^2$  and therefore the top quark gives the dominating fermionic contribution. The overall largest contribution to the decay is provided by the  $W$  boson loop. However, it interferes destructively with the top quark loop, resulting in a small branching ratio of  $BR(H \rightarrow \gamma\gamma) = 2.27 \cdot 10^{-3}$ . Like the  $H \rightarrow 4l$  decay, the diphoton final state is experimentally very clean, which made it the second of the two discovery channels. Other possible loop-induced decays are the  $H \rightarrow Z\gamma$  and  $H \rightarrow gg$  channels. While the former has a smaller branching ratio than  $H \rightarrow \gamma\gamma$ , the latter has the third largest partial decay width due to the absence of the destructively interfering  $W$  boson loop. Still, the  $H \rightarrow gg$  channel is experimentally not relevant, since the huge QCD background makes it very hard to detect.

### 3.3.2 Current experimental constraints

#### Signal strengths

Over the last years since the discovery of the Higgs boson, its production and decay channels have been studied intensively and with varying precision. A summary of the production mode measurements performed by the CMS experiment can be found in Fig. 3.6(a). The signal strength of  $ggH$  production is the best constrained ( $< \pm 10\% @ 1\sigma$ ). The error is dominated by the systematic uncertainties, as there is a lot of statistics from the high cross section of this channel. The second-best constrained is the VBF signal strength, with a 15% uncertainty but a higher statistical error. The  $VH$  and  $t\bar{t}H$  production modes have larger uncertainties of around 20% – 25% which are all dominated by the comparably low statistics. All measurements agree with the SM prediction within  $2\sigma$ , except for the  $tH$  production cross section. However, this channel suffers from very low statistics as well as challenging separation from the  $t\bar{t}H$  process, which poses as the main background.

For the decay modes, the best constrained signal strengths are in the  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow ZZ$ ,  $H \rightarrow WW$  and  $H \rightarrow \tau\tau$  channels, as can be seen in Fig. 3.6(b). They all have a precision at  $1\sigma$  of around 10%. The  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ$  channels benefit from a low systematical uncertainty, as they were also the two discovery channels, while the  $H \rightarrow WW$  and  $H \rightarrow \tau\tau$  decays profit from a large branching ratio and low QCD background. The  $H \rightarrow b\bar{b}$  decay, despite its large branching ratio, is constrained to a precision of only 20% due to the difficulties of separating it from the QCD background. Finally, the  $H \rightarrow \mu\mu$  and  $H \rightarrow Z\mu$  signal strengths exhibit large uncertainties due to their low statistics, although recent results

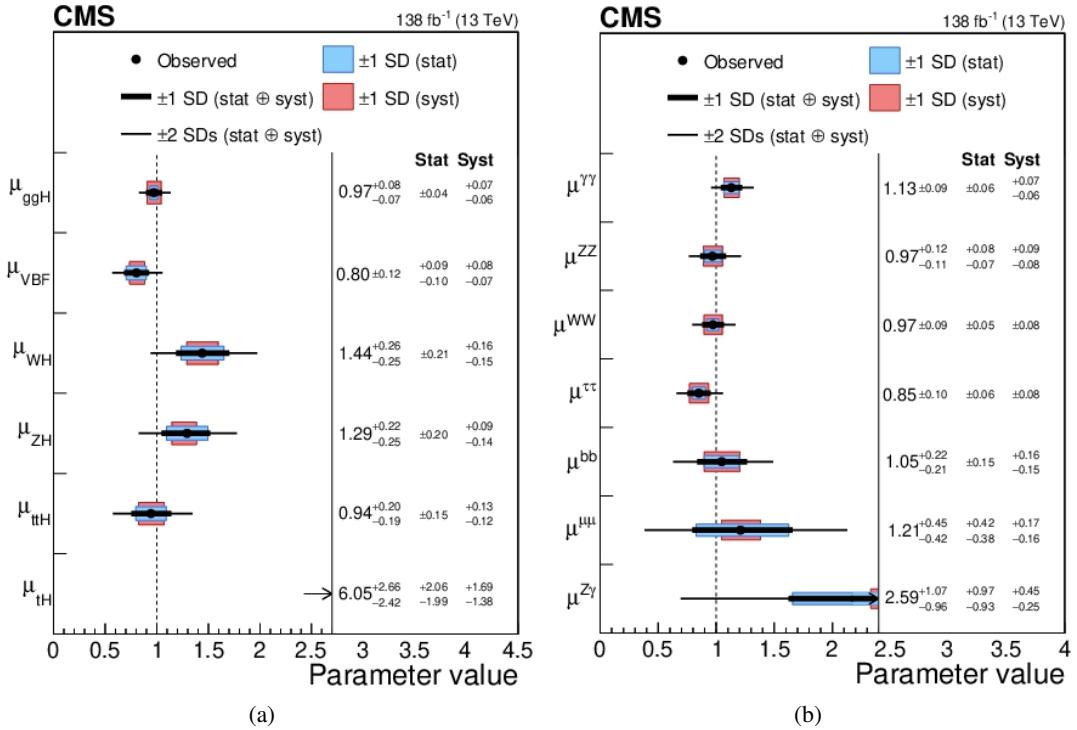


Figure 3.6: Combined cross section measurements from CMS in (a) the Higgs production channels and (b) the Higgs decay channels, including the standard deviations ( $\text{SD} \equiv \sigma$ ). Figures taken from [168].

show significant improvements [177]. All measurements agree with the SM value within  $2\sigma$ .

The production and decay channel measurements can be used to extract constraints on the size of the Higgs couplings to the SM particles. The results of the ATLAS experiment for the various Higgs coupling measurements are shown in Fig. 3.7, where  $\kappa_i = \mu_i / \mu_i^{\text{SM}}$  is the ratio of the measured signal strength  $\mu$  and the SM prediction. The best constrained couplings are the (effective) Higgs couplings to the vector bosons, as well as the  $\tau$  lepton, due to the respective well-constrained production or decay channels. Constraints on the Higgs couplings to the quarks of the third generation are slightly weaker as they originate from  $t\bar{t}H$  production or the  $H \rightarrow b\bar{b}$  decay, respectively. Only very weak constraints can be put on  $\kappa_\mu$  and  $\kappa_{Z\gamma}$ . Again, no notable deviation from the SM can be found. Only in the case that invisible and undetectable decays of the Higgs boson are allowed as free parameters do the bottom- and  $\tau$ -Yukawa couplings show very slight deviations.

### The spin and CP states

Apart from its couplings to other SM particles, the quantum numbers of the Higgs boson  $J^{CP}$  can also be probed at the LHC. Due to the Higgs particle being a boson, only integer spins  $J$  are allowed. A spin of  $J = 0$  corresponds to a scalar particle as predicted by the SM and is experimentally favored. The  $J = 1$  hypothesis would make the Higgs a vector boson. However, it can be confidently excluded by the observation of the  $H \rightarrow \gamma\gamma$  decay according to the Landau-Yang theorem [178]. While this is not the case for a  $J = 2$  Higgs boson, the resulting differing kinematics of such a particle allow the exclusion of this hypothesis with very high confidence. It can be probed in various channels such as  $VH$  production,

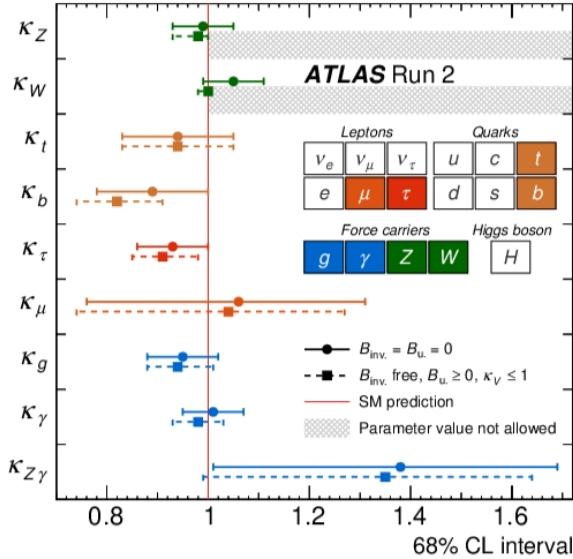


Figure 3.7: ATLAS measurements of the (effective) coupling strengths of all SM particles to the Higgs boson, including the  $1\sigma$  level (= 68% CL). Figure taken from [167].

or the  $H \rightarrow ZZ$  and  $H \rightarrow WW$  decays [33–35].

The  $\mathcal{CP}$  state of the Higgs boson is predicted to be even in the SM, and the results in Figs. 3.6 and 3.7 are also only valid under this hypothesis. Charge conjugation alone is very hard to test experimentally, and there are no studies that set strong limits on the  $C$  state of the Higgs boson. Instead, experimental analyses concentrate on  $\mathcal{P}$ -odd and  $C$ -even observables, which can be used to define genuine  $\mathcal{CP}$ -odd observables (see e.g. the discussion in [179]). The parity nature has a large impact on the kinematics of the Higgs decay products. The parity-odd case  $\mathcal{P} = -1$  is excluded with high precision, translating to an exclusion of the pure  $\mathcal{CP}$ -odd hypothesis of the Higgs boson [33–35].

Since  $C$  and  $\mathcal{P}$  are internal symmetries of the Higgs boson, it could also be a mixed  $\mathcal{CP}$  state. The  $\mathcal{CP}$ -mixed couplings for a spin-zero Higgs boson are parameterized in Eqs. (2.53) and (2.61) to (2.63). The tightest experimental constraints are currently placed on the  $\mathcal{CP}$  nature of the interaction between the Higgs and the weak gauge bosons, which are obtained in the  $H \rightarrow WW$  and  $H \rightarrow 4l$  channels [36–39]. Constraints are set on the three Wilson coefficients  $C_{\Phi\widetilde{W}}$ ,  $C_{\Phi\widetilde{W}B}$ , and  $C_{\Phi\widetilde{B}}$ , which give rise to anomalous  $HVV$  interactions. All of them are observed to have an absolute value smaller than one at the  $1\sigma$  level. Furthermore,  $\mathcal{CP}$  violation in these couplings is loop-suppressed for simple extensions of the SM, due to the non-existence of a tree-level coupling between a pseudoscalar Higgs and heavy vector bosons.

$\mathcal{CP}$  violation in the Yukawa couplings is much less constrained at the LHC. In the case of the  $\tau$ -Yukawa coupling, information can be obtained from the  $H \rightarrow \tau^+\tau^-$  decay with a subsequent decay of the  $\tau$  leptons [180]. A scalar and pseudoscalar Higgs boson give different  $\tau$  spin dependences on the differential decay width in the Higgs boson rest frame. In terms of the  $\mathcal{CP}$ -mixing angle from Eq. (2.54), it can be expressed as

$$d\Gamma(\phi \rightarrow \tau^+\tau^-) \propto 1 - s_z^+ s_z^- + |s_T^+| |s_T^-| \cos(\varphi_{s_T} - 2\alpha_\tau) \quad (3.6)$$

where  $s_z$  and  $s_T$  are the longitudinal and transverse spin components of the  $\tau$ , respectively, and  $\varphi_{s_T}$  is the angle between  $s_T^+$  and  $s_T^-$ . The spin correlation information of the  $\tau^+$  and  $\tau^-$  can be accessed from the charged  $\tau$  decay products by measuring their momentum. This simplifies Eq. (3.6) to

$$d\Gamma(\phi \rightarrow \tau^+ \tau^-) \approx 1 - b(E_+)b(E_-) \frac{\pi^2}{16} \cos(\varphi^* - 2\alpha_\tau) \quad (3.7)$$

where  $\varphi^*$  is now the angle between the two  $\tau$  decay planes, spanned by the  $\tau$  momentum and the charged decay product momentum.  $b(E_+), b(E_-) \in [-1, 1]$  are spectral functions providing the spin analyzing power of a specific  $\tau$  decay product [181]. Charged pions have a maximal spin analyzing power of 1. Limits on  $\alpha_\tau$  have been placed by ATLAS (CMS) to be  $\alpha_\tau \in [-7^\circ, 25^\circ]([-20^\circ, 18^\circ])$  @  $1\sigma$  [182, 183].

While bottom quarks in the  $H \rightarrow b\bar{b}$  decay also follow the distribution in Eq. (3.6), the QCD radiation and hadronization wash out the spin information before a measurement in the detector is possible. An exception to this is the fragmentation of  $b$  quarks into  $\Lambda_b$  baryons, which keep the polarization information and decay weakly [184]. However, the uncertainties of such a measurement are too high at the LHC to put meaningful constraints on  $\alpha_b$  [57, 58]. Therefore, the only information about the  $\mathcal{CP}$  state of the bottom-Yukawa coupling can be obtained indirectly, for example via the total rate of the decay [27].

For the top quark, Eq. (3.6) is not valid due to the heavily suppressed decay mode. While indirect constraints on the  $\mathcal{CP}$  state of the top-Yukawa coupling are obtainable from ggF and the  $H \rightarrow \gamma\gamma$  decay, as further discussed below, a direct probe is only possible using top-associated Higgs production. Like in the  $\tau$ -Yukawa case, the top momentum and top polarization are needed to construct  $\mathcal{CP}$ -odd observables, which take the form of triple products (see e.g. [185–189]). In contrast to bottom quarks, top quarks decay before hadronization, preserving the polarization information in the  $W$  bosons. However, reconstructing and obtaining the polarization information of both top quarks is very challenging, and the interference term in the  $t\bar{t}H$  cross section is small, making a measurement at the LHC very challenging [190]. Instead, studies constraining  $\alpha_t$  make use of  $\mathcal{CP}$ -sensitive but  $\mathcal{CP}$ -even observables. Apart from rate information, such observables are sensitive to the  $\mathcal{CP}$ -even parts of Eq. (2.59) via their kinematics but are insensitive to the interference term. The most sensitive study from ATLAS in the  $H \rightarrow \gamma\gamma$  decay channel sets a limit of  $|\alpha_t| < 43^\circ$  at the  $2\sigma$  level, while a similar study from CMS reaches  $|\alpha_t| < 59^\circ$  [191, 192].

The sign of the top-Yukawa coupling is expected to be probed much more efficiently in the  $pp \rightarrow tHj$  and  $pp \rightarrow tHW$  channels. These processes offer two advantages over  $t\bar{t}H$ . First, there are large interference effects between diagrams where the Higgs is emitted from a top quark and where it is emitted from a  $W$  boson. Second, there is only one top quark that needs to be reconstructed. Still, the low statistics of the  $tHj$  and  $tHW$  processes and the difficulties of disentangling them from  $t\bar{t}H$  make a  $\mathcal{CP}$ -sensitive study in these channels alone not feasible at the moment [193].

Finally,  $\mathcal{CP}$  violation can appear in the (effective) coupling of the Higgs boson to gluons and photons. There are two possible ways of parameterizing this. The loop amplitudes of these processes can either be described via the SM prediction with modified Yukawa- and  $HVV$  couplings, or effective couplings, such as the ones in Eqs. (2.61), (2.62) and (2.64). In the former case, the rate measurement of ggF provides the most stringent bounds on the  $\mathcal{CP}$  state of the top-Yukawa coupling, while the rate measurement of  $H \rightarrow \gamma\gamma$  gives sensitivity towards the sign of the top-Yukawa coupling modifier due to interference with the  $W$  boson loop [174]. These results, however, assume that no further BSM physics runs in the loops.

In the latter and more general case, limits on the top-Yukawa coupling are obtained solely from  $t\bar{t}H$ , and the  $\mathcal{CP}$ -mixing in the effective Higgs-gluon and Higgs-photon couplings has to be tested independently. Experimental constraints on an effective mixing angle  $\alpha_g$  are still relatively weak. The best current results were obtained by CMS and constrain  $|\alpha_g| \lesssim 62^\circ$  at the  $2\sigma$  level, while the most recent ATLAS results are not competitive yet [39, 130].<sup>5</sup> There are currently no limits on an effective mixing angle for the Higgs-photon interaction  $\alpha_\gamma$  available owing to the difficulties of measuring the photon polarization at the LHC [56, 194].

### 3.3.3 The Simplified Template Cross Sections framework

Data taking during Run 1 of the LHC not only led to the discovery of the Higgs boson, but also to the first measurements of production and decay channel signal strengths. For measurements starting from Run 2, the so-called *Simplified Template Cross Section* (STXS) framework introduced in [172, 195] was implemented. The concept of the STXS framework is to bin the Higgs production cross sections into specific regions of the phase space. This allows for maximizing the sensitivity of the measurement as well as minimizing the dependence on theoretical uncertainties in the following ways:

- The bins are defined for the specific SM production modes, reducing uncertainties in their fractions. On top of that, it allows for a combination of all possible decay channels as well as a combination of ATLAS and CMS results.
- Within a chosen bin, the theoretical uncertainties as well as drastic changes in the selection efficiency and signal sensitivity are kept under control. For example, production modes can be split into bins with an exclusive number of jets, avoiding uncertainties in the theoretical prediction of other jet multiplicities.
- Phase space regions that are expected to have a high sensitivity towards BSM effects can be isolated in a bin. An example of this is the  $p_T$  shape of the Higgs boson or jets, where in the high- $p_T$  tail, a good signal-to-background ratio of BSM effects is expected. The low number of SM events in this region also reduced the uncertainty of the SM distribution.
- Finally, the STXS framework aims to find the minimal number of bins for which the experimental sensitivity is still maximal to avoid the need to combine a large number of low-statistical measurements. The number of bins can grow with the available luminosity.

Over the years, the STXS framework has undergone multiple updates. Stage 0 refers to the bin-less evaluation of signal strengths during Run 1 of the LHC. Stage 1.0 corresponds to the first introduction of STXS bins in 2016 [172, 195]. During this stage, the ggF process was divided into bins corresponding to jet multiplicity and  $p_T$  of the Higgs boson. The joint VBF and hadronic  $VH$  process was mainly split into low and high  $p_T$  of the leading jet to isolate BSM effects.  $VH$  production with leptonic decays was first divided according to the initial and final state, more precisely into  $q\bar{q} \rightarrow WH$ ,  $q\bar{q} \rightarrow ZH$ , and  $gg \rightarrow ZH$  events, and subsequently into  $p_T$  bins of the vector boson. No binning was specified for  $t\bar{t}H$  or other production modes with lower cross sections, due to the low statistics in these channels at the time.

The STXS framework was first revised in 2019 for Stage 1.1, which mainly included a different BSM

---

<sup>5</sup> It should be noted that the constrain on  $\alpha_g$  assumes the parameterization of Eq. (2.62), where the top quark is integrated out. This differs from the results provided in [130].

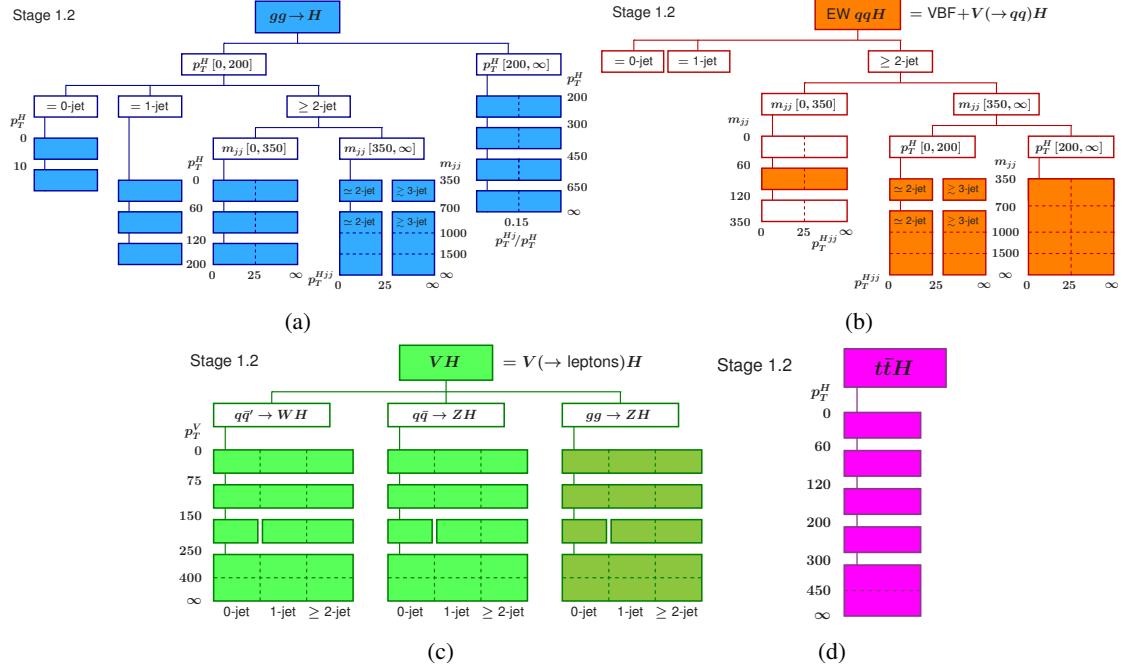


Figure 3.8: Overview of the STXS bins defined in (a) ggF, (b) VBF and hadronic  $VH$ , (c) leptonic  $VH$ , and (d)  $t\bar{t}H$ . Figures taken from [60].

isolation treatment in ggF, overall finer binnings in  $p_T$ , and more precise splits in the jet multiplicity in VBF and hadronic  $VH$  [196]. The current standard for LHC is Stage 1.2, which includes additional bins at high  $p_{T,H}$  in ggF. The most important update, however, is a splitting of the  $t\bar{t}H$  process into 5 + 1 bins in  $p_{T,H}$ , which follow the  $p_{T,H}$  bins in gluon fusion to allow the possibility of combining both channels [59, 60]. The current bins for all major production channels are depicted in Fig. 3.8. An update to Stage 1.3 of the STXS framework for future analyses is currently being finalized.

# CHAPTER 4

---

## Statistical learning in high energy physics

---

*This chapter is based in parts on the discussions in [197]. For reviews, see [198–201].*

The LHC is the biggest science experiment in the world and produces a massive amount of data from particle collisions. CERN stores up to several tens of petabytes of data every year during LHC runs [202]. During Run 2, it is estimated that about  $8 \cdot 10^6$  Higgs bosons were produced [168]. However, the statistics massively decrease when specific channels are considered. For example, in an analysis of the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel with a luminosity of  $\mathcal{L} = 139 \text{ fb}^{-1}$ , the expected number of events amounts to  $N_{\text{exp}} = \sigma_{t\bar{t}H} \cdot \mathcal{L} \cdot \text{BR}(H \rightarrow \gamma\gamma) \approx 160$ . This number will further be reduced when taking into account the selection efficiency of the analysis. This makes data analysis in such a channel a very challenging task. A few signal events have to be sampled from a vast amount of data based on the kinematic information obtained from the detectors.

To tackle such issues, huge progress has been made in the field of statistical learning, which also resulted in the physics Nobel Prize of 2024 [203]. Today, many advanced *machine learning* (ML) techniques are the standard for analyzing data from events at the LHC. This chapter provides a brief introduction and overview of the techniques used. The general idea of supervised learning, along with some simple algorithms, is presented in Section 4.1. Afterwards, decision trees and neural networks will be discussed in Section 4.2 and Section 4.3, respectively. A special focus lies on the interpretability of such models, especially using Symbolic Regression, for which two different approaches used in this work are discussed in Section 4.4.

### 4.1 Supervised learning

In a supervised learning approach, the data is often divided into some input features  $X$  and some known output features  $Y$ . The goal is to develop a model that can learn to predict the correct value of  $Y$  from the information provided in  $X$ . Further discrimination is made between *classification* problems, where the output takes on discrete values, most usually in the form of a one-hot encoded vector, and *regression* tasks, where the output is continuous with a notion of closeness.

A very simple but famous problem for a regression task is linear regression, in which the target values  $y_i$  of some data points  $x_i$  are estimated using a linear function  $\hat{Y} = \beta_1 X + \beta_0$ . In a noisy experiment,

there will be a residual

$$\epsilon_i = y_i - \hat{y}_i \quad (4.1)$$

that measures the error between the prediction and the true data. Noise in an experiment is usually well approximated by a normal distribution, so that the likelihood function for  $N$  independent data points is given by

$$L = \prod_i^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right) \quad . \quad (4.2)$$

The *maximum likelihood estimator* (MLE) is defined as the set of arguments that maximize the (log-)likelihood function

$$\hat{\theta} = \arg \max_{\theta} \log L(\theta) \quad . \quad (4.3)$$

In the case of a normal distribution,

$$\log L = -\frac{1}{2\sigma^2} \sum_i^N (y_i - \hat{y}_i)^2 + \text{const} \quad (4.4)$$

is maximized when

$$\text{LS} = \sum_i^N (y_i - \hat{y}_i)^2 \quad (4.5)$$

is minimized (for fixed  $\sigma$ ). The minimization of Eq. (4.5) is known as the method of *least squares* (LS), which provides the optimal statistical test for linear regression if the errors are normally distributed.

In classification tasks, a similar method can be used. For the simple task that an event  $x_i$  can have two outcomes  $y_i = \{0, 1\}$ , the likelihood function is given by a Bernoulli distribution

$$L = \hat{y}_i^{y_i} (1 - \hat{y}_i)^{1-y_i} \quad (4.6)$$

with the corresponding log-likelihood

$$\log L = \sum_i^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad . \quad (4.7)$$

In the context of ML, discussed in the next sections, Eqs. (4.4) and (4.7) define the two most widely used *loss functions*, the *mean squared error* (MSE) loss and the *cross-entropy* loss.

There are many different algorithms available to solve the minimization problems introduced above. The least squares problem for  $N$  data points and  $M$  input features can be solved exactly by writing

$$\text{LS} = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \quad (4.8)$$

where  $\mathbf{y}$  and  $\beta$  are now  $M$ -vectors and  $\mathbf{X}$  is the  $N \times M$  input matrix. The MLE can be found by differentiating with respect to  $\beta$  which results in

$$\hat{\theta} \equiv \hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad . \quad (4.9)$$

This is known as a *normal equation* and it involves calculating the inverse of the input matrix (if it exists).

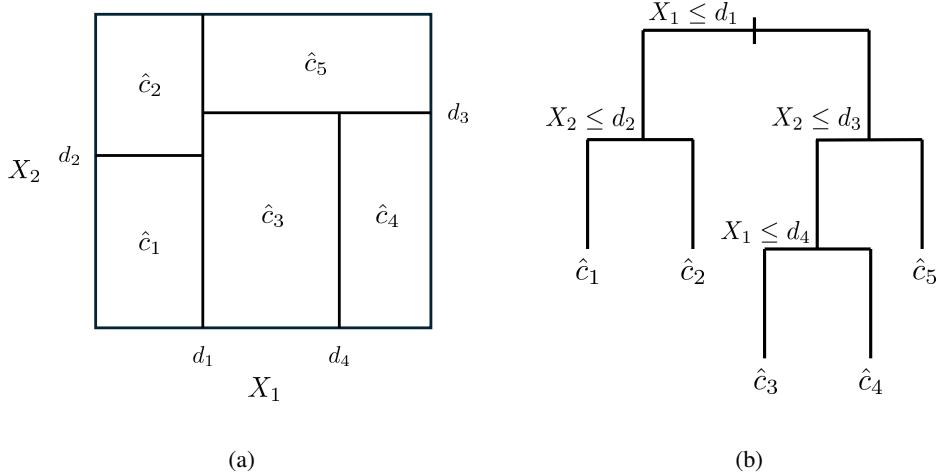


Figure 4.1: Sketch of (a) a two-dimensional parameter space alongside its decision points and boundaries and (b) the corresponding binary decision tree.

Although it provides an exact solution, this method is not efficient for large datasets and in cases where a least squares fit might not provide the best results. In such cases, iterative algorithms can be used. A common example is *stochastic gradient descent* (SGD), which starts with randomly initialized values for  $\beta$  and iteratively updates the parameters via

$$\beta \leftarrow \beta + \eta \frac{\partial \mathcal{L}}{\partial \beta} \quad . \quad (4.10)$$

Here,  $\mathcal{L} \propto -\log L$  is the objective function to minimize and  $\eta$  is the tunable *learning rate*.

## 4.2 Boosted decision trees

So far, only solutions to linear regression problems have been discussed. However, in many cases, the dependence of the target labels on the input is not linear and can also not be modeled by a fit to a simple function. Then, dedicated algorithms are needed that can introduce non-linearities into the model. A very simple method for this is to build so-called (*binary*) *decision trees*. They split the parameter space of the input variables  $\mathbf{X}$  into regions where constant values  $\hat{c}$  are predicted for the output  $Y$ . This is conceptually shown in Fig. 4.1, where the decision boundaries in a 2-dimensional parameter space are depicted in Fig. 4.1(a) and the corresponding decision tree is shown in Fig. 4.1(b).

Binary decision trees are grown by making subsequent splits that are classified by the variable  $X_i$  to split and by the splitting point  $d_i$ . In a regression problem, the optimal split corresponds to minimizing

$$\hat{\theta} = \arg \min_{\theta} \left[ \min_{c_1} \sum_{i \in R_1} (y_i - c_1)^2 + \min_{c_2} \sum_{i \in R_2} (y_i - c_2)^2 \right] \quad (4.11)$$

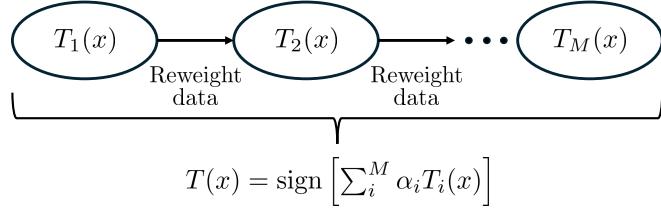


Figure 4.2: Schematic overview of the Adaboost algorithm. Several decision trees are trained based on weighted data from the previous tree. The final output is a weighted sum of the individual trees.

with  $\theta = (X_i, d_i)$  for the exemplary case of two regions  $R_1, R_2$ . The best minimization of the constant prediction  $c_i$  in area  $R_i$  corresponds to

$$\hat{c}_i = \frac{1}{n} \sum_j^n (y_j \mid x_j \in R_i) \quad (4.12)$$

where  $n$  is the number of data points in region  $R_i$ . Trees can also be built for classification problems, in which the arguments of the sums in Eq. (4.11) are replaced by the cross-entropy and the  $\hat{c}_i$  from Eq. (4.12) are taken as the majority value of all  $y_j$  instead. The total number of *leaves* (amount of regions  $R_i$ ) a tree can acquire is either determined by applying a threshold for the amount of improvement a new branch should give, or by building a large tree of fixed size and removing branches via a threshold. The maximum number of decisions a tree can make before reaching its leaves is called the *depth*.

Apart from their simplicity, decision trees offer the advantage of being easily interpretable. Since predictions for the output are made based on several classifications of the input variables, the "physical meaning" of the tree can be well understood. On the other hand, the training process of a single decision tree can be unstable with respect to small changes in the data. This is mainly due to early splits having a major influence on decisions at larger depths. Additionally, decision trees can have issues with modeling parameter spaces that are best described by smooth functions or that are additive in nature.

To avoid the issues that regular decision trees encounter, more complex techniques that combine multiple decision trees were invented. Arguably the most successful of these is the concept of *boosting*, which was first demonstrated using the Adaboost algorithm in classification problems [204, 205]. The concept of building a *boosted decision tree* (BDT) is visualized in Fig. 4.2 for the Adaboost algorithm. First, a basic decision tree is trained on unweighted data samples. Afterwards, the data samples are reweighted based on the error in their predictions, and a new tree is trained on these samples. After  $m$  iterations, the final model is computed as a weighted sum of the individual trees, where the weights  $\alpha_i$  are proportional to the accuracy of the tree  $T_i$ .

While the Adaboost algorithm was initially designed for classification tasks and used an exponential loss function, it can be adapted to work for a wider range of statistical problems. For an arbitrary loss function  $\mathcal{L}$  the parameters of the current tree  $T_m$  are determined by

$$\hat{\theta}_m = \arg \min_{\theta} \sum_{i=1}^N \mathcal{L} \left( y_i, \sum_{j=1}^{m-1} T_j(x_i) + T_m(x_i | \theta) \right) \quad (4.13)$$

which is based on the previous model. Eq. (4.13) can be solved by methods of *gradient boosting* [206]. The model  $f_m(x) = \sum_i^m T_i$  at step  $m$  is found as follows.

$$\begin{aligned}
 (1) \quad g_{im} &= \left[ \frac{\partial \mathcal{L}(y_i, f(x_i))}{\partial f(x_i)} \right] \Big|_{f=f_{m-1}} && \text{Compute the gradient } g_{im}. \\
 (2) \quad \hat{\theta} &= \arg \min_{\theta} \sum_{i=1}^N (-g_{im} - T_m(x_i|\theta))^2 && \text{Fit a tree } T_m \text{ to mimic } g_{im}. \\
 (3) \quad f_m(x) &= f_{m-1}(x) + T_m(x|\hat{\theta}) && \text{Update the model.}
 \end{aligned} \tag{4.14}$$

Step (2) includes finding the regions  $R_i$  and constants  $c_i$  of the tree so that it resembles the negative gradient, which is the computationally most expensive step of the sequence. Gradient boosting works with every differentiable loss function suited to the problem.

### 4.2.1 XGBoost

Among the most successful gradient boosting algorithms is **XGBoost** (eXtreme Gradient Boosting) [207], which won the "Higgs machine learning challenge" in 2014 [208]. Due to its high effectiveness in tackling classification and regression tasks, as well as its acceleration of the training process thanks to parallelization, it has already been implemented in experimental analyses, also in searches for  $\mathcal{CP}$  violation (see e.g. [191, 192]). XGBoost extends the generic gradient boosting algorithm by adding a *regularization term* to the loss function

$$\mathcal{L}_{\text{tot}} = \sum_i^N \mathcal{L}(y_i, \hat{y}_i) + \sum_k \omega(T_k), \quad \omega(T) = \gamma N_{\text{leaves}} + \frac{1}{2} \lambda \sum_{j=1}^{N_{\text{leaves}}} c_j^2 \tag{4.15}$$

where  $\gamma, \lambda$  are tunable parameters and  $c_j$  is the prediction of the leaf, also called the *score*. This prevents the model from *overfitting* to the data during the training and is common among most statistical learning models. Furthermore, alongside the gradient, updates to the model also take the Hessian of the loss function into account. This is achieved by approximating the loss to second order in a Taylor expansion and dropping constants

$$\begin{aligned}
 \hat{\theta}_m &= \arg \min_{\theta} \sum_{i=1}^N \mathcal{L} \left( g_{im} T_m(x_i|\theta) + \frac{1}{2} h_{im} T_m(x_i|\theta)^2 \right) + \omega(T_m), \\
 h_{im} &= \left[ \frac{\partial^2 \mathcal{L}(y_i, f(x_i))}{\partial f(x_i)^2} \right] \Big|_{f=f_{m-1}}.
 \end{aligned} \tag{4.16}$$

Again, like in Eq. (4.14), the minimization of the loss has to be captured by the decision tree  $T_m$ . A tree can be defined as

$$\sum_{i=1}^N T_m(x_i) = \sum_{j=1}^{N_{\text{leaves}}} \sum_{i \in R_j} c_j \tag{4.17}$$

where  $R_j$  contains all data points that get mapped to leaf  $j$ . Inserting Eq. (4.17) into Eq. (4.16) yields

$$\hat{\theta}_m = \arg \min_{\theta} \sum_{j=1}^{N_{\text{leaves}}} \left[ G_{jm} c_j + \frac{1}{2} (H_{jm} + \lambda) c_j^2 \right] + \gamma N_{\text{leaves}} \quad (4.18)$$

with  $G_{jm} = \sum_{i \in R_j} g_{im}$  and  $H_{jm} = \sum_{i \in R_j} h_{im}$ . Eq. (4.18) is a generalized form of Eq. (4.11). The according generalization of Eq. (4.12) is

$$\hat{c}_j = \frac{G_j}{H_j + \lambda} \quad . \quad (4.19)$$

Finally, the structure of the tree that minimizes the loss function is determined by analyzing the gain of a new split. The threshold criterion is already defined by the regularization term. In detail,

$$\text{Gain} = \left( \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda} \right) - \gamma > 0 \quad (4.20)$$

is required where the  $L$  and  $R$  indices label the two respective regions after the split.

### 4.3 Neural networks

Methods of machine learning are based on the idea of mimicking human decision-making by learning from data and being able to make decisions based on that. *Neural networks* (NNs) represent a special class of models within ML. Their name originates from the idea of artificially recreating the structure of the human brain. The nodes and connections of an NN, as shown below, are inspired by the concept of neurons and synapses. Neural networks come in a large variety of specialized structures based on their usage. As such, they are utilized in a wide range of applications, some of which are mentioned below.

The most basic and well-known form of a neural network is a *multi-layer perceptron* (MLP), also sometimes called a hidden-layer perceptron. It is based on the single-layer perceptron, which was first introduced in 1958 and later extended to include non-linearities [209, 210]. The sketch of a simple perceptron can be found in Fig. 4.3(a). Two input features  $X_1$  and  $X_2$  are linearly combined with *weights*  $w_1$  and  $w_2$ , as well as an additional *bias*  $w_0$ , and are passed to a non-linear *activation function*  $a$ . The output is

$$\hat{Y} = a \left( w_0 + \sum_{i=1} w_i X_i \right) \quad (4.21)$$

and  $a$  is generally taken to be a function that maps to the range  $[0, 1]$ . Common examples are a sigmoid function, a step function, or a *rectified linear unit* (ReLU) function.

A more complex MLP is depicted in Fig. 4.3(b) where each of the circles represents a *node*. A collection of vertically stacked nodes forms a *layer*. The left- and rightmost layers are the input and output layers, while the middle one is dubbed the *hidden layer*. It gets its name from the fact that each hidden node is an internal representation of the input features that does not necessarily have a clear interpretation. In principle, there is no limit to the number of hidden layers an NN can have. Here, for

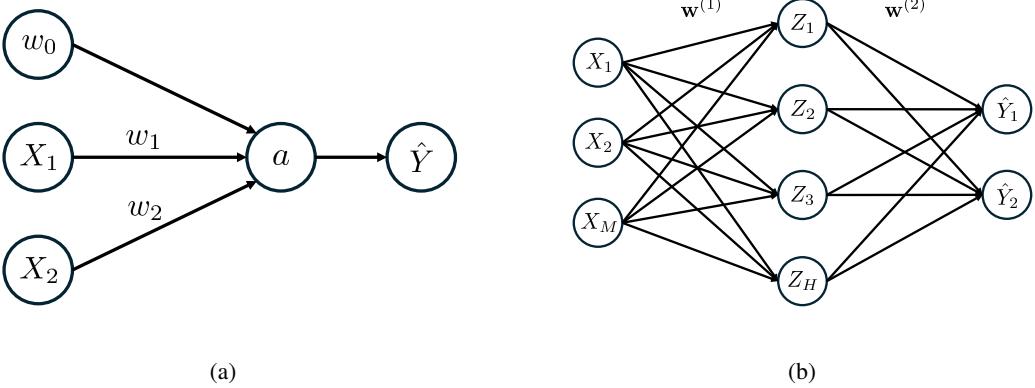


Figure 4.3: (a) The concept scheme for a single basic perceptron, where for clarity the bias  $w_0$  has been included as a node and  $a$  is a non-linear activation function. (b) A multi-layer perceptron with one hidden layer  $Z$ , where the weights and biases are included in the  $w^{(i)}$  and the activation functions are included in each node.

one hidden layer  $Z$ , each node consists of a linear combination of the input layers, as well as a bias

$$Z_i = a^{(1)} \left( w_{0,i}^{(1)} + \sum_{j=1}^M w_{j,i}^{(1)} X_j \right) . \quad (4.22)$$

The nodes of the hidden layers are then further processed to yield the output nodes

$$\hat{Y}_i = a^{(2)} \left( w_{0,i}^{(2)} + \sum_{j=1}^H w_{j,i}^{(2)} Z_j \right) . \quad (4.23)$$

An NN with hidden layers is called a *deep neural network* (DNN). The layers of an MLP are called *fully-connected* or *dense* because every incoming node is connected to every outgoing node.

The main challenge in training an NN is to tune the weights and biases so that  $\hat{Y}$  mimics the true target. This is a very challenging task since the number of trainable parameters for one layer is

$$\#(w^{(1)}) = M_2 \cdot (M_1 + 1) \quad (4.24)$$

for the number of incoming (outgoing) nodes  $M_1$  ( $M_2$ ). The weights and biases are usually trained using gradient descent. Calculating each gradient of the loss function with respect to every single network weight is costly. Therefore, they are calculated through *backpropagation* in which computations are made for each layer going "backward" through the NN, that is, from the output to the input layer [211]. Backpropagation makes use of the fact that the loss function is a composite function of the layers

$$\mathcal{L}(\mathbf{Y}, \hat{\mathbf{Y}}) = \mathcal{L}\left(\mathbf{Y}, f^{(l)}\left(f^{(l-1)}\left(\dots f^{(1)}(\mathbf{X})\dots\right)\right)\right) . \quad (4.25)$$

To obtain the change in loss from a single weight, the chain rule can be applied by calculating

$$\frac{\partial \mathcal{L}}{\partial w_{j,i}^{(l)}} = \frac{\partial \mathcal{L}}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial w_{j,i}^{(l)}} . \quad (4.26)$$

The gradient of a single node can be obtained straightforwardly from the gradient of the activation function and the nodes in the previous layer that influence the current node

$$\frac{\partial z_j^{(l)}}{\partial w_{j,i}^{(l)}} = \frac{\partial a^{(l)}}{\partial w_{j,i}^{(l)}} z_i^{(l-1)} . \quad (4.27)$$

The first term on the right hand side of Eq. (4.26) can be solved making use of Eq. (4.25)

$$g_j^{(l)} = \frac{\partial \mathcal{L}}{\partial z_j^{(l)}} = \sum_{k=1}^{M_k} \frac{\partial \mathcal{L}}{\partial z_k^{(l+1)}} \cdot \frac{\partial z_k^{(l+1)}}{\partial z_j^{(l)}} = \sum_{k=1}^{M_k} g_k^{(l+1)} \left( \frac{\partial a^{(l+1)}}{\partial z_j^{(l)}} w_{j,k}^{(l+1)} \right) . \quad (4.28)$$

Calculating the loss in the *forward pass* and using the gradients  $g_k^{(l+1)}$  in the computation of  $g_j^{(l)}$  in the *backward pass* is the key concept of backpropagation.

The final step of the training process is to update the weights using gradient descent. A common strategy for this is to distribute the training data into small *batches* and update

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{1}{B} \sum_{i=0}^B \frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial \mathbf{W}} \quad (4.29)$$

where  $B$  is the number of batches, depending on the chosen *batch size*, and  $\mathbf{W}$  includes all weights and biases. Using a fixed batch size is a compromise between updating the weights for every single data point independently, which can cause large fluctuations, and using the entire data set, which is computationally costly. Small fluctuations in the training process are wanted to avoid getting stuck in local minima. Apart from the batch size, a suitable learning rate  $\eta$  has to be determined. This is a challenging task because the parameter space of an NN is usually high-dimensional, and  $\eta$  needs to be tuned to deal with local minima or saddle points of the loss function. Furthermore, each weight may have a unique optimal learning rate. Several techniques have become standard for solving this problem, the earliest one being SGD with momentum [212] and more recent ones being Adagrad [213], Adadelta [214], and Adam [215].

In recent decades, a vast amount of novel NN techniques suited for specific problems have been developed. Notable examples include *Convolutional Neural Networks* (CNNs), used for image classification, *Graph Neural Networks* (GNNs), used for connected data, *Variational Autoencoders* (VAEs), used for anomaly detection, and *Generative Adversarial Networks* (GANs), used for data generation. NNs have been increasingly used in all kinds of fields in recent years, and particle physics is no exception. Apart from the previously mentioned 2014 "Higgs machine learning challenge", in 2020 the "LHC Olympics" were hosted, featuring a large number of weakly- or unsupervised algorithms for anomaly detection [216]. An increasingly growing collection of studies employing machine learning techniques in particle physics can be found in the "HEPML Living Review" [201].

## 4.4 Interpretability of machine learning models

Although gradient boosting techniques, deep neural networks, or other complex models significantly outperform linear models and single decision trees, they come at the price of reduced interpretability. While the general goal of a predictive model is to fit some function  $f(X) = Y$ , understanding what a complex model does to approximate  $f(X)$  can be very challenging. DNNs are often associated with "black boxes" for this reason. Not understanding the learning process of an ML algorithm can be a considerable issue for further improving the model. Interpretability can also be helpful for answering questions like whether the model "understands" the physics behind a problem or whether it detects some unknown structure within the data.

### 4.4.1 Shapley values

First insights into a model can be provided by examining the relative importance of the input parameters  $X$  on the model output  $Y$ . One way of doing this is by classifying the input variables via so-called *Shapley* values, which form the basis for the *SHAP* (SHapley Additive exPlanations) framework [217, 218]. Shapley values originate from game theory. The statistical model is interpreted as the "game" in which all input variables participate as "players". For a game with  $|M|$  players, the "worth" (Shapley value) of player  $x_i$  is

$$\phi_i = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|! (|M| - |S| - 1)!}{|M|!} [f(S \cup \{i\}) - f(S)] \quad (4.30)$$

where  $M$  is the set of all players and the sum runs over all subsets  $S$  of  $M$  not containing  $x_i$ . The worth is estimated by training two models  $f$  with and without  $x_i$  and comparing their scores for each subset. The difference in the scores is multiplied by a binomial factor equivalent to the probability of the specific subset appearing.

Summing over all possible subsets of the input features guarantees fairness in determining the worth of all  $x_i$ . However, for large  $|M|$  Eq. (4.30) becomes intractable as it requires training the model  $2^{|M|} - 1$  times. Therefore, the SHAP framework defines a linear explanation model

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{|M|} \phi_i x'_i, \quad x = h_x(x') \quad (4.31)$$

where  $x_i \in \{0, 1\}$  are simplified input features that are mapped to the original input features by  $h_x$ . The SHAP values are then the solutions to

$$\phi_i = \sum_{z' \subseteq x'} \frac{|z'|! (|M| - |z'| - 1)!}{|M|!} [f(h_x(z')) - f(h_x(z' \setminus i))] \quad . \quad (4.32)$$

This allows them to be defined for a wide range of statistical models. The SHAP framework offers several methods for approximating  $f(h_x)$  to evaluate  $\phi_i$  efficiently.

#### 4.4.2 Symbolic regression

While the SHAP values allow assessing the relative importance of the input parameters, they are far from providing a fully interpretable model. This can be achieved by *symbolic regression* (SR). The concept of SR is similar to any other regression model in that it aims to find

$$\hat{f} = \arg \min_{f \in \mathcal{F}} \mathcal{L}(Y, f(\mathbf{X})) \quad (4.33)$$

where  $\mathcal{F}$  is the function space and  $\hat{f}$  is the mathematical expression that minimizes the objective function.  $\mathcal{F}$  is spanned by a set of functions  $\mathcal{S}$  that can be arbitrarily defined. This gives a lot of power to SR because it is possible to fit an expression to data without prior knowledge about its structure. Other regression problems often require a fixed function and only fit its constants, greatly reducing the available search space. However, minimizing an objective in an arbitrarily complex function space is a very challenging and computationally expensive task. In fact, it has recently been shown to be an NP-hard problem [219]. Therefore, prior knowledge about  $\mathcal{F}$  can greatly benefit an SR algorithm.

Optimizing Eq. (4.33) can quickly result in overfitting when the predicted function becomes arbitrarily complex. In the case of DNNs or BDTs, overfitting is usually controlled by adding a regularization term to the loss function. In the case of SR, it can be handled by limiting the complexity of  $\hat{f}$ . This approach allows, e.g., the discovery of the equation for the gravitational force or Kepler's first law from astrophysical data [220, 221]. In particle physics, SR has been used only sparsely so far. However, recent studies show that there is a large number of possible applications (see e.g. [222–228]). The complexity of SR requires advanced algorithms to obtain results efficiently. In this thesis, two of them will be used, which are introduced below.

#### 4.4.3 Symbolic regression with PySR

PySR [229] is an evolutionary (also called genetic) algorithm providing a Python implementation of the Julia package `SymbolicRegression.jl`. Equations are built and displayed as binary trees. Two examples of such trees can be seen in Fig. 4.4. The leaves of the trees consist of constants and variables, which are connected via mathematical operators. In evolutionary algorithms, the performance of an equation tree is called its *fitness*. A large number of trees are being built at the same time, being dubbed a *population*. The population evolves via the Darwinian "survival of the fittest" principle, hence the name evolutionary algorithm. Precisely, updates to trees are performed via:

1. sample a subset  $n_s$  of the population;
2. make a tournament selection by selecting the fittest tree in  $n_s$  with probability  $p$ , else discard it and try with the next fittest;
3. copy the selected tree and apply a simplification, constant optimization, mutation, or crossover with the second fittest tree (see Fig. 4.5 for an example of a crossover);
4. replace the least fit tree with the new tree.

This sequence corresponds to the *inner loop* of PySR. It alternates between phases in which either modifications, simplifications, or optimizations to trees are performed in step 3. This allows the algorithm to explore the entire function space while keeping complexities low and therefore improving performance.

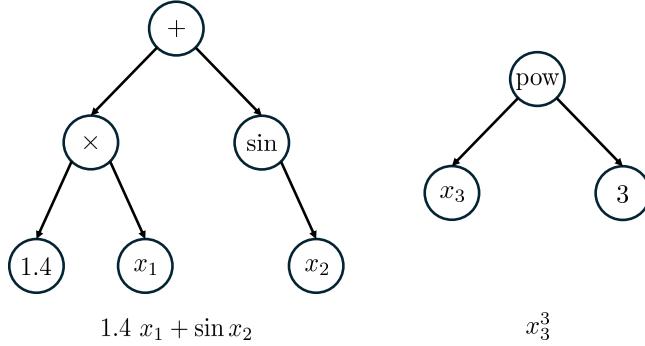


Figure 4.4: Exemplary function trees displaying the functions  $1.4 x_1 + \sin x_2$  and  $x_3^3$  in PySR. Figure taken from [230].

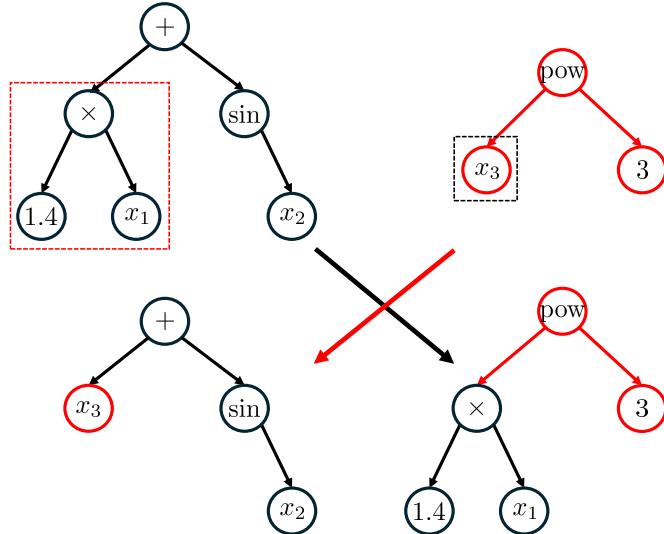


Figure 4.5: A crossover operation in which the red-dashed part of the original tree has been exchanged with the black-dashed part of another tree. As a result the equations are modified to  $x_3 + \sin x_2$  and  $(1.4 x_1)^3$ . Figure taken from [230].

The *outer loop* involves training a set of populations, dubbed *islands*. Since islands evolve independently, the inner loop can be parallelized. After a tunable number of epochs, trees can migrate between islands. Furthermore, the fittest trees are stored in the so-called *hall of fame*. Trees can also migrate to islands from the hall of fame.

In order to avoid overfitting and building arbitrarily large trees, PySR keeps track of the complexity of the individual trees. The complexity corresponds to the number of nodes and can be limited via a maximum complexity parameter. Additionally, PySR allows specifying a *parsimony*, which acts as a regularization term and punishes the complexity of the expressions. Alternatively, an *adaptive parsimony* can be used, which penalizes frequently appearing complexities and allows the algorithm to explore

more of the function space. Finally, PySR allows the use of *simulated annealing*, where mutations of trees can be rejected. In this work, a modified acceptance probability based on [222]

$$p = \exp \frac{L_f - L_i}{\alpha T L_i} \quad (4.34)$$

is used, with the fitness  $L_f$  and  $L_i$  after and before the mutation, the temperature  $T \in [0, 1]$  and a hyperparameter  $\alpha$ . This modification prevents rejecting formulas that are better in shape but not yet optimized.

PySR is highly customizable, allowing the specification of not only the default hyperparameters, but also the implementation of custom operators and custom loss functions. It also allows for flexible constraints on the form of the equation. After running for a fixed number of iterations, PySR returns the hall of fame, where expressions are defined for each complexity.

#### 4.4.4 Symbolic regression with SymbolNet

Genetic algorithms for SR can be very powerful, but they often struggle when the function space is too large, especially in cases involving a high number of input variables or a very complex final equation. An alternative approach to SR is to use a DNN-like structure, as employed by SymbolNet [231], which is based on Tensorflow [232]. Mathematical expressions are built as follows. The typical activation functions are replaced with operators specified via the set of functions  $\mathcal{S}$ . During the training process of the DNN, thresholds can be trained alongside the usual model weights, allowing for the pruning of input features, operators, and connections between nodes. This is called *sparsity training* and serves as a form of regularization. The use of backpropagation to update weights and thresholds enables fast SR, even for a large number of input variables. SymbolNet therefore provides a strong contrast to PySR by starting with the most general equation via fully connected layers and then reducing its complexity. An example of a SymbolNet architecture with one symbolic layer and  $\mathcal{S}_i = \{\sin(\cdot), \text{cube}(\cdot), \text{mult}(\cdot, \cdot)\}$  can be seen in Fig. 4.6.

The mathematical operators and the sparsity training are implemented in SymbolNet via an *input layer* and *symbolic layers*. Both consist of weights and corresponding thresholds, which are listed in Table 4.1 along with their dimension. For this work, SymbolNet has been modified compared to [231] in several ways. First, the training process is split into three steps in the following order:

1. default training  $D$ , which corresponds to the usual training of a DNN without any thresholds;
2. mixed training  $M$ , where both weights and thresholds are trainable, and which corresponds to the implementation in [231];
3. sparsity training  $S$ , where only thresholds are trainable, and which serves as a simplification.

This change is made to stabilize the training. Furthermore, a vector dimension  $l$  is added to SymbolNet such that it supports 4-vectors of the form  $(E, p_x, p_y, p_Z)$ . This can help preserve physical information, although care has to be taken since certain operations can render the 4-vectors unphysical. It is also an advantage compared to PySR, where (binary) trees sit at the core of the algorithm, and a modification using vectors is not straightforward to implement. Adding the vector dimension gives rise to different types of symbolic layers, as further discussed below. The modifications to SymbolNet are reflected in the dimensions of the parameters and their trainability in Table 4.1.

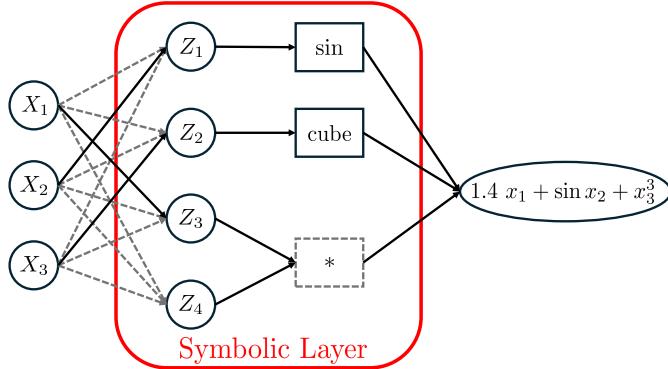


Figure 4.6: Exemplary structure of SymbolNet producing the function  $1.4x_1 + \sin x_2 + x_3^3$ . First, linear combinations of the input features are built. They are subsequently given to the predefined mathematical operators in the symbolic layer. The output of the symbolic layer is passed through another linear transformation to reach the output. Pruned weights and operators are depicted via gray dashed lines. Figure taken from [230].

Layer	Description	Label	Dim.	Values	Trainable	Dim. of trafo
Input layer	Input weights	$\mathbf{W}_{\text{in}}$	$n_{\text{in}}$	1	$\times$	$n_{\text{in}} \times l$
	Input thresholds	$\mathbf{T}_{\text{in}}$	$n_{\text{in}}$	$[0, 1]$	$M, S$	$\rightarrow n_{\text{in}} \times l$
Linear trafo in symbolic layer	Weights	$\mathbf{W}$	$n \times m \times l$	$\mathbb{R}$	$D, M$	
	Weight thresholds	$\mathbf{T}_W$	$n \times m \times l$	$\mathbb{R}^+$	$M, S$	$n \times l$
	Biases	$\mathbf{B}$	$m \times l$	$\mathbb{R}$	$D, M$	$\rightarrow m \times l$
	Bias thresholds	$\mathbf{T}_B$	$m \times l$	$\mathbb{R}^+$	$M, S$	
Symbolic trafo in symbolic layer	Unary weights	$\mathbf{W}_{\text{unary}}$	$n_f$	1	$\times$	
	Unary thresholds	$\mathbf{T}_{\text{unary}}$	$n_f$	$[0, 1]$	$M, S$	$m \times l$
	Binary weights	$\mathbf{W}_{\text{binary}}$	$n_g$	1	$\times$	$\rightarrow k \times l'$
	Binary thresholds	$\mathbf{T}_{\text{binary}}$	$n_g$	$[0, 1]$	$M, S$	

Table 4.1: Definition of all weights and thresholds as used for the training of SymbolNet (see Chapter 7 and [231]), split into the layers in which they appear. Also shown are their dimension, possible values, and the phases they are trainable in.  $m = n_f + 2n_g$  and  $k = n_f + n_g$  are defined via the numbers of unary and binary operators.  $l, l' = \{(1, 3), 1\}$  specify the vector dimension.  $D, M$ , and  $S$  indicate the default, mixed, and sparsity training, respectively. Table taken from [230].

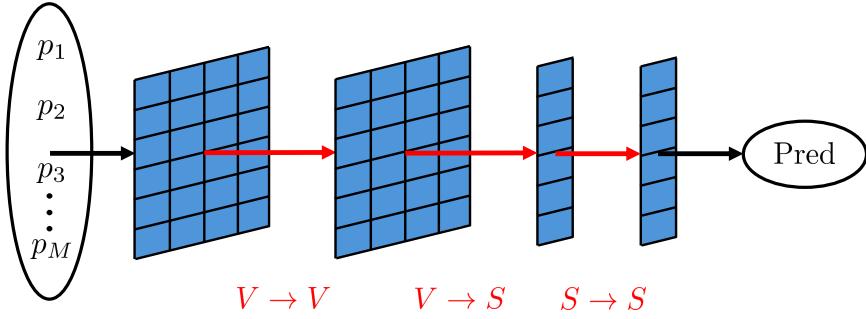


Figure 4.7: Schematic of the **SymbolNet** architecture consisting of one of each layer type and including the input data and output. Figure taken from [230].

The input layer of **SymbolNet** serves as a means to prune input features that are not needed for training. Each input feature is assigned an auxiliary weight fixed at 1. Furthermore, they are assigned a threshold  $t_{in,i} \in [0, 1]$ . Pruning of input features is performed via

$$\mathbf{X}_{in} \rightarrow \mathbf{X}_{in} \theta(\mathbf{1} - \mathbf{T}_{in}) \quad (4.35)$$

where  $\theta$  is the step function. Consequently, if  $t_{in,i} = 1$ , the input feature  $x_i$  is pruned.

As mentioned, the additional vector dimension leads to three types of symbolic layers, as the output of **SymbolNet** must be scalar-valued:

- $V \rightarrow V$  layers, that transform the input according to  $\mathbb{R}^{(n,1,3)} \rightarrow \mathbb{R}^{(k,1,3)}$  (where  $\mathbb{R}^{(1,3)}$  signals the Minkowski space). The layer can contain scalar operators that are applied component-wise. Furthermore, it can contain additional functions that are only defined for 4-vectors, such as a Lorentz boost.
- $V \rightarrow S$  layers, that remove the vector dimension by applying functions that map  $\mathbb{R}^{(n,1,3)} \rightarrow \mathbb{R}^{(k,)}$ . Each network that takes 4-vectors as input must contain exactly one of these layers. The set of functions that defines this layer is

$$\mathcal{S}_{V \rightarrow S} = \underbrace{\{p_0, p_x, p_y, p_z, \|p\|_3, \|p\|_4, \langle p_i \times p_j \rangle_3, \langle p_i \times p_j \rangle_4\}}_{\text{unary operators}} \quad \underbrace{\{ \}}_{\text{binary operators}} \quad .$$

The  $p_i$  are projection operators,  $\|p\|_3$  and  $\|p\|_4$  are the Euclidian and the Minkowskian norm, and  $\langle p_i \times p_j \rangle_3$  and  $\langle p_i \times p_j \rangle_4$  are vector products.

- $S \rightarrow S$  layers, that are the symbolic layers used in [231] and transform  $\mathbb{R}^{(n,)} \rightarrow \mathbb{R}^{(k,)}$ .

A schematic of a **SymbolNet** setup using one of each type of symbolic layer is shown in Fig. 4.7.

Symbolic layers generally consist of two steps. First, the incoming features are linearly transformed, and second, the operators are applied. The linear transformation is similar to other DNNs as it contains a weight matrix  $\mathbf{W}$  and a bias vector  $\mathbf{B}$ . For the sparsity training, they are associated with weight and bias

thresholds  $\mathbf{T}_W$  and  $\mathbf{T}_B$  of the same dimension. Pruning works similarly to the input layer

$$\begin{aligned} \mathbf{W} &\rightarrow \mathbf{W} \theta(|\mathbf{W}| - \mathbf{T}_W), \\ \mathbf{B} &\rightarrow \mathbf{B} \theta(|\mathbf{B}| - \mathbf{T}_B) \end{aligned} \quad . \quad (4.36)$$

The mathematical operators replacing the activation functions are divided into unary operators  $f(p)$ , which transform a single feature, and binary operators  $g(p_i, p_j)$ , which combine two features. Again, they are assigned a weight and threshold vector each, where the weight is fixed to 1 and the threshold is clipped to  $[0, 1]$ . The pruning is defined as

$$\begin{array}{ll} V \rightarrow V \& S \rightarrow S: & \begin{aligned} f_i(a) &\rightarrow f_i(a) \theta(1 - t_{\text{unary},i}) + a \left(1 - \theta(1 - t_{\text{unary},i})\right) \\ g_i(a, b) &\rightarrow g_i(a, b) \theta(1 - t_{\text{binary},i}) + (a + b) \left(1 - \theta(1 - t_{\text{binary},i})\right) \end{aligned} \\ \hline & V \rightarrow S: \\ & \begin{aligned} f_i(a) &\rightarrow f_i(a) \theta(1 - t_{\text{unary},i}) \\ g_i(a, b) &\rightarrow g_i(a, b) \theta(1 - t_{\text{binary},i}) \end{aligned} \end{array} \quad (4.37)$$

for the different types of symbolic layers. In the  $V \rightarrow V$  and  $S \rightarrow S$  layers, a pruned unary operator returns the identity, while a pruned binary operator returns an addition. Pruned operators in the  $V \rightarrow S$  layer return zero.

To ensure a dynamic pruning in the backward pass of the training process, a smooth function is needed instead of the step function. Here,

$$\frac{d\theta(x)}{dx} \simeq \frac{\kappa e^{-\kappa x}}{(1 + e^{-\kappa x})^2} \quad (4.38)$$

is used with the default value of  $\kappa = 5$ .

The objective function consists of

$$\mathcal{L} = \mathcal{L}_{\text{base}} + \mathcal{L}_{\text{sparse}} \quad (4.39)$$

where the MSE or cross-entropy loss is used as  $\mathcal{L}_{\text{base}}$  for regression or classification tasks, unless mentioned otherwise.  $\mathcal{L}_{\text{sparse}}$  controls the sparsity training by punishing equations of high complexity. Specifically,  $\mathcal{L}_{\text{sparse}}$  consists of

$$\begin{aligned} \mathcal{L}_{\text{sparse}} &= \sum_{\text{obj}} \mathcal{L}_{\text{sparse}}^{\text{obj}} & \text{obj} \in \{\text{in}, W, B, \text{unary}, \text{binary}\} \\ \mathcal{L}_{\text{sparse}}^A &= \mathcal{L}_{\text{base}} f D(S_A; \alpha_A, d) \frac{1}{n_A} \sum_{i=1}^{n_A} \exp(-t_{A,i}) & A \in \{W, B\} \\ \mathcal{L}_{\text{sparse}}^B &= \mathcal{L}_{\text{base}} f D(S_B; \alpha_B, d) \exp\left(-\frac{1}{n_B} \sum_{i=1}^{n_B} t_{B,i}\right) & B \in \{\text{in}, \text{unary}, \text{binary}\} \end{aligned} \quad (4.40)$$

where  $f = 0.1$  is a constant scaling factor,  $n_{\text{obj}}$  counts the number of weights for a specific object, and

$t_{\text{obj},i}$  are the thresholds. The term

$$D(S; \alpha, d) = \exp \left[ - \left( \frac{\alpha}{\alpha - \min(S, \alpha)} \right)^d + 1 \right] \quad (4.41)$$

represents a decay factor as a function of the sparsity  $S \in [0, 1]$ , which is defined as the ratio of pruned objects to the total number of objects.  $\alpha \in [0, 1]$  is the target sparsity and  $d$  is a decay rate which is set to the default value of  $d = 0.01$ . For  $S \geq \alpha$ , the decay factor approaches zero and disables the sparsity loss punishment.

## Classifying the $\mathcal{CP}$ properties of the $ggH$ coupling in $H + 2j$ production

---

*This chapter is based on the work in [233] which was done in collaboration with Henning Bahl, Elina Fuchs, and Marc Hannig. In comparison to [233], the evaluation of the confidence levels is slightly different here (for details see Appendix C). Furthermore, a minor correction to the amplitude of the VBF Higgs background has been applied. The qualitative results and conclusions remain unchanged. All results taken from [233] have been produced by the author of this thesis unless stated otherwise.*

The Higgs coupling to gluons is an interesting candidate for studies about the  $\mathcal{CP}$  character of the Higgs boson and its interactions, since gluon fusion is the primary Higgs production mode at the LHC. It is related to the top-Yukawa coupling due to its loop-induced nature, which is important for  $\mathcal{CP}$  studies as well, because it is the only Yukawa coupling of order 1. A direct method to probe the top-Yukawa coupling is via top-associated Higgs production [27, 43, 129, 174, 186–188, 190, 193, 234–260], which will be further examined in this work in Chapters 6 and 7. However, the ggF process is also rich in information about the top-Yukawa coupling, albeit some assumptions are necessary. Without additional hard jets, ggF production is sensitive to the  $\mathcal{CP}$  nature of the top-Yukawa interaction only via the total rate of the process (see e.g. [174]). If ggF production in association with additional jets is considered, kinematic variables can be exploited as well. In particular,  $\mathcal{CP}$ -odd variables can be built already if only one additional jet is considered; however, jet substructure information is needed in this case [261]. A more straightforward way is to consider gluon fusion in association with two jets (ggF2j), where a  $\mathcal{CP}$ -odd observable  $\Delta\phi_{jj}$  can be easily constructed by measuring the azimuthal angles of the two jets [239, 260, 262–274].

As already mentioned in Section 3.3.2, the ggF2j production process only provides an indirect probe of the  $\mathcal{CP}$  state of the top-Yukawa coupling. Instead, it probes an effective coupling between the Higgs boson and gluons.  $\mathcal{CP}$  violation in this coupling can stem either from  $\mathcal{CP}$  violation in the top-Yukawa interaction or the interaction of the Higgs boson with unknown colored BSM particles. Candidates for the latter are, for example, leptoquarks or supersymmetric particles, which can be searched for at the LHC. However, no signs of such particles have been found so far (for overview plots see e.g. [275, 276]), implying huge BSM masses or tiny couplings. Furthermore, the contribution of other quarks to the ggF loop is negligible compared to the top quark contribution. Therefore, any sign of  $\mathcal{CP}$  violation in an

effective Higgs-gluon coupling measurable via ggF2j also hints towards  $\mathcal{CP}$  violation in the Higgs-top coupling. Consequently, the results in this chapter will not only be used to constrain the coupling modifiers of an effective Higgs-gluon coupling but also to put limits on the top-Yukawa coupling.

Eq. (2.59) shows how parameterizing  $\mathcal{CP}$ -violating effects with two coupling modifiers  $c_g$  and  $\tilde{c}_g$  will lead to three terms in the matrix element, two of which are  $\mathcal{CP}$ -even. Most kinematic variables measured at the LHC are  $\mathcal{CP}$ -even, but are sensitive to  $\mathcal{CP}$  violation due to changes in their shape. Such variables can therefore be used to constrain the two  $\mathcal{CP}$ -even terms. To gain sensitivity towards the interference term, dedicated  $\mathcal{CP}$ -odd variables, such as  $\Delta\phi_{jj}$ , have to be built. These offer the additional advantage of providing a clear sign of  $\mathcal{CP}$ -violation if a deviation from the (symmetric) SM-distribution is measured. However, measuring a  $\mathcal{CP}$ -odd observable for a Higgs coupling at the LHC can be challenging since four independent momenta from the Higgs production or decay are needed [179]. As stated in Section 3.3.2, such measurements in  $t\bar{t}H$  are not feasible with the current luminosity. Therefore,  $\mathcal{CP}$ -sensitive observables are either used primarily or to provide complementary information.

As discussed in Section 3.3.2, the existing constraints on  $\alpha_g$  from LHC measurements are currently relatively weak, with the most sensitive study yielding  $|\alpha_g| \lesssim 62^\circ$  [130]. Limits from EDM measurements can yield similar results, but they rely on the assumption that other couplings are SM-like [277]. Therefore, in this chapter, possible improvements of these limits are investigated. A special focus is put on the comparison of two potential kinematic regions in which the  $\mathcal{CP}$  state is tested (a VBF-like and a ggF2j-like kinematic region). Furthermore, machine learning techniques are used first to identify the Higgs production process and then to determine which part of the matrix element in Eq. (2.59) the events originate from.

The chapter is organized as follows. First, the ggF2j process and its main backgrounds, as well as data generation for each process, are discussed in Section 5.1. Next, the setup and results of the classifier training are presented in Section 5.2. Based on these results, exclusion limits can be obtained that are shown in Section 5.3. Section 5.4 deals with the possibility of additional  $\mathcal{CP}$  violation in the  $HVV$  coupling and examines the impact on the previous results. An interpretation of the obtained exclusion limits in terms of the  $\mathcal{CP}$  state of the top-Yukawa coupling is given in Section 5.5. Lastly, Section 5.6 presents the conclusions and describes the prospects for future work.

## 5.1 The topology and generation of $H + 2j$ events

Fig. 3.3 demonstrates how gluon fusion is the dominant Higgs production channel at the LHC, with a cross section that exceeds those of all other production modes by at least an order of magnitude at the collision energies at the LHC. Even if an STXS-like cut of  $N_{\text{jets}} \geq 2$  is placed on ggF events, the cross section remains higher than for VBF, the second most likely production mode. More specifically,  $\sigma_{\text{ggF}}^{j \geq 2} = 7.88 \text{ pb}$  and  $\sigma_{\text{VBF}} = 3.78 \text{ pb}$  at  $\sqrt{s} = 13 \text{ TeV}$  [172]. The topology of the ggF2j process differs depending on its initial state, which can be divided into  $gg$ ,  $gq$ , and  $qq$ . Representative Feynman diagrams for each initial state are shown in Fig. 5.1, with effective interactions represented by a red dot. Initial states involving gluons contribute more significantly due to the higher gluon densities in the proton's parton distribution functions (PDFs). More precisely, in the SM case, the  $gg$  initial state accounts for  $\sim 72\%$  of the total ggF2j events, followed by the  $gq$  initial state, which contributes  $\sim 26.5\%$ . The smallest contribution with  $\sim 1.5\%$  is given by the  $qq$  initial state.

Due to the two additional jets, the background processes of ggF2j include other Higgs production

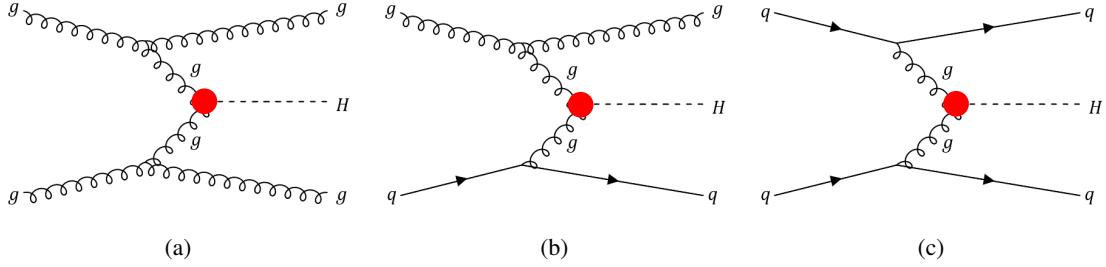


Figure 5.1: Exemplary Feynman diagrams for Higgs production via ggF in association with two jets in the (a)  $gg$  initial state, (b)  $gq$  initial state, and (c)  $qq$  initial state. Effective couplings are marked by a red circle. Diagrams modified from [233].

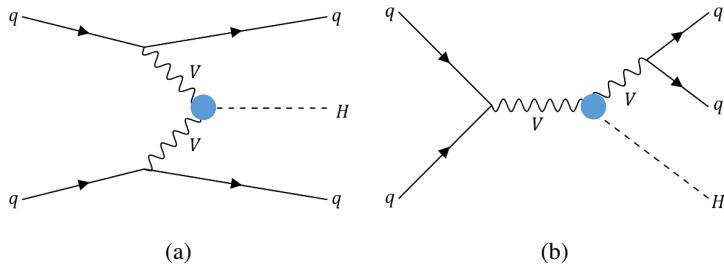


Figure 5.2: Exemplary Feynman diagrams for the Higgs background processes of ggF2j, namely (a) VBF, and (b)  $VH$ . Effective couplings are marked by a blue circle. Diagrams modified from [233].

processes, mainly VBF and  $VH$  production. While the corresponding non-Higgs background based on the Higgs decay channel typically has a larger cross section, the Higgs background is interesting as it may contain similar  $\mathcal{CP}$ -violating effects as the ggF2j signal. Feynman diagrams for the background considered, including a potential  $\mathcal{CP}$ -violating coupling, are shown in Fig. 5.2. Although interference between the  $qq$ -induced ggF2j production and Higgs production via VBF exists, it was found to be negligible and will therefore not be considered here [278, 279].

The parameterization of  $\mathcal{CP}$ -violating effects in the  $Hgg$  and  $HVV$  couplings follows Section 2.2.2. First, only a modified  $Hgg$  coupling will be considered. To enable a comparison between the Higgs-top and Higgs-gluon coupling later, Eq. (2.62) is defined such that there is a direct correspondence of the two when no other particles besides the top quark are considered in the loop. While the absence of colored BSM particles can only be assumed based on current experimental results, the infinite top mass limit allows to ignore the effects of lighter quarks. Here, this is accounted for by fixing the total rate to the SM and limiting the transverse momentum of the Higgs, as discussed in more detail below.

In addition to a modification of the Higgs-gluon interaction, Section 5.4 will consider  $\mathcal{CP}$  violation in the interaction between the Higgs boson and the weak vector bosons. This is parameterized using the SMEFT operator  $\mathcal{O}_{\Phi\widetilde{W}}$ . Technically, non-zero values of the  $\mathcal{O}_{\Phi\widetilde{B}}$  and  $\mathcal{O}_{\Phi\widetilde{W}B}$  SMEFT operators also result in  $\mathcal{CP}$  violation in the  $HVV$  coupling (see e.g. [280]). However, the Lorentz structure of the resulting  $\mathcal{CP}$ -violating couplings from all three operators is equal, and they are expected to yield similar results (see e.g. [281]).

Datasets are generated for the ggF2j, VBF, and  $VH$  Higgs production modes, with the Higgs decaying

to a pair of photons in all three cases. The continuous background originating from the  $qq \rightarrow qq$  process with two photons emitted via final state radiation is not included here and is left for future investigations. However, it can be assumed to be  $\mathcal{CP}$ -even and would therefore not influence the classification of the  $\mathcal{CP}$  states, as detailed below. The analysis strategy presented here is adaptable to alternative Higgs decay channels. Parton-level events are generated at LO with `MadGraph5_aMC@NLO` (version 3.4.0) [143] for  $\sqrt{s} = 13$  TeV. They are then scaled to NLO by using flat K-factors from [239] for ggF2j, [282] for VBF, and [283] for  $VH$ . The events are subsequently given to `Pythia8` (version 8.306) [153] and `Delphes3` (version 3.4.2) [155] to simulate the effects of the parton showering, hadronization, and limited detector resolution. For the latter, the ATLAS detector card was used, where the parameter for the reconstructed jet radius has been set to  $\Delta R = 0.4$ . This agrees with the definition commonly used in experimental analyses. The reconstruction of the jets is done by employing the anti- $k_t$  algorithm [158]. Furthermore, a threshold is put on the minimal transverse momentum of the jets and photons by requiring  $p_{T,j} \geq 20$  GeV and  $p_{T,\gamma} \geq 0.5$  GeV.

For ggF2j production, the events are categorized according to the terms in Eq. (2.59), resulting in separate datasets for each of the three contributing components: the pure  $\mathcal{CP}$ -even amplitude squared ( $\propto c_g^2$ ), the pure  $\mathcal{CP}$ -odd amplitude squared ( $\propto \tilde{c}_g^2$ ), and the interference term ( $\propto c_g \tilde{c}_g$ ).  $1.5 \cdot 10^6$  ( $3 \cdot 10^6$ ) events were generated at parton level for the squared (interference) terms. Datasets for VBF and  $VH$  are generated separately, both with and without  $\mathcal{CP}$  violation from  $O_{\Phi\widetilde{W}}$ , and consist of  $1 \cdot 10^6$  parton-level events each. If not mentioned otherwise, the  $\mathcal{CP}$ -even Higgs background events are used. In total, five event classes are used for the analysis chain. The events are further split into training and evaluation datasets, as detailed below. To ensure that the classifiers learn the  $\mathcal{CP}$  state and not exploit statistical fluctuations, all generated events are  $\mathcal{CP}$ -flipped and added to the respective dataset. Additional information about the event generation, including the `MadGraph` syntax used, is provided in Appendix B.

To obtain a realistic selection of events, several cuts are applied to the events obtained at the detector level. They are closely aligned to cuts set by ATLAS and CMS analyses in the same Higgs decay channel (see e.g. [191, 284]). Table 5.1 shows the cuts in order of their application alongside the percentage of surviving events at each step. The cutflow table is further divided into the five data classes. The initial selection requires a minimum of two jets and two photons per event to perform the Higgs reconstruction and identify ggF2j events. Events are kept only if the diphoton invariant mass lies within the window  $110 \text{ GeV} \leq m_{\gamma\gamma} \leq 140 \text{ GeV}$ , which captures a majority of the Higgs events. Additional thresholds on the transverse momenta of photons  $p_{T,\gamma}$  and jets  $p_{T,j}$  help eliminate contributions from objects not originating from the hard scattering process. Finally, a cut on the pseudorapidity of both jets and photons  $|\eta_j| \leq 2.5$  and  $|\eta_\gamma| \leq 2.5$  mimics the coverage of the inner detector of both ATLAS and CMS.

As previously stated, interpreting the effective Higgs-gluon coupling in terms of a modified top-Yukawa coupling requires the infinite top quark mass limit. To ensure the validity of this assumption, an additional cut is used that limits the transverse momentum of the Higgs boson to  $p_{T,H} \leq 200$  GeV (see [285–287]). Here, the Higgs boson is taken as the diphoton system. The cut on  $p_{T,H}$  can be seen as conservative, as studies on mass effects of the top quark suggest that relaxations of this cut are possible. For instance, an approximation of the Full Theory (FT<sub>approx</sub>) at NLO suggests that events up to  $p_{T,H} \leq 300$  GeV can be used without deviating significantly from the results in the infinite top quark mass limit [288]. A more detailed study of this effect could result in improved limits, but is left for future work.

Applied cut	Fraction of accepted events				
	$ggF2j \,  \mathcal{M}_{\text{even}} ^2$	$ggF2j \, \text{Interf.}$	$ggF2j \,  \mathcal{M}_{\text{odd}} ^2$	VBF	$VH$
Initial events	100%	100%	100%	100%	100%
$N_j \geq 2; N_\gamma \geq 2$	48.1%	50.8%	48.1%	62.6%	49.8%
$110 \, \text{GeV} \leq m_{\gamma\gamma}$	47.8%	50.5%	47.9%	62.0%	49.4%
$m_{\gamma\gamma} \leq 140 \, \text{GeV}$					
$p_{T,\gamma_1}/m_{\gamma\gamma} \geq 0.35$	39.4%	40.9%	39.8%	50.0%	40.5%
$p_{T,\gamma_2}/m_{\gamma\gamma} \geq 0.25$					
$p_{T,j_1} \geq 30 \, \text{GeV}$	38.6%	40.2%	38.6%	49.7%	39.9%
$p_{T,j_2} \geq 20 \, \text{GeV}$					
$ \eta_j  \leq 2.5$	22.9%	21.5%	22.7%	39.8%	31.2%
$ \eta_\gamma  \leq 2.5$					
$p_{T,H} \leq 200 \, \text{GeV}$	18.6%	18.4%	18.3%	34.4%	26.8%

Table 5.1: List of cuts applied to select events at the detector level. For each cut, the percentage of events surviving it is listed. Table taken from [233].

## 5.2 Analysis strategy

The outline of the analysis is shown as a flowchart in Fig. 5.3. In a first step, two binary classifiers are trained on the samples for  $ggF2j$ , VBF, and  $VH$  production. One of the classifiers obtains events that label  $ggF2j$  events as the signal and the rest as background. The trained classifier is then used to define a so-called  $ggF2j$  signal region ( $ggF2j$ -SR). The other classifier is used to define a VBF-SR by labeling the VBF events as signal instead. Although the  $ggF2j$  process is considered the "true" signal in this analysis, events originating from the  $qq$  initial state share a topology with the VBF events. The VBF-SR is therefore used to filter out such events. An excess of  $qq$  gluon fusion events could result in a boosted sensitivity, as such events have been shown to carry more  $\mathcal{CP}$  information in a similar analysis [268]. On the other hand, the  $ggF2j$ -SR is expected to obtain a boost in sensitivity due to the much higher number of signal events.<sup>1</sup> Therefore, both signal regions are well motivated.

In a second step, two additional classifiers are trained to differentiate between the  $\mathcal{CP}$ -states of the

<sup>1</sup> The  $ggF2j$ -SR also has a higher number of expected non-Higgs background events, although these events are not considered here.

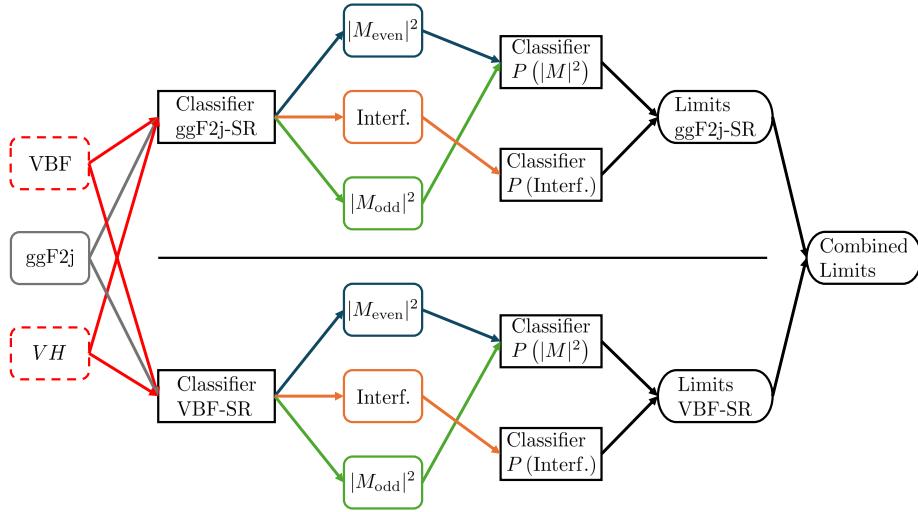


Figure 5.3: Flowchart demonstrating the analysis strategy employed. The datasets for the  $ggF2j$ ,  $VBF$ , and  $VH$  processes are given to two separate classifiers to define two signal regions. Then, two additional classifiers are trained in each signal region, learning a  $\mathcal{CP}$ -even and a  $\mathcal{CP}$ -odd observable, respectively. From these classifiers, limits are obtained in the individual signal regions, as well as combined limits. Figure modified from [233].

events. Only  $ggF2j$  events are used for the training. One classifier learns a  $\mathcal{CP}$ -even observable, while the other one learns a  $\mathcal{CP}$ -odd observable. For the final limits, a test dataset is used that first passes through the signal-background separation and then through the  $\mathcal{CP}$  classifiers. Four distributions are obtained this way, which can be used to define the exclusion limits, either individually or combined across the signal regions.

For training, validating, and testing the data are prepared in the following way:  $ggF2j$  events are split into five sets with 25%, 5%, 40%, 10%, and 20% of the total data for the training and validation sets of the signal-background and  $\mathcal{CP}$  state classifiers, as well as one final test set.  $VBF$  and  $VH$  events are split into three sets with 50%, 10%, and 40% of the events, respectively, since they are not needed to train the  $\mathcal{CP}$  classifiers. This results in approximately  $6 \cdot 10^5$  ( $4 \cdot 10^5$ ) events used for training the signal-background ( $\mathcal{CP}$ ) classifiers. The number of signal and background events in each dataset is approximately equal to ensure that the classifiers are not biased.

### 5.2.1 Defining the signal regions

For the classification of the signal and background events in the two signal regions, two sequential models are built using PyTorch [289]. They are multi-layer perceptrons with two hidden layers, each having 100 nodes and 10 nodes, respectively. After each layer (apart from the output layer), a batch normalization layer is inserted to keep all learned representations at order 1. The ReLU function is used for activation. Between the two hidden layers, a dropout layer with  $p = 0.3$  is used to prevent overtraining. For the  $VBF$ -classifier, an additional dropout layer with  $p = 0.1$  is inserted after the second hidden layer. The default binary cross-entropy loss is used as the objective function in both cases.

The classifiers are trained with the following kinematic information obtained from the `Delphes` files:

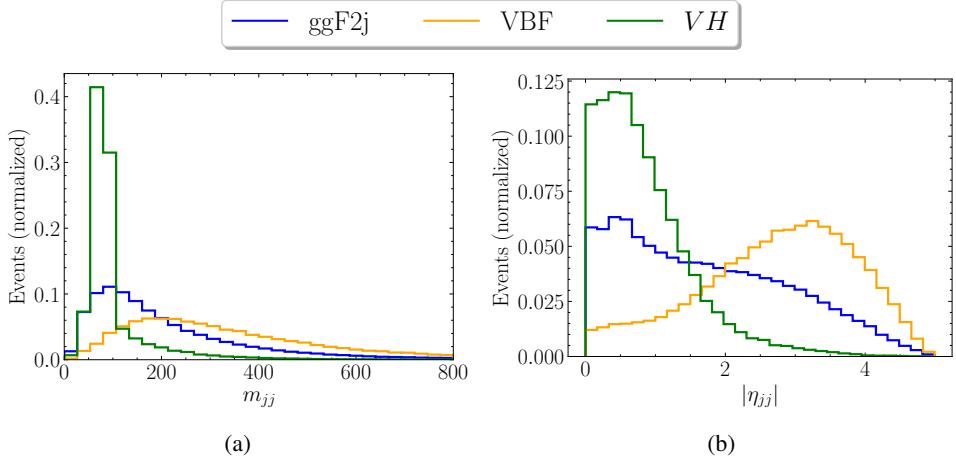


Figure 5.4: The commonly used features (a)  $m_{jj}$  and (b)  $|\eta_{jj}|$  to separate the Higgs production processes considered.

- The full 4-momenta of the Higgs boson and the two jets leading in  $p_T$ , split into the energy  $E$ , transverse momentum  $p_T$ , pseudorapidity  $\eta$  and azimuthal angle  $\phi$ .
- The invariant mass of the dijet system  $m_{jj}$ , as well as the absolute pseudorapidity difference  $|\eta_{jj}|$ , and azimuthal angle difference  $\Delta\phi_{jj} = \phi_{j1} - \phi_{j2}$ ,  $\eta_{j1} > \eta_{j2}$ . These are higher-level variables which are known to separate well the Higgs production processes in association with two jets (see e.g. [239]).
- The number of jets  $N_j$  present in the event file alongside the energy of all jets that were not already accounted for.

All variables are taken in the laboratory frame. Two of the commonly used higher-level variables,  $m_{jj}$  and  $|\eta_{jj}|$ , are shown in Fig. 5.4, while distributions for the other variables are provided in Appendix D. As can be seen, the three signal processes can already be well separated by placing cuts on these two variables. The separation of VBF from the other two processes is easier, since only one cut is needed for each of the two variables to achieve good separation. This is also reflected in the performance of the classifiers. The metrics from the ggF2j-classifier (VBF-classifier) training are shown in Fig. 5.5 (Fig. 5.6). Each model is trained for 200 epochs, and the model with the lowest validation loss is chosen after training, indicated by the black arrows in the respective figures. The ggF2j-classifier reaches  $\text{Acc} \approx 71\%$ , while the VBF-classifier reaches  $\text{Acc} \approx 80\%$ .

The distributions obtained from the test dataset after classification are shown in Fig. 5.7(a) for the ggF2j-classifier and in Fig. 5.7(b) for the VBF-classifier. In both cases, the classifier successfully separates the majority of events labeled as signal from the other contributions. For a score of  $P(\text{ggF2j}) \gtrsim 0.5$ , or  $P(\text{VBF}) \gtrsim 0.5$ , the events labeled as signal make up the relative majority of events in each bin. Therefore, a cut  $P(\text{ggF2j}) \geq 0.5$  ( $P(\text{VBF}) \geq 0.5$ ) is used on the test dataset to define the ggF2j-SR (VBF-SR). The ggF2j events in Fig. 5.7 are split into the three initial states. As previously discussed, the  $qq$  initial state ggF2j events share a topology with VBF and are expected to be enriched in the VBF-SR. This is validated by the distributions obtained.

The same test dataset is given to both classifiers. Since their training is performed independently, the two resulting datasets defining the ggF2j-SR and VBF-SR may contain events that are classified as both.

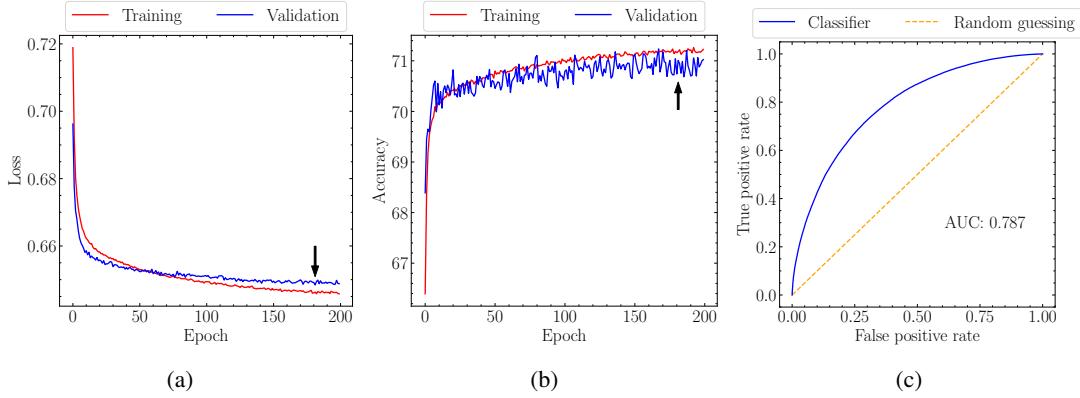


Figure 5.5: (a) Loss per epoch, (b) Accuracy per epoch, and (c) ROC curve of the classifier used to define the ggF2j-SR. The black arrow indicates the epoch, at which the final model is chosen.

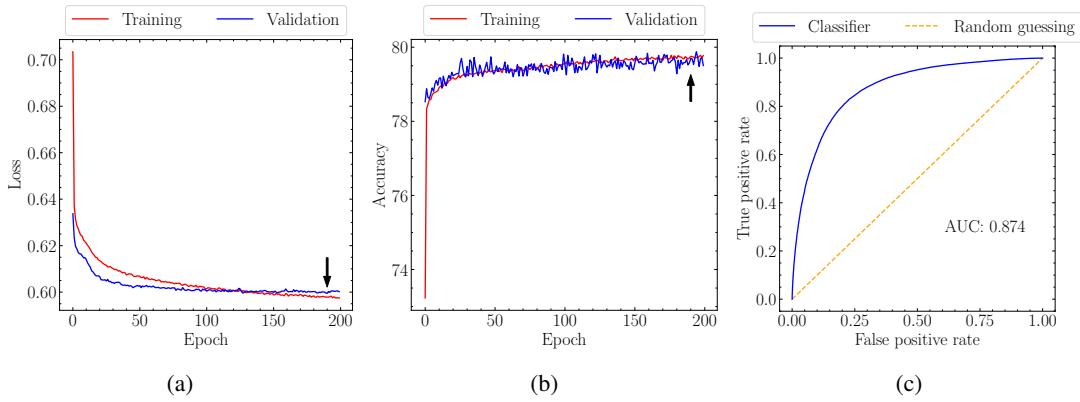


Figure 5.6: Same as Fig. 5.5 but for the VBF-SR.

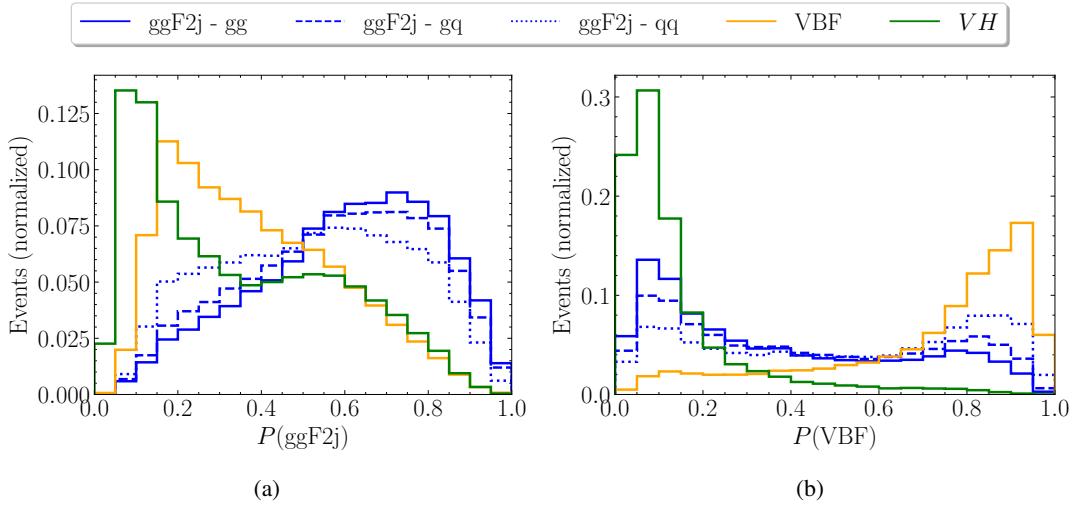


Figure 5.7: Scores for the (a) ggF2j-classifier and (b) VBF-classifier. The ggF2j events are split into three distributions, showcasing how the different initial states are classified. Figure modified from [233].

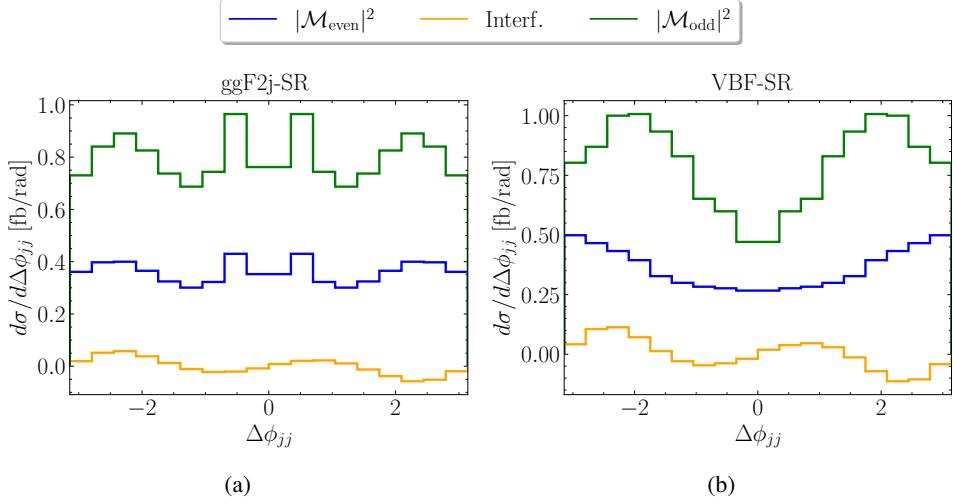


Figure 5.8: Differential cross section of the  $\mathcal{CP}$ -odd variable  $\Delta\phi_{jj}$  obtained from signal events in (a) the ggF2j-SR and (b) VBF-SR. Figure modified from [233].

Here, the amount of signal (ggF2j) events that end up in both signal regions is about 8%. Therefore, it can be concluded that the two signal regions are approximately orthogonal to each other. About 11% of the signal events do not pass either event selection from the classifiers and are therefore not used further.

### 5.2.2 Classification of the ggF2j $\mathcal{CP}$ states

After the two signal regions have been defined, the next task is to separate the different  $\mathcal{CP}$  contributions to the total ggF2j amplitude in Eq. (2.59). As discussed above, sensitivity to the squared terms can be obtained from the total rate and shape differences in  $\mathcal{CP}$  sensitive observables. For the interference term, a  $\mathcal{CP}$ -odd observable such as  $\Delta\phi_{jj}$  is needed. The differential cross section of  $\Delta\phi_{jj}$  in the two respective signal regions is shown in Fig. 5.8, while other variables are shown in Appendix D. It can be seen that the distributions vary across the signal regions. The shape differences are stronger in the VBF-SR and, most importantly, the interference term is more pronounced. This agrees with the hypothesis that the  $q\bar{q}$  initial state events carry more  $\mathcal{CP}$  information. In the ggF2j-SR, two peaks can be seen in the  $\mathcal{CP}$ -even distributions close to  $\Delta\phi_{jj} = 0$ . These are the result of events with a low  $m_{jj}$ .

Two classifiers are trained in each signal region to gain sensitivity towards the  $\mathcal{CP}$  state of the Higgs-gluon coupling. Both classifiers use similar input data as the signal-background classifiers, that is, the 4-vectors of the Higgs boson and the two jets, as well as  $m_{jj}$ ,  $|\eta_{jj}|$ , and  $\Delta\phi_{jj}$ . The total number of jets and the energy of the subleading jets are not used here, because they carry no  $\mathcal{CP}$  information. It is evident from the distributions in Appendix D that the kinematic differences between the two  $\mathcal{CP}$ -even contributions to the ggF2j amplitude are very small. To better exploit these differences, the `GradientBoostingClassifier` from the `scikit-learn` [290, 291] package is used. The classifiers are given different training data to learn two observables:

- The first classifier is designed to build a  $\mathcal{CP}$ -even observables by differentiating between the  $c_g^2$  and  $\tilde{c}_g^2$  distributions. It will be dubbed the  $\mathcal{CP}$ -even classifier, and its output score is  $P(c_g^2)$ .

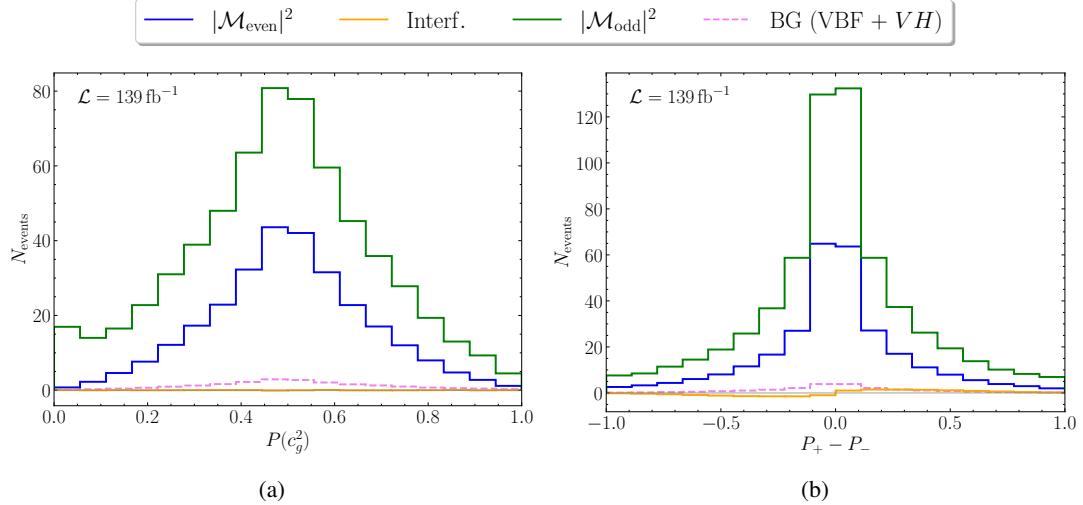


Figure 5.9: Differential distributions of the individual contributions to the ggF2j cross section, as well as the background, after running through the (a)  $\mathcal{CP}$ -even classifier and (b)  $\mathcal{CP}$ -odd classifier in the ggF2j-SR. Only the classifier resulting in the best exclusion limits out of 100 trained classifiers is shown. Figure modified from [233].

- The second classifier is built to exploit information about the interference term. Specifically, it learns to differentiate between positive and negative interference events and builds the observable  $P_+ - P_-$ . This approach to defining a  $\mathcal{CP}$ -odd observable using a classifier is based on [292]. The classifier is dubbed the  $\mathcal{CP}$ -odd classifier.

Combining the output of the classifiers allows for building observables that are closely related to the output of machine learning models used to approximate the parton-level optimal observable (see the discussion in Section 7.1 and e.g. [293, 294]).

The results of the  $\mathcal{CP}$  classification in the ggF2j-SR can be seen in Fig. 5.9. The test dataset is split into the contributions originating from the three ggF2j terms and the background events. The contributions are then given to the two classifiers individually. 100 classifiers were trained in total to assess uncertainties due to the training process, but only the results for the best classifiers in terms of the exclusion limits are shown here. The training uncertainty is discussed in more detail in Appendix E. The left-hand plot shows the results of the  $\mathcal{CP}$ -even classifier. Low values of  $P(c_g^2)$  should correspond to the  $|\mathcal{M}_{\text{odd}}|^2$  contribution after training and high values to the  $|\mathcal{M}_{\text{even}}|^2$  one. As can be seen, some separation is achieved by slight differences in the shape of the two distributions. Especially the left-most bin shows an enrichment of events stemming from the distribution proportional to  $\tilde{c}_g^2$ . The background events are strongly suppressed from the signal-background classifier, and the interference contribution cancels out. The right-hand plot of Fig. 5.9 shows the output of the  $\mathcal{CP}$ -odd classifier. The two  $\mathcal{CP}$ -even contributions of ggF2j, as well as the background, are approximately symmetric around zero, as expected. The interference term shows a slight asymmetry but is very small in magnitude.

Fig. 5.10 shows the distributions obtained in the VBF-SR. Again, only the distributions of the best classifiers are shown. For the  $\mathcal{CP}$ -even classifier in the left-hand plot, the two squared contributions are again separated via slight differences in their shape. The visual statistical fluctuations average out over multiple training runs of the classifier, as further discussed in Appendix E. The background is much more visible here compared to the ggF2j-SR due to the higher amount of VBF events. Interestingly,

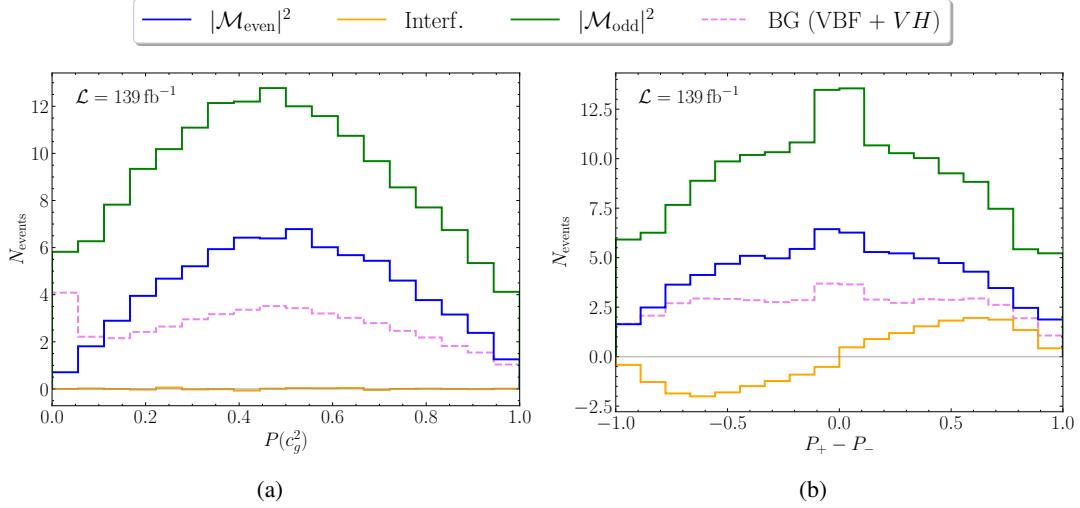


Figure 5.10: Same as Fig. 5.9 but in the VBF-SR. Figure modified from [233].

the background also exhibits a peak in the left-most bin. For the  $\mathcal{CP}$ -odd classifiers, all  $\mathcal{CP}$ -even distributions are again symmetric around zero. However, the interference term is much more pronounced here.

### 5.3 Sensitivity to the Higgs-gluon coupling

The distributions obtained in the previous section can be used to place limits on the Higgs-gluon coupling modifiers  $c_g$  and  $\tilde{c}_g$ , and correspondingly the mixing angle  $\alpha_g$ . The exclusion limits are obtained by comparing the BSM data to the SM prediction at each point in the parameter space. It is assumed that the measured data corresponds to the SM unless stated otherwise. Details on the calculation of the likelihood and the exclusion limits can be found in Appendix C.

#### 5.3.1 Constraints in the ggF2j-SR

First, the limits on  $c_g$  and  $\tilde{c}_g$  in the ggF2j-SR will be discussed. The results can be found in Fig. 5.11, where the  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  levels are shown in the  $(c_g, \tilde{c}_g)$  parameter plane. The limits are derived with the SM as the truth, and therefore the best-fit point in all cases corresponds to the SM value of  $c_g = 1, \tilde{c}_g = 0$ . A luminosity of  $\mathcal{L} = 139 \text{ fb}^{-1}$  is assumed, which corresponds to the luminosity of the full Run 2 dataset.

The first limits in Fig. 5.11(a) are shown for the  $\Delta\phi_{jj}$  variable, which has been used in previous  $\mathcal{CP}$  analyses and is  $\mathcal{CP}$ -odd by definition. The limits are based on the distributions in Fig. 5.8(a) for the ggF2j process, although the Higgs background processes are also taken into account for the calculation of the  $\Delta\chi^2$  values. All  $\sigma$  levels have the form of a closed ellipse which contains the SM value and the sign-flip  $c_g = \pm 1, \tilde{c}_g = 0$ , as well as the parameter points  $c_g = 0, \tilde{c}_g = \pm 2/3$ . This is a result of the rate information, since Eq. (2.62) yields the same prefactors for the  $\mathcal{CP}$ -even and  $\mathcal{CP}$ -odd term when comparing the parameter points listed above. Consequently, all parameter combinations within the

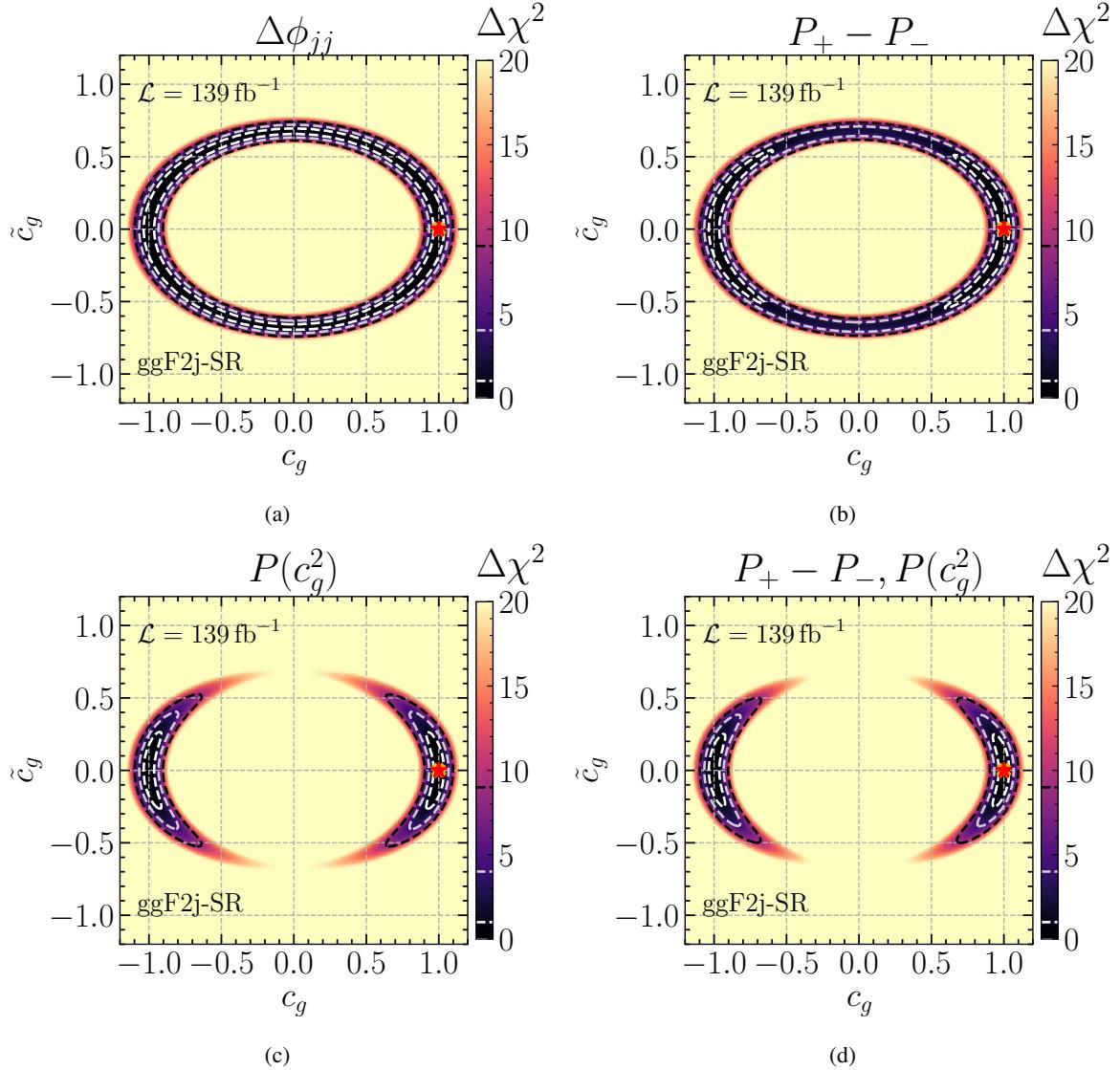


Figure 5.11: Exclusion limits on the coupling modifiers  $c_g$  and  $\tilde{c}_g$  of the effective Higgs-gluon coupling. Limits are shown for (a)  $\Delta\phi_{jj}$ , (b) the output of the  $\mathcal{CP}$ -odd classifier, (c) the output of the  $\mathcal{CP}$ -even classifier, and (d) their combinations. The  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  are depicted as white, light grey, and dark grey dashed lines, respectively. The SM is shown as an orange cross, and the best-fit point as a red star. Figure modified from [233].

ellipse also yield (approximately) the same rate.

The constraints in the lower left and upper right plots of Fig. 5.11 show the constraints obtained from the  $\mathcal{CP}$ -even and  $\mathcal{CP}$ -odd classifier, which are based on Fig. 5.9(a) and Fig. 5.9(b), respectively. In both cases, the constraints are stronger than the ones from using the  $\Delta\phi_{jj}$  variable alone. The  $P_+ - P_-$  observable allows to split up the  $1\sigma$  level into two parts, which are constrained within the ellipse to  $|c_g| > 0.6$ ,  $|\tilde{c}_g| < 0.54$ . A slight tilt of the  $1\sigma$  level due to the suppressed interference term is noticeable. However, the exclusion limits from the  $P(c_g^2)$  observable are much stronger. In this case, even the  $3\sigma$  region is split up into two parts, which are mirrored at the  $c_g = 0$  line. The  $\mathcal{CP}$ -odd part of the effective Higgs gluon coupling is constrained to  $\tilde{c}_g \in [-0.26, 0.26]$  at the  $1\sigma$  level.

Finally, in Fig. 5.11(d) the limits from the combination of the  $\mathcal{CP}$ -even and  $\mathcal{CP}$ -odd classifiers are shown. They are based on a 2-dimensional binning of the distributions of the individual classifiers. The constraints are dominated by the ones from the  $\mathcal{CP}$ -even classifier in this case. Only a slight improvement is noticeable, constraining  $\tilde{c}_g \in [-0.25, 0.25]$  at the  $1\sigma$  level.

### 5.3.2 Constraints in the VBF-SR

Next, the same scenarios will be examined in the VBF-SR. Fig. 5.12 shows the limits obtained in this kinematic region for the same variables as before. While in all cases the constraints again appear in the form of an ellipse, this ellipse is much wider compared to the ggF2j-SR. This is a consequence of the significantly lower statistics due to the signal-background separation. On the other hand, the  $1\sigma$  level for a measurement of  $\Delta\phi_{jj}$  is split up in contrast to the ggF2j-SR. Therefore, while the rate information is lower, the shape information is better. This is evident by comparing the two distributions in Fig. 5.8, where especially the two squared contributions show much larger differences in their shape in the VBF-SR.

The limits obtained from the individual classifiers are worse than the ones obtained from  $\Delta\phi_{jj}$ . The  $P_+ - P_-$  observable does not manage to split up the ellipse that the  $1\sigma$  level forms. However, the impact of the interference term is visible by looking at the regions where both  $c_g$  and  $\tilde{c}_g$  contribute with similar strength. In the regions of the ellipse where  $c_g \cdot \tilde{c}_g$  is negative, the constraints are tighter than in the regions where the product is positive. No asymmetry is visible in the results from the  $P(c_g^2)$  observable, as is expected for a  $\mathcal{CP}$ -even observable. The  $\mathcal{CP}$ -odd part of the Higgs-gluon coupling is constrained to  $\tilde{c}_g \in [-0.6, 0.6]$  at the  $1\sigma$  level.

The fact that  $\Delta\phi_{jj}$  alone performs better than the two individual classifiers is not a contradiction, as the limits from  $\Delta\phi_{jj}$  are obtained using all contributions to the ggF2j process, while the  $\mathcal{CP}$ -even and  $\mathcal{CP}$ -odd classifiers are only trained on the squared distributions or the interference term. The full information is included in the combined limits, which are shown in Fig. 5.12(d). Here, the constraints are tighter compared to  $\Delta\phi_{jj}$  and yield  $\tilde{c}_g \in [-0.42, 0.42]$  at the  $1\sigma$  level. However, comparing Fig. 5.12(a) and Fig. 5.12(d) makes it clear that the classifiers do not achieve much improvement. This agrees with previous findings, in which  $\Delta\phi_{jj}$  was shown to carry most of the  $\mathcal{CP}$  sensitivity in VBF Higgs production (see Chapter 7 and e.g. [222]).

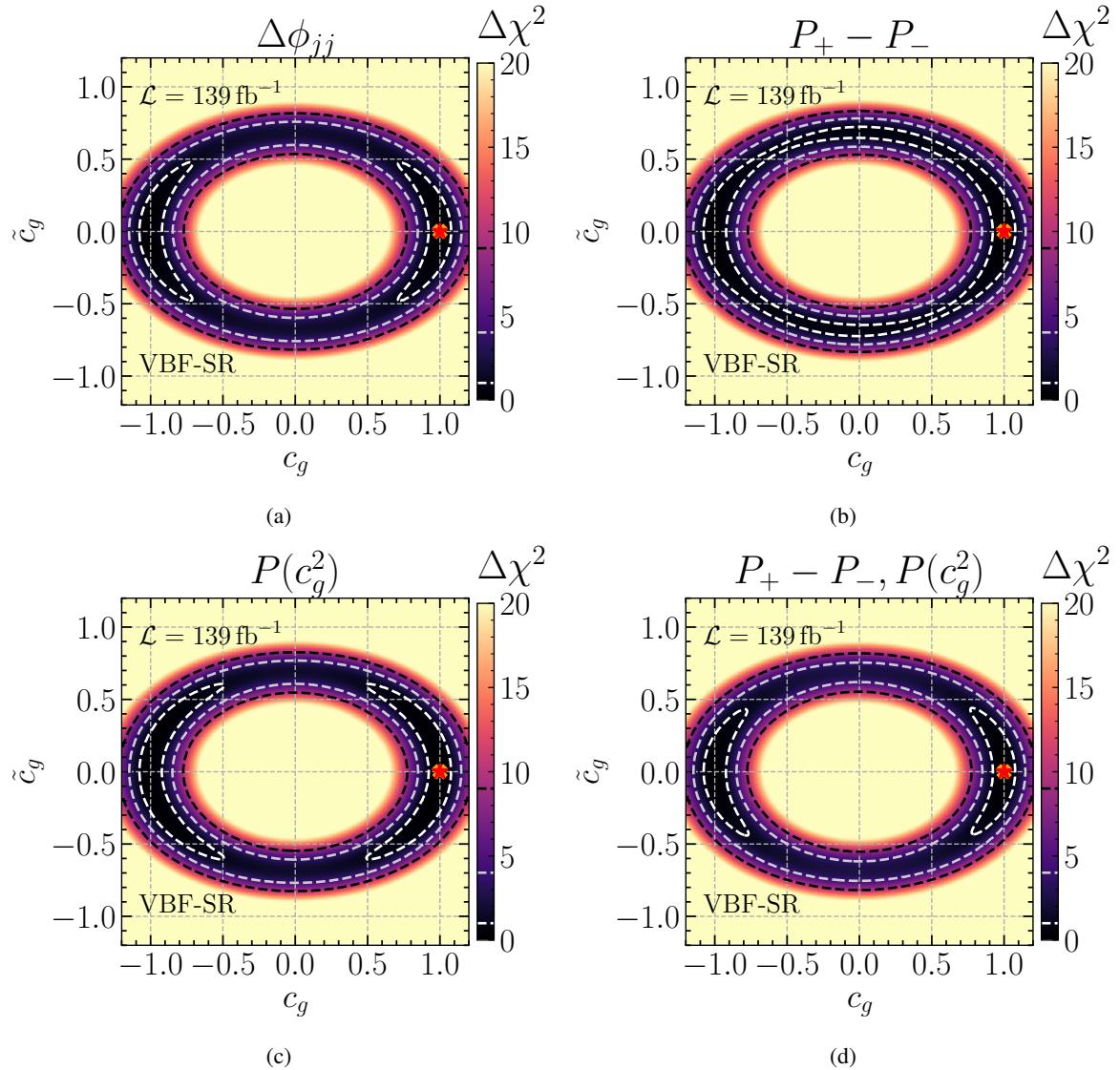


Figure 5.12: Same as Fig. 5.11 but in the VBF-SR. Figure modified from [233].

### 5.3.3 Combination

After obtaining limits in the two individual signal regions, it is possible to combine the ggF2j-SR and VBF-SR into one joint kinematic region due to their approximate orthogonality. The constraints obtained from this combined signal region can be found in Fig. 5.13. Here, the  $1\sigma$  and  $2\sigma$  levels are shown for the individual signal regions as well as their combinations. Fig. 5.13(a) depicts the limits obtained from the  $\mathcal{CP}$ -odd classifier. While the constraints from the ggF2j-SR are stronger than the ones from the VBF-SR individually, their combination allows for setting much tighter constraints. The  $2\sigma$  level from the combined signal region (dotted black) lies on top of the  $1\sigma$  region in the ggF2j-SR (dashed red). Overall, the limits improve to  $\tilde{c}_g \in [-0.36, 0.36]$  at the  $1\sigma$  level.

In the  $\mathcal{CP}$ -even classifier and the combined classifier, the situation is different. The constraints from the ggF2j-SR alone are already strong, and no improvement is achieved by adding the events from the VBF-SR. Instead, the limits become slightly weaker from  $\tilde{c}_g \in [-0.25, 0.25]$  to  $\tilde{c}_g \in [-0.27, 0.27]$  at the  $1\sigma$  level. These two opposite effects are explainable by looking at the types of events present in the VBF-SR. The relative number of ggF2j interference events, as well as VBF background events, is enhanced compared to the ggF2j-SR. The former leads to improved limits for the  $\mathcal{CP}$ -odd classifier, while the latter worsens the overall limits.

All previous plots were created under the assumption that SM-like data have been measured. Fig. 5.13(d) shows the scenario in which the measured data corresponds to the  $\alpha_g = 45^\circ$  hypothesis. In this case, the VBF events lose some of their sensitivity, as neither the SM case, nor the  $\alpha_g = 90^\circ$  case can be excluded at the  $2\sigma$  level. However, they show a slight preference for a BSM scenario where  $c_g = 0$ . The  $1\sigma$  level from the ggF2j-SR splits into four regions and allows for the exclusion of the SM in this hypothetical scenario. For the combination of both signal regions, these four regions are moved further away from the SM point due to the added VBF events. In this case, the SM can be excluded even at the  $2\sigma$  level.

Finally, Fig. 5.14 shows the exclusion limits at the  $2\sigma$  level projected to different luminosities.  $\mathcal{L} = 139 \text{ fb}^{-1}$  (black dashed) corresponds to the cases discussed above for the full Run 2 data.  $\mathcal{L} = 300 \text{ fb}^{-1}$  and  $\mathcal{L} = 3000 \text{ fb}^{-1}$  correspond to the expected luminosity at the end of Run 3 of the LHC and the HL-LHC, respectively. The limits are plotted using  $|c_g|$  since there is little to no sensitivity towards the sign of the coupling modifiers. The limits shown are naive projections, as only the total number of events is taken into account, while improvements in the reconstruction algorithms or detector resolution are not considered. The improvement of the constraints scales slightly worse than  $1/\sqrt{\mathcal{L}}$  because the shape information is reduced the closer a tested point in the parameter space is to the SM. The limits at the  $2\sigma$  level show an improvement from  $\tilde{c}_g \in [-0.39, 0.39]$  at  $\mathcal{L} = 139 \text{ fb}^{-1}$  to  $\tilde{c}_g \in [-0.32, 0.32]$  at  $\mathcal{L} = 300 \text{ fb}^{-1}$  and  $\tilde{c}_g \in [-0.17, 0.17]$  at  $\mathcal{L} = 3000 \text{ fb}^{-1}$ . Loosened constraints are expected for a full experimental analysis, where systematic uncertainties and especially the impact of the non-Higgs background are taken into account.

### 5.3.4 Relative importance of the observables

In this section, the impact of the individual variables on the classifier output is studied to improve the interpretability of the results. For this, the SHAP framework as introduced in Section 4.4.1 is used [217, 218]. Every variable  $x_i$  is assigned a worth by calculating Eq. (4.30), which compares the performance of the classifier in each subset of variables with and without  $x_i$ .

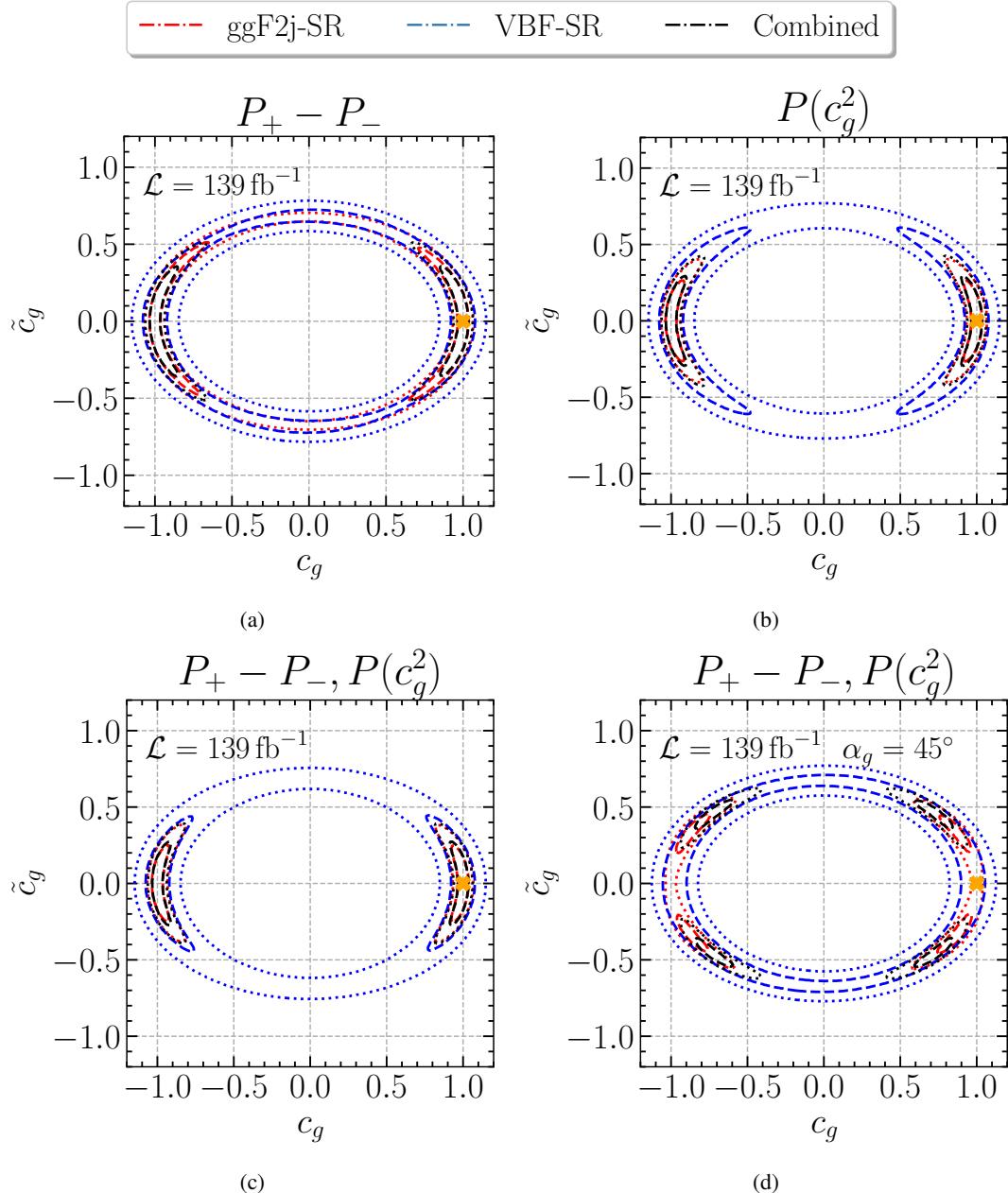


Figure 5.13: Exclusion limits obtained in the combined signal region for (a) the output of the  $\mathcal{CP}$ -odd classifier, (b) the output of the  $\mathcal{CP}$ -even classifier, (c) their combination, and (d) their combination for a measurement of  $\alpha_g = 45^\circ$ .  $1\sigma$  and  $2\sigma$  levels are shown as dashed and dotted lines, respectively. Figure modified from [233].

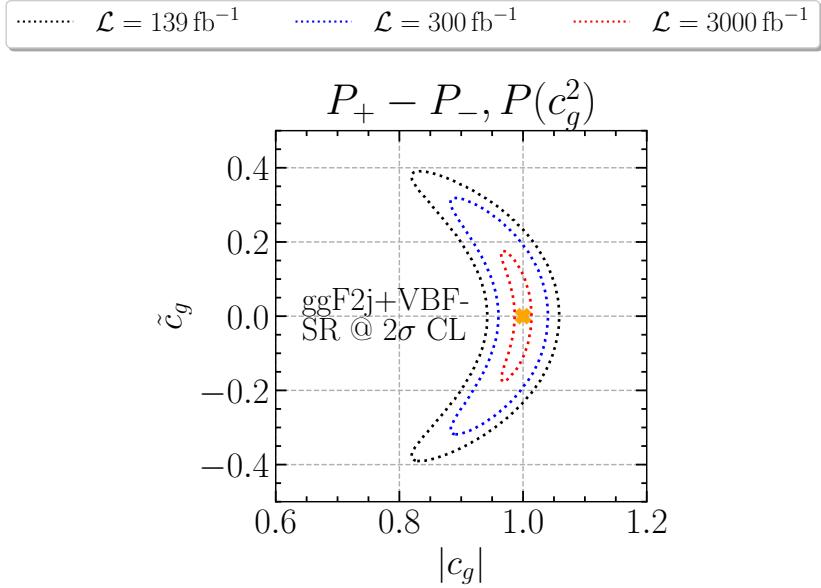


Figure 5.14: Projected  $2\sigma$  levels for the combined signal region at the expected luminosity after the end of Run 3 ( $\mathcal{L} = 300 \text{ fb}^{-1}$ ) and the HL-LHC ( $\mathcal{L} = 3000 \text{ fb}^{-1}$ ). Figure modified from [233].

The results are shown for the  $\mathcal{CP}$ -odd classifiers in Fig. 5.15 and for the  $\mathcal{CP}$ -even classifiers in Fig. 5.16. The input variables on the left y-axis in each plot are ranked by their relative importance. Variables that rank higher are more influential on the classifier score. The x-axis shows the respective SHAP value of each variable, which is calculated for every single event in the dataset. The sign of the SHAP value indicates the pull on the classifier score towards high or low values if the variable is included. The importance of a feature corresponds to the mean of the absolute SHAP values. Finally, the colors of the individual points correspond to the value of the respective feature, normalized over its entire range.

First, the results of the classifier separating positive and negative interference are discussed. In both signal regions,  $\Delta\phi_{jj}$  reaches the highest SHAP values by far. This is expected because it is the only dedicated  $\mathcal{CP}$ -odd observable used in the classifier training. Since  $\Delta\phi_{jj} \in [-\pi, \pi]$ , red (blue) points correspond to positive (negative) values of the angle. They show a clear distinction in their respective SHAP values, indicating sensitivity to the sign via the  $\mathcal{CP}$  angle. While all other features are  $\mathcal{CP}$ -even, they can still be combined with  $\Delta\phi_{jj}$  to boost the sensitivity of the classifiers. In the VBF-SR (see Fig. 5.15(b)), the second and third highest impact features are  $p_{T,j_1}$  and  $p_{T,j_2}$ . Since many VBF-like events are expected in this signal region, this compares well with previous findings, which determine  $p_{T,j_1} \cdot p_{T,j_2} \cdot \sin(\Delta\phi_{jj})$  as a near-optimal observable for VBF [222]. The ggF2j-SR generally shows much lower SHAP values. This is reflected in Figs. 5.11 and 5.12 where the interference classifier has little to no importance in the ggF2j-SR but improves the constraints in the VBF-SR.

The classifiers separating the  $|\mathcal{M}_{\text{even}}|^2$  and  $|\mathcal{M}_{\text{odd}}|^2$  contributions yield the SHAP values displayed in Fig. 5.16. Overall, the SHAP values are lower than for the  $\mathcal{CP}$ -odd classifiers, as indicated by the smaller scale on the x-axis. This is not surprising, as the  $P(c_g^2)$  variables in Figs. 5.9 and 5.10 do not show major shape differences. Furthermore, there is no dominating observable, which means that the full kinematic

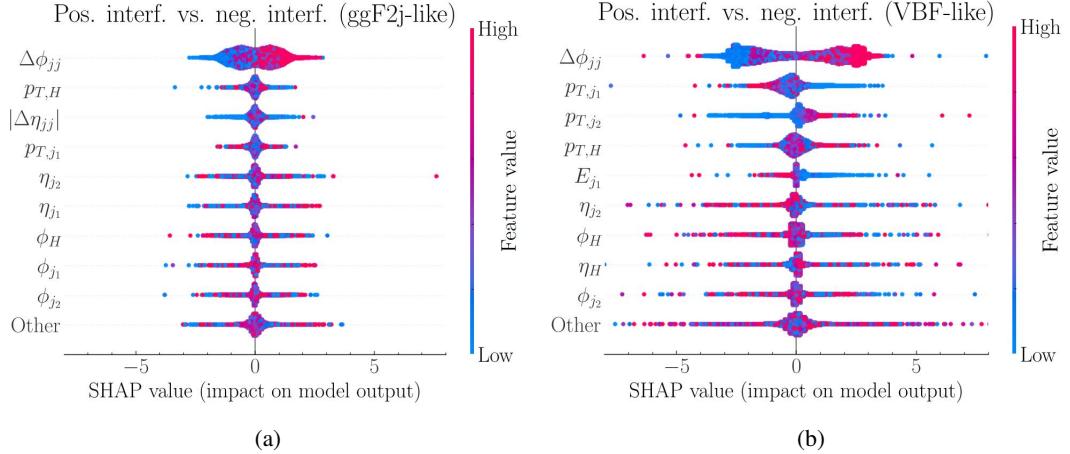


Figure 5.15: SHAP values for the variables used in the training of the  $\mathcal{CP}$ -odd classifiers in the (a) ggF2j-SR and (b) VBF-SR. Variables are ranked from top to bottom by their relative importance. The 9 best variables are shown while the combined SHAP values of all other variables are grouped in "Other". Figure modified from [233].

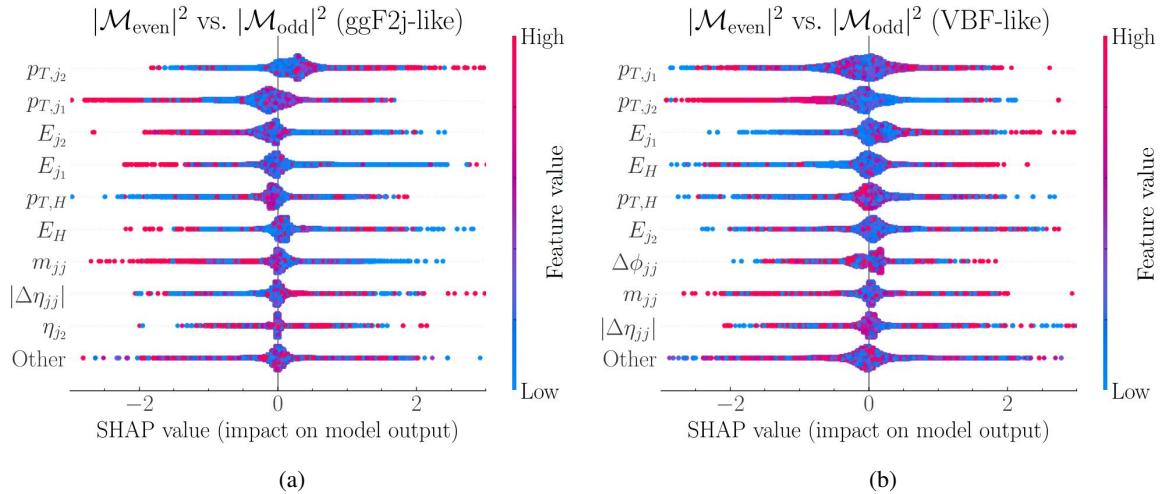


Figure 5.16: Same as Fig. 5.15 but for the  $\mathcal{CP}$ -even classifiers. Figure modified from [233].

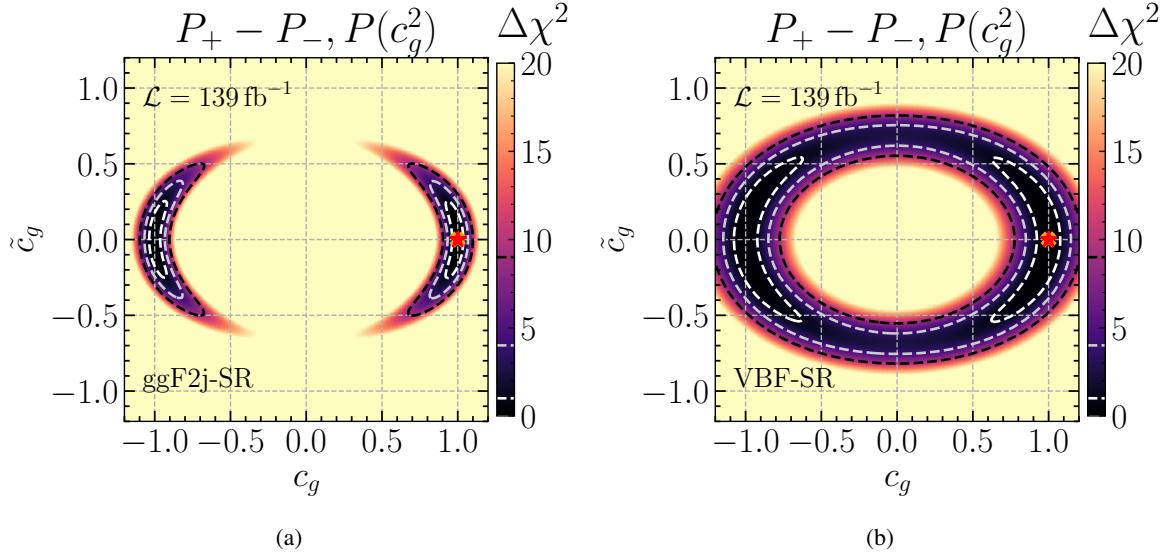


Figure 5.17: Exclusion limits for the combination of both classifiers in the (a) ggF2j-SR and (b) VBF-SR when the possibility of a  $\mathcal{CP}$ -violating  $HVV$  coupling is considered. Figure modified from [233].

information is needed to achieve a separation of the distributions. Still, the transverse momentum and energy of all final-state particles are consistently identified as the most important features. Specifically,  $p_{T,j_1}$  and  $p_{T,j_2}$  rank as the two best variables in both signal regions. In the ggF2j-SR, they pull the classifier output in different directions, suggesting an interplay between the two variables. This effect is less pronounced in the VBF-SR.

## 5.4 Impact of $\mathcal{CP}$ violation in the $HVV$ coupling

Up to now,  $\mathcal{CP}$  violation was only assumed to be present in the effective Higgs-gluon coupling. However, the considered Higgs background could also contain  $\mathcal{CP}$ -violating interactions from the  $HVV$  couplings. While it was discussed in Section 3.3.2 that the  $\mathcal{CP}$  nature of this coupling is strongly constrained,  $\mathcal{CP}$ -violating effects could still appear and affect the limits obtained in the previous sections. Various BSM scenarios introduce  $\mathcal{CP}$  violation in more than one Higgs coupling, which would make the assumption of  $\mathcal{CP}$ -even Higgs background processes in the ggF2j channel invalid. As previously mentioned, additional datasets are generated for the VBF and  $VH$  channels where the amount of  $\mathcal{CP}$  violation is controlled via the  $c_{\Phi\widetilde{W}}$  Wilson coefficient in the SMEFT. Details on this are listed in Appendix B.

To assess the robustness of the results in Section 5.3, the classifiers are not trained again with the  $\mathcal{CP}$ -violating Higgs background data. Instead, the newly created datasets are added to the test dataset and passed through the pretrained signal-background and  $\mathcal{CP}$  classifiers. For the subsequent evaluation of the constraints on  $c_g$  and  $\tilde{c}_g$ , the amount of  $\mathcal{CP}$  violation in the  $HVV$  coupling is allowed to vary. This is done by freely floating  $c_{\Phi\widetilde{W}} \in [-1, 1]$  during the fit, corresponding to the current experimental constraints.

The results of the combined classifiers in the ggF2j-SR and the VBF-SR are shown in Fig. 5.17.

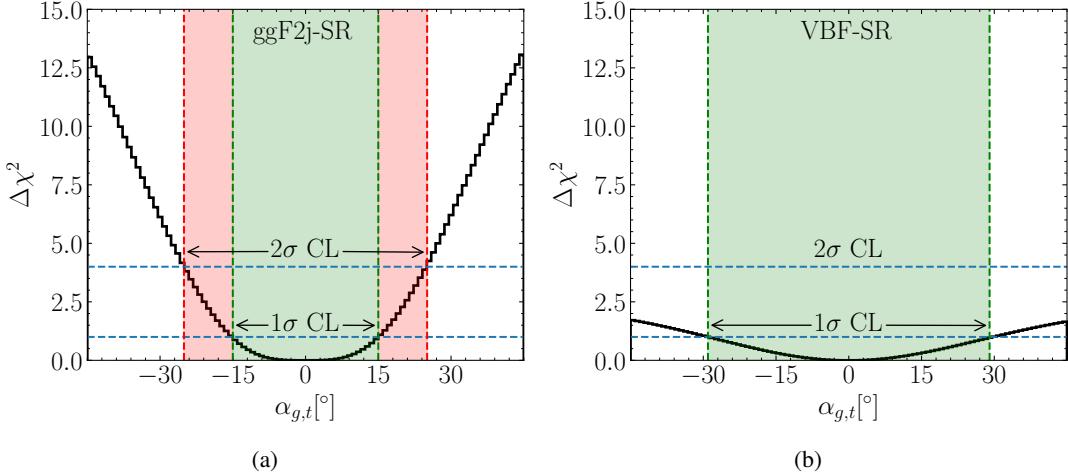


Figure 5.18: Constraints on the  $\mathcal{CP}$ -mixing angle  $\alpha_{g,t}$  in the (a) ggF2j-SR and (b) VBF-SR obtained by profiling over the rate modifier  $g_{g,t}$ . The  $1\sigma$  and  $2\sigma$  levels are shown as shaded green and red regions, if available.

For each point in the parameter plane, the limits are obtained by profiling over  $c_{\Phi\widetilde{W}}$ . Comparing the exclusion limits in Fig. 5.17(a) with the previously obtained limits in Fig. 5.11(d) for the ggF2j-SR makes evident that the additional  $\mathcal{CP}$  violation does not impact the results. Instead, the same limits as in the  $c_{\Phi\widetilde{W}} = 0$  are recovered, indicating a good robustness of the results in this signal region. This is a consequence of the strong suppression of the Higgs background.

For the VBF-SR, the picture is different. The VBF-SR shows a much higher relative number of  $\mathcal{CP}$ -violating VBF events than the ggF2j-SR. This means that the classifier still labels such events as signal, even when trained only on SM-like data in this process. Comparing Fig. 5.17(b) to Fig. 5.12(d) yields visibly weaker constraints at the  $1\sigma$  level. Specifically, the limits are loosened from  $\tilde{c}_g \in [-0.42, 0.42]$  to  $\tilde{c}_g \in [-0.54, 0.54]$ . The weakening of the limits solely stems from the  $\mathcal{CP}$ -odd  $P_+ - P_-$  variable, while no changes are observed in the  $P(c_g^2)$  variable. Therefore, it can be concluded that the additional free parameter affects the results and has to be considered in an analysis of this signal region.

## 5.5 Evaluation in terms of the top-Yukawa coupling

As previously discussed, the parameterization of the effective Higgs-gluon coupling in Eq. (2.62) allows for a direct relation between  $c_g$ ,  $\tilde{c}_g$  and  $c_t$ ,  $\tilde{c}_t$ . Two additional assumptions are needed for this, the first one being the infinite top-mass approximation. This is accounted for by excluding events with  $p_{T,H} > 200$  GeV throughout this chapter. Second, no colored heavy BSM state is allowed to contribute to the ggF loop. Constraints on the energy scale of such states are constantly increasing, and therefore, their potential impact is decreasing. For example, the impact of a new colored fermion with a Yukawa-coupling strength similar to the top quark is suppressed by  $v^2/\Lambda^2$  with  $\Lambda \gtrsim 1$  TeV.

Limits on the coupling modifiers of the top-Yukawa coupling are obtained by replacing the respective coupling modifiers of the Higgs-gluon coupling. Consequently, also the  $\mathcal{CP}$ -mixing angle  $\alpha_{g,t}$  and rate modifier  $g_{g,t}$  can be evaluated via Eq. (2.55). The limits obtained on  $\alpha_{g,t}$  are shown for both signal regions in Fig. 5.18. The results are obtained by profiling the two-dimensional exclusion plots over  $g_{g,t}$ .

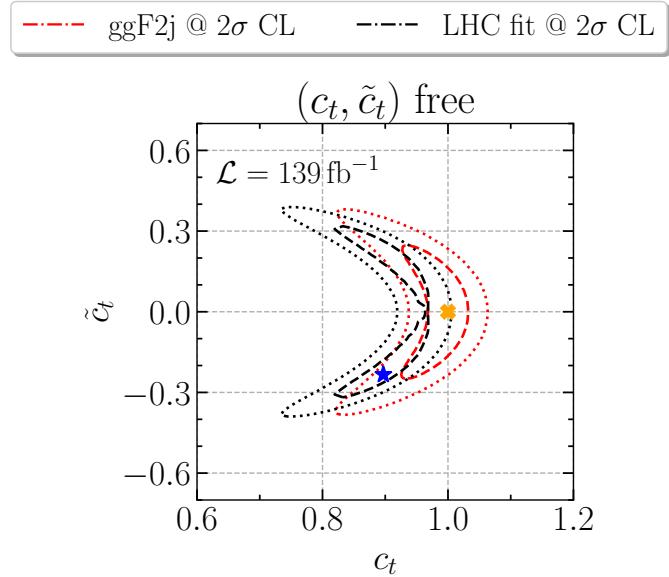


Figure 5.19: Comparison of the limits obtained in the ggF2j-SR with the current limits obtained from a global fit to LHC data using `HiggsTools`. The  $1\sigma$  and  $2\sigma$  levels are shown as dashed and dotted lines, respectively. The best-fit point of the global fit is shown as a blue star. Figure modified from [233].

In the ggF2j-SR, the  $\mathcal{CP}$ -mixing angle can be constrained to

$$\alpha_{g,t}^{\text{ggF2j-SR}} \in [-15^\circ, 15^\circ] @ 1\sigma \text{ CL} \quad ([ -25^\circ, 25^\circ ] @ 2\sigma \text{ CL}) \quad (5.1)$$

and the confidence levels are shown as green and red shaded regions in Fig. 5.18(a). For the VBF-SR, only a limit at the  $1\sigma$  level is found. This is shown in Fig. 5.18(b) and yields

$$\alpha_{g,t}^{\text{VBF-SR}} \in [-29^\circ, 29^\circ] @ 1\sigma \text{ CL} \quad . \quad (5.2)$$

Apart from the ggF2j process, the top-Yukawa coupling can also be probed by indirect rate measurements from the ggF process without additional jets, the  $H \rightarrow \gamma\gamma$  decay, and top-associated Higgs production processes. The information of these channels can be combined in global fits, as performed in [27, 174]. To compare with the results obtained from the ggF2j analysis, an updated global fit is performed using `HiggsTools` [295–297]. Constraints from both methods are shown for a luminosity of  $\mathcal{L} = 139 \text{ fb}^{-1}$  in Fig. 5.19. The red dashed and dotted lines show the constraints of the ggF2j-SR at the  $1\sigma$  and  $2\sigma$  level, while the respective black lines show the constraints from `HiggsTools`. The dominant contribution to the constraints of the latter comes from indirect rate measurements. The shift of the exclusion limits and the best-fit point to lower values of  $c_t$ , compared to the ggF2j constraints, is caused by experimental measurements in the ggF and  $t\bar{t}H$  processes [182, 298] (also visible in Figs. 3.6 and 3.7). In general, both methods yield a comparable sensitivity. However, for a full comparison, the non-Higgs background should be taken into account in the ggF2j-SR, which would weaken the limits. At the same time, the ggF2j channel offers the advantage of a more model-independent analysis.

## 5.6 Conclusions

This chapter dealt with the question of how the  $\mathcal{CP}$  character of the Higgs coupling to gluons can be constrained optimally. So far, the most stringent constraints are indirect and stem from rate measurements of the ggF process. This also allows for setting tight indirect constraints on the  $\mathcal{CP}$  state of the top-Yukawa coupling. Here, the ggF2j process was examined, which offers additional  $\mathcal{CP}$  information via the shape of  $\mathcal{CP}$  sensitive observables. This can be used to set more model-independent bounds on the effective Higgs-gluon coupling as well as the top-Yukawa coupling and assess whether improvements in experimental analyses of this channel are possible.

Events were generated for ggF2j production with the Higgs boson decaying into two photons. In addition, the relevant Higgs background processes, VBF and  $VH$ , were simulated, while the non-Higgs background was neglected in this study. Subsequently, two classifiers were trained to separate the three Higgs production processes. They were used to define a ggF2j-SR and a VBF-SR by training them to recognize either ggF2j- or VBF-like events. Both kinematic regions offer possible advantages: The ggF2j-SR has a higher number of signal events, while the VBF-SR carries the events with the most  $\mathcal{CP}$  information.

Afterwards, two classifiers were trained in each signal region to separate the events into the  $\mathcal{CP}$  contributions to the ggF2j process. A  $\mathcal{CP}$ -even observable, based on the squared contributions, and a  $\mathcal{CP}$ -odd observable, based on the interference term, were defined from the output of the classifiers. Combining both observables allowed for placing limits on the  $\mathcal{CP}$ -mixing angle of the Higgs-gluon coupling. The ggF2j-SR gave stronger constraints with  $|\alpha_g| < 15^\circ$  while the VBF-SR yielded  $|\alpha_g| < 29^\circ$  (both at the  $1\sigma$  level and for a luminosity of  $\mathcal{L} = 139 \text{ fb}^{-1}$ ). These results are much stronger than current experimental constraints, but neglect the non-Higgs background and systematic uncertainties, which would degrade the constraints. Still, the superior constraints in the ggF2j-SR due to the higher number of signal events indicate that improvements in the current experimental constraints may be achievable. Further improvements to the obtained limits might also be possible by relaxing the cut on  $p_{T,H}$ , or by employing more powerful analysis techniques based on the matrix-element approach or machine-learning based inference [129, 299–312].

To interpret the results of the classifiers, SHAP values were determined for each feature to rank their relative importance. For the  $\mathcal{CP}$ -odd classifier, it is not surprising that the  $\mathcal{CP}$ -odd variable  $\Delta\phi_{jj}$  was shown to be the most important. However, the findings also agree with [222] where it was shown that the  $p_T$  of the jets can help to boost the sensitivity in the VBF-SR. For the  $\mathcal{CP}$ -even classifiers, there is no dominant observable. Still, the transverse momenta of the jets were identified as the features that most influenced the classifier output.

The two signal regions were then examined in terms of their model dependency. Concretely,  $\mathcal{CP}$  violation was introduced via a SMEFT operator in the Higgs background processes. The ggF2j-SR showed good robustness against the  $\mathcal{CP}$ -character of the Higgs background, as an additional free parameter in the fit did not impact the results. Contrarily, the constraints in the VBF-SR were weakened by this effect.

Finally, the results obtained for the effective Higgs-gluon coupling were reinterpreted in terms of the top-Yukawa coupling, which assumes that no unknown BSM states affect the gluon fusion loop. It was revealed that constraints on the  $\mathcal{CP}$  state of the top-Yukawa coupling from a measurement of the ggF2j channel might be competitive with the current constraints from global fits.

# CHAPTER 6

---

## **$\mathcal{CP}$ -sensitive simplified template cross-sections for $t\bar{t}H$**

---

*This chapter is based on the work in [313] which was done in collaboration with Henning Bahl, Alberto Carnelli, Frédéric Déliot, Elina Fuchs, Anastasia Kotsokechagia, Tanguy Lucas Marsault, Laurent Schoeffel, and Matthias Saimpert. All results taken from [313] have been produced by the author of this thesis unless stated otherwise.*

In Chapter 5 it was shown how the top-Yukawa coupling can be constrained using the ggF2j process. Specifically, Fig. 5.19 provided a comparison of the results using kinematic information in the ggF2j channel to the results from pure rate information from the ggF production and  $H \rightarrow \gamma\gamma$  decay channel without the association of additional jets. Using the results obtained from `HiggsTools`, the most stringent constraints on the  $\mathcal{CP}$  state of the top-Yukawa coupling are  $\alpha_t \lesssim 28^\circ$  at the 95% confidence level (CL). Like in the case of the Higgs-gluon coupling, EDM measurements can also indicate the amount of  $\mathcal{CP}$  violation present in the top-Yukawa coupling. However, the effect is much more drastic here, as current limits on the eEDM constrain  $\tilde{c}_t < 10^{-3}$  at the 90% CL ( $\alpha_t < 0.06^\circ$  at the SM rate) when assuming all other couplings to be SM-like [25, 42, 43]. As shown in [27], these stringent constraints can completely vanish if the electron-Yukawa coupling deviates from its SM predictions.

While the ggF process via its rate information or its ggF2j subprocess, as well as measurements of the eEDM, provide valuable information about the  $\mathcal{CP}$  state of the top-Yukawa coupling, they do not provide direct probes. In the ggF process, other unknown states could affect the loop and introduce  $\mathcal{CP}$  violation. In the EDM case, precise information about the Yukawa couplings of the first generations is needed, but they are only very weakly constrained so far [44–55]. Therefore, a direct probe of the Higgs-top quark coupling is required that does not rely on other couplings. This is possible via the measurement of top-associated Higgs production at the LHC. Three channels are classified as such, consisting of  $t\bar{t}H$ ,  $tHq$ , and  $tWH$ , although only  $t\bar{t}H$  provides the necessary amount of data for precise differential measurements. Such measurements are needed because constraints from the total rate of  $t\bar{t}H$  only yield very weak constraints of  $\alpha_t \lesssim 72^\circ$  [174].

Like in the ggF2j process, kinematic information of the final state momenta can be exploited for  $\mathcal{CP}$  analyses. An unambiguous probe comes from defining  $\mathcal{CP}$ -odd observables to probe the interference term of  $t\bar{t}H$ . However, as already discussed in Section 3.3.2, such studies are very challenging at the LHC.

Instead,  $\mathcal{CP}$ -sensitive observables are used to compromise between rate and  $\mathcal{CP}$ -odd measurements. To maximize the  $\mathcal{CP}$  sensitivity, such observables can be constructed by using gradient boosting algorithms or neural networks [250, 255, 256, 314]. ATLAS obtains the best current experimental constraints in the  $H \rightarrow \gamma\gamma$  channel with  $|\alpha_t| < 43^\circ$  [191], while CMS combines information from the  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow 4l$  and multilepton decay channels, yielding  $|\alpha_t| < 45^\circ$  [36, 192, 284]. Weaker results are obtained in the  $H \rightarrow b\bar{b}$  channel [315].

Generally, the combination of several decay channels and the combination of results across experiments is a complex task, since each analysis depends on assumptions and optimization procedures that must be considered in the combination. This includes the modeling of the signal and the background in the specific channel, assumptions about other couplings or channels, theoretical uncertainties, and cuts that are set to maximize the sensitivity towards the measured quantity. The STXS framework, which was introduced in Section 3.3.3, was invented to prevent these difficulties. However, the current Stage 1.2 of the framework only contains a binning of the  $t\bar{t}H$  process in  $p_{T,H}$  and is not optimized towards  $\mathcal{CP}$  measurements. Therefore, an extension of the STXS binning in a second dimension is examined in this chapter, taking into account 30  $\mathcal{CP}$ -sensitive variables from the literature. A concrete proposal for an STXS extension is made, focusing on the  $\mathcal{CP}$  sensitivity.

This chapter is structured as follows. First, the event generation, as well as the definition of all considered observables based on previous studies, is discussed in Section 6.1. Afterwards, Section 6.2 describes experimental considerations, such as detector resolution effects or channel efficiencies, and how they affect the sensitivity of the observables towards  $\mathcal{CP}$  measurements. The results are used to study the sensitivities of 1- and 2-dimensional distributions in comparison with a multivariate analysis in Section 6.3. The results are also presented as exclusion plots. Section 6.4 subsequently discusses a possible extension of the STXS framework using the previous findings and complementing them with information about background processes. Conclusions are presented in Section 6.5.

## 6.1 $\mathcal{CP}$ -sensitive observables in $t\bar{t}H$ events

The  $t\bar{t}H$  production process is not only challenging to study due to its comparably low cross section but also due to the number of particles in the final state. The top quark has a leptonic and a hadronic decay mode, depending on the subsequent  $W$  boson decay, and ends up in a three-particle final state. This results in a  $t\bar{t}H$  event consisting of 8 (or more) particles at the parton level at the LHC, of which up to two can be neutrinos, escaping the detector. The complex final state provides the opportunity to define a rich number of observables suited for  $\mathcal{CP}$  analyses in this channel. A large number of observables have been proposed in prior work [129, 187, 190, 234, 238, 239, 241–244, 246, 247, 253–256, 259]. Some of them were already used successfully in experimental analysis [315]. However, it is not clear which observable is optimally suited for  $\mathcal{CP}$ -studies independent of the exact analysis, such as is needed for an extension of the STXS framework.

### 6.1.1 The generation of $t\bar{t}H$ events

The events for the  $t\bar{t}H$  process are generated at LO using `MadGraph5_aMC@NLO` (version 3.3.2) [143] with a COM energy of  $\sqrt{s} = 13$  TeV. They are subsequently scaled to NLO in QCD by using flat K-factors taken from [239]. Datasets are generated for two  $\mathcal{CP}$  hypotheses corresponding to  $\alpha_t = 0^\circ$

(the SM) and  $\alpha_t = 90^\circ$  (a full pseudoscalar Higgs coupling). In both cases, the SM rate  $g_t = 1$  is used. The two datasets consist of  $10^6$  events each at the parton level.

Throughout this chapter, the sensitivities of the observables will be tested at two benchmark scenarios. The first is  $\alpha_t = 45^\circ$ , which corresponds to the current experimental constraints as previously discussed. The second is  $\alpha_t = 35^\circ$ , which is a realistic value to be probed by analyses using the full Run 3 data. From the datasets at  $0^\circ$  and  $90^\circ$ , signal yields for any  $\mathcal{CP}$  hypothesis can be built via

$$N(g_t, \alpha_t) = g_t^2 \left[ N_{\text{SM}} \cos^2 \alpha_t + N_{\text{odd}} \sin^2 \alpha_t \right], \quad (6.1)$$

assuming that the interference term cancels out. Consequently,  $N_{\text{SM}} = N(1, 0)$  and  $N_{\text{odd}} = N(1, 90)$  are the signal yields for the two generated data samples, respectively. As a cross-check, two independent datasets were generated with  $\alpha_t = 35^\circ$  and  $\alpha_t = 45^\circ$  and compared to the corresponding predictions from Eq. (6.1). The distributions of the dedicated datasets agreed with the calculated yields within 1%, indicating the correctness of the scaling.

To test the performance of observables across multiple channels, three decay modes are considered in this chapter. The decays are not simulated using a decay routine such as `MadSpin`. Instead, their effect is approximated by smearing the final state variables and applying efficiency cuts to emulate the detector response and reconstruction process, as detailed in Section 6.2.1. Additional datasets are generated for the common non-Higgs background in the three decay channels. This includes the  $t\bar{t}\gamma\gamma$ , the  $t\bar{t}b\bar{b}$ , and the  $t\bar{t}W$  processes.  $5 \cdot 10^5$  events are generated for each non-Higgs background process using a similar setup as for the signal events. More details, such as the exact `MadGraph` syntax, are provided in Appendix B.

The cuts applied to events in the three decay channels can differ based on the analysis. Therefore, only a cut on the pseudorapidity of the Higgs and the tops  $|\eta_H| \leq 2$  and  $|\eta_t| \leq 2$  is used. While this is tighter than the pseudorapidity coverage of the ATLAS and CMS inner detectors of  $|\eta| \leq 2.5$ , it accounts for the decay products of heavy particles that escape the placed cuts. The selection efficiencies are taken directly from experimental analyses in the respective decay channel, instead of imposing additional cuts.

### 6.1.2 Observable definitions

The observables defined in the literature for  $\mathcal{CP}$  measurements of  $t\bar{t}H$  are defined in different rest frames. Moreover, choosing a specific rest frame can be important because it simplifies the description of certain effects. For example, polarization information about the top quark can best be obtained in its rest frame by decaying it and measuring the angular distribution of the charged lepton. This motivates a study of the observables in different rest frames in which they are defined.

In this work, the rest frame  $X$  is reached by performing subsequent Lorentz transformations, as Fig. 6.1 shows. The upper left graphic shows the laboratory (lab) frame, which is the default frame in which the events are generated. The interaction point (IP) of the colliding beams forms the origin of the coordinate system, while the beams themselves travel along the  $z$ -axis. The  $x$ -and  $y$ -axes are defined as pointing towards the center of the LHC and upwards, respectively. The first Lorentz transformation is a rotation around the  $z$ -axis, where the angle is chosen such that  $p_{y,X} = 0$  and  $p_{T,X} \parallel x$  after the rotation. This is shown in the lower left graphic of Fig. 6.1. The lower right sketch demonstrates the results of a Lorentz boost to the rest frame of  $p_{T,X}$ , where now  $p_X$  is parallel to the  $z$ -axis. The proton momenta no longer align with the  $z$ -axis in this case. The final step is the boost to the rest frame of  $p_X$ . This procedure

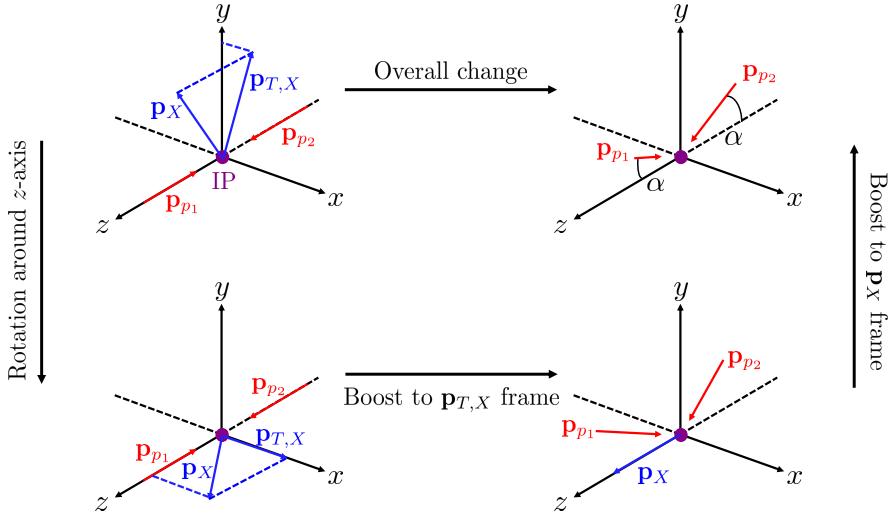


Figure 6.1: The procedure for boosting events to a rest frame  $X \in [H, t\bar{t}, t\bar{t}H]$ . Boosting happens via three individual steps. Figure modified from [313], original figure created by Alberto Carnelli.

guarantees that the two proton momenta have the same angle with the  $z$ -axis, which is dubbed the new beam axis. Here, three rest frames are considered besides the lab frame, specifically,  $X \in [H, t\bar{t}, t\bar{t}H]$ . They are defined by

$$\sum_i \mathbf{p}_i = 0, \quad i \in X \quad (6.2)$$

where  $\mathbf{p}$  corresponds to the 3-momentum.

An overview of all the observables considered here is given in Table 6.1. First, some basic kinematic observables are included.  $p_{T,H}$  is included because it corresponds to the current STXS binning in  $t\bar{t}H$ , if taken in the lab frame. It is also defined in the  $t\bar{t}$  and  $t\bar{t}H$  rest frames. Next, angular information about the tops is included via the differences  $|\Delta\eta_{t\bar{t}}|$  and  $|\Delta\phi_{t\bar{t}}|$ .  $|\Delta\phi_{t\bar{t}}|$  is not defined in the  $t\bar{t}$  frame, as it is zero per definition. Furthermore, the Lorentz-invariant masses of the  $t\bar{t}$  and  $t\bar{t}H$  systems are used.

Next, several higher-level observables are used, which are motivated by the literature. The first one is the so-called Collins-Soper angle  $\theta^*$ , which was first defined in [316]. It is only defined in the  $t\bar{t}$  frame and describes the angle between the axis of the  $t\bar{t}$  system and the beam axis. The variables  $b_1, \dots, b_4$  are taken from [234] and exploit different kinematic regions of the top quarks, which are sensitive to  $\mathcal{CP}$  violation. Although they were initially defined in the lab frame, they are straightforwardly extended to other frames (see e.g. [317]). Finally, the  $\phi_C$  variable is motivated in [253] and is defined as the angular difference between the planes spanned by the protons and top quarks in the Higgs rest frame. All of the variables in Table 6.1 are  $\mathcal{CP}$ -sensitive, but  $\mathcal{CP}$ -even. Most importantly, all observables are invariant under an exchange of the top and anti-top quark, avoiding the need to distinguish between them in an analysis.

Observable	Definition	Frame	Reference
$p_{T,H}$	–	lab, $t\bar{t}$ , $t\bar{t}H$	–
$ \Delta\eta_{t\bar{t}} $	$ \eta_t - \eta_{\bar{t}} $	all	–
$ \Delta\phi_{t\bar{t}} $	$ \phi_t - \phi_{\bar{t}} $	lab, $H$ , $t\bar{t}H$	–
$m_{t\bar{t}}$	$(p_t + p_{\bar{t}})^2$	frame-invariant	–
$m_{t\bar{t}H}$	$(p_t + p_{\bar{t}} + p_H)^2$	frame-invariant	–
$ \cos\theta^* $	$\frac{ \mathbf{p}_t \cdot \mathbf{n} }{ \mathbf{p}_t   \mathbf{n} }$	$t\bar{t}$	[247, 316]
$b_1$	$\frac{(\mathbf{p}_t \times \mathbf{n}) \cdot (\mathbf{p}_{\bar{t}} \times \mathbf{n})}{p_{T,t} p_{T,\bar{t}}}$	all	[234]
$b_2$	$\frac{(\mathbf{p}_t \times \mathbf{n}) \cdot (\mathbf{p}_{\bar{t}} \times \mathbf{n})}{ \mathbf{p}_t   \mathbf{p}_{\bar{t}} }$	all	[234]
$b_3$	$\frac{p_t^x p_{\bar{t}}^x}{p_{T,t} p_{T,\bar{t}}}$	all	[234]
$b_4$	$\frac{p_t^z p_{\bar{t}}^z}{ \mathbf{p}_t   \mathbf{p}_{\bar{t}} }$	all	[234]
$\phi_C$	$\arccos\left(\frac{ (\mathbf{p}_{p_1} \times \mathbf{p}_{p_2}) \cdot (\mathbf{p}_t \times \mathbf{p}_{\bar{t}}) }{ \mathbf{p}_{p_1} \times \mathbf{p}_{p_2}   \mathbf{p}_t \times \mathbf{p}_{\bar{t}} }\right)$	$H$	[253]

Table 6.1: Definitions of all observables used for the analysis, including a list of frames they are defined in, as well as references with further information about them. Table taken from [313].

### 6.1.3 Parton-level distributions

The events are generated at the parton level to avoid the need for channel-specific cuts, as motivated by the STXS framework. Here, some selected observables are shown at the parton level. The distributions of all other particles can be found in Appendix D. Selection efficiencies do not affect the distribution since they are all normalized to unity, but the channel-specific resolution may impact the shape (see Section 6.2 for a discussion).

In Fig. 6.2(a), the distribution of  $p_{T,H}$  is shown for the four different  $\mathcal{CP}$  hypotheses  $\alpha_t \in [0^\circ, 35^\circ, 45^\circ, 90^\circ]$ . Additionally, the current STXS bins are depicted as dashed gray lines in the plot. The bins with larger values of  $p_{T,H}$  are most sensitive to  $\mathcal{CP}$  as they show an excess of events over the SM hypothesis in a kinematic region where the number of background (SM) events is low. Furthermore, Fig. 6.2 shows distributions for  $|\Delta\eta_{t\bar{t}}|$ ,  $|\Delta\phi_{t\bar{t}}|$ ,  $m_{t\bar{t}}$ , and  $m_{t\bar{t}H}$ . The impact of the cut  $|\eta_t| \leq 2$  is best visible in the distribution of  $|\Delta\eta_{t\bar{t}}|$  which is constrained to  $[0, 4]$ .  $|\Delta\phi_{t\bar{t}}|$  shows potential for good sensitivity towards lower values, while all other variables have excesses in the non-SM distributions at higher values.

Next, the distributions of the variables  $b_1, \dots, b_4$  is shown in Fig. 6.3. The  $b_1$  variable exhibits an

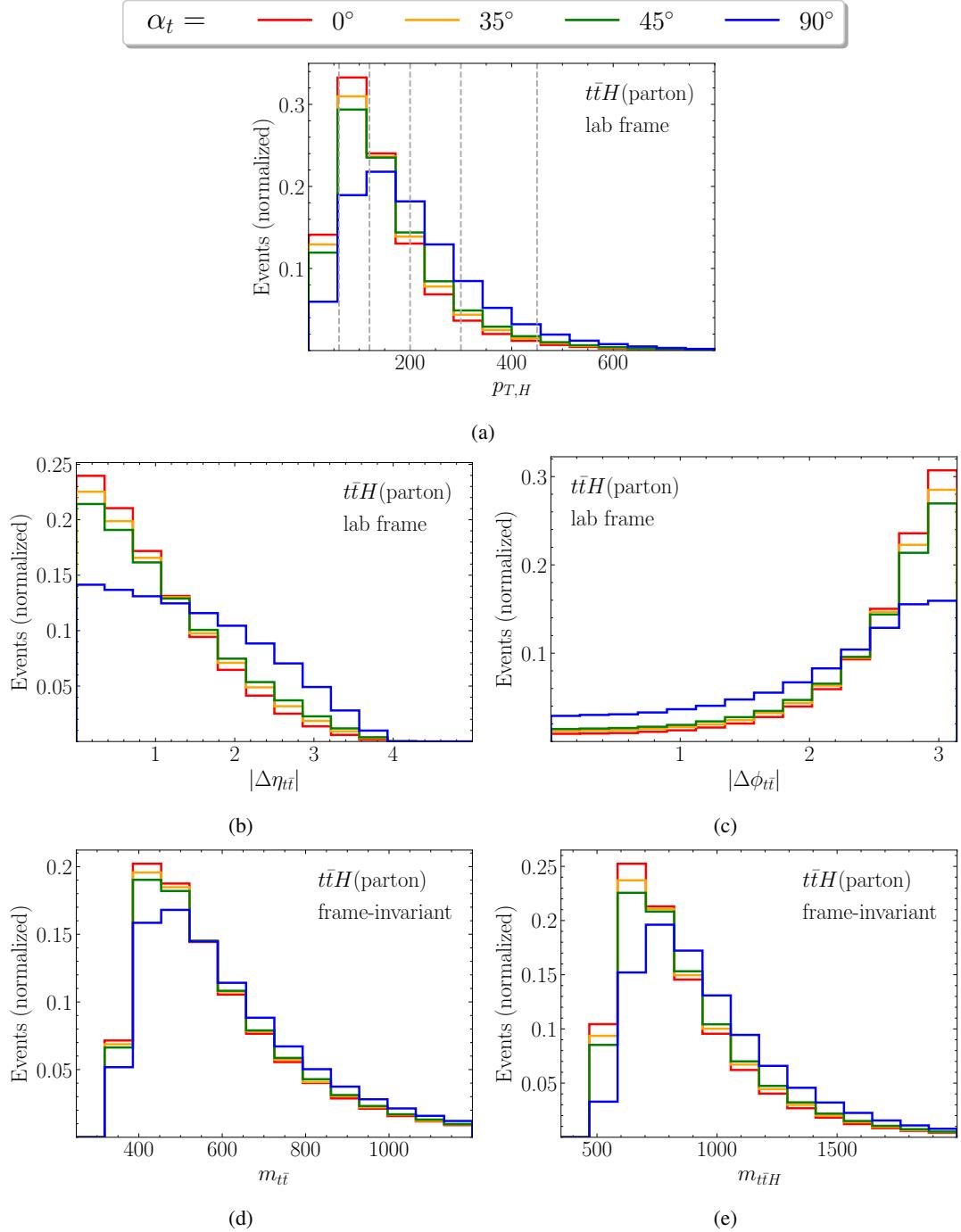


Figure 6.2: Parton-level distributions for (a)  $p_{T,H}$ , (b)  $|\Delta\eta_{t\bar{t}}|$ , (c)  $|\Delta\phi_{t\bar{t}}|$ , (d)  $m_{t\bar{t}}$ , and (e)  $m_{t\bar{t}H}$  for four different  $\mathcal{CP}$  hypotheses. Figure modified from [313].

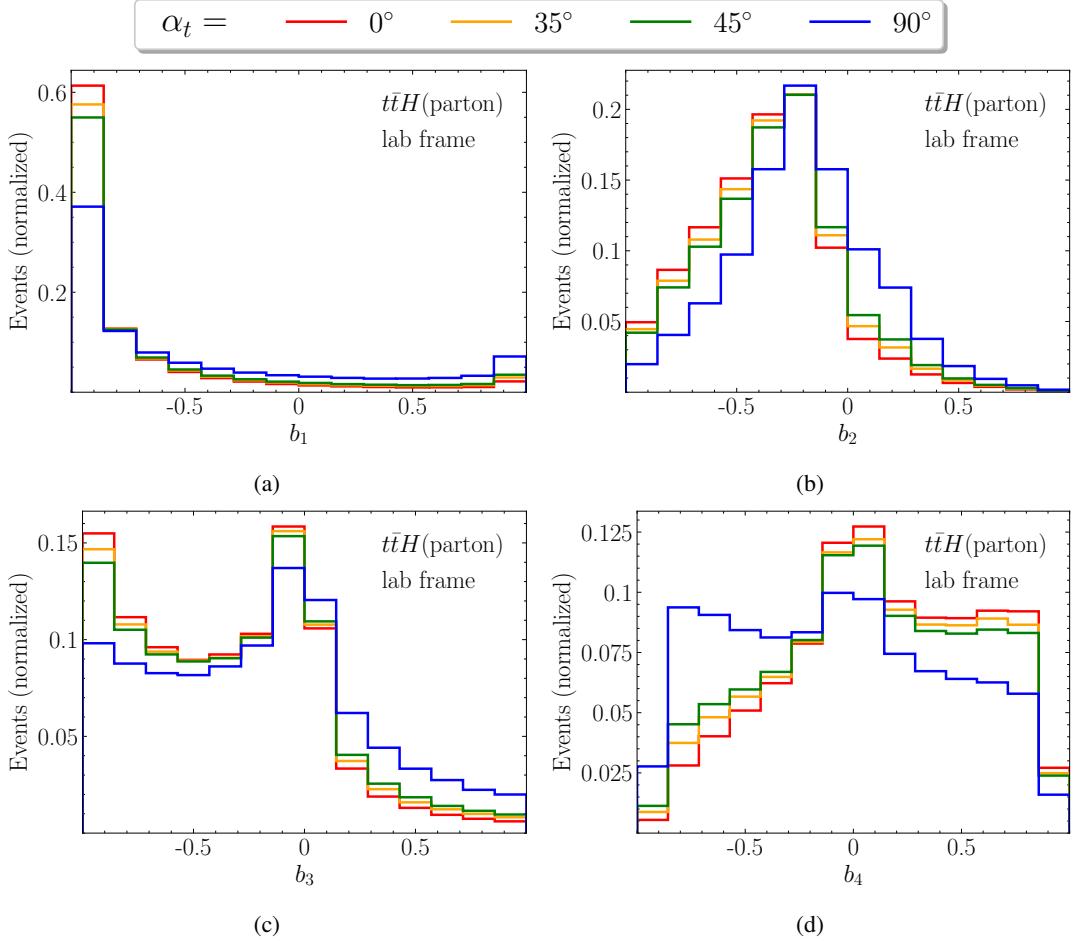


Figure 6.3: Parton-level distributions for (a)  $b_1$ , (b)  $b_2$ , (c)  $b_3$ , and (d)  $b_4$  for four different  $\mathcal{CP}$  hypotheses. Figure modified from [313].

interesting behavior, as most of the events are contained within a single bin near  $b_1 = -1$ . This is a consequence of the  $b_1$  variable capturing the angle between the transverse momenta of the top, or equivalently  $b_1 = \cos(\Delta\phi_{t\bar{t}})$ . Finally, the  $\phi_C$  and  $|\cos\theta^*|$  variables are depicted in Fig. 6.4. For larger allowed values of  $|\eta_t|$ , they appear as very flat distributions. Here, the cut  $|\eta_t| < 2$  introduces a drop in the events towards one end of the distribution, depending on the  $\mathcal{CP}$  angle.

## 6.2 Evaluation of the observable performance

With the observables defined, the next step is to assess their performance in separating a BSM  $\mathcal{CP}$  state from the SM prediction. Since the goal of this chapter is to provide suitable candidates for an STXS extension, the observables have to be evaluated in multiple channels. Specifically, this requires estimating the precision of reconstructing parton-level objects and separating background events, and including these effects in the sensitivity assessment.

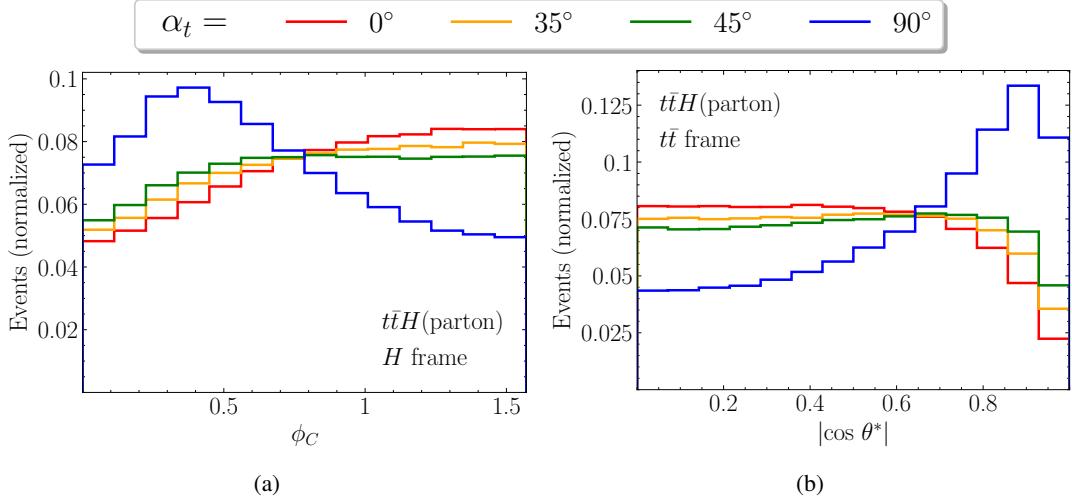


Figure 6.4: Parton-level distributions for (a)  $\phi_C$  and (b)  $|\cos \theta^*|$  for four different  $\mathcal{CP}$  hypotheses. Figure modified from [313].

### 6.2.1 Detector effects and selection efficiency

For the total estimation of the observable performance, three channels are considered here. These are the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel, the  $t\bar{t}H(\rightarrow b\bar{b})$  channel, and the  $t\bar{t}H(\text{multilep.})$  channel. The first two channels describe the respective Higgs boson decay, while the latter includes final states in which multiple leptons appear. All of these channels have been explored experimentally by ATLAS or CMS already [36, 130, 191, 192, 284, 315]. For each channel, several effects have to be accounted for independently. Specifically, the efficiency of both the event selection and reconstruction of the parton-level objects, as well as the smearing of variables due to limited detector resolution, must be considered. These are very challenging tasks that are beyond the scope of this work and are left for dedicated experimental analyses. To still provide a first assessment of the importance of the three channels, a simplified approach is taken here. The initial number of events in each channel is calculated from the production cross section, luminosity, and branching ratio. Then, previous results from ATLAS and CMS in the respective channels are used to obtain realistic acceptance factors of the event selection and efficiency factors of the reconstruction algorithms. Therefore, the final yield is

$$N(Y, \alpha_t) = \sigma_{t\bar{t}H}(\alpha_t) \cdot K_{\text{NLO}} \cdot \mathcal{L} \cdot \text{BR}_{\text{SM}}(Y) \cdot F(Y, \alpha_t) \quad (6.3)$$

for the channel  $Y$  at a  $\mathcal{CP}$ -mixing angle  $\alpha_t$ . The scaling factors  $F$  are listed in Table 6.2 and  $K_{\text{NLO}} = 1.14$  is used [239]. Additionally, observables are smeared to account for the limited detector resolution. This is done by smearing the kinematics of the parton-level objects using a Gaussian distribution. The mean of the distribution is the parton-level value, while the width is fixed and extracted from experimental analyses. The smearing is then propagated to all considered observables.

Table 6.2 shows the branching ratios, scaling factors, smearing factors, and final event yields for each channel. The  $t\bar{t}H(\rightarrow \gamma\gamma)$  has by far the smallest BR of the channels considered here. However, this is compensated for by two effects. First, the photons leave tracks in the inner detector, resulting

	$t\bar{t}H(\rightarrow \gamma\gamma)$	$t\bar{t}H(\text{multilep.})$	$t\bar{t}H(\rightarrow b\bar{b})$
BR	$2.27 \cdot 10^{-3}$	$6.79 \cdot 10^{-2}$	$5.81 \cdot 10^{-1}$
Acceptance/efficiency scaling factors			
$\alpha_t = 0^\circ$	$2.5 \cdot 10^{-1}$	$3.6 \cdot 10^{-2}$	$5.0 \cdot 10^{-3}$
$\alpha_t = 35^\circ$	$2.5 \cdot 10^{-1}$	$3.6 \cdot 10^{-2}$	$5.2 \cdot 10^{-3}$
$\alpha_t = 45^\circ$	$2.7 \cdot 10^{-1}$	$3.8 \cdot 10^{-2}$	$5.4 \cdot 10^{-3}$
$\alpha_t = 90^\circ$	$3.2 \cdot 10^{-1}$	$4.2 \cdot 10^{-2}$	$6.5 \cdot 10^{-3}$
Smearing factors			
$\Delta p_{T,H}$ [GeV]	4	120	80
$\Delta p_{T,t}$ [GeV]	40	70	70
$\Delta \eta_t$	0.5	0.8	0.8
$\Delta \phi_t$ [ $^\circ$ ]	--	20	20
Final $t\bar{t}H$ event yields at $300 \text{ fb}^{-1}$			
$\alpha_t = 0^\circ$	86	372	442
$\alpha_t = 35^\circ$	70	302	373
$\alpha_t = 45^\circ$	67	281	341
$\alpha_t = 90^\circ$	47	185	245

Table 6.2: The relevant factors that determine the final event yields expected for each channel. Displayed are the branching ratios, acceptance factors at each  $\mathcal{CP}$  angle, smearing factors for the parton-level kinematics, and final event yields. The initial number of events is determined via the  $t\bar{t}H$  cross section at each  $\mathcal{CP}$  angle times the K-factor and the luminosity of  $\mathcal{L} = 300 \text{ fb}^{-1}$ . Table taken from [313].

in a very precise Higgs reconstruction. This also yields a good separation of the signal from the non-Higgs background. Second, the photons from the Higgs decay cannot negatively influence the top reconstruction. This leads to very good scaling factors, which are estimated from the signal regions in [191]. In the SM case, 0.25 of all events end up in the signal region, and this value slightly improves to 0.32 in the  $\alpha_t = 90^\circ$  case. The improvement could be an effect of the Higgs being more boosted or the top quarks being more separated. The smearing is applied to the  $p_{T,H}$  variable using a Gaussian with a width of 4 GeV, which is obtained from the energy resolution for photons in ATLAS [318]. The smearing for  $p_{T,t}$  and  $\eta_t$  is again taken from [191] and corresponds to 40 GeV and 0.5, respectively.

In the  $t\bar{t}H(\rightarrow b\bar{b})$  channel, the situation is quite different. With over 50%, it is the Higgs decay with the largest branching ratio by far. However, this comes at the price of a much worse reconstruction of the Higgs boson, since two  $b$ -jets have to be reconstructed from the detected objects. Combining this with the large non-Higgs background containing top and bottom quarks makes the signal-to-background ratio in this channel significantly worse than in the previous one. Furthermore, four  $b$ -jets are expected at the final state, which allows for some combinatorics when reconstructing the Higgs boson and top quarks. These effects are reflected in the much worse scaling and smearing factors. The scaling factors slightly improve from  $5 \cdot 10^{-3}$  at  $\alpha_t = 0^\circ$  to  $6.5 \cdot 10^{-3}$  at  $\alpha_t = 90^\circ$  and are obtained from [315]. The transverse momentum of the Higgs boson is smeared using a width of 80 GeV to match [298]. The top quarks are smeared in  $p_{T,t}$  by 70 GeV, in  $\eta_t$  by 0.8, and in  $\phi_t$  by  $20^\circ$ , based on the reconstruction performed in [319, 320].

The  $t\bar{t}H(\text{multilep.})$  channel combines several decay modes that yield multiple charged leptons in the final state. Specifically, decay channels of the Higgs and tops with two leptons of the same sign (2ISS) or at least three leptons ( $>3l$ ) are considered, where  $l = e, \mu$ . With about 7%, the sum of the individual branching ratios forms a middle ground with the other two channels. The same is true for the signal-background ratio and consequently the scaling factors, which are calculated from the event yields in [192]. The scaling factors again exhibit the same trend of slightly improving when moving away from the SM. Precisely, they are within 0.036 for  $\alpha_t = 0^\circ$  and 0.042 for  $\alpha_t = 90^\circ$ . When it comes to the reconstruction and detector effects, the  $t\bar{t}H(\text{multilep.})$  channel has the disadvantage of containing a high possible number of energetic neutrinos. These can stem from both the Higgs decays and leptonic top quark decays. They are expected to complicate the reconstruction of  $p_{T,H}$ , which is therefore smeared using a width of  $\Delta p_{T,H} = 120$  GeV. The smearing of the top quarks is expected to be similar to the one in the  $t\bar{t}H(\rightarrow b\bar{b})$  channel, and therefore the same values are taken.

### 6.2.2 Evaluation of the sensitivity

Observables and their combinations are compared based on the significance with which they can exclude a certain BSM hypothesis. For this, it is assumed that events have been measured in each channel that follow the SM expectation. The resulting distributions are normalized to the final event yield in Table 6.2 for  $\alpha_t = 0^\circ$ . BSM expectations are obtained from the simulated samples and by normalizing them according to Eq. (6.1) and Table 6.2. Details of the significance evaluation can be found in Appendix C. The relevant equation is [321]

$$S = \sqrt{-2 \sum_{i=1}^{N_{\text{bins}}} \left( \lambda_{\text{SM},i} \ln \left[ \frac{\lambda_{\text{BSM},i}(\lambda_{\text{SM},i} + \sigma_i^2)}{\lambda_{\text{SM},i}^2 + \lambda_{\text{BSM},i}\sigma_i^2} \right] - \frac{\lambda_{\text{SM},i}^2}{\sigma_i^2} \ln \left[ 1 + \frac{\sigma_i^2(\lambda_{\text{BSM},i} - \lambda_{\text{SM},i})}{\lambda_{\text{SM},i}(\lambda_{\text{SM},i} + \sigma_i^2)} \right] \right)} \quad (6.4)$$

with the predicted number of SM (BSM) events per bin  $\lambda_{\text{SM},i}$  ( $\lambda_{\text{BSM},i}$ ) and the uncertainty parameter  $\sigma_i$ .

Eq. (6.4) produces accurate results if the number of events in each bin is at least 2 [321]. To account for this, a merging procedure is implemented when significances are calculated for binned variables. In an iterative approach, if the number of events in a single bin is less than 2, it is merged with a neighboring bin until each bin contains at least 2 events. In the case of 2-dimensional distributions, merging is preferred in the first dimension, although the final results were found not to depend on which dimension is preferred. The results are also stable with respect to the number of bins. Specifically, significances only deviate by a few percent if at least 5 events are demanded in each bin.

The uncertainties  $\sigma_i$  that appear in Eq. (6.4) are heavily dependent on the channel considered. In a real  $t\bar{t}H$  analysis, background events will enter the signal region and pollute the sensitivity. The amount of background events that manage to escape the signal-background separation depends not only on how well the background can be reduced via cuts, but also on uncertainties arising from the limited detector resolution or theoretical predictions. The  $\sigma_i$  can be split into statistical and systematic uncertainties

$$\sigma_i = \sqrt{(\sigma_i^{\text{stat}})^2 + (\sigma_i^{\text{syst}})^2} \quad (6.5)$$

and will be estimated for each channel as follows.

The statistical uncertainty  $\sigma_i^{\text{stat}}$  is based on the limited knowledge about the number of background events  $n_{\text{BG}}$  that mistakenly enter the signal region. Since the background events also follow a Poisson distribution, the corresponding uncertainty is  $\sigma_i^{\text{stat}} = \sqrt{n_{\text{BG}}}$ . This can be estimated from the signal-to-background ratios in the experimental analyses. Precisely

$$\sigma_i^{\text{stat}} = \sqrt{\lambda_{\text{SM},i} \left( \frac{S}{B} \right)^{-1}} \quad (6.6)$$

where  $S/B$  is taken to be constant across bins for simplification.

The systematical uncertainty  $\sigma_i^{\text{syst}}$  contains two effects. First, there is an uncertainty in the shape of the observable distributions, again stemming from uncertainties in the background events. This is directly proportional to the number of events in each bin and defines the  $\sigma_i^{\text{syst}}$  parameter. Second, the overall background rate is uncertain. This is a constant effect over all bins. In the  $t\bar{t}H(\rightarrow b\bar{b})$  and  $t\bar{t}H(\text{multilep.})$  channels, the background normalization uncertainty is expected to be larger than the rate difference between  $\lambda_{\text{SM},i}$  and  $\lambda_{\text{BSM},i}$ . Therefore, in these two channels, the rate information is removed by rescaling  $\lambda_{\text{BSM},i} \rightarrow \lambda_{\text{BSM},i}^*$ , where  $\sum \lambda_{\text{BSM},i}^* = \sum \lambda_{\text{SM},i}$ .

In Table 6.3, the final statistical and systematic uncertainties are listed for each channel, alongside the analyses they are based on. The uncertainties in the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel are statistically dominated and therefore  $\sigma_i^{\text{syst}}$  is set to zero. A value of  $S/B = 1$  is assumed due to the excellent energy resolution of the photons and the corresponding background suppression. The  $t\bar{t}H(\text{multilep.})$  and  $t\bar{t}H(\rightarrow b\bar{b})$  channels have worse uncertainties.  $S/B$  values of 0.4 and 0.1 are taken for the statistical uncertainty, respectively. Furthermore, the systematic uncertainty enters as 0.2 and 0.5 of the total events in each bin, respectively. Additionally, the rate information for all bins is removed, as stated above.

Channel	$\sigma_i^{\text{stat}}$	$\sigma_i^{\text{syst}}$	$\sigma_i$	Reference
$t\bar{t}H(\rightarrow \gamma\gamma)$	$\sqrt{\lambda_{\text{SM},i}}$	0	$\sqrt{\lambda_{\text{SM},i}}$	[191]
$t\bar{t}H(\text{multilep.})$	$\sqrt{2.5 \lambda_{\text{SM},i}}$	$0.2 \lambda_{\text{SM},i}$	$\sqrt{2.5 \lambda_{\text{SM},i} + 0.04 \lambda_{\text{SM},i}^2}$	[192]
$t\bar{t}H(\rightarrow b\bar{b})$	$\sqrt{10 \lambda_{\text{SM},i}}$	$0.5 \lambda_{\text{SM},i}$	$\sqrt{10 \lambda_{\text{SM},i} + 0.25 \lambda_{\text{SM},i}^2}$	[315]

Table 6.3: The statistical and systematic uncertainties, as well as the total uncertainties entering Eq. (6.4) in the respective channels.

## 6.3 Sensitivity to the top-Yukawa coupling

The current best experimental limit on the  $\mathcal{CP}$ -mixing angle of the top-Yukawa coupling is  $\alpha_t = 43^\circ$  [191], assuming the SM rate and using a luminosity of  $\mathcal{L} = 139 \text{ fb}^{-1}$ . In this work, analyses are targeted for the end of Run 3, where approximately  $\mathcal{L} = 300 \text{ fb}^{-1}$  of data will be available. With the expected improvements,  $\alpha_t = 35^\circ$  is chosen as a realistic benchmark scenario, for which the observables are evaluated. For a first test, all observables listed in Table 6.1 as well as all possible 2-dimension combinations are considered and their significances for excluding the  $\alpha_t = 35^\circ$  hypothesis are evaluated. Subsequently, the most promising variables are chosen and further examined, also in comparison to results from a multivariate analysis.

### 6.3.1 Significance evaluation

The significances are evaluated by employing Eq. (6.4) with the uncertainties in each channel provided in Table 6.3. 30 variables are considered in total. Their 1-dimensional significance is calculated using 14 evenly spaced bins. In addition, all possible 435 2-dimensional combinations (without permutations) of these observables are evaluated using an evenly spaced  $6 \times 6$  binning. While this binning is most likely too fine for an experimental analysis, it is used here to capture the full shape information. Experimental validity is ensured via the bin merging procedure detailed above. The full significance tables with all 465 significance estimations for each decay can be found in Appendix F.

For visual clarity, only a selected number of results are shown here. Specifically, the variables  $p_{T,H}$  and  $|\Delta\phi_{t\bar{t}}|$  are chosen, as  $p_{T,H}$  forms the basis of the current STXS binning and 2-dimensional combinations with  $|\Delta\phi_{t\bar{t}}|$  yield the highest overall significances. The results for excluding the  $\alpha_t = 35^\circ$  hypothesis are displayed in Fig. 6.5. The table shows the significances obtained from 1-dimensional distributions (indicated by the black bordered columns) and 2-dimensional combinations of variables. Only the combinations of  $p_{T,H}$  in Fig. 6.5(a) and  $|\Delta\phi_{t\bar{t}}|$  in Fig. 6.5(b) with variables in the lab and  $t\bar{t}$  frame are shown. Variables in the  $H$  and  $t\bar{t}H$  rest frames are discarded here as they consistently yield slightly lower significances.

Fig. 6.5 demonstrates that the by far highest significances are obtained in the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel. They show slight variations between 1.48 and 1.59 depending on the combination of variables. Therefore, they are relatively stable, which is a consequence of the rate information dominating the significance. In

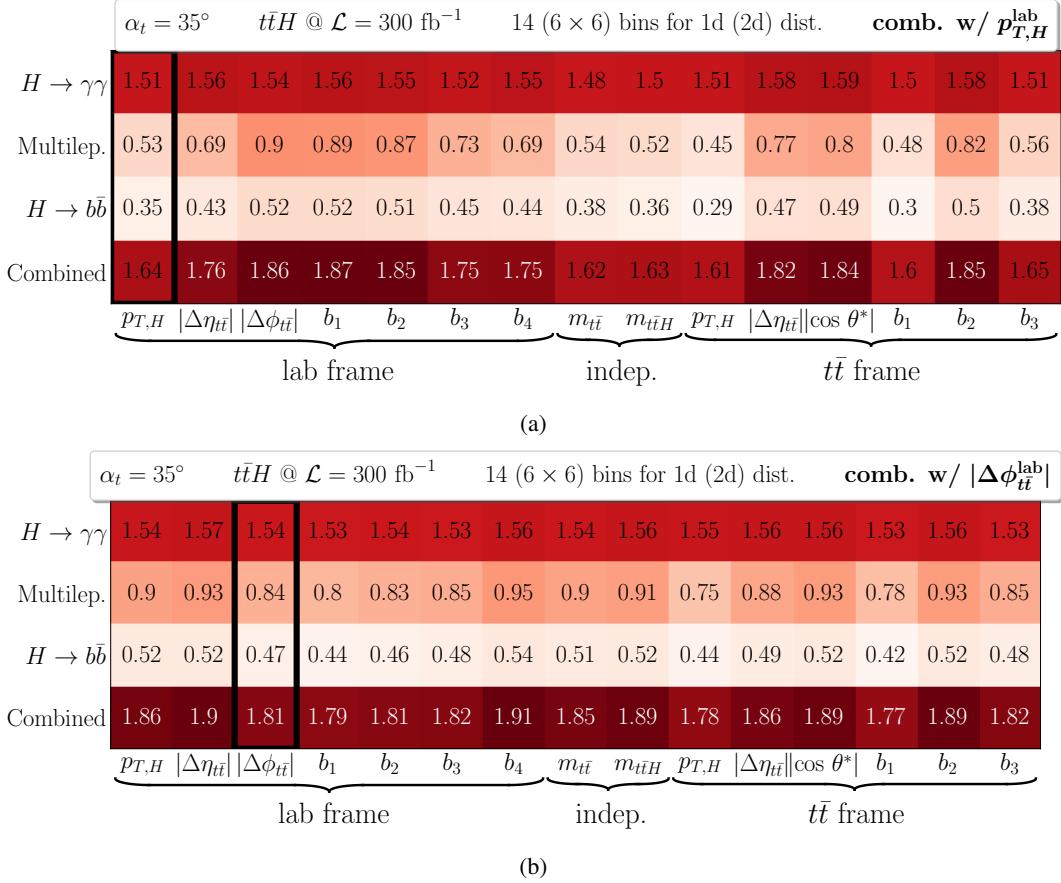


Figure 6.5: Significances for excluding the  $\alpha_t = 35^\circ$  BSM hypothesis in the three decay channels and their combination. Shown are combinations with (a)  $p_{T,H}$  and (b)  $\Delta\phi_{t\bar{t}}$ , both in the lab frame. Significances from 1-dimensional distributions are indicated by black borders. Figure taken from [313].

the other two channels, where the rate information is removed due to the uncertainty in the background normalization, the relative differences due to the shape information are much higher. The  $t\bar{t}H$ (multilep.) channel is more sensitive than the  $t\bar{t}H(\rightarrow b\bar{b})$  channel in all cases. The last column in both plots shows the significance obtained for a combination of the three channels. The highest overall significance of 1.91 is reached when combining  $|\Delta\phi_{t\bar{t}}|$  and  $b_4$ , both in the lab frame. However, the best 2-dimensional combinations involving  $p_{T,H}$  reach up to  $S = 1.87$ . This shows that an extension of the STXS  $p_{T,H}$  binning in  $t\bar{t}H$  is well possible while retaining  $\mathcal{CP}$  sensitivities that are nearly optimal for 2-dimensional distributions.

In Fig. 6.6, the same combinations of variables are shown, but with calculated significances to exclude the  $\alpha_t = 45^\circ$  hypothesis. A comparison with Fig. 6.5 shows that, while overall significances are of course higher, the ranking of the variable combinations is approximately the same. The highest significance is still obtained from the  $(|\Delta\phi_{t\bar{t}}|, b_4)$  pairing with  $S = 2.8$ , while multiple combinations with  $p_{T,H}$  yield  $S = 2.72$ . This shows that the results are robust to changes in the  $\mathcal{CP}$ -mixing angles that can be probed in the near future at the LHC. It also motivates the selection of a few candidates for an STXS extension to examine more closely. Specifically, variables are chosen that yield the highest significance when

$\alpha_t = 45^\circ$ $t\bar{t}H @ \mathcal{L} = 300 \text{ fb}^{-1}$ 14 (6 $\times$ 6) bins for 1d (2d) dist. <b>comb. w/ <math>p_{T,H}^{\text{lab}}</math></b>															
$H \rightarrow \gamma\gamma$	2.06	2.17	2.13	2.16	2.15	2.08	2.14	2.03	2.04	2.06	2.2	2.22	2.03	2.21	2.06
Multilep.	0.9	1.15	1.46	1.44	1.43	1.21	1.16	0.91	0.87	0.77	1.28	1.32	0.79	1.35	0.94
$H \rightarrow b\bar{b}$	0.59	0.71	0.83	0.83	0.82	0.73	0.74	0.62	0.59	0.5	0.77	0.8	0.49	0.81	0.63
Combined	2.32	2.55	2.72	2.72	2.71	2.51	2.55	2.3	2.3	2.26	2.66	2.7	2.24	2.72	2.35
	$p_{T,H}$	$ \Delta\eta_{t\bar{t}} $	$ \Delta\phi_{t\bar{t}} $	$b_1$	$b_2$	$b_3$	$b_4$	$m_{t\bar{t}}$	$m_{t\bar{t}H}$	$p_{T,H}$	$ \Delta\eta_{t\bar{t}}  \cos\theta^* $	$b_1$	$b_2$	$b_3$	
	lab frame				indep.				$t\bar{t}$ frame						

(a)

$\alpha_t = 45^\circ$ $t\bar{t}H @ \mathcal{L} = 300 \text{ fb}^{-1}$ 14 (6 $\times$ 6) bins for 1d (2d) dist. <b>comb. w/ <math> \Delta\phi_{t\bar{t}}^{\text{lab}} </math></b>															
$H \rightarrow \gamma\gamma$	2.13	2.18	2.11	2.1	2.11	2.1	2.17	2.0	2.17	2.14	2.16	2.16	2.1	2.17	2.08
Multilep.	1.46	1.51	1.37	1.3	1.36	1.38	1.54	1.45	1.48	1.23	1.43	1.52	1.25	1.5	1.39
$H \rightarrow b\bar{b}$	0.83	0.84	0.75	0.7	0.74	0.77	0.87	0.82	0.84	0.72	0.8	0.85	0.66	0.84	0.79
Combined	2.72	2.78	2.62	2.57	2.62	2.63	2.8	2.6	2.76	2.57	2.71	2.77	2.53	2.77	2.62
	$p_{T,H}$	$ \Delta\eta_{t\bar{t}} $	$ \Delta\phi_{t\bar{t}} $	$b_1$	$b_2$	$b_3$	$b_4$	$m_{t\bar{t}}$	$m_{t\bar{t}H}$	$p_{T,H}$	$ \Delta\eta_{t\bar{t}}  \cos\theta^* $	$b_1$	$b_2$	$b_3$	
	lab frame				indep.				$t\bar{t}$ frame						

(b)

 Figure 6.6: Same as Fig. 6.5 but for  $\alpha_t = 45^\circ$ .

paired with  $p_{T,H}$ . This results in  $|\Delta\phi_{t\bar{t}}|$ ,  $b_1$ , and  $b_2$  in the lab frame being chosen, as well as  $|\Delta\eta_{t\bar{t}}|$  and  $|\cos\theta^*|$  in the  $t\bar{t}$  frame.  $b_2$  in the  $t\bar{t}$  frame is discarded because it is approximately equal in sensitivity to  $b_2$  in the more intuitive lab frame.

### 6.3.2 Binning optimization

To extend the STXS framework for  $t\bar{t}H$  in a second dimension, a fixed binning of the second variable is needed. The  $p_{T,H}$  binning is adopted from the current STXS stage and is shown in the first row of Table 6.4. The other five variables considered are also split into six bins. The bins are chosen such that they all are expected to contain a reasonable number of events for all  $\alpha_t$ . Therefore, the bins in the sparsely populated regions of the phase space are wider. The so-obtained binning for each variable is dubbed the optimized binning and is presented in Table 6.4.

In Fig. 6.7, the normalized distributions with the optimized binning are shown for all lab frame variables, while Fig. 6.8 shows the corresponding distributions for variables defined in the  $t\bar{t}$  frame. In all cases, the distributions are shown for  $\alpha_t = 0^\circ$  (black solid lines) and  $\alpha_t = 35^\circ$  (orange dashed lines). Furthermore, the distributions are shown at the parton level and for each of the decays. Differences in

Variable	Bin borders						
	1	2	3	4	5	6	7
$p_{T,H}$ [GeV]	0	60	120	200	300	450	$+\infty$
$ \Delta\phi_{t\bar{t}}^{\text{lab}} $ [rad.]	0	$\pi/4$	$\pi/2$	$2\pi/3$	$5\pi/6$	$11\pi/12$	$\pi$
$b_1^{\text{lab}}$	-1	-0.95	-0.8	-0.2	0.3	0.8	1.0
$b_2^{\text{lab}}$	-1	-0.6	-0.4	-0.2	0.	0.3	1.0
$ \Delta\eta_{t\bar{t}}^{t\bar{t}} $	0	0.5	1	1.5	2	3	5
$ \cos\theta^* $	0	0.2	0.4	0.55	0.7	0.85	1

Table 6.4: Bin borders for each variable in the optimized binning. The binning in  $p_{T,H}$  corresponds to the one in the current STXS framework.

the distributions are the result of the applied smearing.

From these normalized distributions, the significance for excluding the  $\alpha_t = 35^\circ$  hypothesis is calculated in the same way as above. The results are presented for combinations of variables with  $p_{T,H}$  and  $|\Delta\phi_{t\bar{t}}|$  in Fig. 6.9(a) and Fig. 6.9(b), respectively. A comparison with the corresponding columns in Fig. 6.5 shows that the significance of each 2-dimensional combination is either kept or improves slightly. The 1-dimensional binning in only  $p_{T,H}$  or  $|\Delta\phi_{t\bar{t}}|$  is less fine compared to Fig. 6.5 and therefore leads to slightly reduced significances. Overall, the results are stable with respect to the different binning. In the following, the optimized binning will be used exclusively.

### 6.3.3 Comparison to a multivariate analysis

So far, only combinations of two variables have been considered for calculating the  $\mathcal{CP}$  sensitivity. However, these do not capture the full kinematic information of the  $t\bar{t}H$  events. To achieve optimal  $\mathcal{CP}$  sensitivity, multivariate analysis techniques can be employed. Despite their better performance in a specific analysis, common approaches are not well-suited for an STXS-like binning due to their model dependency and difficulty in interpretability. Still, the loss in information when using a simple, 2-dimensional binning over a machine learning approach leveraging all kinematic variables should be assessed. This is done in the following.

`XGBoost` is used to perform a multivariate analysis in the three  $t\bar{t}H$  channels with boosted decision trees, as it has already been used in previous analyses [191, 192]. The training of `XGBoost` is performed independently in each of the three decay channels with all variables listed in Table 6.1. Accordingly, the  $t\bar{t}H$  data used for the training is smeared according to the resolution in each channel. The training process is a classification task of differentiating between the SM at  $\alpha_t = 0^\circ$ , and the full  $\mathcal{CP}$ -odd case at

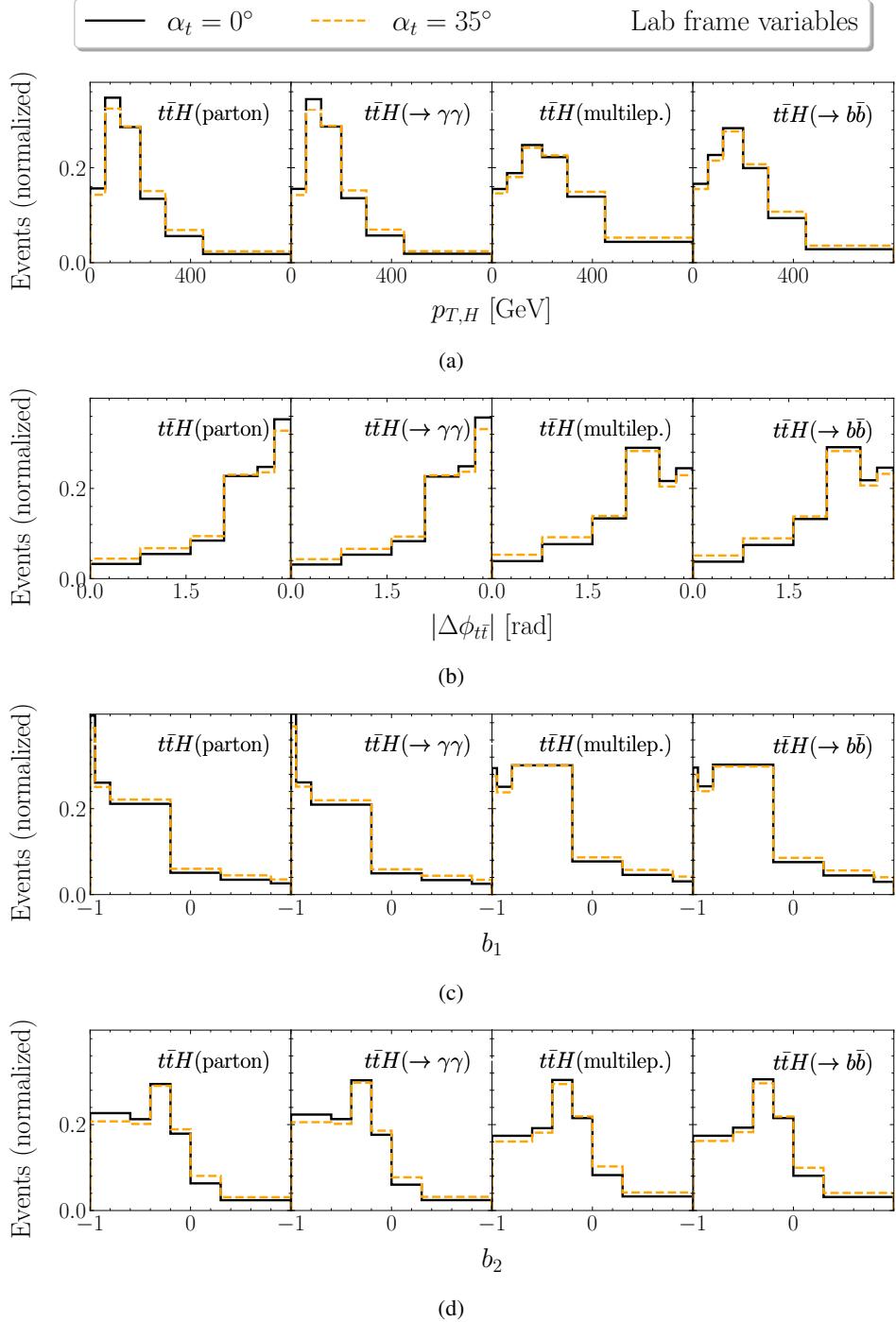


Figure 6.7: Normalized distributions at  $\alpha_t = 0^\circ$  and  $\alpha_t = 35^\circ$  for (a)  $p_{T,H}$ , (b)  $|\Delta\phi_{t\bar{t}}|$ , (c)  $b_1$ , and (d)  $b_2$  (all in the lab frame). For each variable, distributions are shown at the parton-level and for each of the three decay channels. Figure taken from [313].

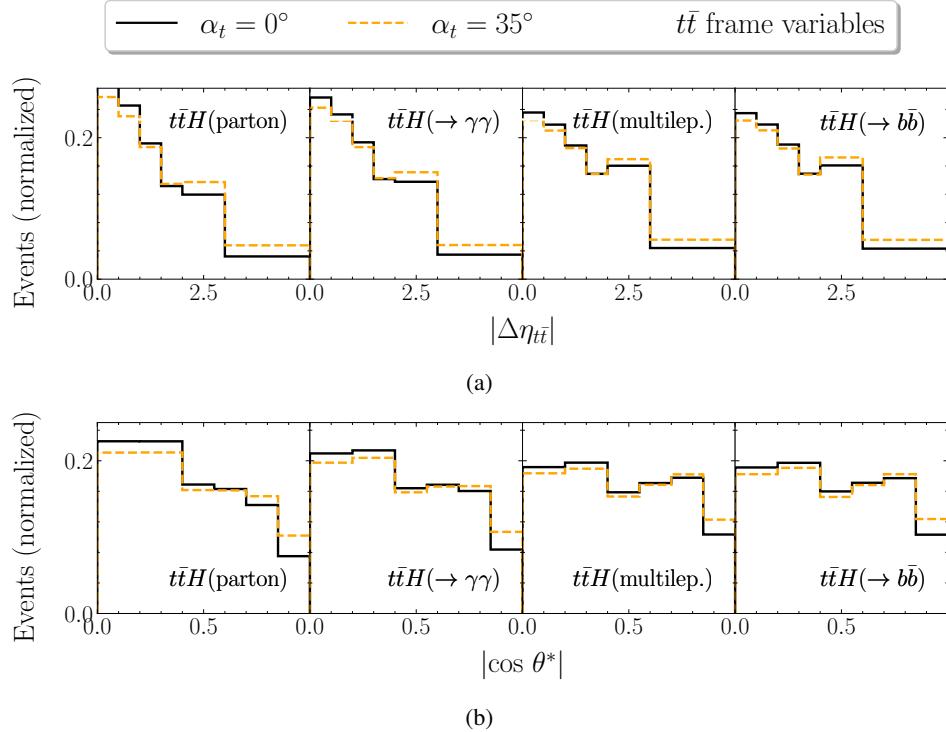


Figure 6.8: Same as Fig. 6.7 but for (a)  $|\Delta\eta_{t\bar{t}}|$  and (b)  $|\cos\theta^*|$  in the  $t\bar{t}$  frame. Figure taken from [313].

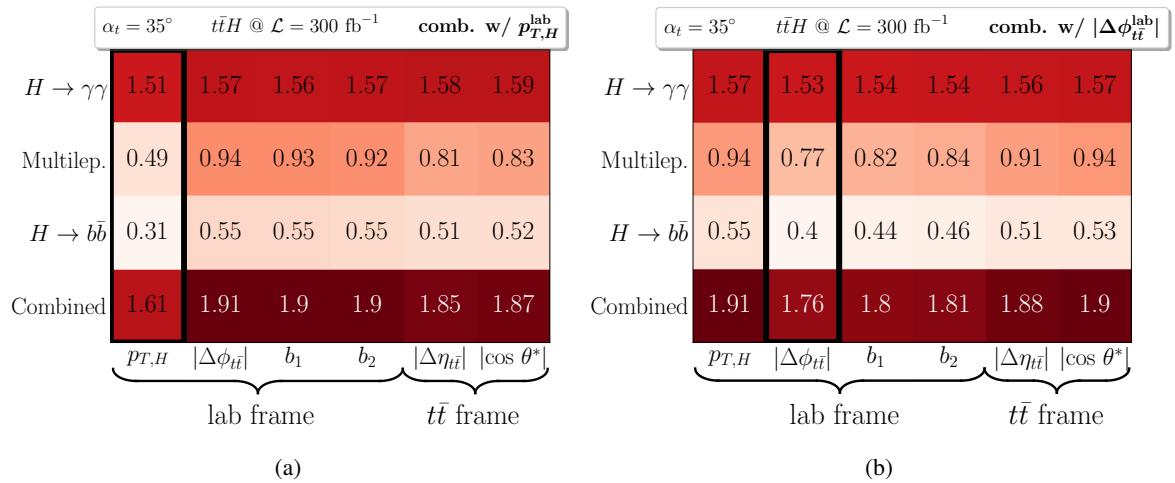


Figure 6.9: Same as Fig. 6.5 but for a subset of variables best performing in combination with  $p_{T,H}$  and including optimized binning. Figure taken from [313].

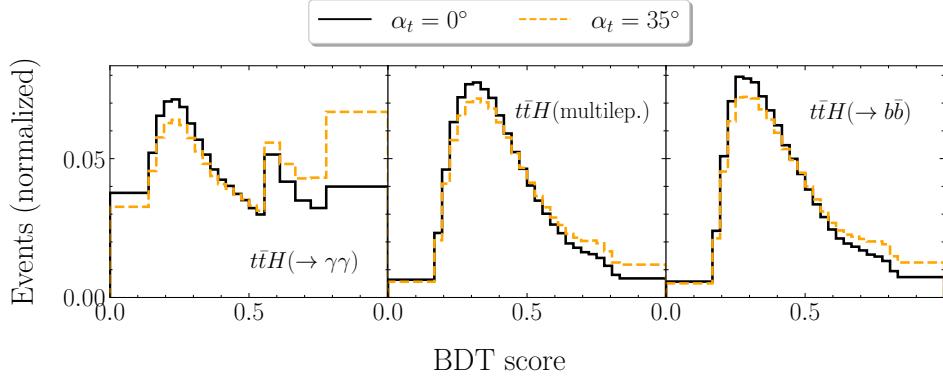


Figure 6.10: Distributions of the  $\alpha_t = 0^\circ$  and  $\alpha_t = 35^\circ$  events after passing through the trained XGBoost model and merging bins in all three decay channels.

$\alpha_t = 90^\circ$ . Other  $\mathcal{CP}$  angles can be extrapolated since sensitivity to the interference term is not expected. The  $\alpha_t = 90^\circ$  samples are used during training because the final results generalize better compared to training with, e.g., the  $\alpha_t = 35^\circ$  events. Furthermore, the performance at  $\alpha_t = 35^\circ$  is not reduced when training with the  $\alpha_t = 90^\circ$  data. The  $k$ -fold cross validation is used with  $k = 5$ , subsequently splitting the training and validation data into 80% and 20% of the events and iterating over the 5 possible permutations of the splitting (see e.g. [322]).

After training, pure test samples corresponding to  $\alpha_t = 0^\circ$  and  $\alpha_t = 35^\circ$  are given to XGBoost. The resulting BDT score is first divided into 36 evenly spaced bins. Then, the distributions are normalized to the expected events in the respective channel, and neighboring bins are merged according to the procedure detailed in Section 6.2.2. The distributions with the merged bins are shown in Fig. 6.10, normalized to unity for a better comparison between the three decay channels. The widest bins in each case are found in the regions close to a BDT score of 0 and 1, indicating a low number of events in these regions. However, after merging, the bins close to a BDT score of 1 are expected to drive the total significance due to the suppressed SM-like  $t\bar{t}H$  events.

In Table 6.5, the significances obtained per channel from the BDT score are compared to the ones obtained from the  $(p_{T,H}, |\Delta\phi_{t\bar{t}}|)$  combination. In the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel, the significance due to the BDT usage increases by  $\sim 11\%$ , whereas in the  $t\bar{t}H(\text{multilep.})$  and  $t\bar{t}H(\rightarrow b\bar{b})$  channels it increased by  $\sim 24\%$  and  $\sim 25\%$ , respectively. The latter two channels see much larger improvements since the rate information is not available, and the entire sensitivity stems from the shape of the distributions. Furthermore, the higher statistics in these channels additionally boost the shape information. The usage of the BDT boosts the significance from  $S = 1.91$  to  $S = 2.21$  compared to the 2-dimensional binning and consequently indicates an improvement of  $\sim 16\%$ . Therefore, while not optimal in comparison to a multivariate analysis, the 2-dimensional binning still captures most of the  $\mathcal{CP}$  information. This is ensured since the statistically limited  $t\bar{t}H(\rightarrow \gamma\gamma)$  dominates the constraints, which is not expected to change for future analyses using  $\mathcal{L} = 300 \text{ fb}^{-1}$  of data.

Channel	Significance (BDT)	Significance (2d)
$t\bar{t}H(\rightarrow \gamma\gamma)$	1.75	1.57
$t\bar{t}H(\text{multilep.})$	1.17	0.94
$t\bar{t}H(\rightarrow b\bar{b})$	0.69	0.55
Combined	2.21	1.91

Table 6.5: Comparison of the significances for excluding  $\alpha_t = 35^\circ$  obtained from the BDT score and from the 2-dimensional binning in  $p_{T,H}$  and  $|\Delta\phi_{t\bar{t}}|$ . Table taken from [313].

### 6.3.4 Limits on the $\mathcal{CP}$ state of the top-Yukawa coupling

So far, only fixed values of the  $\mathcal{CP}$ -mixing angle  $\alpha_t$  at the SM rate  $g_t = 1$  have been considered for the significance evaluation. In Figs. 6.11 and 6.12, exclusion plots are shown for the full  $(\alpha_t, g_t)$  parameter plane. The dashed lines correspond to the  $2\sigma$  levels in each of the three decay channels, as well as their combination. BSM samples at each parameter point are obtained via Eq. (6.1) and tested against the SM hypothesis by assuming a SM-like measurement. Further details are provided in Appendix C.

In all plots, the limits from the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel appear as closed contours, while the  $t\bar{t}H(\text{multilep.})$  and  $t\bar{t}H(\rightarrow b\bar{b})$  channels show only slight variations in  $\alpha_t$  when  $g_t$  is varied. This is a consequence of the unavailable rate information in these channels, which results in almost vertical lines in the exclusion plot. The deviation from a vertical line stems from the merging of the bins. The lower  $g_t$  is, the more bins are merged, which removes shape information and consequently results in less stringent limits on  $\alpha_t$ .

The exclusion limits coming from a 1-dimensional binning in  $p_{T,H}$  show that the constraints are mostly dominated by the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel. The  $2\sigma$  contour from the  $t\bar{t}H(\rightarrow b\bar{b})$  channel is barely visible and has only a minor impact on the combined results. At the SM rate, also the  $t\bar{t}H(\text{multilep.})$  channel has only minimal impact, as the constraints improve from  $|\alpha_t| < 45^\circ$  (for  $H \rightarrow \gamma\gamma$ ) to  $|\alpha_t| < 43^\circ$ . On the other hand, at higher  $g_t$  the impact of the multi-lepton final state appears more clearly as the  $H \rightarrow \gamma\gamma$  channel constrains  $g_t < 1.36$ , which improves to  $g_t < 1.28$  in the combined limit. For a 1-dimensional binning in  $|\Delta\phi_{t\bar{t}}|$  (see Fig. 6.11(b)), the improvements in the limits compared to  $p_{T,H}$  largely stem from the higher sensitivity in the  $t\bar{t}H(\text{multilep.})$  and  $t\bar{t}H(\rightarrow b\bar{b})$  channels. Here, the constraints improve from  $|\alpha_t| < 45^\circ$  to  $|\alpha_t| < 39^\circ$  at the  $2\sigma$  level and for a SM-like rate. The latter also improves from  $g_t < 1.36$  to  $g_t < 1.26$ .

In Figs. 6.11(c) and 6.11(d), the binning in  $p_{T,H}$  and  $|\Delta\phi_{t\bar{t}}|$  is extended by a binning in  $|\cos\theta^*|$  as a second dimension. The exact choice of the second variable makes only a minor difference in the exclusion limits. The constraints of all three channels get tighter compared to the 1-dimensional cases. A comparison between the  $2\sigma$  contour of the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel with the combination of channels yields an improvement of  $|\alpha_t| < 44^\circ$  to  $|\alpha_t| < 37^\circ$  at  $g_t = 1$  for both variable combinations. The deviation of the  $t\bar{t}H$  rate from the SM is allowed within  $g_t \in [0.85, 1.33]$  before combining the channels and  $g_t \in [0.85, 1.23]$  afterwards.

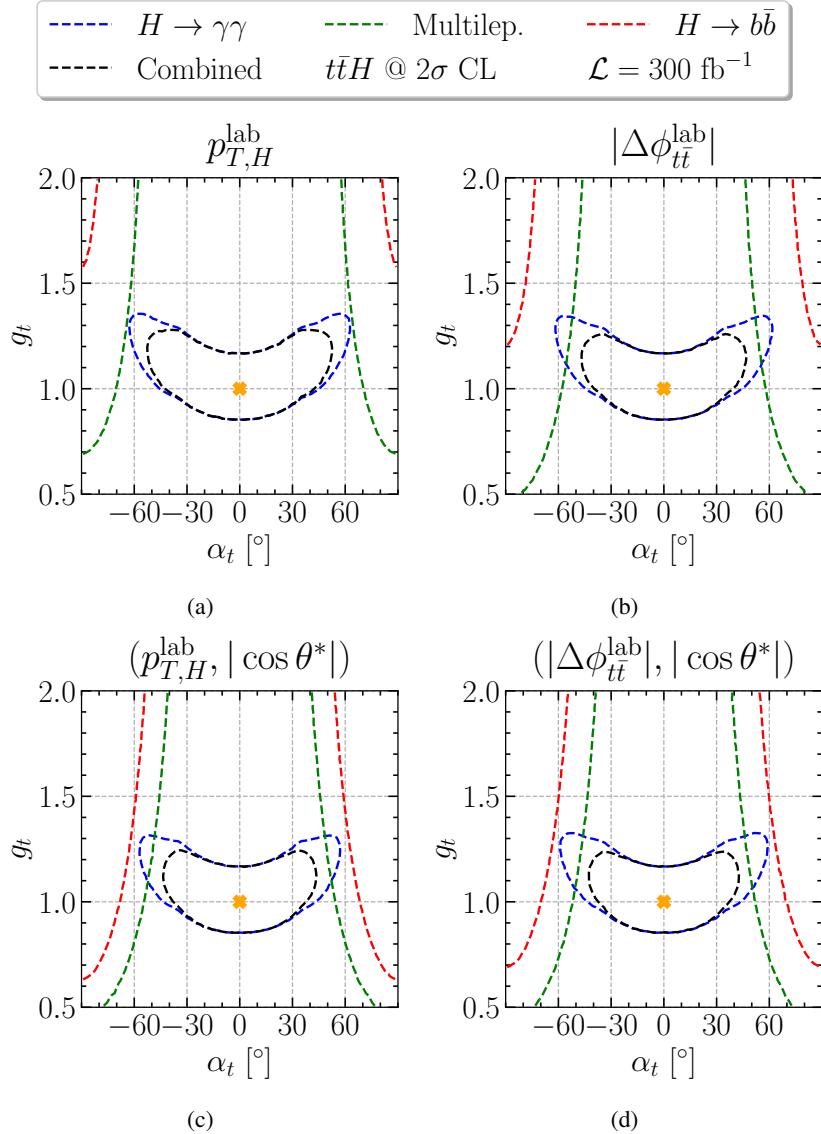


Figure 6.11: Exclusion limits for the top-Yukawa coupling when binning in (a) only  $p_{T,H}$ , (b) only  $|\Delta\phi_{t\bar{t}}|$ , (c)  $p_{T,H}$  and  $|\cos\theta^*|$ , and (d)  $|\Delta\phi_{t\bar{t}}|$  and  $|\cos\theta^*|$ . The  $t\bar{t}H(\rightarrow \gamma\gamma)$ ,  $t\bar{t}H(\text{multilep.})$ , and  $t\bar{t}H(\rightarrow b\bar{b})$  channels are shown as red, green, and blue dashed lines, respectively. Their combination is shown in black dashed lines. The orange cross corresponds to the SM point. Figure modified from [313].

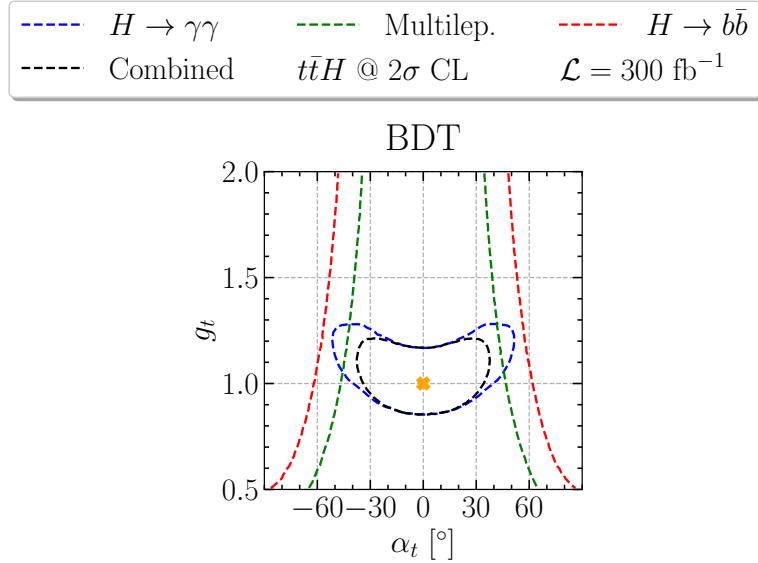


Figure 6.12: Same as Fig. 6.11 but for the BDT score. Figure modified from [313].

As shown in Section 6.3.3, a multivariate analysis reaches a higher sensitivity towards the  $\mathcal{CP}$ -state of the top-Yukawa coupling in a specific channel. While XGBoost was trained independently in all three channels, a naive combination can be made to compare its results with the 2-dimensional binning. The corresponding exclusion limits are shown in Fig. 6.12. As expected, the constraints from all three channels are tighter because the BDT has access to the full kinematic information. Here, already the  $t\bar{t}H$ (multilep.) allows to constrain  $|\alpha_t| < 47^\circ$  at the  $2\sigma$  level for  $g_t = 1$ . The  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel yields  $|\alpha_t| < 40^\circ$  and the combination  $|\alpha_t| < 34^\circ$ . It should be noted here that (for the SM rate) the extension of the STXS binning by a second variable already offers a potential improvement in the measurement of  $\alpha_t$  by about  $6^\circ$  compared to the current binning (at the  $2\sigma$  level). The further improvement from the XGBoost score with full kinematic information is about  $3^\circ$ .

## 6.4 Extending the STXS framework

In the previous section, five candidate observables have been identified to serve as an extension of the  $t\bar{t}H$  STXS binning. All of them show only minor differences in their sensitivity towards the  $\mathcal{CP}$ -state of the top-Yukawa coupling. However, so far, the effects of the Higgs-background processes were taken into account only indirectly via the acceptance factors and uncertainties in Section 6.2. In the following, the background will be studied in more detail to further restrict the variable choices.

### 6.4.1 Impact of the background distributions

Up to now, the signal-to-background ratios extracted from experimental measurements have been assumed to be constant. For a more realistic scenario, the differential cross section of the background

is taken into account. The shape of the background distributions can have strong implications for the significance evaluation. For example, in  $p_{T,H}$  a constant  $S/B$  assumption yields only conservative values for the significance because the high- $p_T$  bins that are enriched in BSM events are simultaneously almost free of background events. In contrast, if the background peaks in a kinematic region that was previously expected to drive the sensitivity, the significance might be overestimated. Here, this behavior will be studied for each of the five candidate variables.

### Higgs background processes

The main Higgs background processes to  $t\bar{t}H$  are associated with a single top quark, namely  $tHq$  and  $tHW$ . The light quark or  $W$  boson can be wrongly reconstructed as an additional top quark. However, as discussed in Section 3.3.1, the cross section of these processes is comparably low in the SM. Specifically,

$$\frac{\sigma_{tHq}^{\text{SM}}}{\sigma_{t\bar{t}H}^{\text{SM}}} \approx 0.15, \quad \frac{\sigma_{tHW}^{\text{SM}}}{\sigma_{t\bar{t}H}^{\text{SM}}} \approx 0.03 \quad . \quad (6.7)$$

Interestingly, while the cross section of  $t\bar{t}H$  decreases with increasing  $\mathcal{CP}$ -mixing angle, the opposite is true for the  $tH$  processes. At the benchmark point of  $\alpha_t = 35^\circ$  the ratios become

$$\frac{\sigma_{tHq}^{\alpha_t=35^\circ}}{\sigma_{t\bar{t}H}^{\alpha_t=35^\circ}} \approx 0.24, \quad \frac{\sigma_{tHW}^{\alpha_t=35^\circ}}{\sigma_{t\bar{t}H}^{\alpha_t=35^\circ}} \approx 0.05 \quad (6.8)$$

and therefore  $tH$  is still suppressed. Recent experimental studies have shown that the impact of  $tH$  for constraining  $\alpha_t$  is small compared to  $t\bar{t}H$  due to the low yields and large uncertainties (see e.g. [315]). With the additional reduction in  $tH$  events due to the signal selection, the Higgs background processes are not expected to significantly influence the results. While  $\mathcal{CP}$ -sensitivity might also be extracted from these channels in future analyses, this is beyond the scope of this study.

### Non-Higgs background processes

The picture changes for the non-Higgs background processes, as these have much higher cross sections. The main background in each of the considered channels consists of a  $t\bar{t}$  system and particles that can be misidentified as a Higgs boson. For the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel, two photons with  $m_{\gamma\gamma} \approx m_H$  may be produced in the  $t\bar{t}\gamma\gamma$  process via FSR. Similarly, the non-Higgs background to  $t\bar{t}H(\rightarrow b\bar{b})$  is  $t\bar{t}b\bar{b}$ , where  $m_{b\bar{b}} \approx m_H$  could fake a Higgs boson. Finally,  $t\bar{t}W$  is the dominant background of the  $t\bar{t}H$ (multilep.) channel due to the number of leptons in the final state. For all non-Higgs background channels, events are generated as described in Section 6.1.1 with additional details provided in Appendix B. Distributions for the variables are obtained by reconstructing a fake Higgs boson out of the respective parton-level objects.

The results are shown in Fig. 6.13 for each of the five variables. Each plot showcases a single variable and is split into two panels. The upper panel shows the distribution of the three main non-Higgs background channels, while the lower panel illustrates the combined significance of the  $t\bar{t}H$  channel per bin. This allows for estimating whether the constant  $S/B$  assumption holds and whether the sensitivity estimation was conservative or too optimistic. The significances are presented for excluding the  $\alpha_t = 35^\circ$  hypothesis at  $g_t = 1$  and  $\mathcal{L} = 300 \text{ fb}^{-1}$  given a measurement of SM-like data. For this, the non-Higgs

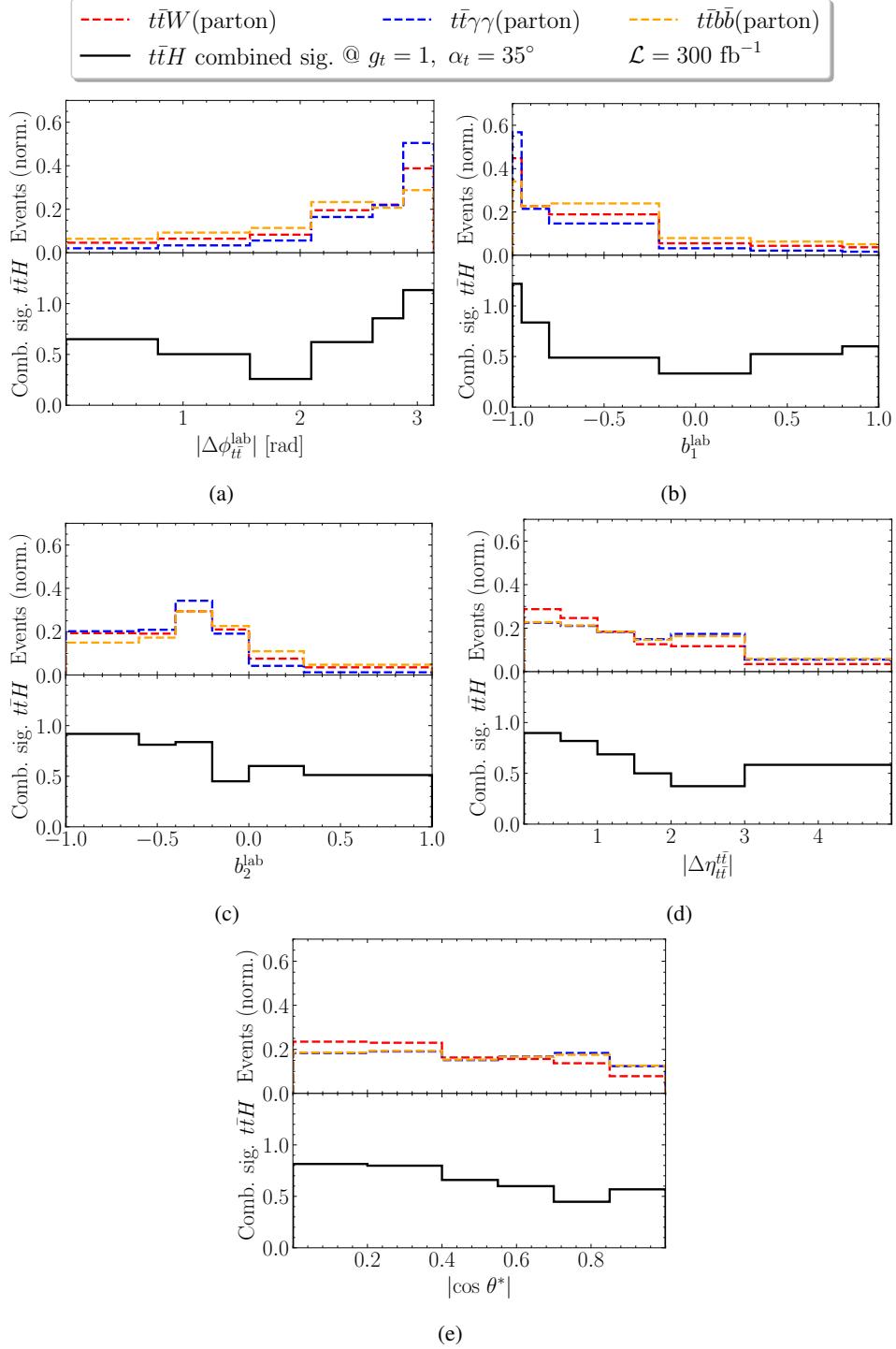


Figure 6.13: Distributions of the relevant non-Higgs background processes for the three decay channels in (a)  $|\Delta\phi_{t\bar{t}}|$ , (b)  $b_1$ , (c)  $b_2$ , (d)  $|\Delta\eta_{t\bar{t}}|$ , and (e)  $|\cos\theta^*|$ . Additionally, the significance of the combined decay channels per bin is shown. Figure taken from [313].

background is assumed to be SM-like.

In the upper row, the variables  $|\Delta\phi_{t\bar{t}}|$  and  $b_1$  are shown, which are connected via the cosine. Consequently, both exhibit a similar behavior. There is an outer bin that drives the overall sensitivity; however, all three non-Higgs background processes also peak in that bin. Therefore, it must be assumed that the sensitivity of this bin is washed out in an experimental measurement and that the initial estimation of  $S$  was too optimistic. The  $b_2$  variable shows a smaller peak of the background in the third bin, but this is not the only bin that provides a high significance. For the  $|\Delta\eta_{t\bar{t}}|$  and  $|\cos\theta^*|$  variables, the shape of the background is similar to the shape of the significances, but without any distinct peak. Overall, the constant  $S/B$  assumption appears to work best for the Collins-Soper angle. In conclusion, since  $b_2$ ,  $|\Delta\eta_{t\bar{t}}|$ , and  $|\cos\theta^*|$  do not show any sign of an overestimation of  $S$  due to the background, they are selected as the final candidate observables. On the other hand,  $|\Delta\phi_{t\bar{t}}|$  and  $b_1$  are discarded.

### 6.4.2 Final $t\bar{t}H$ STXS extension proposal

As of now, the 2-dimensional variable combinations allow the usage of up to 36 bins. This can generally lead to better shape information compared to the 6 bins from the 1-dimensional binning. For a fair comparison, the option of extending the current STXS binning into a 1-dimensional one containing 36  $p_{T,H}$  bins needs to be tested. This is shown in Fig. 6.14 for  $\mathcal{L} = 300 \text{ fb}^{-1}$  and a naive but conservative projection to  $\mathcal{L} = 3000 \text{ fb}^{-1}$  for the HL-LHC.<sup>1</sup> The dotted lines show the current STXS binning, the dashed lines a binning in  $p_{T,H}$  with 36 bins, and the solid lines the 2-dimensional binning in  $p_{T,H}$  and  $|\cos\theta^*|$  with  $6 \times 6$  bins. It is evident that the 2-dimensional combination of variables outperforms the 1-dimensional distribution with the same number of bins. This is true both for the luminosity after Run 3 has ended and at the HL-LHC.

Using all previous findings, a concrete proposal for extending the STXS framework in  $t\bar{t}H$ , focusing on the  $\mathcal{CP}$  structure of the top-Yukawa coupling, can be made. The following findings of this study are the most important for supporting the suggestion:

- For backwards compatibility, an extension of the STXS Stage 1.2 should be based on the current binning in  $p_{T,H}$ . It was shown that combinations of a second variable with  $p_{T,H}$  yield close to optimal sensitivity in comparison to all possible 2-dimensional combinations of variables.
- The variables that are best suited for a combination with  $p_{T,H}$  were shown to be  $|\Delta\phi_{t\bar{t}}|$ ,  $b_1$ , and  $b_2$  in the lab frame, as well as  $|\Delta\eta_{t\bar{t}}|$  and  $|\cos\theta^*|$  in the  $t\bar{t}$  frame. They all yield similar constraints in the  $(g_t, \alpha_t)$  parameter plane and are robust with respect to the concrete choice of the binning.
- While multivariate analysis techniques will consistently outperform 2-dimensional variable combinations, the gain in sensitivity going from the current STXS binning to a 2-dimensional one is higher than from a 2-dimensional one to an analysis containing all kinematic variables with XGBoost (at  $g_t = 1$ ). On the other hand, the gain from simply adding bins in the 1-dimensional distribution is much smaller.
- Many non-Higgs background events are expected in the regions of the  $|\Delta\phi_{t\bar{t}}|$  and  $b_1$  that are projected to be the most important to differentiate between SM and BSM  $t\bar{t}H$  events. This makes them not suitable for an STXS extension focusing on  $\mathcal{CP}$ .

---

<sup>1</sup> The projection to  $\mathcal{L} = 3000 \text{ fb}^{-1}$  only scales up the expected number of events but does not take into account improvements from a higher detector resolution or better acceptance factors.

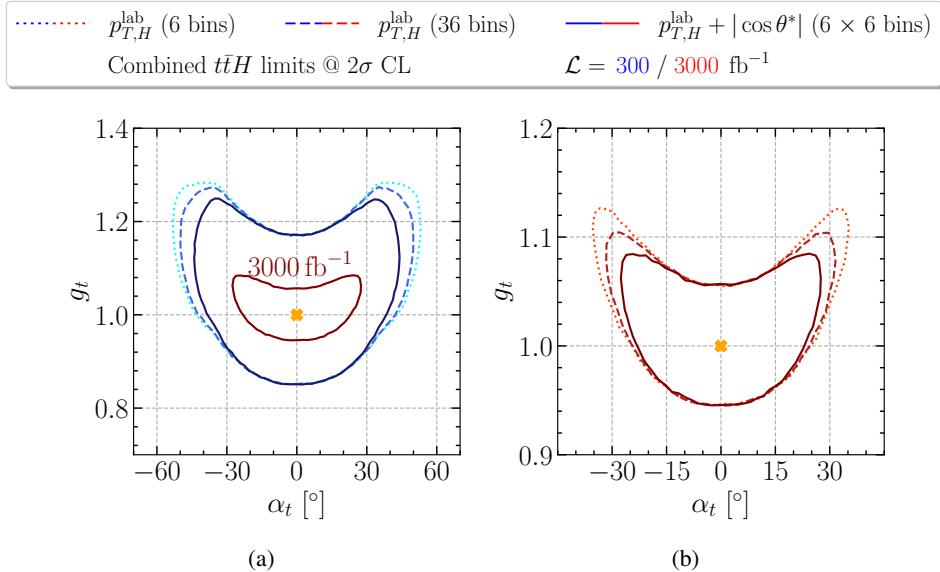


Figure 6.14: Projected  $2\sigma$  exclusion limits for the combination of channels at (a)  $\mathcal{L} = 300 \text{ fb}^{-1}$  and (b)  $\mathcal{L} = 3000 \text{ fb}^{-1}$ . Dotted and dashed contours correspond to a binning in  $p_{T,H}$  with 6 and 36 bins, respectively, while the solid contours show results from a 2-dimensional ( $p_{T,H}, |\cos \theta^*|$ ) binning. Figure modified from [313].

- The final selection of variables includes  $b_2$  in the lab frame, and  $|\Delta\eta_{t\bar{t}}|$  and  $|\cos \theta^*|$  in the  $t\bar{t}$  frame. There is no clear preference for any of these variables based on the findings of this study.

The proposed extensions of the STXS framework for  $t\bar{t}H$  are depicted in Fig. 6.15. Each  $p_{T,H}$  bin is split into six additional bins in one of the three candidate variables. The bin borders are determined by the optimized binning detailed in Section 6.3.2.

The three candidate observables all depend on kinematic information about the top quarks. To ensure that different decay channels and results across experiments can be combined, a common definition of the top quark is needed. While a pseudo-top quark definition exists for the leptonic and semi-leptonic  $t\bar{t}$  channel [323], it has not been studied whether a similar definition can be applied to  $t\bar{t}H$ . Such a study is left for future work.

## 6.5 Conclusions

In this chapter, the  $\mathcal{CP}$  character of the top-Yukawa coupling was studied via the  $t\bar{t}H$  process to propose an extension of the current STXS framework. Currently, the STXS binning in  $t\bar{t}H$  consists of six bins in  $p_{T,H}$ , which has been shown not to be optimal for future analyses with  $\mathcal{L} = 300 \text{ fb}^{-1}$ . An extension of the STXS framework must ensure backward compatibility and allow combinations of different channels and across experiments. Therefore, while machine learning approaches yield the most sensitivity towards the  $\mathcal{CP}$  nature of the top-Yukawa coupling, extensions to a 2-dimensional STXS binning were considered.

30 different ( $\mathcal{CP}$ -even) variables which have been shown in the literature to be  $\mathcal{CP}$ -sensitive to the top-Yukawa coupling were considered over four different rest frames. To evaluate their sensitivity, three decay channels of the  $t\bar{t}H$  system were considered: the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel, the  $t\bar{t}H(\text{multilep.})$  channel,

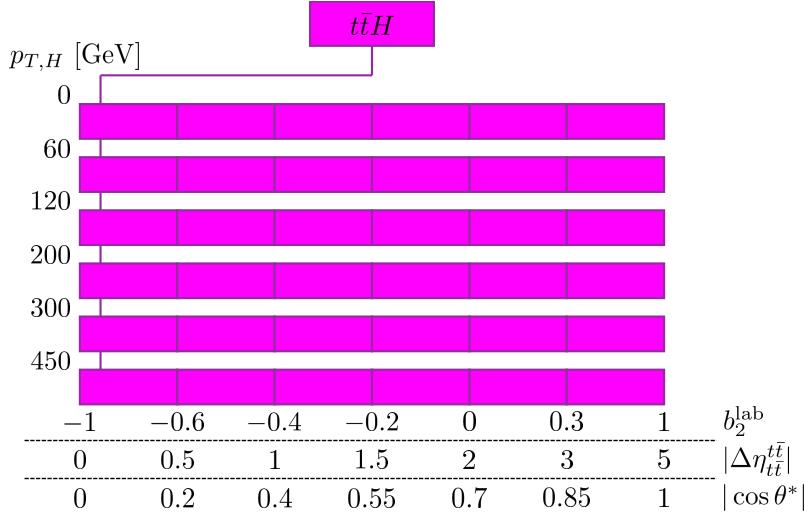


Figure 6.15: Three proposals for an extension of the  $t\bar{t}H$  STXS binning.  $6 \times 6$  bins are suggested, combining  $p_{T,H}$  with either  $b_2$ ,  $|\Delta\eta_{t\bar{t}}|$ , or  $|\cos\theta^*|$ . Figure modified from [313].

and the  $t\bar{t}H(\rightarrow b\bar{b})$  channel. For each of them, realistic yields were obtained by investigating the efficiency of the reconstruction and the acceptance of the event selection in former experimental analyses. Additionally, channel-dependent resolution effects were taken into account via Gaussian smearing of the kinematic variables.

It was shown that 2-dimensional combinations involving  $p_{T,H}$  can reach close to the highest significance among all possible combinations of variables. Consequently, five variables were selected as candidates for an STXS extension, and their binning was optimized so that each bin contains a minimum number of events. The distribution of the non-Higgs background in each variable revealed that  $|\Delta\phi_{t\bar{t}}|$  and  $b_1$  are negatively impacted by the background shape. They were therefore discarded from the list of candidates.

The final selection of variables to be paired with  $p_{T,H}$  was  $b_2$ ,  $|\Delta\eta_{t\bar{t}}|$ , and  $|\cos\theta^*|$ . The results obtained showed no clear preference for any of the three. A comparison with a multivariate BDT approach, trained with `XGBoost`, revealed that at  $g_t = 1$ , the improvement in sensitivity on  $\alpha_t$  from the 1-dimensional to the 2-dimensional STXS binning is higher than from the 2-dimensional STXS binning to the BDT. On the other hand, more bins in  $p_{T,H}$  corresponding to the number of bins in the 2-dimensional binning showed less strong improvements. It can therefore be concluded that, once a common top quark definition for  $t\bar{t}H$  has been agreed on, an extension of the STXS framework in a second dimension by  $b_2$ ,  $|\Delta\eta_{t\bar{t}}|$ , or  $|\cos\theta^*|$  is optimal for the  $\mathcal{CP}$  sensitivity. This holds for analyses after the end of Run 3 and is expected to hold also for the HL-LHC.

## Symbolic regression for Higgs $\mathcal{CP}$ analyses

---

*This chapter is based on the work in [230] which was done in collaboration with Henning Bahl, Elina Fuchs, and Tilman Plehn. All results taken from [230] have been produced by the author of this thesis unless stated otherwise.*

The continuously growing luminosity of the data collected at the LHC allows testing the SM with increasing precision, but also requires advanced techniques for the data analysis. Especially, the algorithms used need to sample large amounts of data efficiently, but are also required to identify rare processes from a few events against the statistically dominant background. A differential measurement, as required for direct tests of  $\mathcal{CP}$  violation via a  $\mathcal{CP}$ -odd variable, is even more challenging. It is necessary to achieve a good signal-to-background separation, as well as optimal sensitivity to the  $\mathcal{CP}$  state of the coupling under examination. For this, ML algorithms have been used with great success, as shown in Chapter 5 and many other studies (see e.g. [179, 258, 292, 294, 324]).

While ML techniques are powerful tools for experimental analyses, it has also been established in Chapter 6 that their usage can pose challenges, for example, when combining measurements across experiments. In this case, binning the results in easily accessible observables is helpful, as done in the STXS framework. Ideally, experiments should be able to combine the best of both worlds: obtain optimal results by leveraging the full kinematic information available, and achieve high interpretability, such that the results can be reused in other analyses.

Some interpretability of ML algorithms can be achieved using SHAP values, as introduced in Section 4.4.1 and demonstrated in Section 5.3.4. However, this only allows for ordering the input variables by their relative importance. Instead, a much better interpretability can be achieved by using SR, as discussed in Section 4.4.2. SR approximates the result of an ML model in a single equation, which can be used like any other observable. This is especially important in the context of  $\mathcal{CP}$  violation, where only dedicated  $\mathcal{CP}$ -odd observables provide an unambiguous test for it. Whether the obtained result is fully  $\mathcal{CP}$ -odd or not can be directly read from the formula. Apart from interpretability, SR also enables a fast implementation of results into experimental analyses.

Most common SR approaches can be split into two types of algorithms: genetic programs and NN-based approaches. Two examples for this that will be used in this chapter are PySR [229] and SymbolNet [231], which have been introduced in Sections 4.4.3 and 4.4.4. Other examples of genetic programming can be found in [325–328], while DNN-like implementations of SR are found in [329–332].

PySR has been shown to outperform many other algorithms for SR [229], while `SymbolNet` was designed to scale well in high-dimensional parameter spaces [231], motivating their use in this study.

The focus of this chapter is to learn observables that are optimal for measuring  $\mathcal{CP}$  violation at the detector level. Therefore, Section 7.1 first discusses the definition of optimal observables in the context of an ML approach. This definition is then used in Sections 7.2 and 7.3 to learn  $\mathcal{CP}$ -odd observables for VBF and  $t\bar{t}H$  production. Since analyses in  $t\bar{t}H$  are usually performed using  $\mathcal{CP}$ -sensitive, but  $\mathcal{CP}$ -even observables, a reconstruction of the Collins-Soper angle using SR is performed in Section 7.4. The findings are concluded in Section 7.5.

## 7.1 Optimal $\mathcal{CP}$ -odd observables

The only strict proof of  $\mathcal{CP}$  violation at the LHC is to measure an asymmetry in a dedicated  $\mathcal{CP}$ -odd observable. One possibility is to define

$$\epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu p_3^\rho p_4^\sigma \quad (7.1)$$

where the  $p_i$  can be momenta or polarization vectors.<sup>1</sup> This variable can be defined for every process where at least four independent 4-vectors are available. Sensitivity to the  $\mathcal{CP}$ -odd part of the  $HVV$  coupling is evident from the term proportional to  $a_3$  in Eq. (2.64).

In VBF, spin information is not accessible, and therefore, the momenta of the two partons and two tagging jets are used. In  $t\bar{t}H$ , spin information is preserved due to the short lifetime of the top quark decay. Therefore, the two reconstructed top quarks, as well as the momenta of their decay products, can be used. In all cases, Eq. (7.1) can be reduced to a triple product (TP) by choosing a fitting rest frame. Taking the momenta in the lab frame for VBF production, Eq. (7.1) reduces to the commonly used  $\Delta\phi_{jj}$  (see Chapter 5 and [179, 262, 333]). For  $t\bar{t}H$ , a similar observable obtained in the  $t\bar{t}$  frame is

$$\Delta\phi_{ll}^{t\bar{t}} = \text{sgn} [\vec{p}_t (\vec{p}_{l^+} \times \vec{p}_{l^-})] \arccos \left[ \frac{\vec{p}_t \times \vec{p}_{l^+}}{|\vec{p}_t \times \vec{p}_{l^+}|} \cdot \frac{\vec{p}_t \times \vec{p}_{l^-}}{|\vec{p}_t \times \vec{p}_{l^-}|} \right]. \quad (7.2)$$

This requires a fully leptonic decay of the  $t\bar{t}$  system, which is often chosen, since leptons are easy to tag and have a spin analyzing power of 1 [334, 335]. However, many more TPs can be defined that are often used in analyses targeting the interference term of  $t\bar{t}H$  production [185–189, 193, 247, 252, 257].

While these observables are  $\mathcal{CP}$ -odd by definition, they are in general not the *optimal observables* (*OO*). These can be defined at the parton level via the matrix elements. Starting with either Eq. (2.59) or Eq. (2.66), events can be sampled from

$$p(x|\theta) = \frac{1}{\sigma(\theta)} \frac{d^d \sigma(x|\theta)}{dx^d} = p_e(x|\theta) + p_o(x|\theta), \quad (7.3)$$

which can be split into a  $\mathcal{CP}$ -even part  $p_e(x|\theta)$  and a  $\mathcal{CP}$ -odd part  $p_o(x|\theta)$ .  $x$  is the phase space point, and  $\theta$  includes all parameters that affect the  $\mathcal{CP}$  nature of the interaction. The Neyman-Pearson lemma

<sup>1</sup> These observables are often defined with an additional sign factor to define a fixed direction for the momenta.

[336] then yields [337]

$$\omega_{\mathcal{CP}\text{-odd}} = \frac{2 \operatorname{Re} (\mathcal{M}_{\mathcal{CP}\text{-even}} \mathcal{M}_{\mathcal{CP}\text{-odd}}^*)}{|\mathcal{M}_{\mathcal{CP}\text{-even}}|^2 + |\mathcal{M}_{\mathcal{CP}\text{-odd}}|^2} = \frac{p_o}{p_e} \quad (7.4)$$

for the  $OO$ . However, when the effects of parton showering, hadronization, and the limited detector resolution are taken into account, the  $OO$  cannot be written down analytically. Instead, it can be learned by ML techniques as follows. If a classifier is trained on two samples  $p_e + p_o$  and  $p_e - p_o$ , it will approach (again via the Neyman-Pearson lemma)

$$P(x) = \frac{p_e(x|\theta) + p_o(x|\theta)}{p_e(x|\theta) + p_o(x|\theta) + p_e(x|\theta) - p_o(x|\theta)} = \frac{1 + \omega_{\mathcal{CP}\text{-odd}}}{2} \quad . \quad (7.5)$$

Therefore, the  $OO$  is easily obtained from the classifier score

$$\omega_{\mathcal{CP}\text{-odd}} = 2P(x) - 1 \quad . \quad (7.6)$$

Using SR offers the possibility of obtaining an analytic equation for  $\omega_{\mathcal{CP}\text{-odd}}$  and confirming that it is  $\mathcal{CP}$ -odd.

The differential cross section for a  $\mathcal{CP}$ -odd observable is antisymmetric in the presence of  $\mathcal{CP}$  violation, while a  $\mathcal{CP}$ -even observable is symmetric by definition. This can be leveraged to define the asymmetry

$$\mathcal{A}_i = \frac{N_i^+ - N_i^-}{N_i^+ + N_i^-} \quad (7.7)$$

where  $N_i^+$  ( $N_i^-$ ) are the number of events in bin  $i$  for positive (negative) values of  $\omega_{\mathcal{CP}\text{-odd}}$ . While not exploiting rate information, the asymmetry includes a statistical uncertainty. Therefore, a higher luminosity increases the sensitivity of  $\mathcal{A}$ . For the results presented here,  $\mathcal{L} = 300 \text{ fb}^{-1}$  is used, corresponding to the end of Run 3. Details about the asymmetry can be found in Appendix C.

## 7.2 $\mathcal{CP}$ -odd observables in VBF

The first goal is to find the  $OO$  for the VBF production process. As discussed in Section 5.1,  $\mathcal{CP}$  violation in this channel can be introduced by the three Wilson coefficients  $c_{\Phi\widetilde{W}}$ ,  $c_{\Phi\widetilde{B}}$ , and  $c_{\Phi\widetilde{WB}}$ . It has also been justified that using a non-zero value for  $c_{\Phi\widetilde{W}}$  is sufficient to cover the relevant effects [280, 281]. A simple analytic expression close to the optimal result near the SM  $c_{\Phi\widetilde{W}} \approx 0$  is known from previous studies [179, 222, 262, 338]

$$\frac{1}{m_H^2} p_{T,j_1} p_{T,j_2} \sin \Delta\phi_{jj} \quad . \quad (7.8)$$

### 7.2.1 Event generation and training

Datasets are generated at the detector level using `MadGraph5_aMC@NLO` [143], `Pythia8` [153], and `Delphes3` [155]. In total, 11 datasets are generated using

$$c_{\Phi\widetilde{W}} \in \{0, \pm 0.1, \pm 0.25, \pm 0.5, \pm 0.75, \pm 1\} \quad (7.9)$$

with  $c_{\Phi\widetilde{W}} = 0$  corresponding to the SM. The  $H \rightarrow \gamma\gamma$  decay is considered here, which is also influenced by the choice of  $c_{\Phi\widetilde{W}}$ . The Higgs boson is reconstructed from the diphoton system after applying the cuts

- $m_{\gamma\gamma} \in [110 \text{ GeV}, 140 \text{ GeV}]$ ,
- $p_{T,\gamma_1}/m_{\gamma\gamma} > 0.35$ , and  $p_{T,\gamma_2}/m_{\gamma\gamma} > 0.25$ .

Jets are clustered using `FastJet` [157] and are preselected using

- $p_{T,j} > 30 \text{ GeV}$ ,
- $|\eta_j| < 4.4$ , and
- $\Delta\eta_{jj} > 2$ .

For the training of the PySR and `SymbolNet`, the following input variables are used:

- $x_{j_{1,2}} = \frac{p_{T,j_{1,2}}}{m_h}$ ,  $\eta_{j_1}$ , and  $\phi_{j_1}$  for variables depending on a single jet;
- $x_H = \frac{p_{T,H}}{m_H}$ ,  $\eta_H$ , and  $\phi_H$  for variables depending on the Higgs;
- $\Delta\eta_{jj}$ ,  $\Delta\phi_{jj}$ , and  $x_{jj} = \frac{m_{jj}}{m_H}$  for higher-level variables of the dijet system.

The momenta and masses are normalized via the Higgs boson mass  $m_H = 125 \text{ GeV}$  to obtain dimensionless variables of order 1 and improve the training process. Only  $\eta_{j_1}$  and  $\phi_{j_1}$  are used because  $\eta_{j_2}$  and  $\phi_{j_2}$  are redundant when using  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$ .

PySR is given the set of functions

$$\mathcal{S} = \underbrace{\{\sin, \cos, \text{abs}, \exp, \log, \text{sinh}, \cosh, +, *, /\}}_{\text{unary}} \underbrace{,}_{\text{binary}} \quad (7.10)$$

to build equations.  $2.5 \cdot 10^5$  events are used for training for each value of  $c_{\Phi\widetilde{W}}$ , and  $1 \cdot 10^5$  events are used for validation. The number of iterations is set to 500, and the maximum complexity of the equations is set to 60. Simulated annealing is used with the modified acceptance probability defined in Eq. (4.34) with  $\alpha = 100$ . Otherwise, the recommended settings for the hyperparameters of PySR are used. For the objective function, a cross-entropy loss is used, but with an additional check for the  $\mathcal{CP}$ -state of the current equation. If an equation is proposed that is not  $\mathcal{CP}$ -odd, it will be discarded during the training.

`SymbolNet` is trained in the default scalar setup because there are only 10 input variables. Therefore, the model is only constructed from  $S \rightarrow S$  layers. Two layers are used with the function sets

$$\begin{aligned} \mathcal{S}_1 &= \{\sin, \cos, \text{abs}, \sqrt[2]{\phantom{x}}, \sqrt{\phantom{x}}, *, /\} \\ \mathcal{S}_2 &= \underbrace{\{\sin, \cos, \text{abs}, \sqrt[2]{\phantom{x}}, \sqrt{\phantom{x}}, *\}}_{\text{unary}} \underbrace{,}_{\text{binary}} \end{aligned} \quad (7.11)$$

The training is performed over 2000 iterations in the mixed setup with a learning rate of  $5 \cdot 10^{-4}$  and a batch size of 512. The target sparsities for the respective terms in Eq. (4.41) are set to  $\alpha_{\text{in}} = 0.6$ ,  $\alpha_W = 0.6$ , and  $\alpha_{\text{unary}} = \alpha_{\text{binary}} = 0.3$ . The cross-entropy loss is used again, but this time without any additional checks for the  $\mathcal{CP}$ -state of the output. While this can be implemented, it complicates the training process, since two separate contributions to the loss must be minimized. Still, `SymbolNet` is able to learn  $\mathcal{CP}$ -odd observables, as shown below.

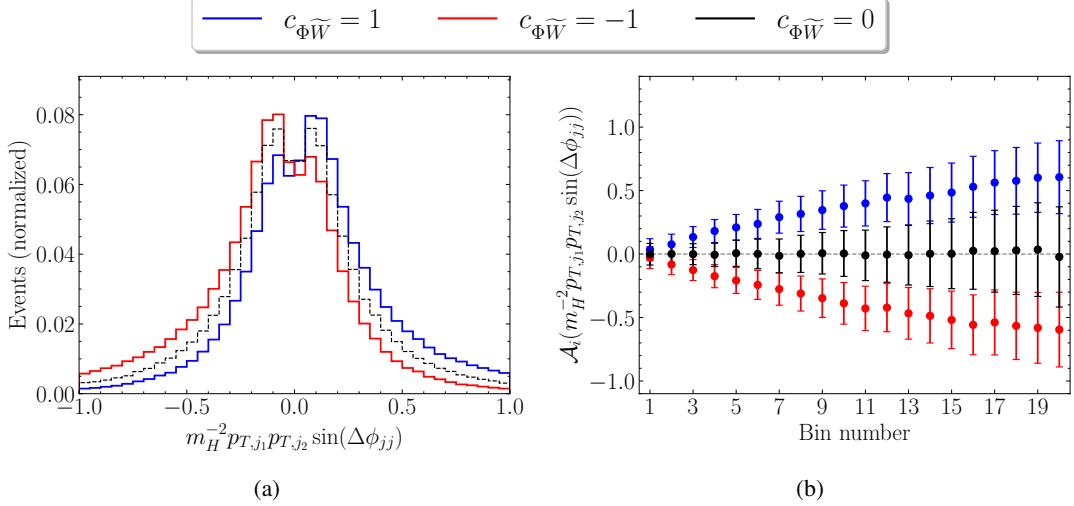


Figure 7.1: The (a) distributions and (b) asymmetries of the variable  $m_H^{-2} p_{T,j_1} p_{T,j_2} \sin(\Delta\phi_{jj})$  for  $c_{\Phi\widetilde{W}} \in \{-1, 0, +1\}$ . Figure modified from [230], original figure created by Henning Bahl.

### 7.2.2 Learned equations and sensitivities

The variable defined in Eq. (7.8) will be used as a benchmark to assess the performance of the classifiers. The distribution of this variable for  $c_{\Phi\widetilde{W}} \in \{-1, 0, +1\}$  and the corresponding asymmetry is shown in Fig. 7.1. The plots show the expected behavior of a  $\mathcal{CP}$ -odd observable. The SM distribution is symmetric around zero, while the asymmetry is positive (negative) for a positive (negative) interference. The asymmetry between the two  $\mathcal{CP}$ -violating hypotheses increases towards the outer bins, where the number of events becomes lower. Consequently, the statistical uncertainty of the asymmetry grows, as visible in Fig. 7.1(b).

Next, Fig. 7.2 shows the distributions obtained from the formula found by PySR. In contrast to the distribution in Fig. 7.1(a), the events for each value of  $c_{\Phi\widetilde{W}}$  peak in the bins closest to zero. At the same time, the number of events drops off less strongly towards higher absolute values of  $\omega_{\mathcal{CP}\text{-odd}}$ . The result is a higher asymmetry in the outer bins compared to the variable  $m_H^{-2} p_{T,j_1} p_{T,j_2} \sin(\Delta\phi_{jj})$ . Still, for the outermost bins, the uncertainty of the SM distribution increases. The bins with the most separation power are therefore expected to be those slightly before the end of the tails.

The equation found by PySR is [230]

$$P^{\text{PySR}} = \frac{1.857 \sin \Delta\phi_{jj}}{\left| \frac{0.308 x_{j_1} \log \Delta\eta_{jj} + \log \Delta\eta_{jj} \sinh(x_{j_2} - 2.598) + 0.308 \sinh x_H}{x_{j_1} \log \Delta\eta_{jj} + \sinh x_H} \right| + 0.605} \quad (7.12)$$

where the  $\mathcal{CP}$ -odd observable  $\sin \Delta\phi_{jj}$  is colored in blue. As can be seen, it is the only observable in the numerator and it is normalized by a  $\mathcal{CP}$ -even term, validating that the learned equation is  $\mathcal{CP}$ -odd.

The results obtained from SymbolNet can be found in Fig. 7.3. They are very similar to the results from PySR, with the only visible difference being a slightly worse performance of the outermost bins, as

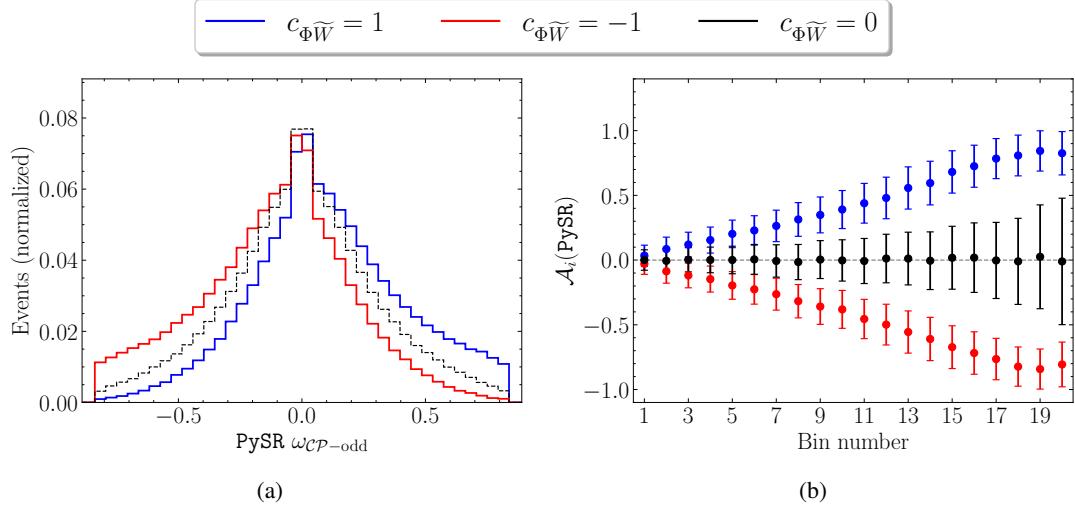


Figure 7.2: Same as Fig. 7.1 but for the formula learned by PySR. Figure modified from [230], original figure created by Henning Bahl.

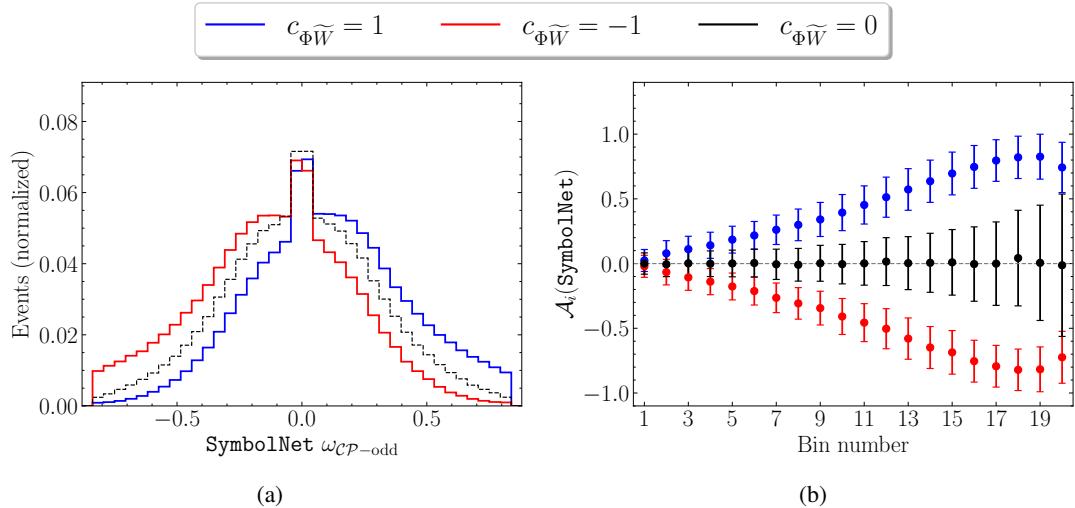


Figure 7.3: Same as Fig. 7.1 but for the formula learned by SymbolNet. Figure modified from [230], original figure created by Henning Bahl.

Classifier	Trained on	$\sigma(c_{\Phi\widetilde{W}} = 1 \text{ vs. SM})$	$\sigma(c_{\Phi\widetilde{W}} = 0.25 \text{ vs. SM})$
$m_H^{-2} p_{T,j_1} p_{T,j_2} \sin \Delta\phi_{jj}$	–	6.76	2.43
PySR		6.98	2.47
SymbolNet	$c_{\Phi\widetilde{W}} = \pm 1$	7.07	2.49
BDT		6.71	2.36
PySR		7.07	2.43
SymbolNet	$c_{\Phi\widetilde{W}} = \pm 0.1$	1.67	0.82
BDT		3.27	1.26

Table 7.1:  $\sigma$  levels for excluding the BSM hypotheses of  $c_{\Phi\widetilde{W}} = 1$  and  $c_{\Phi\widetilde{W}} = 0.25$  based on the classifiers and the data they have been trained on. Table taken from [230], original table created by Henning Bahl.

the error bars of the asymmetry overlap. The corresponding formula is [230]

$$\begin{aligned}
 P^{\text{SymbolNet}} = 0.715 \Delta\phi_{jj} & \left[ -0.348(0.561x_{j_2} - 0.25\Delta\eta_{jj} + 0.032x_{jj} + 0.746x_H) - 0.27 \right] \\
 & \cdot \left[ 0.049(0.603\Delta\eta_{jj} - 0.081x_{jj} - x_H)^2 \right. \\
 & \quad - 0.654|0.463x_{j_2} + 0.477\Delta\eta_{jj} + 0.373x_H|^{0.5} \\
 & \quad - 0.134 \sin(-0.555\Delta\eta_{jj} + 0.345x_{jj} + 0.443x_H) \\
 & \quad \left. - 1.82 \cos(0.642\Delta\phi_{jj}) \right], \tag{7.13}
 \end{aligned}$$

where, again,  $\Delta\phi_{jj}$  is colored in blue. Despite being slightly more complex, the SymbolNet equation shows the same behavior as the PySR equation, i.e., a simple  $\mathcal{CP}$ -odd observable is multiplied by a purely  $\mathcal{CP}$ -even term. As mentioned above, SymbolNet does not contain a contribution to the loss term that guarantees the output to be  $\mathcal{CP}$ -odd. In some training runs, this led to observables that provided good separation between  $c_{\Phi\widetilde{W}} = 1$  and  $c_{\Phi\widetilde{W}} = -1$ , but were not purely  $\mathcal{CP}$ -odd, emphasizing the importance of interpretability.

An interesting question is how the classifiers, and consequently the significances of differentiating between positive and negative interference, behave when dealing with data where  $|c_{\Phi\widetilde{W}}| < 1$ . The significances are obtained from the asymmetries, and Table 7.1 shows the  $\sigma$  levels obtained for the  $m_H^{-2} p_{T,j_1} p_{T,j_2} \sin \Delta\phi_{jj}$  observable, PySR, and SymbolNet. Additionally, as a comparison, a BDT is trained with XGBoost [207] and using the default cross-entropy loss. The results are shown for two BSM hypotheses of  $c_{\Phi\widetilde{W}} = 1$  and  $c_{\Phi\widetilde{W}} = 0.25$ . Each of the three classifiers is trained with two different data samples. One with  $c_{\Phi\widetilde{W}} = \pm 1$  and the other with  $c_{\Phi\widetilde{W}} = \pm 0.1$ . As can be seen, for the classifiers trained on the  $c_{\Phi\widetilde{W}} = \pm 1$  samples, PySR and SymbolNet perform slightly better than the simple observable and the BDT. The hypothesis of  $c_{\Phi\widetilde{W}} = 1$  ( $c_{\Phi\widetilde{W}} = 0.25$ ) can be excluded at the  $7\sigma$  ( $2.5\sigma$ ) level. Training on samples with  $c_{\Phi\widetilde{W}} = \pm 0.1$  is a lot harder because the difference in the distributions of positive

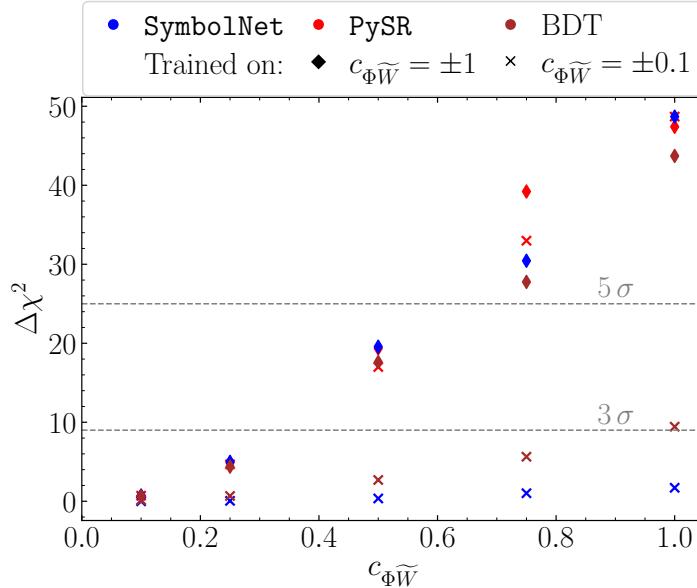


Figure 7.4: Comparison of the significances obtained from `SymbolNet`, `PySR`, and a BDT when being trained on a  $|c_{\Phi\widetilde{W}}| = \pm 1$  and a  $|c_{\Phi\widetilde{W}}| = \pm 0.1$  sample, and evaluated for various values of  $c_{\Phi\widetilde{W}}$ . Figure modified from [230], original figure created by Henning Bahl.

and negative interference events is much smaller. Consequently, `SymbolNet` and the BDT drastically drop in performance and yield much lower significances. In contrast, `PySR` manages to reach the same significance as before, indicating the robustness of its training.

These results are also visualized in Fig. 7.4, where the  $\Delta\chi^2$  values of `PySR`, `SymbolNet`, and the BDT are shown for  $c_{\Phi\widetilde{W}} \in \{0.1, 0.25, 0.5, 0.75, 1\}$ , again separated into training samples with  $c_{\Phi\widetilde{W}} = \pm 1$  (diamonds) and  $c_{\Phi\widetilde{W}} = \pm 0.1$  (crosses). When evaluating the classifiers on  $c_{\Phi\widetilde{W}} > 0.25$ , a clear separation between `PySR` and the others can be seen when training with  $c_{\Phi\widetilde{W}} = \pm 0.1$ . For training on  $c_{\Phi\widetilde{W}} = \pm 1$  data, the BDT shows an overall slightly worse performance than `SymbolNet` and `PySR`. At  $c_{\Phi\widetilde{W}} = 0.75$ , an outlier can be seen where `PySR` achieved an exceptionally good result.

On top of the good performance of `PySR` for small differences in the training data, it also performs well when training with a small number of events, as demonstrated in Fig. 7.5. Here, `PySR` needs about  $10^3$  to reach good results and only about  $10^4$  events to achieve its maximal separation power. This is most likely a consequence of the "bottom-up" approach from `PySR`, where it starts with a simple equation and then makes it more complex. In contrast, the BDT struggles with lower amounts of training data. For `SymbolNet`, including an additional loss term to guarantee a  $\mathcal{CP}$ -odd output makes a large difference. When this loss term is not included, `SymbolNet` achieves similar separation to `PySR` when using  $10^4$  events. If it is included, more events are needed in the training, as the objective is more complex to minimize.

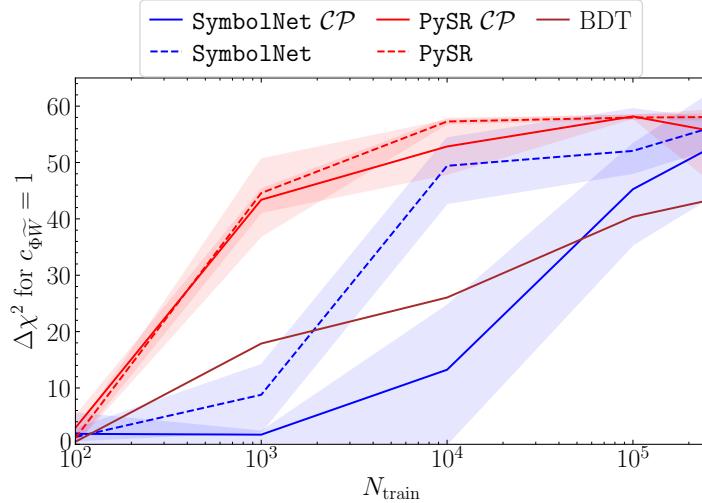


Figure 7.5: Comparison of the significances obtained from PySR, SymbolNet, and a BDT when using different amounts of training data. Additionally, the performance of PySR and SymbolNet is compared when using an additional loss that forces the output to be  $\mathcal{CP}$ -odd (indicated by  $\mathcal{CP}$ ). The curves show the average of four training runs, and the shaded regions indicate the best and worst performance at each  $N_{\text{train}}$ . Figure modified from [230], original figure created by Henning Bahl.

### 7.3 $\mathcal{CP}$ -odd observables in $t\bar{t}H$

An optimal  $\mathcal{CP}$ -odd observable can also be defined for the  $t\bar{t}H$  process, although  $\mathcal{CP}$ -even observables are usually used for analyses at the LHC, as discussed in Chapter 6. This is due to the suppression of the interference term in  $t\bar{t}H$  [190, 255]. Still, it should be assessed how sensitive information from a  $\mathcal{CP}$ -odd observable can be. While Eq. (7.6) still holds, many potential TPs can be built out of the final state particles, in contrast to the single TP  $\Delta\phi_{jj}$  in VBF.

#### 7.3.1 Event generation and training

Events are generated analogously to Section 6.1.1 for three values of the  $\mathcal{CP}$ -mixing angle  $\alpha_t \in \{-45^\circ, 0^\circ, 45^\circ\}$ . The top quarks decay via the di-leptonic channel, simulated with `MadSpin` [150]. For the Higgs decay, the  $H \rightarrow \gamma\gamma$  channel is assumed and the corresponding smearing and efficiency factors from Table 6.2 are applied. For each  $\mathcal{CP}$  angle,  $2 \cdot 10^5$  events are generated, of which half are used for training. The rest are equally split into a validation and a test dataset.

To build TPs, the version of `SymbolNet` modified to work with 4-vectors is used. However, it needs to be modified further, as Eq. (7.1) is a function of four 4-vectors, and so far, only unary and binary operators are defined. Consequently, the  $\epsilon$ -tensor is implemented as a special function usable in the  $V \rightarrow S$  layer. The setup then corresponds to the one shown in Fig. 7.6. Two instances of `SymbolNet` are created. One contains only  $\mathcal{CP}$ -even observables and therefore produces a  $\mathcal{CP}$ -even observable  $P_{\text{even}}(x)$ . The other one contains a single  $\epsilon$ -tensor in its  $V \rightarrow S$  layer and only odd functions in subsequent layers, so that the score  $P_{\text{odd}}(x)$  is  $\mathcal{CP}$ -odd by definition. The outputs are then multiplied so that  $P(x) = P_{\text{even}}(x) \cdot P_{\text{odd}}(x)$  is again  $\mathcal{CP}$ -odd. This avoids the need to define a (harder to train)  $\mathcal{CP}$ -odd loss while also giving

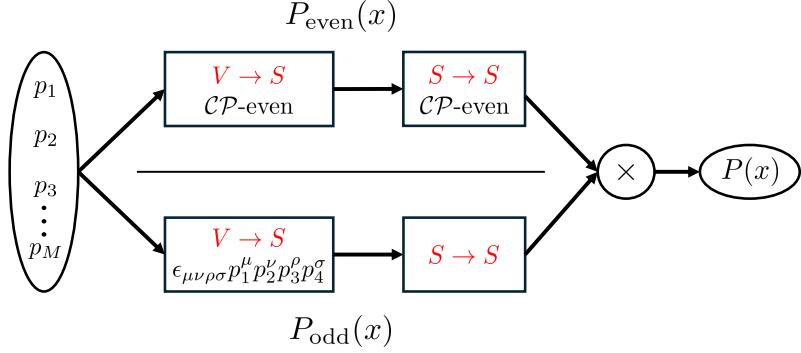


Figure 7.6: Schematic of **SymbolNet** as used in the training for finding a  $\mathcal{CP}$ -odd observable for  $t\bar{t}H$ . The network is split into two independent parts, a  $\mathcal{CP}$ -even and a  $\mathcal{CP}$ -odd part, which are multiplied in a final layer to yield a  $\mathcal{CP}$ -odd variable.

**SymbolNet** the possibility to define TPs itself.

The input variables for **SymbolNet** are the 4-momenta of the Higgs boson, the bottom quarks, and the leptons. **SymbolNet** is trained for 1000 epochs with a learning rate of  $5 \cdot 10^{-3}$  in the default setup, and subsequently for 400 and 50 epochs with a learning rate of  $5 \cdot 10^{-4}$  in the mixed and sparsity setup. A batch size of 256 is used, and the objective function is the cross-entropy loss. The target sparsities are  $\alpha_W = 0.9$  and  $\alpha_{\text{unary}} = \alpha_{\text{binary}} = 0.5$ . No target sparsity is defined for the input variables and the  $\epsilon$ -tensor. The function sets are

$$\begin{aligned}
 \mathcal{S}_{V \rightarrow S}^{\text{even}} &= \{p_0, p_x, p_y, p_z, \|p\|_3, \langle p_i \times p_j \rangle_3\} \\
 \mathcal{S}_{S \rightarrow S}^{\text{even}} &= \{\cos, ^2, \sqrt, *, /\} \\
 \mathcal{S}_{V \rightarrow S}^{\text{odd}} &= \{\epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu p_3^\rho p_4^\sigma\} \\
 \mathcal{S}_{S \rightarrow S}^{\text{odd}} &= \{\sin, \tanh, ^3, -\} \quad .
 \end{aligned} \tag{7.14}$$

PySR is not trained here, as the implementation of 4-vectors and the  $\epsilon$ -tensor into PySR is beyond the scope of this work.

### 7.3.2 Learned equations and sensitivities

First, two benchmarks are defined in terms of simple TPs. At the parton level, the TP yielding the highest sensitivity is

$$\epsilon_{\text{parton}} = \epsilon_{\mu\nu\rho\sigma} (p_t + p_{\bar{t}})^\mu (p_t - p_{\bar{t}})^\nu (p_t + p_{\bar{t}})^\rho (p_t - p_{\bar{t}})^\sigma \quad . \tag{7.15}$$

As previously discussed, this reduces to Eq. (7.2) in the  $t\bar{t}$  frame. A similar TP

$$\epsilon_{\text{reco}} = \epsilon_{\mu\nu\rho\sigma} (p_b + p_{\bar{b}})^\mu (p_b - p_{\bar{b}})^\nu (p_b + p_{\bar{b}})^\rho (p_b - p_{\bar{b}})^\sigma \tag{7.16}$$

yields the highest sensitivity at the detector level, but with the bottom quarks replacing the top quarks.

Fig. 7.7 shows the distributions obtained from these TPs. As can be seen, the TPs do not achieve a

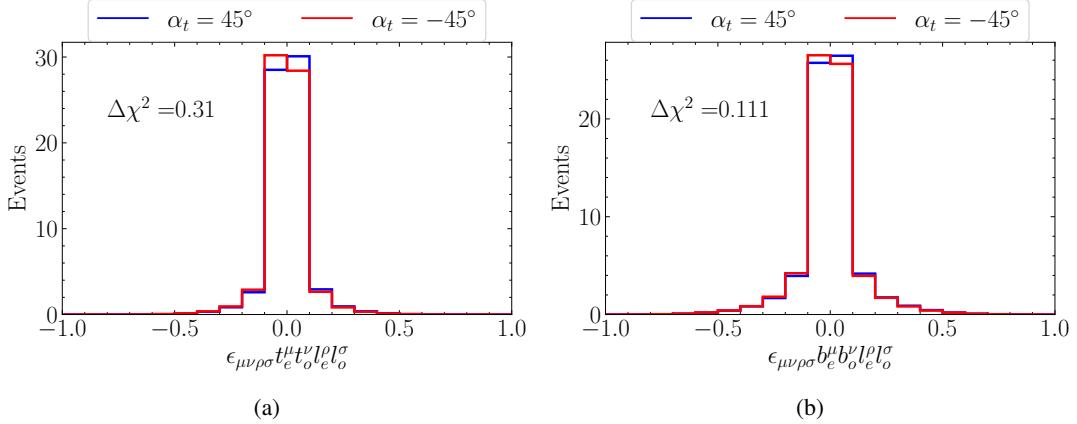


Figure 7.7: The TPs yielding the highest sensitivities for  $t\bar{t}H$  at (a) the parton level and (b) the detector level. Figure taken from [230].

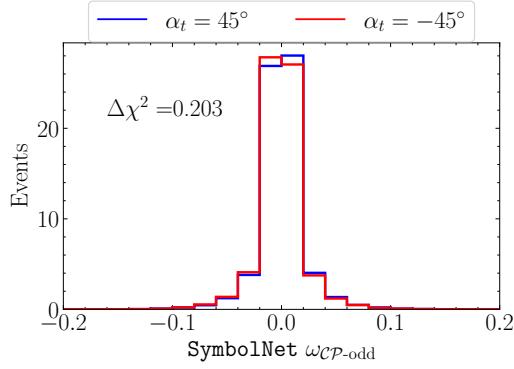


Figure 7.8: Distribution from the equation predicted by SymbolNet with the modified setup including a TP. Figure taken from [230].

high separation of the positive and negative interference terms. The innermost bins give the dominant contribution to the sensitivity. The TP defined at the detector level provides a much lower sensitivity, as expected. However, with  $\Delta\chi^2 = 0.31$ , the TP defined at the parton level also does not yield enough separation power for an experimental analysis.

The obtained distribution from SymbolNet can be found in Fig. 7.8. The corresponding  $\mathcal{CP}$ -odd and  $\mathcal{CP}$ -even parts of the equation are [230]

$$\begin{aligned}
 P_{\text{odd}}(x) = & \epsilon_{\mu\nu\rho\sigma} (0.32p_H + 0.622p_b - 0.164p_{\bar{b}} + 0.634p_l)^\mu \\
 & (0.298p_H + 0.196p_b + 0.54p_{\bar{b}} + 0.551p_{\bar{l}} - 0.216p_l)^\nu \\
 & (+0.259p_H + 0.448p_b - 0.319p_{\bar{b}} + 0.761p_l)^\rho \\
 & (0.176p_H - 0.051p_b + 0.488p_{\bar{b}} + 0.772p_{\bar{l}} - 0.462p_l)^\sigma
 \end{aligned} \tag{7.17}$$

and

$$\begin{aligned}
 P_{\text{even}}(x) = & \text{Abs} \left( (-0.123p_{y,H} + 0.139p_{y,b} + 0.159p_{y,\bar{b}} - 0.375p_{y,l} + 0.394p_{z,H} \right. \\
 & + 1.405p_{z,b} - 0.134p_{z,\bar{b}} - 0.627p_{z,\bar{l}} + 0.803p_{z,l} \\
 & + 0.232 \left\| -2.715p_b - 4.855p_{\bar{b}} - 2.711p_{\bar{l}} \right\|_3 \\
 & + 0.016 \left( (4.888p_H - 3.127p_b + 4.287p_{\bar{b}} - 1.211p_{\bar{l}} - 2.628p_l) \right. \\
 & \left. \times (-2.516p_H + 0.7615p_b - 3.161p_{\bar{b}} + 6.152p_{\bar{l}} + 0.9251p_l) \right) \Big)_3^{1/2}, \quad (7.18)
 \end{aligned}$$

which are multiplied to yield  $P(x)$ . This equation performs much better than the detector-level TP, reaching 65% of the  $\mathcal{CP}$  sensitivity at the parton level, compared to 36%. However, the sensitivities are still very low in comparison to  $\mathcal{CP}$ -even but  $\mathcal{CP}$ -sensitive observables. Furthermore, all detector-level TPs require differentiating between the bottom quark and its antiquark.

## 7.4 Reconstruction of the Collins-Soper angle for $t\bar{t}H$

The results above demonstrate that the sensitivity of  $\mathcal{CP}$ -odd observables in the  $t\bar{t}H$  process is too low to motivate an experimental study. However, SR can still be useful in this channel. A very powerful  $\mathcal{CP}$ -sensitive observable is the Collins-Soper angle  $\theta^*$  [316]. It is one of the candidates for the STXS extension proposed in Chapter 6 and has been shown in other studies to have good sensitivity to  $\mathcal{CP}$  [247, 255, 259]. Additionally, it does not require distinguishing between the bottom quark and its antiquark. As a reminder, it is defined in the  $t\bar{t}$  frame as

$$\cos \theta^* = \frac{\vec{p}_t \cdot \vec{n}}{\|\vec{p}_t\| \cdot \|\vec{n}\|} \quad (7.19)$$

where it forms the angle of the top quark axis with the beam axis. This requires a full reconstruction of both top quarks, which is a challenging task. Here, SR will be used to learn an approximate equation for  $\cos \theta^*$  from the kinematic information at the detector level.

### 7.4.1 Event generation and training

Generation of the parton-level events at LO proceeds in the same way as in the previous section. However, the top quarks decay semi-leptonically here. Furthermore, additional datasets are generated for the  $t\bar{t}Hj$  process and for  $\alpha_t \in \{10^\circ, 25^\circ, 90^\circ\}$ . The input variables for the training are

$$\mathbf{X}_{\text{in}} = \underbrace{\{p_H, p_b, p_{\bar{l}}, p_\nu \text{ or } p^{\text{miss}}, p_{\bar{b}}, p_q, p_{\bar{q}}\}}_{p_t} \cup \underbrace{\{p_{\bar{l}}\}}_{p_{\bar{t}}}, \quad (7.20)$$

where  $p^{\text{miss}}$  is a 4-vector containing the accessible information about the neutrino momentum. It is needed to train `SymbolNet` in the vectorized setup, while  $E_T^{\text{miss}}$  is used for `PySR`. All 4-momenta are

Scenario	Frame	$\nu$ Info	QCD Radiation	Smearing
1	$t\bar{t}$	Full	✗	✗
2	Lab	Full	✗	✗
3	Lab	$E_T^{\text{miss}}$	✗	✗
4	Lab	$E_T^{\text{miss}}$	✓	✗
5	Lab	$E_T^{\text{miss}}$	✗	✓
6	Lab	$E_T^{\text{miss}}$	✓	✓

Table 7.2: Definitions of the six benchmark scenarios for the classifier training. Table taken from [230].

Scenario	$N_{\text{epochs}}^{\text{default}}$	$N_{\text{epochs}}^{\text{mixed}}$	$N_{\text{epochs}}^{\text{sparsity}}$	$LR^{\text{default}}$	$LR^{\text{mixed}}$	$LR^{\text{sparsity}}$	$BS$
1	200	100	20	0.01	0.005	0.001	128
2	500	100	20	0.01	0.005	0.001	128
3 & 4	1000	400	50	0.01	0.002	0.001	64
5 & 6	2000	500	50	0.02	0.002	0.001	64

Table 7.3: The hyperparameters used for SymbolNet in each scenario. Table taken from [230].

normalized by the top quark mass  $m_t = 173$  GeV, although not explicitly stated in the equations. The bottom quarks are assumed to be correctly associated with the hadronic or the leptonic top quark decay.

The classifiers will be trained in six different scenarios, which are listed in Table 7.2. Going from scenario 1 to scenario 6, reconstructing  $\cos\theta^*$  becomes increasingly difficult. In scenarios 1 and 2, all information about the final-state particles is available without radiation or detector smearing. In scenario 1, observables are provided in the  $t\bar{t}$  frame, while in scenario 2, they are taken from the lab frame. Starting from scenario 3,  $E_T^{\text{miss}}$  is used instead of the full neutrino information. Scenarios 4 and 6 use the  $t\bar{t}Hj$  sample, where an additional jet is radiated. Finally, all momenta are smeared in scenarios 5 and 6 to mimic the limited resolution of the detector.

The hyperparameters used for the training depend on the scenario investigated. PySR uses a maximum size for the equation of 35 and trains for 200 iterations in scenario 1. For scenarios 2, 3, and 4, the maximum size and training iterations increase to 50 and 2000, respectively. For scenarios 5 and 6, they increase further to 60 and 3000, respectively. Like in Section 7.2, simulated annealing is used with

Scenario	Operators	$V \rightarrow V$ layer	$V \rightarrow S$ layer	$S \rightarrow S$ layer	$S \rightarrow S$ layer
1	$\mathcal{S}_f$	–	$\{p_z, \ p\ _3\}$	$\{id.\}$	–
	$\mathcal{S}_g$	–	$\{\langle p_i \times p_j \rangle_3\}$	$\{/ \}$	–
2	$\mathcal{S}_f$	$\{id.\}$	$\{p_z, \ p\ _3\}$	$\{id.\}$	–
	$\mathcal{S}_g$	$\{\text{boost}\}$	$\{\langle p_i \times p_j \rangle_3\}$	$\{/ \}$	–
3 & 4	$\mathcal{S}_f$	$\{\tanh\}$	$\{p_z, \ p\ _3\}$	$\{\sqrt[2]{\cdot}, \sin, \cos\}$	–
	$\mathcal{S}_g$	$\{\text{boost}\}$	$\{\langle p_i \times p_j \rangle_3\}$	$\{*, /\}$	–
5 & 6	$\mathcal{S}_f$	$\{\tanh\}$	$\{p_0, p_x, p_y, p_z, \ p\ _3\}$	$\{\sqrt[2]{\cdot}, \sqrt{\cdot}\}$	$\{\sin, \cos\}$
	$\mathcal{S}_g$	$\{\text{boost}\}$	$\{\langle p_i \times p_j \rangle_3, \langle p_i \times p_j \rangle_4\}$	$\{*, /\}$	$\{+, -\}$

Table 7.4: The function sets  $\mathcal{S}_f$  and  $\mathcal{S}_g$  of unary and binary functions, respectively, used for **SymbolNet** in each scenario. Table taken from [230].

$\alpha = 100$  and all other parameters correspond to the recommended settings. The set of functions is

$$\begin{aligned} \mathcal{S} &= \{\sqrt[2]{\cdot}, \sqrt{\cdot}, +, /\} && \text{for scenario 1;} \\ \mathcal{S} &= \{\sin, \cos, \sinh, \sqrt[2]{\cdot}, +, -, *, /\} && \text{else.} \end{aligned} \quad (7.21)$$

For **SymbolNet**, the number of training iterations and the learning rate in each training step, as well as the batch size, are summarized in Table 7.3. The target sparsities are set to

$$\begin{aligned} \alpha_W &= 0.5, \alpha_{\text{unary}} = \alpha_{\text{binary}} = 0.05 && \text{for scenarios 1 and 2;} \\ \alpha_W &= 0.8, \alpha_{\text{unary}} = \alpha_{\text{binary}} = 0.3 && \text{else.} \end{aligned} \quad (7.22)$$

The function sets used in the specific layers can be found in Table 7.4, where they are split into unary ( $\mathcal{S}_f$ ) and binary ( $\mathcal{S}_g$ ) sets.

Both PySR and **SymbolNet** use the same objective functions. In scenarios 1 and 2, the full kinematic information is available, and the MSE loss is used. As discussed in Section 4.1, it is optimal for a Gaussian smearing of the labels. However, in scenarios 3 to 6, this is not expected to be the case due to the missing neutrino information, and the effects of the additional radiation and limited detector resolution. Much better results were achieved using an "Inverse Gaussian" loss

$$\mathcal{L}_{\text{InvGaussian}} = 1 - \exp\left(-\frac{(y - \hat{y})^2}{2\left(\frac{\max(y)}{\sigma} - \frac{\min(y)}{\sigma}\right)^2}\right). \quad (7.23)$$

where  $\sigma = 8$  is used. It approaches the MSE loss for  $y \sim \hat{y}$ , and one for  $|y - \hat{y}| \gg 0$ . Therefore, it is less sensitive to outliers, which are present in scenarios 3 to 6.

### 7.4.2 Predicted equations for the Collins-Soper angle

The distributions for the parton-level Collins-Soper angle, as well as the predictions from `SymbolNet` and `PySR`, are shown for each of the six scenarios in Fig. 7.9. As can be seen, for the easiest reconstruction task, corresponding to scenario 1, both `SymbolNet` and `PySR` match the true distribution. Therefore, they are able to build Eq. (7.19) from the decayed particle momenta. Specifically, they find the equations [230]

$$\cos \theta_{\text{PySR}}^* = \frac{p_{z,b} + p_{z,\bar{l}} + p_{z,\nu}}{\sqrt{\left(p_{x,b} + p_{x,\bar{l}} + p_{x,\nu}\right)^2 + \left(p_{y,b} + p_{y,\bar{l}} + p_{y,\nu}\right)^2 + \left(p_{z,b} + p_{z,\bar{l}} + p_{z,\nu}\right)^2}}, \quad (7.24)$$

$$\begin{aligned} \cos \theta_{\text{SymbolNet}}^* &= \frac{1.006p_{z,b} + 1.001p_{z,\bar{l}} + 1.002p_{z,\nu} - 1.027p_{z,\bar{b}} - 1.027p_{z,q} - 1.031p_{z,\bar{q}}}{\left\|1.034p_b + 1.022p_{\bar{l}} + 1.024p_{\nu} - p_{\bar{b}} - 1.007p_q - 1.009p_{\bar{q}}\right\|_3} \\ &\approx \frac{1}{2} \left( \frac{p_{z,t}}{\|p_t\|_3} - \frac{p_{z,\bar{t}}}{\|p_{\bar{t}}\|_3} \right). \end{aligned} \quad (7.25)$$

In the case of `PySR`, this corresponds to the exact equation, while for `SymbolNet`, some of the numerical constants slightly deviate from 1. However, the MSE loss of the `SymbolNet` equation is slightly lower, which is a consequence of numerical instabilities induced by `MadSpin`. Furthermore, `SymbolNet` does not exclude any of the input features, and therefore calculates  $\cos \theta^*$  from both the hadronic and the leptonic top decay (which are equivalent up to a sign in the  $t\bar{t}$  frame).

In scenario 2, where the full kinematic information is still available but variables are provided in the lab frame, only `SymbolNet` matches the true  $\cos \theta^*$  distribution. This is a consequence of `SymbolNet` being able to process 4-vectors and apply a Lorentz boost to them. `PySR` can only approximate the effect of the boost and therefore shows slight deviations in its prediction.

Starting from scenario 3, the neutrino momentum is removed, and an analytic equation resembling the parton-level  $\cos \theta^*$  is not available anymore. Consequently, `SymbolNet` and `PySR` show larger deviations from the true distribution. Their performance is similar in scenarios 3 and 4. In scenarios 5 and 6, where smearing is added, the overall performance decreases further, but `SymbolNet` shows better results than `PySR` (as demonstrated in more detail below). An interesting observation is that both algorithms consistently predict too few events in the outermost bins across all scenarios. However, `SymbolNet` compensates for this by predicting values outside of the range  $[-1, 1]$ . Clipping its output to this range improves the results. On the other hand, `PySR` always restricts itself to  $[-1, 1]$ , which could be due to it starting from a very simple equation.

The corresponding equations to the predictions in Fig. 7.9 are provided in Appendix G. Although most equations are very complex, they still show some similarities. Across all scenarios, `PySR` predicts equations of the form [230]

$$\cos \theta_{\text{PySR}}^* \sim \sin \frac{\sum_i a_i p_{i,z}}{\sum_i b_i E_i}. \quad (7.26)$$

Interestingly, it favors the  $\sin$  over the  $\cos$ , despite having both functions available. For the normalization,

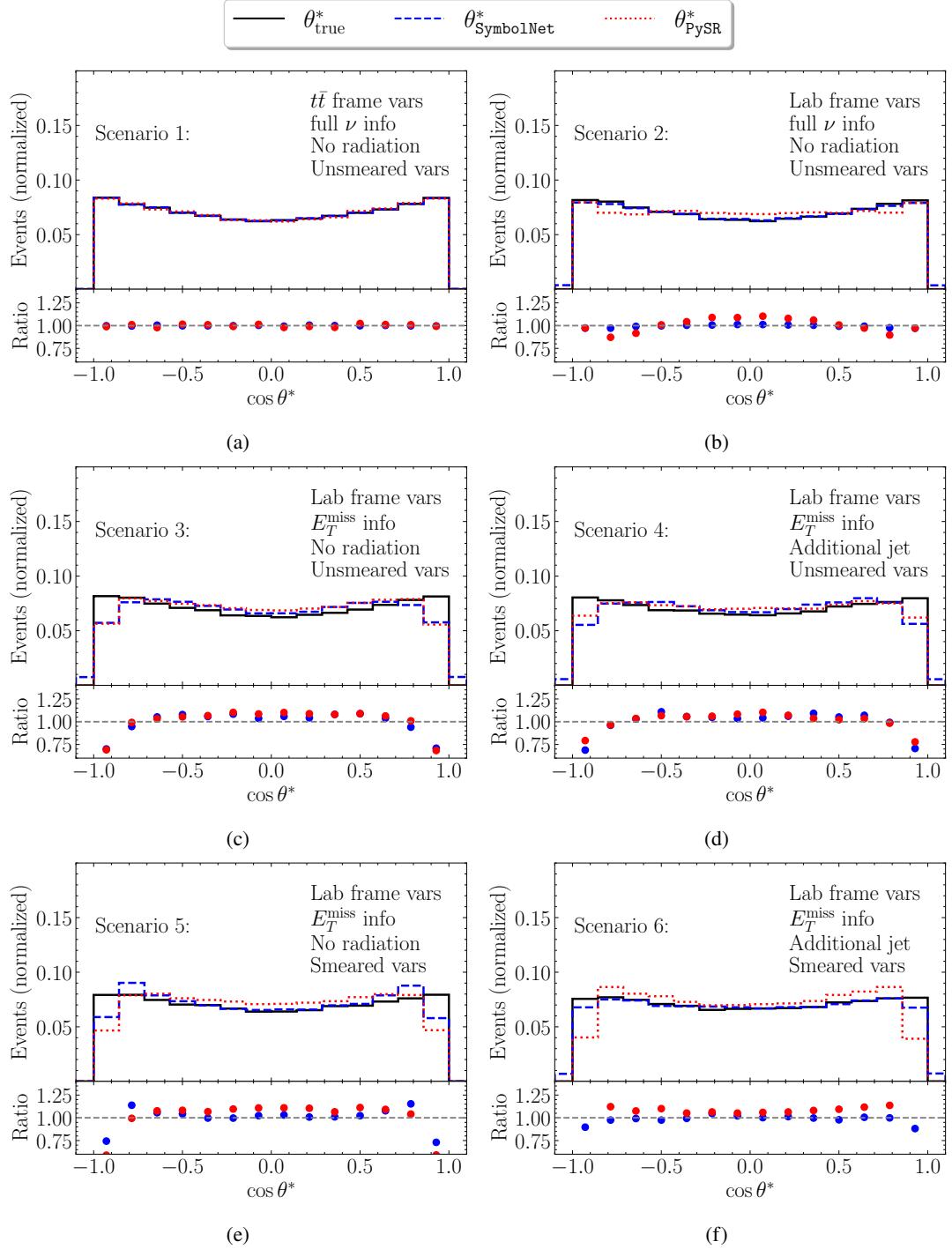


Figure 7.9: Distributions of the parton level Collins-Soper angle (black), and the predictions from SymbolNet (blue) and PySR (red). The distributions shown in (a) to (f) correspond to scenarios 1 to 6, respectively. Figure taken from [230].

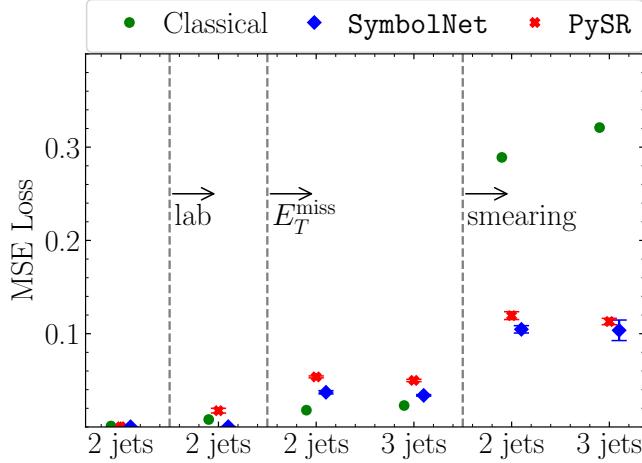


Figure 7.10: Comparison of the MSE losses obtained from a classical reconstruction method, and the predictions from PySR and **SymbolNet** in scenarios 1 to 6. The error bars correspond to the standard deviation in 10 training runs. It should be noted that for scenarios 3 to 6, the Inverse Gaussian loss was used for the training. Figure taken from [230].

it chooses the energy of the particles  $E_i$  as a proxy for the three-momentum norm. The equations learned by **SymbolNet** have a higher complexity due to the different approach to SR. They can be approximated via [230]

$$\cos \theta_{\text{SymbolNet}}^* \sim \frac{\text{boost} \left( \sum_i a_i p_i \mid \sum_j b_j p_j \right) \Big|_z}{\left\| \text{boost} \left( \sum_i a_i p_i \mid \sum_j b_j p_j \right) \right\|_3} . \quad (7.27)$$

This core structure of the equations is a generalized form of the parton-level formula of the Collins-Soper angle. The results show that it can be used as a first proxy even at the detector level.

### 7.4.3 Performance and sensitivity of the equations

Fig. 7.10 shows the MSE loss calculated for **SymbolNet** and PySR for the different scenarios. Although the objective function in scenarios without the full neutrino information is the Inverse Gaussian loss, the MSE loss still allows for assessing the goodness of the fit. The SR results are compared to a classical reconstruction method, which is used in [255]. In this method, the unknown neutrino momentum is reconstructed using the  $W$  boson mass. The MSE loss shows that the classical reconstruction performs slightly better in scenarios 3 and 4. However, it is heavily outperformed by **SymbolNet** and PySR once detector effects are taken into account. Fig. 7.10 confirms the slightly better performance of **SymbolNet** over PySR in all scenarios.

The most realistic scenario for an experimental analysis, scenario 6, is further examined in Fig. 7.11. The plots in the upper row show again the prediction of **SymbolNet** and PySR. It is evident that the prediction obtained from **SymbolNet** reflects the parton-level distribution much better. This can also be

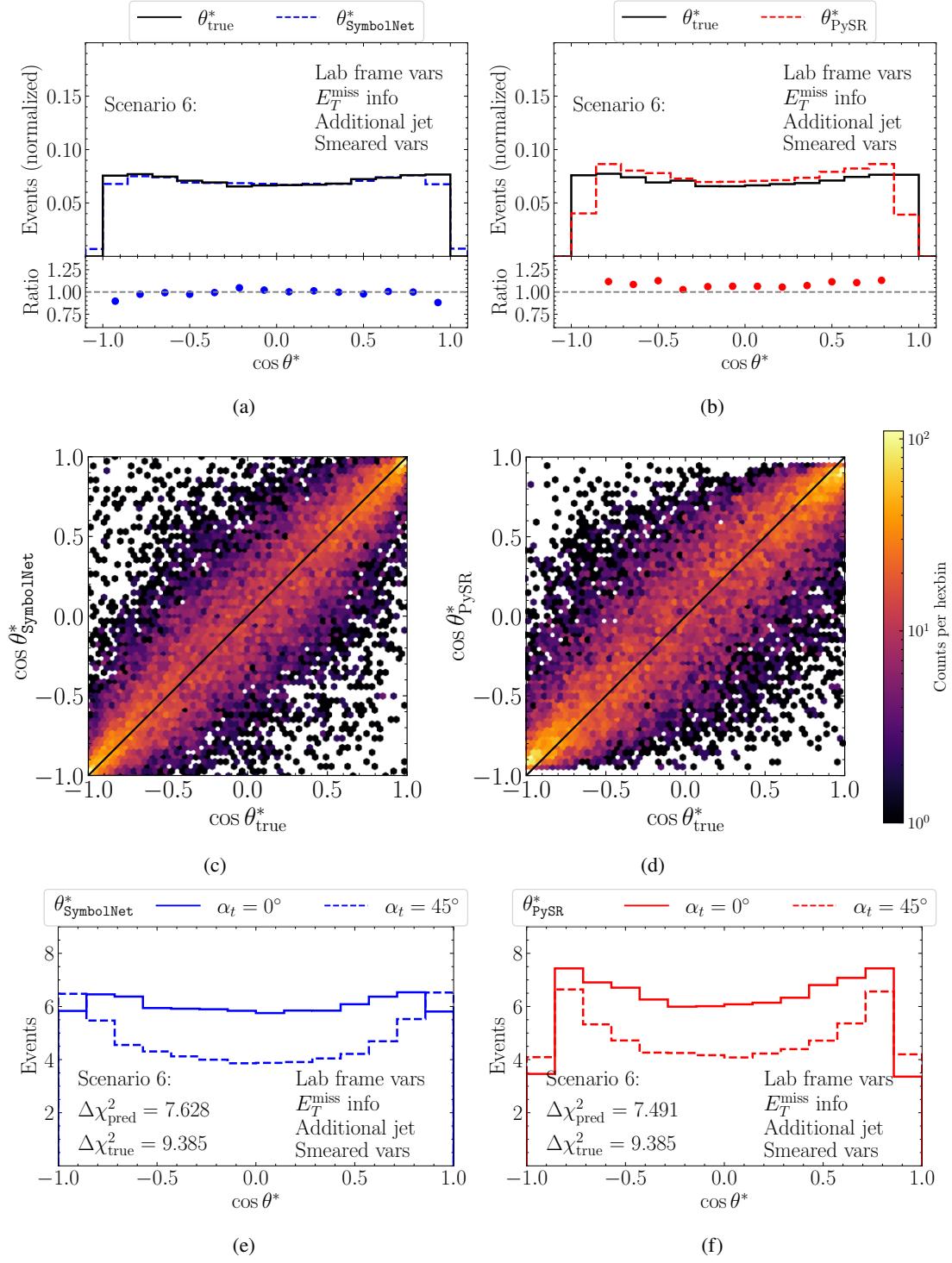


Figure 7.11: (a) + (b) binned predictions, (c) + (d) predictions of individual events, and (e) + (f)  $\mathcal{CP}$  sensitivity for  $\alpha_t = 45^\circ$  obtained from (left column) SymbolNet and (right column) PySR in scenario 6. Figure taken from [230].

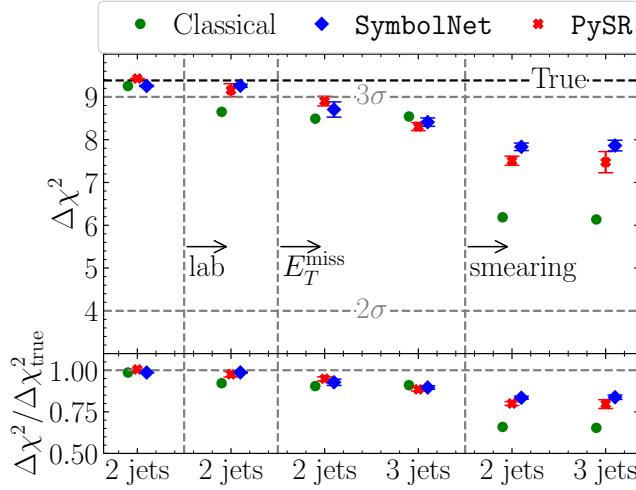


Figure 7.12: Same as Fig. 7.10, but instead of the MSE loss, the  $\Delta\chi^2$  value for excluding the  $\alpha_t = 45^\circ$  hypothesis is shown. Figure taken from [230].

shown by comparing the prediction of individual events to the true values in the middle row. For the SymbolNet prediction, the events follow the diagonal line, although with some spread. For PySR, the bulk of the predictions shows a small rotation away from the diagonal line. Furthermore, the spread of the events is larger, and events do not reach predictions of 1 or  $-1$ . Finally, in the lower row, the  $\mathcal{CP}$  sensitivities for  $\alpha_t = 45^\circ$  are shown. Again, SymbolNet shows slightly better results, reaching  $\Delta\chi^2 = 7.628$ , while the PySR prediction results in  $\Delta\chi^2 = 7.491$ .

The  $\mathcal{CP}$  sensitivity is further examined in Fig. 7.12, where it is shown for all six scenarios. Again, the SR approaches are compared to the classical reconstruction algorithm. A similar behavior to that of the MSE loss can be seen, as the classical algorithm is heavily outperformed in scenarios 5 and 6, and SymbolNet performs slightly better than PySR. Overall, the  $\Delta\chi^2$  decreases with increasing scenario, as expected. Still, SymbolNet and PySR manage to retain up to 85% of the parton-level  $\mathcal{CP}$  sensitivity, even in the most realistic scenario.

Finally, Fig. 7.13 shows a scan of the  $\Delta\chi^2$  values for testing the SM hypothesis against  $\alpha_t \in \{10^\circ, 25^\circ, 45^\circ, 90^\circ\}$  in scenario 6. Comparing  $\alpha_t = 45^\circ$  with  $\alpha_t = 90^\circ$  shows that the percentage of  $\mathcal{CP}$  information that is retained at the detector level stays approximately constant. For smaller  $\mathcal{CP}$ -mixing angles, the performance of all methods drops. This is most likely due to the smaller kinematic differences between the BSM distributions and the SM, as well as the corresponding higher impact of the smearing. SymbolNet and PySR outperform the classical reconstruction algorithms at all  $\alpha_t$ .

## 7.5 Conclusions

The goal of this chapter was to examine how SR can enhance the interpretability of Higgs  $\mathcal{CP}$  analyses at the LHC while maintaining optimal sensitivity. For this, two approaches to SR were considered. First, a genetic algorithm implemented in PySR, and second, a DNN-like approach implemented in SymbolNet. The performance of the two was tested in the VBF and  $t\bar{t}H$  production channels and compared to more

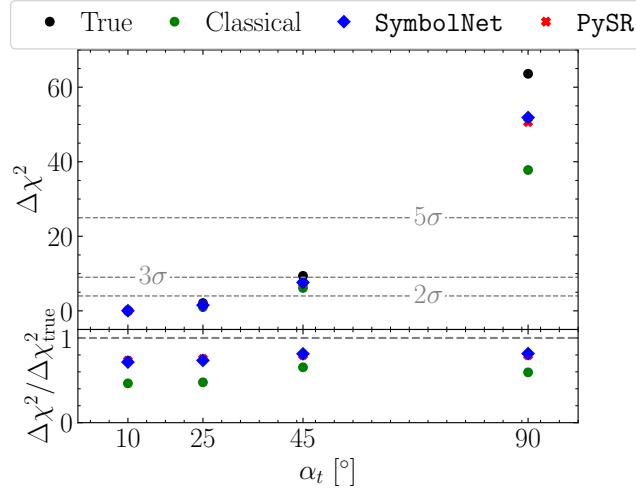


Figure 7.13:  $\Delta\chi^2$  values for excluding different  $\alpha_t$  hypotheses in scenario 6, based on the predictions of the classical reconstruction method, SymbolNet, and PySR. Figure taken from [230].

traditional methods. Furthermore, their data efficiency was assessed, i.e., their performance with few training samples and small differences between the labels, as well as when dealing with high-dimensional parameter spaces.

First, it was examined how optimal observables can be found for VBF. For this, the equations obtained from PySR and SymbolNet were compared to the locally optimal observable  $m_H^{-2} p_{T,j_1} p_{T,j_2} \sin \Delta\phi_{jj}$ . It was shown that both algorithms outperform this observable. Furthermore, they were also able to outperform a BDT that was trained using XGBoost. The equations obtained were investigated, and it was found that they were indeed  $\mathcal{CP}$ -odd. They showed the same structure, i.e.,  $\Delta\phi_{jj}$  was multiplied by a complex  $\mathcal{CP}$ -even term. Overall, PySR showed the best results and maintained high precision, even with sparse training data.

Next, the optimal observables were investigated in  $t\bar{t}H$ . For this, SymbolNet was adapted to construct TPs from the final-state momenta. Furthermore, the network itself was constructed to be  $\mathcal{CP}$ -odd. However, due to the smallness of the interference term and the few statistics in  $t\bar{t}H$ , the projected sensitivities were too weak to motivate a study at the LHC in the near future.

Finally, the  $\mathcal{CP}$ -sensitive Collins-Soper angle in  $t\bar{t}H$  was investigated. Specifically, it was tested how well the parton-level variable can be approximated from the momenta at the detector level. Six different scenarios were tested, becoming increasingly more complex. For the most realistic scenario 6, SymbolNet and PySR strongly outperformed a traditional method for top reconstruction. Overall, SymbolNet performed slightly better than PySR but also yielded more complex equations. Both approaches were able to retain more than 80% of the  $\mathcal{CP}$  sensitivity of the parton-level variable.

SR can be concluded to be a powerful tool that is especially important when testing fundamental symmetries. It was demonstrated how SR can be used to achieve optimal sensitivity while improving interpretability. The two tools, PySR and SymbolNet, complement each other, with PySR being more sensitive to small changes in the data and more data-efficient, and SymbolNet being better at handling more complex parameter spaces. The equations obtained by the two approaches can be easily implemented in an analysis and are fast to evaluate.

## Conclusions

---

Finding  $\mathcal{CP}$  violation in the Higgs interactions would be an explicit sign of BSM physics and could also help understand where the observed baryon asymmetry of the Universe originated from. It is therefore crucial to optimize collider analyses probing the  $\mathcal{CP}$  state of the Higgs boson couplings. Hence, several channels have been studied in this thesis. The gluon fusion process in association with two jets (ggF2j) allows for constraining the  $\mathcal{CP}$  state of an effective Higgs gluon coupling, while the  $t\bar{t}H$  channel yields direct constraints on the top-Yukawa coupling. The VBF and  $VH$  channels assess possible  $\mathcal{CP}$  violation in the  $HVV$  coupling. Additionally, the interplay of all these channels can have interesting effects on the constraints.

The first part of this thesis examined the ggF2j process. As of now, experimental limits on the  $\mathcal{CP}$  state of the effective Higgs-gluon coupling are still relatively weak. Here, a neural network was trained to separate the ggF2j process into two signal regions, the ggF2j-SR and the VBF-SR. Then, the  $\mathcal{CP}$  structure of the coupling was studied using a  $\mathcal{CP}$ -even and a  $\mathcal{CP}$ -odd classifier. The ggF2j-SR yielded significantly stronger constraints than current experimental results for the Higgs-gluon  $\mathcal{CP}$ -mixing angle of  $|\alpha_g| < 25^\circ$  @  $2\sigma$  CL for  $\mathcal{L} = 139 \text{ fb}^{-1}$ , but without taking into account the non-Higgs background. The  $\mathcal{CP}$ -even classifier gave the dominant contribution to this result. The constraints in the ggF2j-SR were shown to be robust when additional  $\mathcal{CP}$ -violating effects in the  $HVV$  coupling were taken into account. Consequently, the ggF2j-SR performed significantly better overall. It should be considered in experimental analyses targeting the Higgs-gluon coupling, despite the VBF-SR containing more of the  $\mathcal{CP}$ -sensitive  $qq$  initial state events. It can also yield competitive, albeit indirect, limits with a global fit for the top-Yukawa coupling.

In the second part of the thesis, a possible STXS extension for the  $t\bar{t}H$  process was examined. In the current STXS framework, the binning is 1-dimensional in the transverse momentum of the Higgs  $p_{T,H}$ , which was shown here not to be optimal. To find the best observable for a  $\mathcal{CP}$ -sensitive extension, 30 different variables across multiple rest frames were evaluated. The  $t\bar{t}H(\rightarrow \gamma\gamma)$ ,  $t\bar{t}H(\text{multilep.})$ , and  $t\bar{t}H(\rightarrow b\bar{b})$  channels were taken into account and matched to recent experimental analyses, to obtain realistic values for the efficiency of the event selection and the detector effects. Furthermore, statistical and systematic uncertainties in the individual channels were considered, as well as the distribution of the non-Higgs background. Three candidates were identified to be optimal for an STXS extension, adding a second dimension to the  $p_{T,H}$  binning. These were the top quark-based variable  $b_2$  in the lab

frame, and the pseudorapidity difference of the top quarks  $|\Delta\eta_{t\bar{t}}|$  and the Collins-Soper angle  $|\cos\theta^*|$  in the  $t\bar{t}$  frame. Using the respective 2-dimensional binning allows to exclude  $|\alpha_t| < 37^\circ$  @  $2\sigma$  CL for  $\mathcal{L} = 300 \text{ fb}^{-1}$  and a SM-like rate. Although less sensitive than a multivariate approach, this offers much higher sensitivity than the 1-dimensional binning and allows a straightforward combination of decay channels and results from different experiments, as well as good interpretability.

Finally, symbolic regression (SR) was used in the last part of this thesis to find analytical equations for optimal observables and for reconstructing parton-level observables, using information at the detector level. For this, two different approaches to SR were considered. PySR is based on a genetic algorithm, while **SymbolNet** implements SR in a deep neural network-like structure. **SymbolNet** was further modified to support 4-vectors. For VBF, the SR approaches were able to outperform a BDT while returning a formula that was easily recognizable as  $\mathcal{CP}$ -odd. It showed a similar structure to the previously known, locally optimal observable, as the azimuthal angle difference of the top quarks  $\Delta\phi_{jj}$  was multiplied by a purely  $\mathcal{CP}$ -even term. Further modifying **SymbolNet** to be manifestly  $\mathcal{CP}$ -odd also yielded optimal observables for  $t\bar{t}H$ . Although they performed better than a simple triple product, the sensitivity of such an approach is too low to be measurable at the LHC in the near future. A reconstruction of the Collins-Soper angle led to much more promising results, as **SymbolNet** and PySR significantly outperformed a classical reconstruction algorithm and retained up to 85% of the parton-level  $\mathcal{CP}$  sensitivity. Overall, **SymbolNet** and PySR showed complementary results. While PySR performed significantly better in scenarios with few data or tiny differences between the two hypotheses, **SymbolNet** excelled in high-dimensional parameter spaces and when using large datasets.

The results presented in this thesis demonstrate how improvements in experimental Higgs  $\mathcal{CP}$  studies might be possible now and in the coming years. They mainly concentrated on the Higgs couplings to top quarks, gluons, and weak vector bosons, utilizing the luminosities achievable with the full Run 2 data or after the completion of Run 3. Furthermore, the results demonstrated that the improvements still hold in projections for the HL-LHC. The techniques utilized in this work are applicable to tests of other symmetries and beyond the Higgs couplings.

Much more is to be explored in the future. First studies show that constraints on the Higgs couplings will become significantly stronger at the proposed future circular collider (FCC) [339]. This also includes the Higgs couplings to light fermions and their  $\mathcal{CP}$  nature, which are not possible to measure with high precision at the LHC. On top of that, low-energy and cosmological probes of the SM become increasingly important. The upper limits of several EDMs are expected to be reduced by orders of magnitude by upcoming experiments [340–343]. Such searches should be used complementarily to the collider measurements, especially in the context of the baryon asymmetry of the Universe. The many open questions guarantee that particle physics will remain an exciting field for many years to come.

## APPENDIX A

---

### Symmetry groups

---

A concept that has been of great importance in building the theoretical foundation of the SM and continues to be of great importance today is the idea of *symmetries*. Symmetries appear as invariances under a set of transformations and can be described mathematically via symmetry groups. Some examples for this are internal symmetries that act on the quantum numbers of fields, and external symmetries that act on spacetime. The easiest non-trivial examples are discrete symmetries, e.g., invariance under a parity flip ( $\mathcal{P}$ ), charge conjugation ( $\mathcal{C}$ ), or time reversal ( $\mathcal{T}$ ), or their combinations, such as  $\mathcal{CP}$ . They are described by the  $Z_2$  symmetry group, which consists of exactly two elements: the identity and the element performing the reflection at the origin.<sup>1</sup>

Most of the symmetries needed for the construction of the SM are based on the concept of *Lie groups* [345]. These are groups that depend on a continuous set of parameters  $\alpha$  with a notion of closeness. The identity element can be defined as

$$g(\alpha)|_{\alpha=0} = e, \quad D(\alpha)|_{\alpha=0} = 1 \quad (\text{A.1})$$

where  $g(\alpha)$  is an element of the group and  $D(\alpha)$  its representation. An element close to the identity can then be reached via an infinitesimal transformation

$$D(d\alpha) = 1 + id\alpha_a X_a + \mathcal{O}(d\alpha^2), \quad X_a = -i \frac{\partial}{\partial \alpha_a} D(\alpha)|_{\alpha=0} \quad (\text{A.2})$$

where the  $X_a$  are the so-called *generators* of the group. A group element far away from the identity is reached by subsequently performing infinitesimal transformations, and therefore

$$D(\alpha) = \exp i\alpha_a X_a . \quad (\text{A.3})$$

The generators in a Lie group fulfill commutator relations

$$[X_a, X_b] = if_{abc} X_c \quad (\text{A.4})$$

---

<sup>1</sup> This is true only for the conventional choices of discrete symmetries. If degenerate fields exist in a theory, it is possible to generalize the symmetries to follow a  $Z_{2^p}$  symmetry where  $p > 1$  is an integer. See e.g. [344].

which is called the *Lie algebra* and is completely defined via the structure constants  $f_{abc}$ .

Some Lie groups are of particular interest for the SM. The first and simplest one is  $U(1)$ , which corresponds to the rotation on a circle. It is a 1-dimensional and therefore an abelian Lie group where  $\alpha$  corresponds to a rotation angle and  $X$  is a real number. In the SM, the  $U(1)$  Lie group is identified with the gauge symmetry of QED, where  $X$  corresponds to the electric charge  $Q$ . Due to the  $U(1)$  group being Abelian, there are no self-interactions of photons.

The second important Lie group is  $SU(2)$  which gives rise to three generators connected via the  $su(2)$  algebra

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad . \quad (\text{A.5})$$

This group can be used to describe the spin of particles as the fundamental representation corresponds to spin-1/2 (with states  $m = \pm 1/2$ ). In the spin-1/2 representation, the generators of the group that follow

$$D(\alpha) = \exp i \sum_{i=1}^3 \alpha_i J_i \quad (\text{A.6})$$

can be identified with the Pauli matrices  $\sigma_i = 2J_i$ . The trivial representation corresponds to spin-0 particles, which are unaffected by transformations under  $SU(2)$ . Particles with higher spin can be expressed by higher representations of  $SU(2)$  via tensor products of the spin-1/2 case, where for a spin  $l$  the dimension of the representation is  $2l + 1$ . Examples are fermions for  $l = 1/2$ , massive vector bosons for  $l = 1$ , and  $\Delta$ -baryons for  $l = 3/2$ . In the SM, the  $SU(2)$  Lie group is also identified with the symmetry of weak isospin, which is important for the electroweak interaction. In this case, the three generators are identified with three gauge fields forming a weak isospin triplet, which mix into physical gauge bosons after the electroweak symmetry breaking.

The final Lie group needed for the completion of the SM is  $SU(3)$ , which is generated by the eight Gell-Mann matrices  $\lambda_1 \dots \lambda_8$ . The Gell-Mann matrices contain the three Pauli matrices acting on a 2-dimensional subspace (and therefore containing  $SU(2)$  as a subgroup). The  $SU(3)$  Lie group has been of importance multiple times throughout the development of particle physics. Before the discovery of the quark model, the  $SU(2)$  subgroup was first used to describe *isospin* by placing the proton and the neutron in an isospin doublet. The discovery of additional particles historically led to the introduction of the strangeness quantum number, from which on particles could be described much better under the full  $SU(3)$  symmetry with the proton, neutron, and  $\Lambda$  baryon forming a triplet. From a modern perspective, this corresponds to the flavor symmetry of the light quarks. In the SM,  $SU(3)$  is realized in the quark model where the fundamental representation describes the *color* of a particle with the three color states red, green, and blue. The eight generators of  $SU(3)$  are associated with the gluons, which represent color states themselves.

Lie algebras also allow the incorporation of the concept of special relativity into a theory. In a  $1 + 3$  dimensional Minkowskian spacetime, there are 10 symmetry transformations (generators) that leave the description of physical processes under

$$x^\mu \rightarrow x'^\mu = \Lambda_\nu^\mu x^\nu + a^\mu \quad (\text{A.7})$$

invariant. They consist of 4 translations, 3 rotations, and 3 boosts and are known as the *Poincaré* group, which itself is therefore a 10 dimensional Lie group. The 3 rotations and 3 boosts with additional requirements  $\Lambda_0^0 \geq 1$  and  $\det \Lambda = 1$  form a  $SO(1, 3)$  subgroup which is referred to as the *(proper*

*orthochronous) Lorentz group.* In contrast to the Lie groups specifying the internal symmetries of the SM, both the Poincaré and Lorentz group are not compact. The algebra of the Lorentz group is isomorphic to the complexified direct sum of two  $su(2)$  algebras. Therefore, there exist two 2-dimensional subspaces, where the fundamental representation in the first (second) subspace is called a *left (right) handed Weyl spinor*. The handedness consequently defines how a Weyl spinor transforms under the Lorentz group. It is also called the *chirality*. A parity transformation flips the handedness of a Weyl spinor. This concept is important for the weak interaction, in which only left-handed particles participate and which is maximally parity-violating. In interactions independent of the chirality, *Dirac spinors* are used, which are combinations of a left- and right-handed Weyl spinor.

## APPENDIX B

---

# Details of the event generation

---

In this appendix, the event generation for Chapters 5 to 7 is discussed in more detail. Specifically, the event syntax for `MadGraph5_aMC@NLO` is provided.

### Event generation for Chapter 5

To filter out the effects from the effective Higgs-gluon coupling, a custom model file employing the `UFO` format is used [346]. In this file, the two effective couplings in Eqs. (2.61) and (2.62) are added to the SM. This allows for the generation of `ggF2j` events with the  $H \rightarrow \gamma\gamma$  decay on tree level. The Higgs coupling to photons is assumed to be SM-like so that the constraints are not dominated by rate information. In total, there are four new coupling parameters included in the file, which are called `QGGH`, `QGGA`, `QGAGAH`, and `QGAGAA`. Enabling one of these parameters (by setting them to 1) forces `MadGraph` to generate events with exactly one scalar or pseudoscalar Higgs boson coupling to gluons or photons. The parameters are specified similarly to the QED and QCD orders during the event generation.

The `ggF2j` events are generated at LO with two exclusive hard QCD jets. The exact syntax is provided below. Afterwards, events are scaled to NLO by using a constant K-factor that is added to the total cross section. To verify that the approach taken here is valid, additional `ggF2j` datasets were created at NLO, including two inclusive jets by employing the `Higgs Characterization` `UFO` model [128]. They were generated for the SM point  $\alpha_g = 0^\circ$  and a BSM benchmark point including a significant interference contribution  $\alpha_g = 35^\circ$ . The inclusive NLO events were then given to the signal-background and  $\mathcal{CP}$  classifiers as detailed in Section 5.2 without retraining the classifiers. The selection efficiency of the signal-background classifiers was found to be slightly worse in the NLO case, leading to a reduced value of the  $\Delta\chi^2$  between the  $\alpha_g = 0^\circ$  and  $\alpha_g = 35^\circ$  hypotheses by about 10%. However, this is primarily an effect of the signal-background classifiers being trained with LO samples. Using the same selection efficiency for the NLO samples, the  $\Delta\chi^2_{\text{LO}}$  and  $\Delta\chi^2_{\text{NLO}}$  agree within 1%. This indicates that the  $\mathcal{CP}$  analysis is stable and that the errors obtained by using a flat K-factor and an exclusive jet generation are small.

The syntax used for generating events at the parton level is listed in Table B.1. For the `ggF2j` processes, the full custom `UFO` model file is used. For the SM events of the Higgs background processes, the effective Higgs-gluon coupling is disabled in the model file. Finally, the events containing  $\mathcal{CP}$  violation in the  $HVV$  coupling employ the `SMEFTsim 3.0` package [146]. An additional diagram filter is used to

Events	Syntax
$ggF2j   \mathcal{M}_{\text{even}} ^2$	$p\ p > a\ a\ j\ j\ QGGH==1\ QGAGAH==1\ QED=4\ QCD=4$
$ggF2j   \mathcal{M}_{\text{odd}} ^2$	$p\ p > a\ a\ j\ j\ QGGA==1\ QGAGAH==1\ QED=4\ QCD=4$
ggF2j Interf.	$p\ p > a\ a\ j\ j\ QGGH^2==1\ QGGA^2==1\ QGAGAH==1\ QED=4\ QCD=4$
VBF SM	$p\ p > a\ a\ j\ j\ \$\$ w+\ w-\ z\ QGAGAH==1$ $p\ p > z > a\ a\ j\ j\ QGAGAH==1$
VH SM	$\text{add process } p\ p > w+ > a\ a\ j\ j\ QGAGAH==1$ $\text{add process } p\ p > w- > a\ a\ j\ j\ QGAGAH==1$
$VBF\ c_{\Phi\tilde{W}} = 1\ \text{TeV}^{-1}$	$p\ p > a\ a\ j\ j\ \$\$ w+\ w-\ z\ /a\ QCD=0\ NP=1\ NP^2==1$ $p\ p > z > a\ a\ j\ j\ /a\ NP=1\ NP^2==1$
$VH\ c_{\Phi\tilde{W}} = 1\ \text{TeV}^{-1}$	$\text{add process } p\ p > w+ > a\ a\ j\ j\ /a\ NP=1\ NP^2==1$ $\text{add process } p\ p > w- > a\ a\ j\ j\ /a\ NP=1\ NP^2==1$

 Table B.1: Syntax for the event generation of all datasets used in Chapter 5. The `generate` command is omitted.

remove any diagrams that do not correspond to VBF or  $VH$ .

### Event generation for Chapter 6

All events are generated using the Higgs Characterization UFO model [128]. This model file replaces the SM Higgs boson with a particle  $x0$  that acts as a Higgs boson with undefined  $\mathcal{CP}$  properties. Although event generation at NLO QCD is possible with the model file used, it introduces unwanted effects. Specifically, in contrast to the  $ggF2j$  case, the top quark cannot be approximated via the infinite top mass limit in  $t\bar{t}H$ . This leads to the problem that an effective Higgs-gluon coupling, as detailed above, cannot be generated. However, a gluon fusion loop that is mainly influenced by the top-quark coupling places much stronger constraints on the  $\mathcal{CP}$  character of this coupling via its rate than the direct  $t\bar{t}H$  measurement itself. It is therefore not possible to disentangle the effects at NLO in this model file. Luckily, scaling the events from LO to NLO via a constant K-factor has been shown to be a valid approximation [174, 239] since the corrections to the differential cross sections are mostly independent of the  $\mathcal{CP}$  state.

The syntax used for generating both the signal events and all considered background events is shown in Table B.2. The signal events at  $\alpha_t = 0^\circ$  and  $\alpha_t = 90^\circ$  are obtained by setting the `cosa` parameter to 1 or 0, respectively. Furthermore, the scalar and pseudoscalar top-Yukawa coupling parameters `khtt` and `katt` are set to 1 in both cases, restoring the SM rate. The  $x0$  does not appear in the s-channel for the background diagrams and therefore does not have to be explicitly excluded in the syntax.

Events	Syntax
$t\bar{t}H$	<code>p p &gt; x0 t t~</code>
$t\bar{t}\gamma\gamma$	<code>p p &gt; t t~ a a</code>
$t\bar{t}b\bar{b}$	<code>p p &gt; t t~ b b~</code>
$t\bar{t}W$	<code>p p &gt; t t~ w+</code> <code>add process p p &gt; t t~ w-</code>

Table B.2: Syntax for the event generation of all datasets used in Chapter 6. The `generate` command is omitted.

### Event generation for Chapter 7

For the study of  $\mathcal{CP}$ -odd observables in VBF, the events are generated using the same syntax as in Table B.1. Again, the SMEFTsim 3.0 package is employed. The different amounts of  $\mathcal{CP}$  violation in the  $HVV$  coupling are obtained by setting the  $c_{\Phi\widetilde{W}}$  parameter to the respective value.

In the case of both  $\mathcal{CP}$ -odd and  $\mathcal{CP}$ -even observables in  $t\bar{t}H$ , the events are generated using the `Higgs Characterization` UFO model and the syntax in Table B.2. Additionally,  $t\bar{t}H$  events associated with an additional hard jet are generated using `p p > x0 t t~ j`. The di-leptonic and semi-leptonic top quark decays are handled by `MadSpin`. Different values for  $\alpha_t$  are obtained by modifying the `cosa` parameter.

## APPENDIX C

---

### Likelihood tests

---

This appendix discusses the details of statistical analyses in high-energy physics. The discussion is mainly based on [347] and the review of statistics in [12]. Specifically, it is discussed how exclusion plots can be obtained.

To test how well a certain BSM scenario, for example with some  $\mathcal{CP}$  mixing angle  $\alpha \neq 0^\circ$ , compares to the measured data, a likelihood has to be defined. For this, the SM case of  $\alpha = 0^\circ$  is considered the *null hypothesis*  $H_0$  while the BSM state to be tested is associated with the *alternative hypothesis*  $H_1$ . When an analysis is conducted at the LHC, likelihood tests are typically performed by conducting a counting experiment. This means that the events follow a Poisson distribution

$$P(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda} \quad (\text{C.1})$$

with the expected number of events  $\lambda$  and the observed number  $n$ . In the corresponding likelihood function, the measured bins factorize

$$L(\lambda) = \prod_{i=1}^{N_{\text{bins}}} \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i} \quad (\text{C.2})$$

According to the Neyman-Pearson lemma, the likelihood ratio provides the most powerful statistical test of  $H_1$  against  $H_0$  [336]. For convenience, the test statistic

$$t_\lambda = -2 \ln \left( \frac{L_{\text{BSM}}}{L_{\text{SM}}} \right) \quad (\text{C.3})$$

is defined, containing the likelihood ratio. When  $H_0$  and  $H_1$  are tested via counting experiments, as done in this work, Eq. (C.3) becomes

$$t_\lambda = -2 \ln \left( \prod_{i=1}^{N_{\text{bins}}} \frac{\lambda_{\text{BSM},i}^{n_i} e^{-\lambda_{\text{BSM},i}}}{\lambda_{\text{SM},i}^{n_i} e^{-\lambda_{\text{SM},i}}} \right) \quad (\text{C.4})$$

The analyses in this thesis concentrate on the question of what BSM parameter space can be excluded

given the measurements at the LHC, which have so far been in good agreement with the SM. For phenomenological studies, this can be answered by assuming the observation of SM-like data  $n_i = \lambda_{\text{SM},i}$ . Then, Eq. (C.4) becomes

$$t_\lambda = -2 \sum_{i=1}^{N_{\text{bins}}} \left[ \lambda_{\text{SM},i} \cdot \ln \left( \frac{\lambda_{\text{BSM},i}}{\lambda_{\text{SM},i}} \right) - \lambda_{\text{BSM},i} + \lambda_{\text{SM},i} \right] \quad . \quad (\text{C.5})$$

According to Wilks' theorem, the test statistic  $t_\lambda$  approaches a  $\chi^2$  distribution [348]. In the case of a counting experiment, the relevant parameter to differentiate between  $H_0$  and  $H_1$  is the predicted number of events  $\lambda$ . Therefore,  $t_\lambda$  is approximately  $\chi^2$ -distributed for one degree of freedom. In this special case, the significance of the test is  $S = \sqrt{\chi^2_{k=1}}$  and subsequently

$$S = \sqrt{-2 \sum_{i=1}^{N_{\text{bins}}} \left[ \lambda_{\text{SM},i} \cdot \ln \left( \frac{\lambda_{\text{BSM},i}}{\lambda_{\text{SM},i}} \right) - \lambda_{\text{BSM},i} + \lambda_{\text{SM},i} \right]} \quad . \quad (\text{C.6})$$

$S$  is a measurement for the confidence level of excluding the BSM hypothesis given a SM measurement. Therefore,  $S = 1$  yields the  $1\sigma$  level,  $S = 2$  yields the  $2\sigma$  level, etc.

Chapter 5 includes a study of the Higgs-background processes VBF and  $VH$ . Since these are independent of the coupling modifiers  $c_g$  and  $\tilde{c}_g$ , the events enter the SM and BSM distributions in the same way. As a result, the expected events entering Eq. (C.6) are

$$\lambda_{\text{SM}} = \lambda_{\text{BG}} + \lambda_{|\mathcal{M}_{\text{even}}|^2} \quad (\text{C.7})$$

and

$$\lambda_{\text{BSM}} = \lambda_{\text{BG}} + c_g^2 \lambda_{|\mathcal{M}_{\text{even}}|^2} + c_g \tilde{c}_g \lambda_{\text{Interf.}} + \tilde{c}_g^2 \lambda_{|\mathcal{M}_{\text{odd}}|^2} \quad . \quad (\text{C.8})$$

The exclusion plots are obtained by performing a scan over  $c_g$  and  $\tilde{c}_g$  and using the prediction at each parameter point as  $H_1$ .

Eq. (C.6) does not take into account uncertainties in the distributions of the expected events. However, if there is some uncertainty  $\sigma$  associated with them, the significance becomes [321]

$$S = \sqrt{-2 \sum_{i=1}^{N_{\text{bins}}} \left( \lambda_{\text{SM},i} \ln \left[ \frac{\lambda_{\text{BSM},i}(\lambda_{\text{SM},i} + \sigma_i^2)}{\lambda_{\text{SM},i}^2 + \lambda_{\text{BSM},i}\sigma_i^2} \right] - \frac{\lambda_{\text{SM},i}^2}{\sigma_i^2} \ln \left[ 1 + \frac{\sigma_i^2(\lambda_{\text{BSM},i} - \lambda_{\text{SM},i})}{\lambda_{\text{SM},i}(\lambda_{\text{SM},i} + \sigma_i^2)} \right] \right)} \quad . \quad (\text{C.9})$$

This is used in Chapter 6 where the expected events for the SM case and the tested BSM scenarios are obtained using Eq. (6.1). In the limit  $\sigma_i \rightarrow 0$ , Eq. (C.6) is restored.

In the case of a  $\mathcal{CP}$ -odd observable, the significances can be obtained in the same way, but it is often more useful to define a bin-wise asymmetry

$$\mathcal{A}_i = \frac{N_i^+ - N_i^-}{N_i^+ + N_i^-} \quad (\text{C.10})$$

as used in Chapter 7. Such an asymmetry has the advantage that most systematic uncertainties cancel

---

## Appendix C Likelihood tests

out. The statistical uncertainty of  $\mathcal{A}_i$  can be obtained by propagating the uncertainties of the bin counts  $N_i$ . Specifically,

$$\sigma_{\mathcal{A}_i} = 2 \sqrt{\frac{\left(N_i^- \sigma_{N_i^+}\right)^2 + \left(N_i^+ \sigma_{N_i^-}\right)^2}{\left(N_i^+ + N_i^-\right)^4}} = 2 \sqrt{\frac{N_i^+ \cdot N_i^-}{\left(N_i^+ + N_i^-\right)^3}}. \quad (\text{C.11})$$

where  $\sigma_{N_i^\pm} = \sqrt{N_i^\pm}$  has been used, because the  $N_i^\pm$  are obtained via a counting experiment and therefore follow a Poisson distribution.

The asymmetry  $\mathcal{A}$  itself can be approximated as a Gaussian distribution

$$L(\mathcal{A}) = \prod_{i=1}^{N_{\text{bins}}} \frac{1}{\sqrt{2\pi\sigma_{\mathcal{A},i}^2}} \exp\left(-\frac{(x_i - \mathcal{A}_i)^2}{2\sigma_{\mathcal{A},i}^2}\right) \quad (\text{C.12})$$

where  $x_i = \mathcal{A}_{\text{SM},i}$  for the observation of SM-like data. Consequently, the test statistic becomes

$$t_{\mathcal{A}} = \sum_{i=1}^{N_{\text{bins}}} \left( \frac{(\mathcal{A}_{\text{BSM},i} - \mathcal{A}_{\text{SM},i})^2}{\sigma_{\mathcal{A}_{\text{BSM},i}}^2} - 2 \log \frac{\sigma_{\mathcal{A}_{\text{SM},i}}}{\sigma_{\mathcal{A}_{\text{BSM},i}}} \right). \quad (\text{C.13})$$

## APPENDIX D

---

# Observable distributions

---

Additional 1-dimensional distributions of the variables used in Chapters 5 and 6 are provided here.

### D.1 ggF2j distributions

First, some of the observables used for the training of the signal-background classifiers from Section 5.2.1 are shown in Fig. D.1. Specifically, the energy,  $p_T$ , and pseudorapidity of the Higgs boson, and the two leading- $p_T$  jets are depicted. Furthermore, the higher level variables  $m_{jj}$ ,  $\Delta\phi_{jj}$ , and  $|\Delta\eta_{jj}|$  are shown. All variables are split into the three Higgs production processes ggF2j, VBF, and VH. Other variables that were used in the training are not shown here due to the limited information they provide.

The same variables are shown in Fig. D.2 but for different  $\mathcal{CP}$  contributions to the ggF2j amplitude, namely the  $|\mathcal{M}_{\text{even}}|^2$ ,  $|\mathcal{M}_{\text{odd}}|^2$ , and interference contributions. The interference events all enter with positive weight in the plots, as they would otherwise cancel each other.

## Appendix D Observable distributions

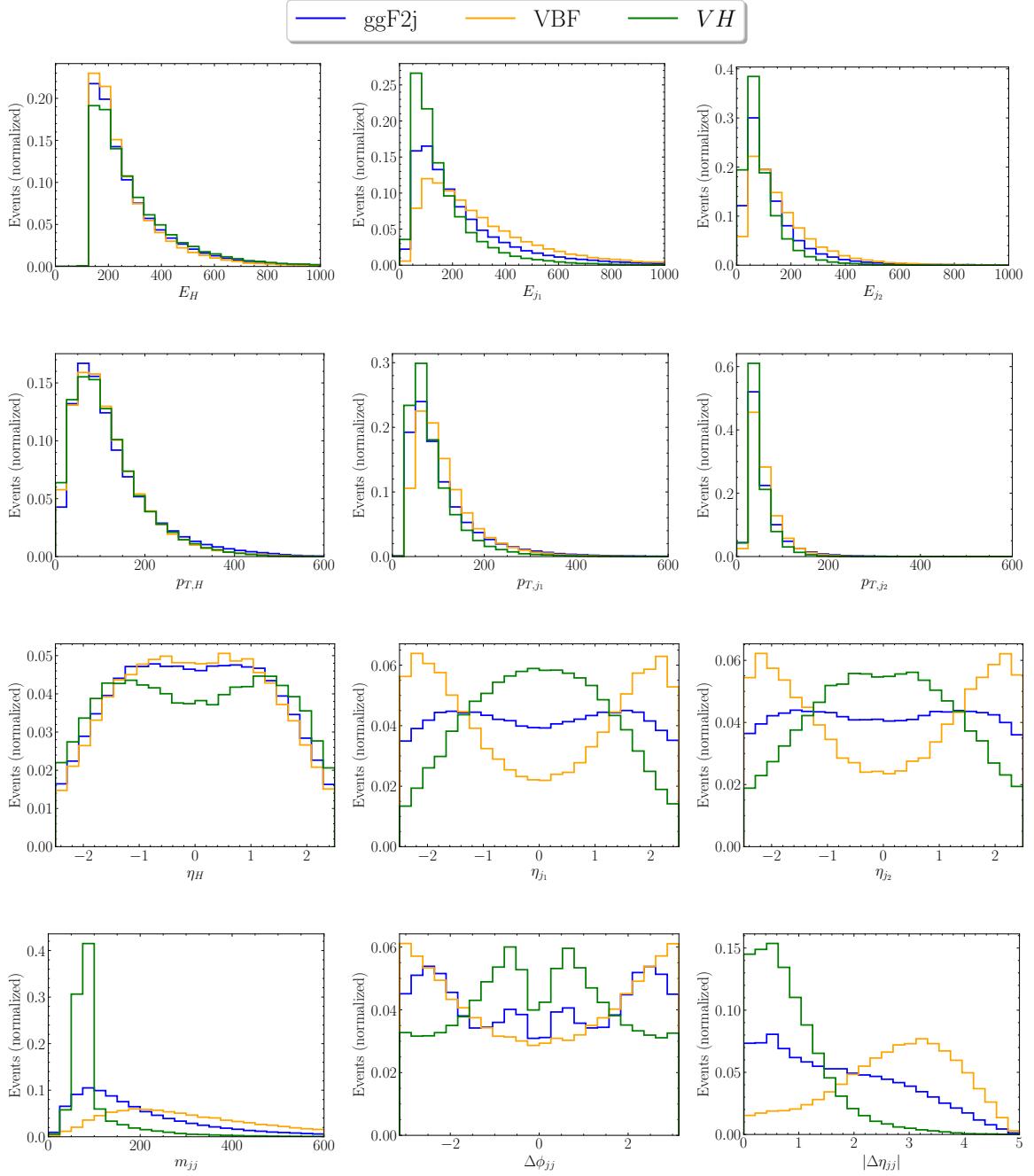


Figure D.1: Distributions of the variables given to the signal-background classifiers in Section 5.2.1, split into the ggF2j, VBF, and  $VH$  contributions. Figure taken from [233].

## Appendix D Observable distributions

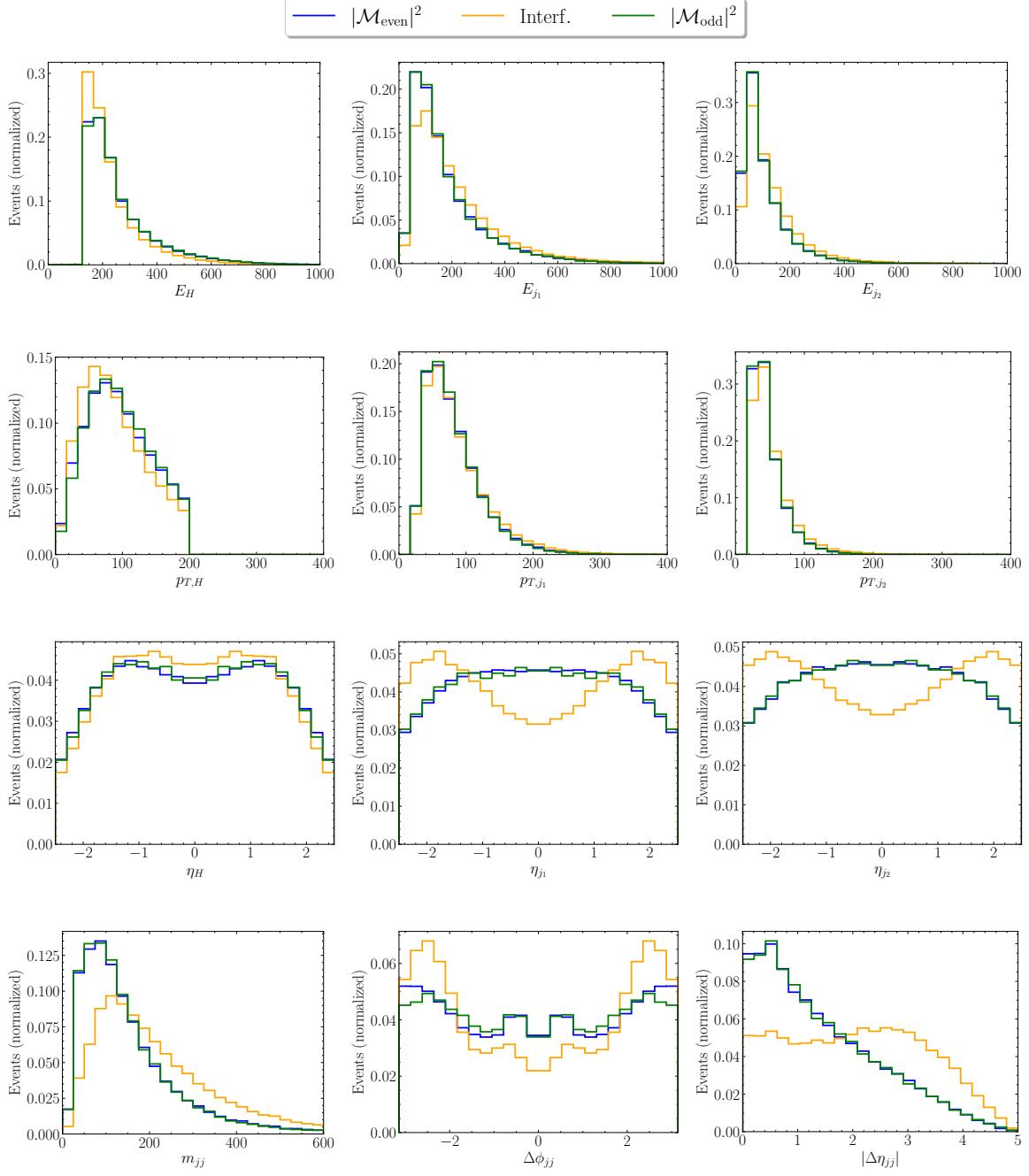


Figure D.2: Distributions of the variables given to the  $\mathcal{CP}$  classifiers in Section 5.2.2, split into the  $|\mathcal{M}_{\text{even}}|^2$ ,  $|\mathcal{M}_{\text{odd}}|^2$ , and interference contributions. Figure taken from [233].

## D.2 $t\bar{t}H$ distributions

In Chapter 6, 30  $C\mathcal{P}$  sensitive variables in  $t\bar{t}H$  are defined across 4 different rest frames. While the distributions in the lab frame are already depicted in Section 6.1.3, the variables in the other rest frames are presented here.

### Higgs rest frame

Variables defined in the rest frame of the Higgs boson are shown in Fig. D.3.

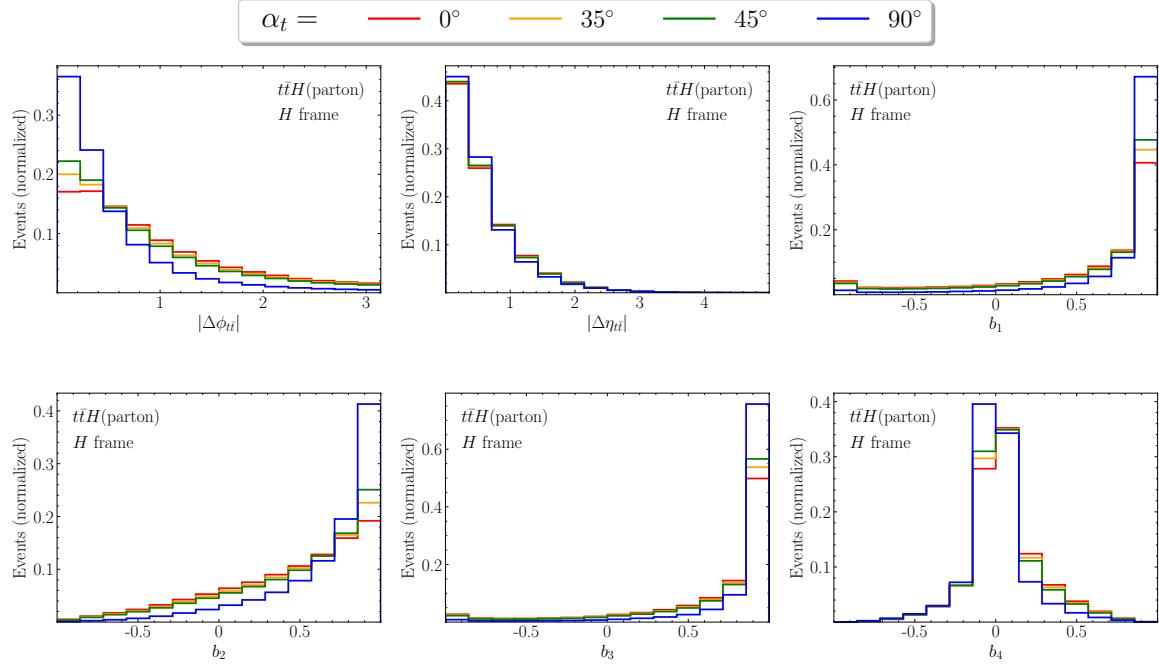


Figure D.3: Parton-level distributions for the variables defined in Section 6.1.2 in the Higgs rest frame at benchmark values of  $\alpha_t$ . Figure taken from [313].

**$t\bar{t}$  rest frame**

Variables defined in the rest frame of the  $t\bar{t}$  are shown in Fig. D.4.

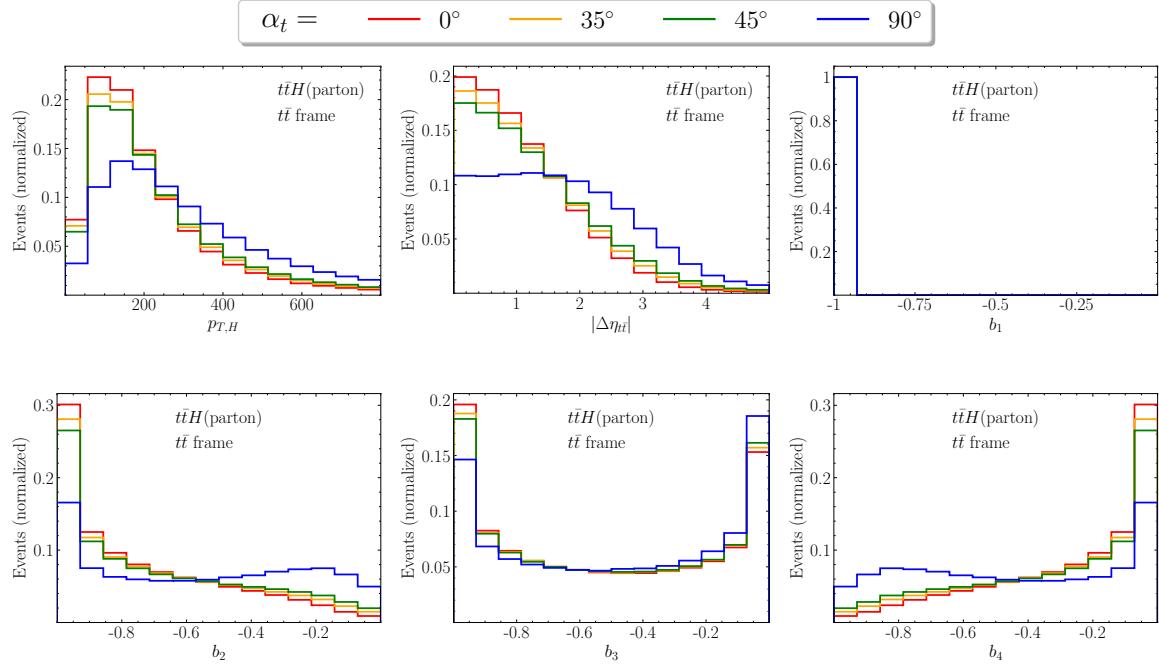


Figure D.4: Same as Fig. D.3 but in the  $t\bar{t}$  rest frame. Figure taken from [313].

**$t\bar{t}H$  rest frame**

Variables defined in the rest frame of the  $t\bar{t}H$  system are shown in Fig. D.5.

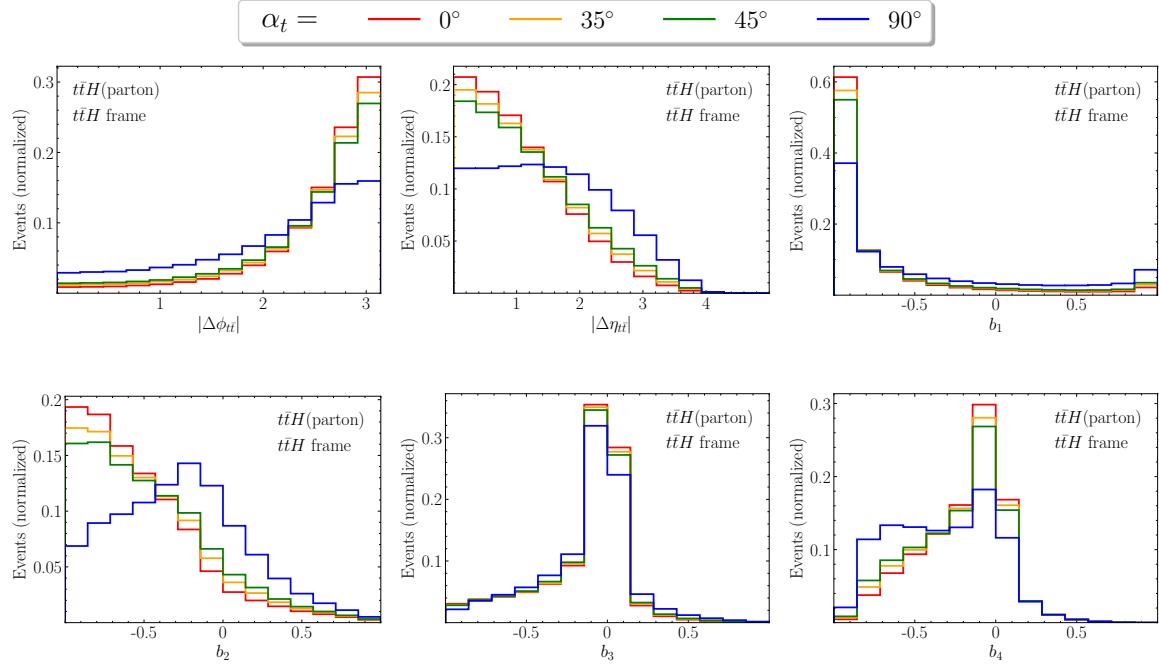


Figure D.5: Same as Fig. D.3 but in the  $t\bar{t}H$  rest frame. Figure taken from [313].

## APPENDIX E

### Training uncertainty

Here, the uncertainties in the training of the  $\mathcal{CP}$  classifiers from Chapter 5 are discussed. The distributions for the  $\mathcal{CP}$ -even and the  $\mathcal{CP}$ -odd classifiers in the ggF2j-SR are shown in Fig. E.1, while the distributions in the VBF-SR are shown in Fig. E.2. For all figures, the classifiers have been trained 100 times with the same input data and hyperparameters. The solid line and shaded region in a single bin correspond to the mean and standard deviation of all runs in this bin, respectively. In the ggF2j-SR, the largest uncertainties appear in the outermost bins of the  $P(c_g^2)$  variable for the  $|\mathcal{M}_{\text{odd}}|^2$  contribution. The best classifiers correspond to the ones with high (low) values in the bin around  $P(c_g^2) = 0$  (1). No uncertainty is visible in the distributions of the  $P_+ - P_-$  variable, indicating very stable training.

For the VBF-SR, the training uncertainties are much larger. Especially in the  $\mathcal{CP}$ -even case, every bin

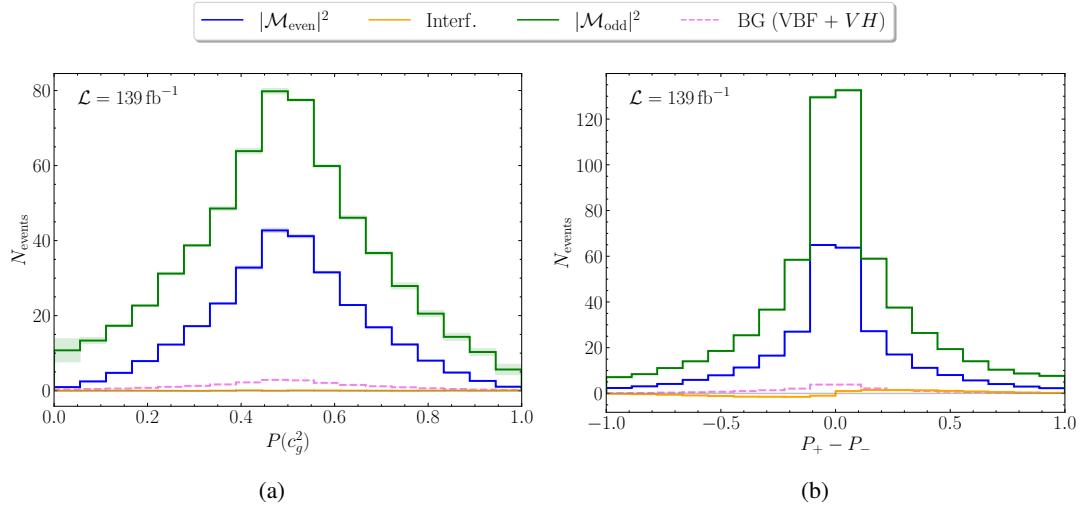


Figure E.1: Differential distributions of the individual contributions to the ggF2j cross section, as well as the background, after running through the (a)  $\mathcal{CP}$ -even classifier and (b)  $\mathcal{CP}$ -odd classifier in the ggF2j-SR. 100 classifiers are trained in total. The mean and standard deviation in each bin are shown as a solid line and dashed region, respectively.

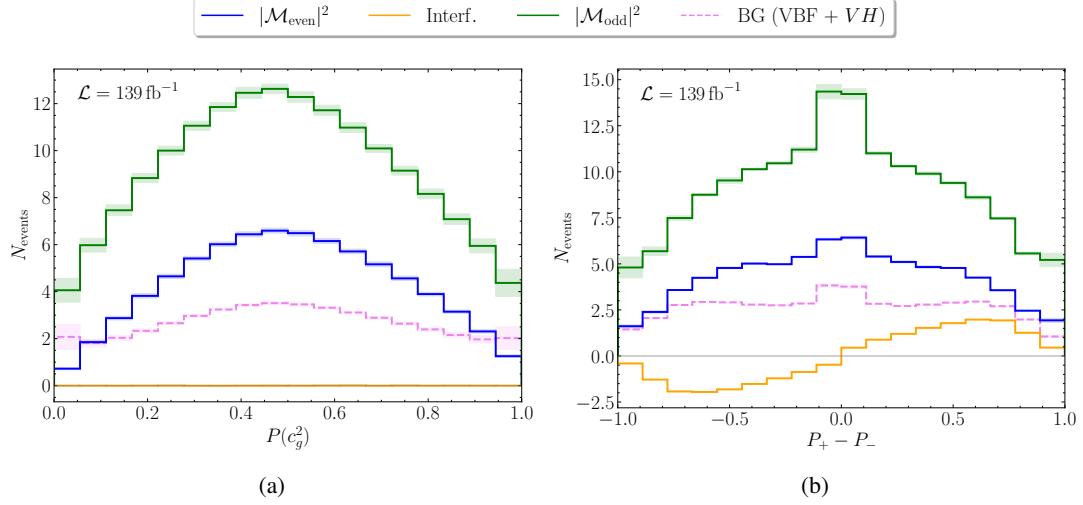


Figure E.2: Same as Fig. E.1 but in the VBF-SR.

of the  $|\mathcal{M}_{\text{odd}}|^2$  contribution has a visible standard deviation. Furthermore, the background distribution has some uncertainty due to the training in the outermost bins. For the  $\mathcal{CP}$ -odd classifier, the training is more stable as uncertainties are only visible in 4 bins. However, in contrast to the classifier in the ggF2j-SR, this still indicates some uncertainty in the training process. The reason for the VBF-SR showing larger uncertainties is the lower number of ggF2j events in this signal region. This comes from the fact that ggF2j events that were used for the training of the  $\mathcal{CP}$ -classifiers were separated into the two kinematic regions before the training. Generally, the  $\mathcal{CP}$ -even observable shows larger deviations across the runs than the  $\mathcal{CP}$ -odd one. This is due to the tiny differences in the kinematic distributions of the  $|\mathcal{M}_{\text{even}}|^2$  and  $|\mathcal{M}_{\text{odd}}|^2$  contributions. The performance of an individual classifier is then mainly decided by the amount of overtraining it exhibits.

The uncertainty of the classifier training process does not propagate to the exclusion limits for the Higgs-gluon coupling. While it can be used to assess difficulties in the learning process, every individual classifier corresponds to some observable after training. The best-performing classifier can therefore be chosen to calculate the final limits without taking the other classifiers into account.

---

## Significance tables

---

This appendix shows the significances for all 1-dimensional distributions and all 2-dimensional variable combinations considered in Chapter 6. The significances are obtained for testing the  $g_t = 1$  and  $\alpha_t = 35^\circ$  hypothesis against the SM case when assuming SM data, as detailed in Appendix C. The entries on the diagonal correspond to a 1-dimensional binning in the respective variable with 14 evenly distributed bins. The 2-dimensional combinations have  $6 \times 6$  evenly distributed bins. Permuting the variables has no influence on the results; therefore, each combination is depicted only once, leaving the lower left half of the matrix empty. Fig. F.1 shows the results in the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel, while Fig. F.2 and Fig. F.3 depict the  $t\bar{t}H(\text{multilep.})$  and the  $t\bar{t}H(\rightarrow b\bar{b})$  channel, respectively. The combination of all three channels is shown in Fig. F.4.

## Appendix F Significance tables

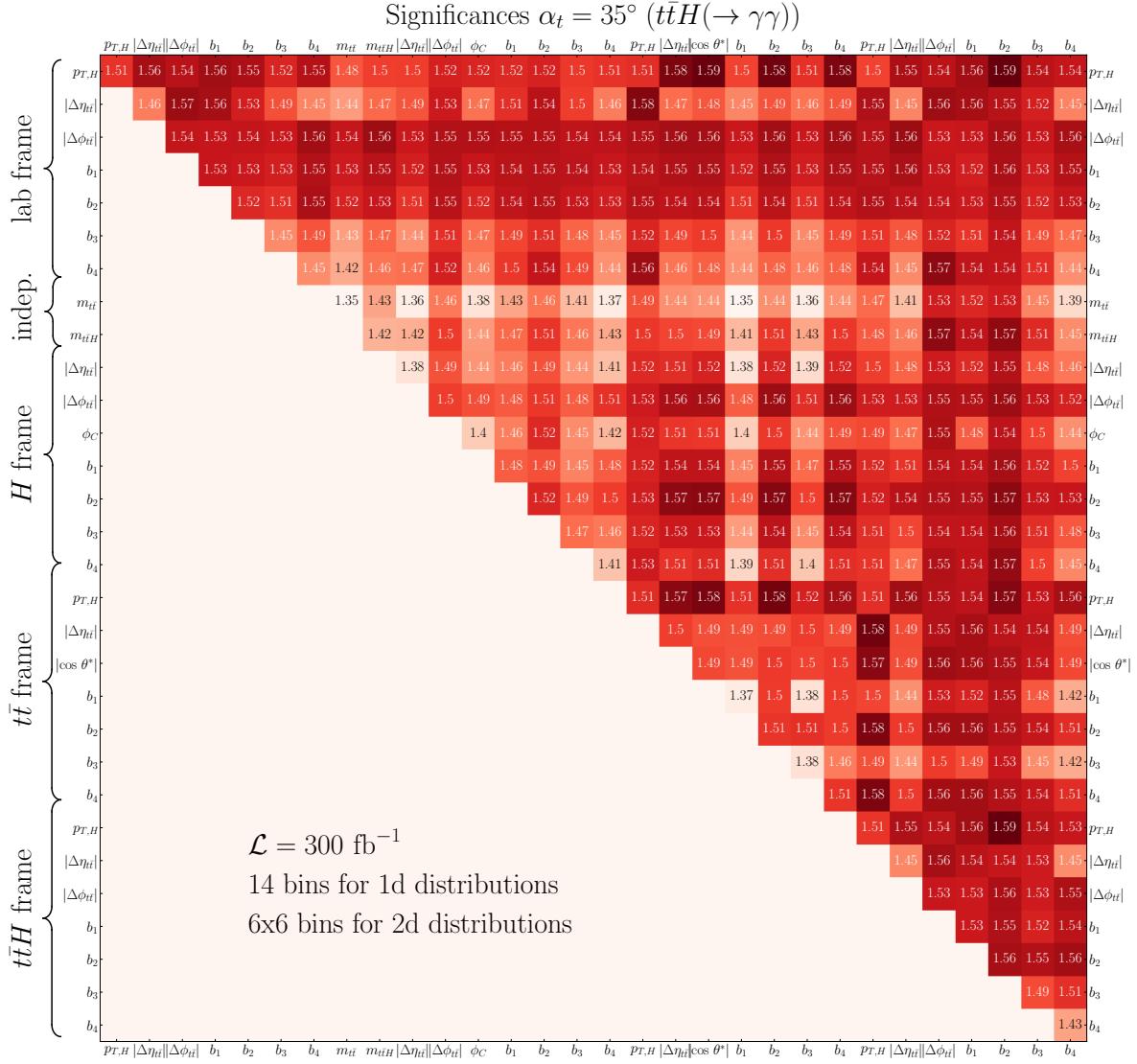


Figure F.1: Significances of all variable combinations for excluding the  $g_t = 1, \alpha_t = 35^\circ$  signal in the  $t\bar{t}H(\rightarrow \gamma\gamma)$  channel. Entries on the diagonal correspond to a 1-dimensional binning. Figure taken from [313].

## Appendix F Significance tables

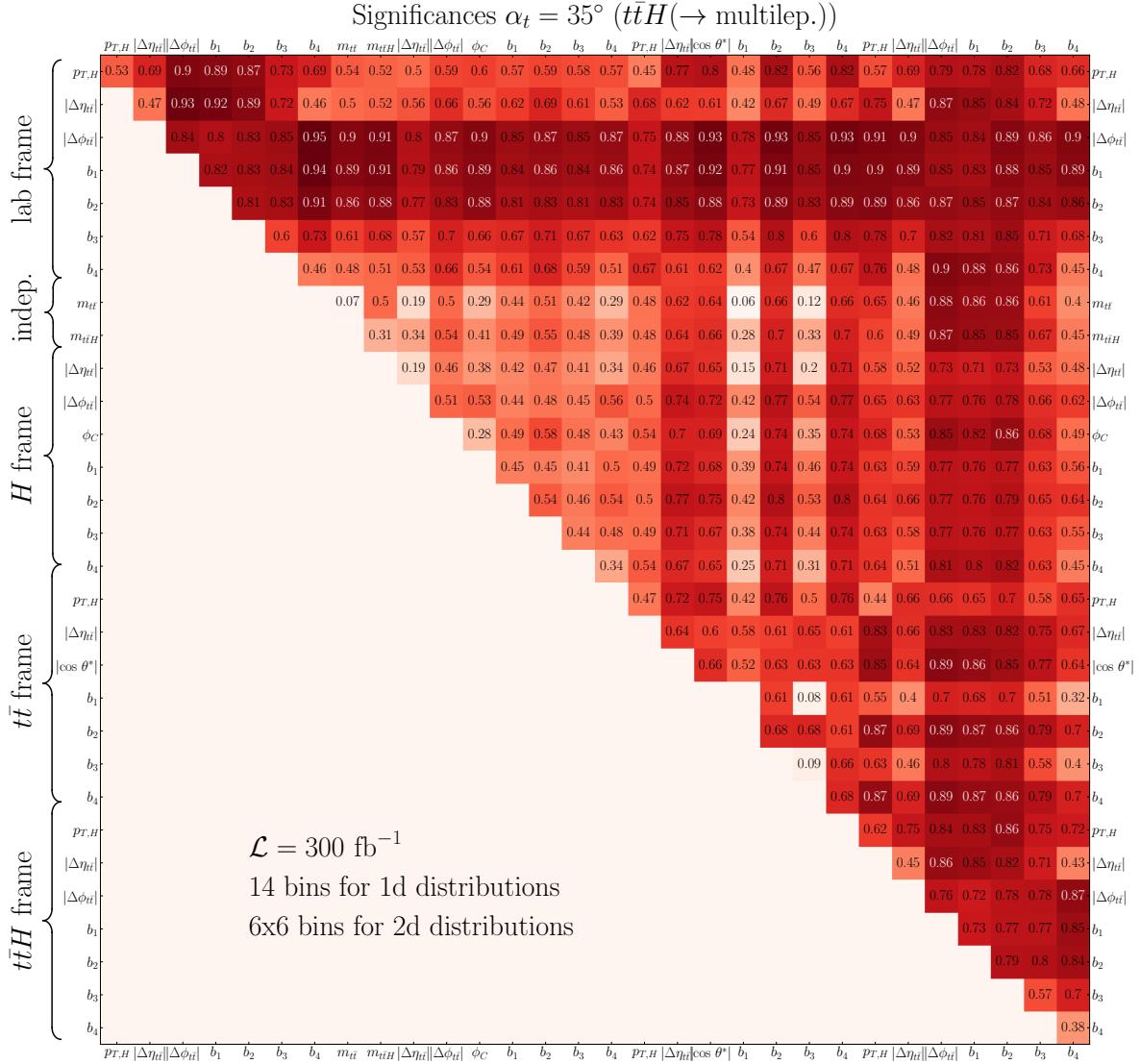


Figure F.2: Same as Fig. F.1 but in the  $t\bar{t}H(\text{multilep.})$  channel. Figure taken from [313].

## Appendix F Significance tables

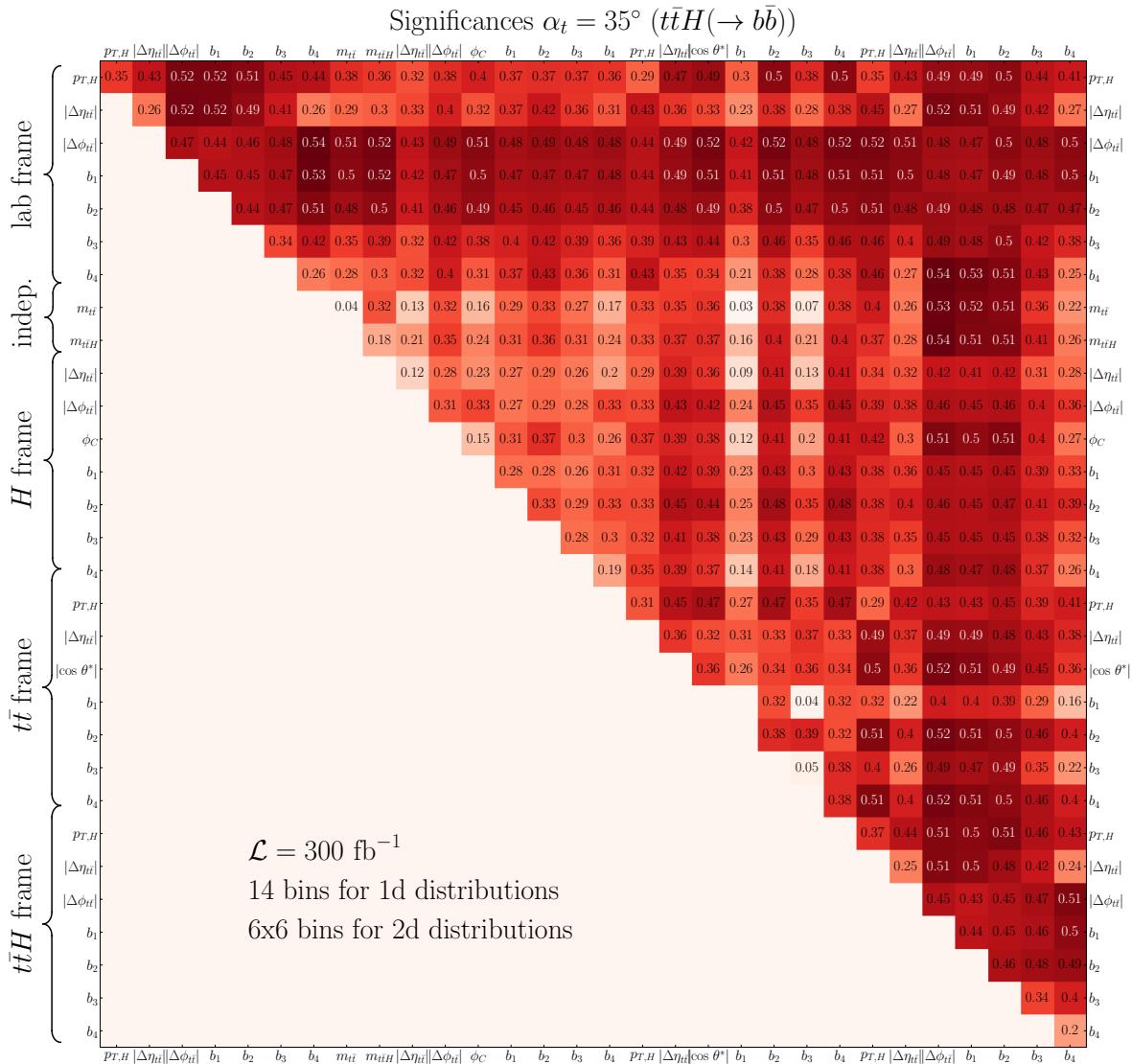


Figure F.3: Same as Fig. F.1 but in the  $t\bar{t}H(\rightarrow b\bar{b})$  channel. Figure taken from [313].

## Appendix F Significance tables

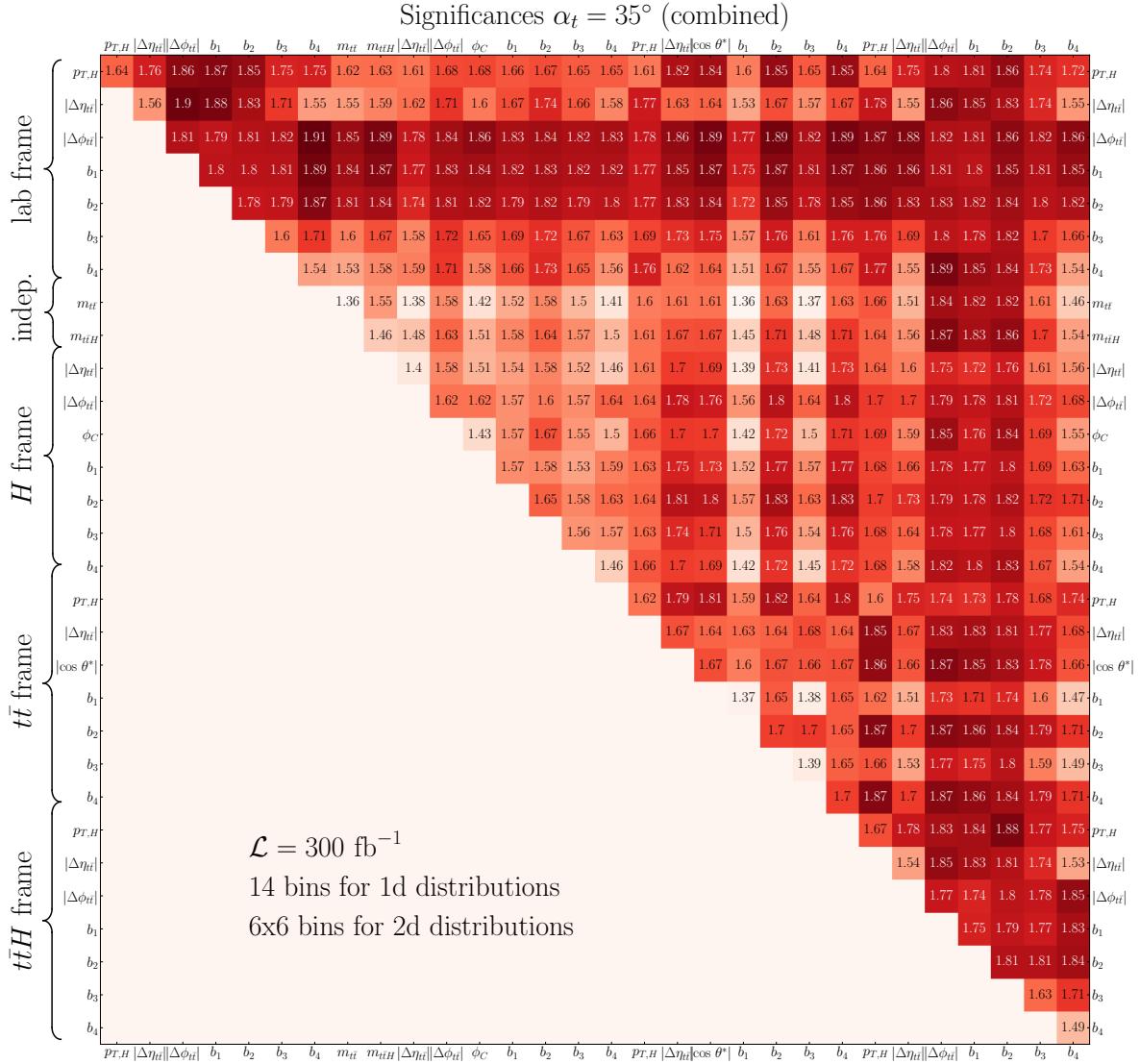


Figure F.4: Same as Fig. F.1 but for the combination of channels. Figure taken from [313].

## APPENDIX G

---

### Equations from PySR and SymbolNet

---

Here, the equations found by PySR and SymbolNet in Section 7.4 for scenarios 1 to 6 are shown. All equations are taken from [230].

#### Scenario 1:

PySR:

$$\cos \theta_{\text{PySR}}^* = \frac{p_{z,b} + p_{z,\bar{l}} + p_{z,\nu}}{\sqrt{\left(p_{x,b} + p_{x,\bar{l}} + p_{x,\nu}\right)^2 + \left(p_{y,b} + p_{y,\bar{l}} + p_{y,\nu}\right)^2 + \left(p_{z,b} + p_{z,\bar{l}} + p_{z,\nu}\right)^2}}$$

SymbolNet:

$$\cos \theta_{\text{SymbolNet}}^* = \frac{1.006p_{z,b} + 1.001p_{z,\bar{l}} + 1.002p_{z,\nu} - 1.027p_{z,\bar{b}} - 1.027p_{z,q} - 1.031p_{z,\bar{q}}}{\left\|1.034p_b + 1.022p_{\bar{l}} + 1.024p_{\nu} - p_{\bar{b}} - 1.007p_q - 1.009p_{\bar{q}}\right\|_3}$$

#### Scenario 2:

PySR:

$$\begin{aligned} \cos \theta_{\text{PySR}}^* = & \sin \left[ 1.41 \left( p_{z,b} + 1.41p_{z,\bar{l}} + 1.41p_{z,\nu} - 1.41p_{z,\bar{b}} - 1.41p_{z,q} - 1.41p_{z,\bar{q}} \right) \right] \\ & \left( E_b + E_{\bar{l}} + E_{\nu} + E_{\bar{b}} + E_q + E_{\bar{q}} \right. \\ & \left. + \sqrt{\left( -p_{y,b} - p_{y,\nu} + p_{y,\bar{b}} + p_{y,\bar{q}} \right)^2 + \left( p_{x,b} + p_{x,\bar{l}} + p_{x,\nu} - p_{x,\bar{b}} - p_{x,\bar{q}} \right)^2} - 1.70 \right] \end{aligned}$$

## SymbolNet:

$$\cos \theta_{\text{SymbolNet}}^* = 1.006 \left( \frac{\left. \text{boost} \left[ -2.793p_b - 2.811p_{\bar{l}} - 2.793p_{\nu} + 2.806p_{\bar{b}} + 2.835p_q + 2.815p_{\bar{q}} \right. \right.}{\left. \left. - 2.443p_b - 2.434p_{\bar{l}} - 2.456p_{\nu} - 2.45p_{\bar{b}} - 2.429p_q - 2.398p_{\bar{q}} \right] \right]_z}{\left\| \text{boost} \left[ -2.793p_b - 2.811p_{\bar{l}} - 2.793p_{\nu} + 2.806p_{\bar{b}} + 2.835p_q + 2.815p_{\bar{q}} \right. \right.} \right. \\ \left. \left. \left. - 2.443p_b - 2.434p_{\bar{l}} - 2.456p_{\nu} - 2.45p_{\bar{b}} - 2.429p_q - 2.398p_{\bar{q}} \right] \right\|_3}$$

### Scenario 3:

PySR:

$$\cos \theta_{\text{PySR}}^* = \sin \left[ \left( -0.300 \sqrt{E_T^{\text{miss}}} \left( p_{z,b} + 2.22 p_{z,\bar{b}} \right) \left( p_{y,\bar{l}} + p_{x,b} + p_{x,\bar{l}} - 1.99 \right) \right. \right. \\ \left. \left. + p_{z,b} + p_{z,\bar{l}} - p_{z,\bar{b}} - p_{z,q} - p_{z,\bar{q}} \right) \right] \\ \left( E_T^{\text{miss}} + 0.949 E_b + 0.949 E_{\bar{b}} + E_{\bar{l}} + E_q + 0.949 E_{\bar{q}} - 1.12 \right)$$

## SymbolNet:

$$\begin{aligned} \cos \theta_{\text{SymbolNet}}^* = & -1.079 \left( \text{boost} \left[ -4.831 p_b - 4.919 p_{\bar{l}} - 0.2614 p^{\text{miss}} \right. \right. \\ & + 4.752 p_{\bar{b}} + 4.827 p_q + 4.711 p_{\bar{q}} \left. \left| 4.506 p_b + 20.27 p_{\bar{l}} + 8.622 p_{\bar{b}} + 10.6 p_q + 8.935 p_{\bar{q}} \right. \right] \\ & - 0.698 \tanh \left( 0.711 p_b + 0.821 p_{\bar{l}} + 2.083 p^{\text{miss}} - 0.717 p_{\bar{b}} - 0.575 p_q - 0.672 p_{\bar{q}} \right) \Big|_z \Big) \\ & \left( \left( \left| \text{boost} \left[ -4.831 p_b - 4.919 p_{\bar{l}} - 0.2614 p^{\text{miss}} + 4.752 p_{\bar{b}} + 4.827 p_q + 4.711 p_{\bar{q}} \right. \right. \right. \right. \\ & \left. \left. \left. \left. 4.506 p_b + 20.27 p_{\bar{l}} + 8.622 p_{\bar{b}} + 10.6 p_q + 8.935 p_{\bar{q}} \right] \right\|_3 \right) \right. \\ & \left. + 1.039 \left\| \tanh \left( 0.711 p_b + 0.821 p_{\bar{l}} + 2.083 p^{\text{miss}} - 0.717 p_{\bar{b}} - 0.575 p_q - 0.672 p_{\bar{q}} \right) \right\|_3^2 + 0.089 \right) \end{aligned}$$

**Scenario 4:**

PySR:

$$\cos \theta_{\text{PySR}}^* = \sin \left[ \left( -p_{z,\bar{b}} - p_{z,q} - p_{z,\bar{q}} + 1.06 \sqrt{0.888 E_T^{\text{miss}} + 1} (p_{z,b} + 1.24 p_{z,\bar{l}}) \right) \right] / \\ \left( E_T^{\text{miss}} + 0.968 E_b + 0.826 E_{\bar{b}} + E_{\bar{l}} + p_{y,\bar{l}}^2 + E_q + E_{\bar{q}} \right. \\ \left. + 0.494 (-p_{x,b} - p_{x,\bar{l}} + p_{x,\bar{b}} + p_{x,\bar{q}}) - 1.48 \right)$$

**SymbolNet:**

$$\cos \theta_{\text{SymbolNet}}^* = 1.074 \left( \text{boost} \left[ -4.341 p_b - 4.71 p_{\bar{l}} - 0.2542 p^{\text{miss}} + 4.087 p_{\bar{b}} + 4.28 p_q + 4.133 p_{\bar{q}} \right] \right. \\ \left. - 4.103 p_b - 14.52 p_{\bar{l}} - 7.11 p_{\bar{b}} - 7.261 p_q - 6.894 p_{\bar{q}} - 0.327 \right] \\ + 0.668 \tanh \left( 0.672 p_b + 0.723 p_{\bar{l}} + 1.949 p^{\text{miss}} - 0.624 p_{\bar{b}} - 0.625 p_q - 0.62 p_{\bar{q}} \right) \Big|_z \\ \left( \left( \text{boost} \left[ -4.341 p_b - 4.71 p_{\bar{l}} - 0.2542 p^{\text{miss}} + 4.087 p_{\bar{b}} + 4.28 p_q + 4.133 p_{\bar{q}} \right] \right. \right. \\ \left. \left. - 4.103 p_b - 14.52 p_{\bar{l}} - 7.11 p_{\bar{b}} - 7.261 p_q - 6.894 p_{\bar{q}} - 0.327 \right] \right. \\ \left. + 0.162 \tanh \left( 0.672 p_b + 0.723 p_{\bar{l}} + 1.949 p^{\text{miss}} - 0.624 p_{\bar{b}} - 0.625 p_q - 0.62 p_{\bar{q}} \right) \right) \Big|_z \\ 0.906 \left( \left( \tanh \left( 0.672 p_b + 0.723 p_{\bar{l}} + 1.949 p^{\text{miss}} - 0.624 p_{\bar{b}} - 0.625 p_q - 0.62 p_{\bar{q}} \right) \right)^2 \Big|_z + 0.154 \right)$$

**Scenario 5:**

PySR:

$$\cos \theta_{\text{PySR}}^* = \sin \left[ \left( 1.156 p_{z,b} - \frac{1.156 p_{z,\bar{b}}}{E_T^{\text{miss}} (0.710 p_{x,\bar{b}} + p_{x,q} + p_{x,\bar{q}}) + 0.892} \right. \right. \\ \left. \left. + 2.06 p_{z,\bar{l}} - 1.156 p_{z,q} - 1.312 p_{z,\bar{q}} \right) \right] / \\ \left( E_b + E_{\bar{b}} - \sqrt{E_{\bar{l}}} + 2 E_{\bar{l}} + p_{x,\bar{l}}^2 + E_q + E_{\bar{q}} + p_{y,\bar{q}}^2 + (-p_{y,\bar{l}} + p_{y,q}) - 0.511 \right)$$

**SymbolNet:**

$$\cos \theta_{\text{SymbolNet}}^* = -0.928 \left( \text{boost} \left[ 1.673p_b + 2.866p_{\bar{l}} - 1.666p_{\bar{b}} - 2.495p_q - 2.294p_{\bar{q}} \right. \right. \\ \left. \left. - 3.822p_b - 5.112p_{\bar{l}} - 3.246p_{\bar{b}} - 3.755p_q - 3.459p_{\bar{q}} \right] \right. \\ \left. - 0.131 \tanh \left( 0.8275p_{\bar{l}} - 0.3036p_{\bar{q}} \right) \right]_z \Bigg/ \\ \left\| \text{boost} \left[ 1.673p_b + 2.866p_{\bar{l}} - 1.666p_{\bar{b}} - 2.495p_q - 2.294p_{\bar{q}} \right. \right. \\ \left. \left. - 3.822p_b - 5.112p_{\bar{l}} - 3.246p_{\bar{b}} - 3.755p_q - 3.459p_{\bar{q}} \right] \right\|_3$$

**Scenario 6:**

PySR:

$$\cos \theta_{\text{PySR}}^* = \sin \left[ \left( 1.114p_{z,b} + 2.143p_{z,\bar{l}} - 0.858p_{z,\bar{b}} - 0.426p_{z,q} - 1.088p_{z,\bar{q}} \right) \right. \\ \left. \left( E_T^{\text{miss}}E_q + E_b + E_{\bar{b}} + 1.85E_{\bar{l}} + E_{\bar{q}} \right. \right. \\ \left. \left. - \left( p_{x,b} + p_{x,\bar{l}} \right) \left( p_{x,\bar{b}} + p_{x,q} + p_{x,\bar{q}} \right) - 0.863 - \frac{0.205p_{z,q}}{\sqrt{E_q}} \right) \right]$$

**SymbolNet:**

$$\begin{aligned}
 \cos \theta_{\text{SymbolNet}}^* = & 0.971 \left( \text{boost} \left[ -1.521p_b - 3.066p_{\bar{l}} + 1.421p_{\bar{b}} + 2.218p_q + 2.043p_{\bar{q}} \right. \right. \\
 & \left. \left. - 2.845p_b - 4.559p_{\bar{l}} - 2.575p_{\bar{b}} - 2.753p_q - 2.676p_{\bar{q}} \right] \right. \\
 & \left. - 0.256 \tanh(0.646p_{\bar{b}} - 0.634p_{\bar{l}} + 0.648p_q) \right) \Big|_z \Bigg/ \\
 & \left( \left\| \text{boost} \left[ -1.521p_b - 3.066p_{\bar{l}} + 1.421p_{\bar{b}} + 2.218p_q + 2.043p_{\bar{q}} \right. \right. \right. \\
 & \left. \left. \left. - 2.845p_b - 4.559p_{\bar{l}} - 2.575p_{\bar{b}} - 2.753p_q - 2.676p_{\bar{q}} \right] \right\|_3 \right. \\
 & \left. + 0.774 \left( 0.071 \text{boost} \left[ -1.521p_b - 3.066p_{\bar{l}} + 1.421p_{\bar{b}} + 2.218p_q + 2.043p_{\bar{q}} \right. \right. \right. \\
 & \left. \left. \left. - 2.845p_b - 4.559p_{\bar{l}} - 2.575p_{\bar{b}} - 2.753p_q - 2.676p_{\bar{q}} \right] \right. \right. \\
 & \left. \left. + \tanh(0.646p_{\bar{b}} - 0.634p_{\bar{l}} + 0.648p_q) \times \right. \right. \\
 & \left. \left. \tanh(0.646p_{\bar{b}} - 0.634p_{\bar{l}} + 0.648p_q) \right) \right)_3 - 0.223 \right)
 \end{aligned}$$

---

## Bibliography

---

- [1] G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214) (cit. on pp. 1, 23, 32).
- [2] S. Chatrchyan et al.,  
*Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*,  
*Phys. Lett. B* **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235) (cit. on pp. 1, 23, 32).
- [3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble,  
*Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (20 1964) 585,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on pp. 1, 11).
- [4] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132  
(cit. on pp. 1, 3, 11).
- [5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,  
*Phys. Rev. Lett.* **13** (16 1964) 508,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on pp. 1, 3, 11).
- [6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,  
*Phys. Rev. Lett.* **13** (9 1964) 321,  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on pp. 1, 3, 11).
- [7] A. Keshavarzi, K. S. Khaw, and T. Yoshioka, *Muon g-2: A review*,  
*Nucl. Phys. B* **975** (2022) 115675, arXiv: [2106.06723 \[hep-ex\]](https://arxiv.org/abs/2106.06723) (cit. on p. 1).
- [8] B. Capdevila, A. Crivellin, and J. Matias, *Review of semileptonic B anomalies*,  
*Eur. Phys. J. ST* **1** (2023) 20, arXiv: [2309.01311 \[hep-ph\]](https://arxiv.org/abs/2309.01311) (cit. on p. 1).
- [9] T. Aaltonen et al., *High-precision measurement of the W boson mass with the CDF II detector*,  
*Science* **376** (2022) 170 (cit. on pp. 1, 18).
- [10] A. A. Aguilar-Arevalo et al., *Updated MiniBooNE neutrino oscillation results with increased data and new background studies*, *Phys. Rev. D* **103** (2021) 052002,  
arXiv: [2006.16883 \[hep-ex\]](https://arxiv.org/abs/2006.16883) (cit. on p. 1).
- [11] N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*,  
*Astron. Astrophys.* **641** (2020) A6, [Erratum: *Astron. Astrophys.* 652, C4 (2021)],  
arXiv: [1807.06209 \[astro-ph.CO\]](https://arxiv.org/abs/1807.06209) (cit. on pp. 1, 17).
- [12] S. Navas et al., *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001  
(cit. on pp. 1, 15, 17, 29–31, 130).

- [13] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32 (cit. on pp. 1, 17).
- [14] M. B. Gavela, P. Hernandez, J. Orloff, and O. Pene, *Standard model CP violation and baryon asymmetry*, *Mod. Phys. Lett. A* **9** (1994) 795, arXiv: [hep-ph/9312215](#) (cit. on pp. 1, 18).
- [15] P. Huet and E. Sather, *Electroweak baryogenesis and standard model CP violation*, *Phys. Rev. D* **51** (1995) 379, arXiv: [hep-ph/9404302](#) (cit. on pp. 1, 18).
- [16] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, *On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe*, *Phys. Lett. B* **155** (1985) 36 (cit. on pp. 1, 18).
- [17] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, *Progress in electroweak baryogenesis*, *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 27, arXiv: [hep-ph/9302210](#) (cit. on pp. 1, 18).
- [18] P. Huet and A. E. Nelson, *CP violation and electroweak baryogenesis in extensions of the standard model*, *Phys. Lett. B* **355** (1995) 229, arXiv: [hep-ph/9504427](#) (cit. on p. 1).
- [19] A. Menon, D. E. Morrissey, and C. E. M. Wagner, *Electroweak baryogenesis and dark matter in the nMSSM*, *Phys. Rev. D* **70** (2004) 035005, arXiv: [hep-ph/0404184](#) (cit. on p. 1).
- [20] M. Carena, G. Nardini, M. Quiros, and C. E. M. Wagner, *The Baryogenesis Window in the MSSM*, *Nucl. Phys. B* **812** (2009) 243, arXiv: [0809.3760](#) [hep-ph] (cit. on p. 1).
- [21] K. Blum, C. Delaunay, M. Losada, Y. Nir, and S. Tulin, *CP violation Beyond the MSSM: Baryogenesis and Electric Dipole Moments*, *JHEP* **05** (2010) 101, arXiv: [1003.2447](#) [hep-ph] (cit. on p. 1).
- [22] J. Shu and Y. Zhang, *Impact of a CP Violating Higgs Sector: From LHC to Baryogenesis*, *Phys. Rev. Lett.* **111** (2013) 091801, arXiv: [1304.0773](#) [hep-ph] (cit. on p. 1).
- [23] I. Baldes, T. Konstandin, and G. Servant, *A first-order electroweak phase transition from varying Yukawas*, *Phys. Lett. B* **786** (2018) 373, arXiv: [1604.04526](#) [hep-ph] (cit. on p. 1).
- [24] P. Basler, M. Mühlleitner, and J. Wittbrodt, *The CP-Violating 2HDM in Light of a Strong First Order Electroweak Phase Transition and Implications for Higgs Pair Production*, *JHEP* **03** (2018) 061, arXiv: [1711.04097](#) [hep-ph] (cit. on p. 1).
- [25] E. Fuchs, M. Losada, Y. Nir, and Y. Viernik, *CP violation from  $\tau$ ,  $t$  and  $b$  dimension-6 Yukawa couplings - interplay of baryogenesis, EDM and Higgs physics*, *JHEP* **05** (2020) 056, arXiv: [2003.00099](#) [hep-ph] (cit. on pp. 1, 2, 20, 76).
- [26] S. Aharony Shapira, *Current bounds on baryogenesis from complex Yukawa couplings of light fermions*, *Phys. Rev. D* **105** (2022) 095037, arXiv: [2106.05338](#) [hep-ph] (cit. on pp. 1, 20).

- [27] H. Bahl et al., *Constraining the  $\mathcal{CP}$  structure of Higgs-fermion couplings with a global LHC fit, the electron EDM and baryogenesis*, *Eur. Phys. J. C* **82** (2022) 604, arXiv: [2202.11753 \[hep-ph\]](#) (cit. on pp. 1, 2, 16, 35, 54, 74, 76, 184).
- [28] Anisha, D. Azevedo, L. Biermann, C. Englert, and M. Mühlleitner, *Effective 2HDM Yukawa interactions and a strong first-order electroweak phase transition*, *JHEP* **02** (2024) 045, arXiv: [2311.06353 \[hep-ph\]](#) (cit. on p. 1).
- [29] E. Camargo-Molina, R. Enberg, and J. Löfgren, *A Catalog of First-Order Electroweak Phase Transitions in the Standard Model Effective Field Theory*, (2024), arXiv: [2410.23210 \[hep-ph\]](#) (cit. on p. 1).
- [30] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, arXiv: [1008.4884 \[hep-ph\]](#) (cit. on pp. 1, 19).
- [31] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1, arXiv: [1106.0034 \[hep-ph\]](#) (cit. on p. 1).
- [32] S. F. King, S. Moretti, and R. Nevzorov, *A Review of the Exceptional Supersymmetric Standard Model*, *Symmetry* **12** (2020) 557, arXiv: [2002.02788 \[hep-ph\]](#) (cit. on p. 1).
- [33] V. Khachatryan et al., *Constraints on the spin-parity and anomalous  $HVV$  couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, *Phys. Rev. D* **92** (2015) 012004, arXiv: [1411.3441 \[hep-ex\]](#) (cit. on pp. 2, 34).
- [34] V. M. Abazov et al., *Constraints on Models for the Higgs Boson with Exotic Spin and Parity in  $VH \rightarrow Vb\bar{b}$  Final States*, *Phys. Rev. Lett.* **113** (2014) 161802, arXiv: [1407.6369 \[hep-ex\]](#) (cit. on pp. 2, 34).
- [35] G. Aad et al., *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 476, [Erratum: *Eur.Phys.J.C* 76, 152 (2016)], arXiv: [1506.05669 \[hep-ex\]](#) (cit. on pp. 2, 34).
- [36] A. M. Sirunyan et al., *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state*, *Phys. Rev. D* **104** (2021) 052004, arXiv: [2104.12152 \[hep-ex\]](#) (cit. on pp. 2, 34, 77, 83).
- [37] G. Aad et al., *Test of  $CP$ -invariance of the Higgs boson in vector-boson fusion production and in its decay into four leptons*, *JHEP* **05** (2024) 105, arXiv: [2304.09612 \[hep-ex\]](#) (cit. on pp. 2, 34).
- [38] A. Hayrapetyan et al., *Constraints on anomalous Higgs boson couplings from its production and decay using the  $WW$  channel in proton–proton collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **84** (2024) 779, arXiv: [2403.00657 \[hep-ex\]](#) (cit. on pp. 2, 34).
- [39] G. Aad et al., *Measurements of Higgs boson production via gluon-gluon fusion and vector-boson fusion using  $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$  decays in  $pp$  collisions with the ATLAS detector and their effective field theory interpretations*, (2025), arXiv: [2504.07686 \[hep-ex\]](#) (cit. on pp. 2, 34, 36).

- [40] T. S. Roussy et al., *An improved bound on the electron's electric dipole moment*, *Science* **381** (2023) adg4084, arXiv: [2212.11841 \[physics.atom-ph\]](https://arxiv.org/abs/2212.11841) (cit. on pp. 2, 16).
- [41] C. Abel et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*, *Phys. Rev. Lett.* **124** (2020) 081803, arXiv: [2001.11966 \[hep-ex\]](https://arxiv.org/abs/2001.11966) (cit. on pp. 2, 15).
- [42] J. Brod, U. Haisch, and J. Zupan, *Constraints on CP-violating Higgs couplings to the third generation*, *JHEP* **11** (2013) 180, arXiv: [1310.1385 \[hep-ph\]](https://arxiv.org/abs/1310.1385) (cit. on pp. 2, 76).
- [43] J. Brod, J. M. Cornell, D. Skodras, and E. Stamou, *Global constraints on Yukawa operators in the standard model effective theory*, *JHEP* **08** (2022) 294, arXiv: [2203.03736 \[hep-ph\]](https://arxiv.org/abs/2203.03736) (cit. on pp. 2, 54, 76).
- [44] V. Khachatryan et al., *Search for a standard model-like Higgs boson in the  $\mu^+\mu^-$  and  $e^+e^-$  decay channels at the LHC*, *Phys. Lett. B* **744** (2015) 184, arXiv: [1410.6679 \[hep-ex\]](https://arxiv.org/abs/1410.6679) (cit. on pp. 2, 29, 76).
- [45] W. Altmannshofer, J. Brod, and M. Schmaltz, *Experimental constraints on the coupling of the Higgs boson to electrons*, *JHEP* **05** (2015) 125, arXiv: [1503.04830 \[hep-ph\]](https://arxiv.org/abs/1503.04830) (cit. on pp. 2, 29, 76).
- [46] Y. Zhou, *Constraining the Higgs boson coupling to light quarks in the  $H \rightarrow ZZ$  final states*, *Phys. Rev. D* **93** (2016) 013019, arXiv: [1505.06369 \[hep-ph\]](https://arxiv.org/abs/1505.06369) (cit. on pp. 2, 29, 76).
- [47] Y. Soreq, H. X. Zhu, and J. Zupan, *Light quark Yukawa couplings from Higgs kinematics*, *JHEP* **12** (2016) 045, arXiv: [1606.09621 \[hep-ph\]](https://arxiv.org/abs/1606.09621) (cit. on pp. 2, 29, 76).
- [48] G. Bonner and H. E. Logan, *Constraining the Higgs couplings to up and down quarks using production kinematics at the CERN Large Hadron Collider*, (2016), arXiv: [1608.04376 \[hep-ph\]](https://arxiv.org/abs/1608.04376) (cit. on pp. 2, 29, 76).
- [49] F. Yu, *Phenomenology of Enhanced Light Quark Yukawa Couplings and the  $W^\pm h$  Charge Asymmetry*, *JHEP* **02** (2017) 083, arXiv: [1609.06592 \[hep-ph\]](https://arxiv.org/abs/1609.06592) (cit. on pp. 2, 29, 76).
- [50] L. Alasfar, R. Corral Lopez, and R. Gröber, *Probing Higgs couplings to light quarks via Higgs pair production*, *JHEP* **11** (2019) 088, arXiv: [1909.05279 \[hep-ph\]](https://arxiv.org/abs/1909.05279) (cit. on pp. 2, 29, 76).
- [51] G. Aad et al., *Search for the Higgs boson decays  $H \rightarrow ee$  and  $H \rightarrow e\mu$  in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **801** (2020) 135148, arXiv: [1909.10235 \[hep-ex\]](https://arxiv.org/abs/1909.10235) (cit. on pp. 2, 29, 76).
- [52] J. A. Aguilar-Saavedra, J. M. Cano, and J. M. No, *More light on Higgs flavor at the LHC: Higgs boson couplings to light quarks through  $h + \gamma$  production*, *Phys. Rev. D* **103** (2021) 095023, arXiv: [2008.12538 \[hep-ph\]](https://arxiv.org/abs/2008.12538) (cit. on pp. 2, 29, 76).
- [53] A. Falkowski et al., *Light quark Yukawas in triboson final states*, *JHEP* **04** (2021) 023, arXiv: [2011.09551 \[hep-ph\]](https://arxiv.org/abs/2011.09551) (cit. on pp. 2, 29, 76).
- [54] N. Vignaroli, *Off-Shell Probes of the Higgs Yukawa Couplings: Light Quarks and Charm*, *Symmetry* **14** (2022) 1183, arXiv: [2205.09449 \[hep-ph\]](https://arxiv.org/abs/2205.09449) (cit. on pp. 2, 29, 76).

[55] E. Balzani, R. Gröber, and M. Vitti, *Light-quark Yukawa couplings from off-shell Higgs production*, *JHEP* **10** (2023) 027, arXiv: [2304.09772 \[hep-ph\]](https://arxiv.org/abs/2304.09772) (cit. on pp. 2, 29, 76).

[56] F. Bishara et al., *Probing CP Violation in  $h \rightarrow \gamma\gamma$  with Converted Photons*, *JHEP* **04** (2014) 084, arXiv: [1312.2955 \[hep-ph\]](https://arxiv.org/abs/1312.2955) (cit. on pp. 2, 36).

[57] M. Galanti et al., *Heavy baryons as polarimeters at colliders*, *JHEP* **11** (2015) 067, arXiv: [1505.02771 \[hep-ph\]](https://arxiv.org/abs/1505.02771) (cit. on pp. 2, 35).

[58] R. Alonso, C. Fraser-Taliente, C. Hays, and M. Spannowsky, *Prospects for direct CP tests of  $hqq$  interactions*, *JHEP* **08** (2021) 167, arXiv: [2105.06879 \[hep-ph\]](https://arxiv.org/abs/2105.06879) (cit. on pp. 2, 35).

[59] S. Amoroso et al., “Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report,” *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 2020, arXiv: [2003.01700 \[hep-ph\]](https://arxiv.org/abs/2003.01700) (cit. on pp. 2, 37).

[60] M. Bonanomi, J. Jovicevic, and F. Tackmann, *Recommended binning Stage 1.2*, [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGfiducialAndSTXS#Recommended\\_binning\\_Stage\\_1\\_2](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGfiducialAndSTXS#Recommended_binning_Stage_1_2), Accessed: 24/04/2025, 2024 (cit. on pp. 2, 37).

[61] J. J. Thomson, *Cathode Rays*, *Philosophical Magazine* **44** (1897) 293 (cit. on p. 3).

[62] M. Planck, *Ueber irreversible Strahlungsvorgänge*, *Annalen der Physik* **306** (1900) 69, eprint: <https://onlinelibrary.wiley.com/doi/10.1002/andp.19003060105> (cit. on p. 3).

[63] A. Einstein, *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*, *Annalen der Physik* **322** (1905) 132 (cit. on p. 3).

[64] P. A. M. Dirac, *The quantum theory of the emission and absorption of radiation*, *Proc. R. Soc. Lond. A* **114** (1927) 243 (cit. on p. 3).

[65] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, *Phys. Rev.* **96** (1 1954) 191, URL: <https://link.aps.org/doi/10.1103/PhysRev.96.191> (cit. on p. 3).

[66] S. L. Glashow, *The renormalizability of vector meson interactions*, *Nuclear Physics* **10** (1959) 107, ISSN: 0029-5582, URL: <https://www.sciencedirect.com/science/article/pii/0029558259901968> (cit. on p. 3).

[67] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, *Nuovo Cim.* **11** (1959) 568 (cit. on p. 3).

[68] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (21 1967) 1264, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264> (cit. on p. 3).

[69] G. Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*, *Nuclear Physics B* **35** (1971) 167, ISSN: 0550-3213, URL: <https://www.sciencedirect.com/science/article/pii/0550321371901398> (cit. on p. 3).

- [70] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, (1961) (cit. on pp. 4, 10).
- [71] Y. Ne’eman, *Derivation of strong interactions from a gauge invariance*, *Nucl. Phys.* **26** (1961) 222, ed. by R. Ruffini and Y. Verbin (cit. on pp. 4, 10).
- [72] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. I*, *Phys. Rev. D* **8** (10 1973) 3633, URL: <https://link.aps.org/doi/10.1103/PhysRevD.8.3633> (cit. on p. 4).
- [73] H. David Politzer, *Asymptotic freedom: An approach to strong interactions*, *Physics Reports* **14** (1974) 129, ISSN: 0370-1573, URL: <https://www.sciencedirect.com/science/article/pii/0370157374900143> (cit. on p. 4).
- [74] J. C. Baez and J. Huerta, *The Algebra of Grand Unified Theories*, 2010, arXiv: [0904.1556 \[hep-th\]](https://arxiv.org/abs/0904.1556), URL: <https://arxiv.org/abs/0904.1556> (cit. on p. 4).
- [75] D. Galbraith and C. Burgard, <https://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/>, Accessed: 2023-12-06 (cit. on p. 4).
- [76] E. Fermi, *An attempt of a theory of beta radiation. I.*, *Z. Phys.* **88** (1934) 161 (cit. on p. 7).
- [77] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, *Phys. Rev.* **105** (4 1957) 1413, URL: <https://link.aps.org/doi/10.1103/PhysRev.105.1413> (cit. on p. 7).
- [78] R. P. Feynman and M. Gell-Mann, *Theory of the Fermi Interaction*, *Phys. Rev.* **109** (1 1958) 193, URL: <https://link.aps.org/doi/10.1103/PhysRev.109.193> (cit. on p. 7).
- [79] E. C. G. Sudarshan and R. E. Marshak, *Chirality Invariance and the Universal Fermi Interaction*, *Phys. Rev.* **109** (5 1958) 1860, URL: <https://link.aps.org/doi/10.1103/PhysRev.109.1860.2> (cit. on p. 7).
- [80] J. J. Sakurai, *MASS REVERSAL AND WEAK INTERACTIONS*, *Nuovo Cim.* **7** (1958) 649 (cit. on p. 7).
- [81] B. W. Lee, C. Quigg, and H. B. Thacker, *Weak interactions at very high energies: The role of the Higgs-boson mass*, *Phys. Rev. D* **16** (5 1977) 1519, URL: <https://link.aps.org/doi/10.1103/PhysRevD.16.1519> (cit. on p. 9).
- [82] H. Yukawa, *On the Interaction of Elementary Particles I*, *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48 (cit. on p. 10).
- [83] M. Breidenbach et al., *Observed behavior of highly inelastic electron-proton scattering*, *Phys. Rev. Lett.* **23** (1969) 935 (cit. on p. 10).
- [84] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, *Phys. Rev.* **117** (3 1960) 648, URL: <https://link.aps.org/doi/10.1103/PhysRev.117.648> (cit. on p. 11).
- [85] J. Goldstone, *Field Theories with Superconductor Solutions*, *Nuovo Cim.* **19** (1961) 154 (cit. on p. 11).

- [86] J. Goldstone, A. Salam, and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (1962) 965 (cit. on p. 11).
- [87] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (12 1963) 531, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531> (cit. on p. 15).
- [88] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652 (cit. on p. 15).
- [89] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson*, Phys. Rev. Lett. **13** (1964) 138 (cit. on p. 15).
- [90] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, Zh. Eksp. Teor. Fiz. **34** (1957) 247 (cit. on p. 15).
- [91] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, Prog. Theor. Phys. **28** (1962) 870 (cit. on p. 15).
- [92] Y. Fukuda et al., *Measurement of the flux and zenith angle distribution of upward through going muons by Super-Kamiokande*, Phys. Rev. Lett. **82** (1999) 2644, arXiv: [hep-ex/9812014](https://arxiv.org/abs/hep-ex/9812014) (cit. on p. 15).
- [93] Y. Yamaguchi and N. Yamanaka, *Large long-distance contributions to the electric dipole moments of charged leptons in the standard model*, Phys. Rev. Lett. **125** (2020) 241802, arXiv: [2003.08195 \[hep-ph\]](https://arxiv.org/abs/2003.08195) (cit. on p. 16).
- [94] Y. Ema, T. Gao, and M. Pospelov, *Standard Model Prediction for Paramagnetic Electric Dipole Moments*, Phys. Rev. Lett. **129** (2022) 231801, arXiv: [2202.10524 \[hep-ph\]](https://arxiv.org/abs/2202.10524) (cit. on p. 16).
- [95] J. P. Archambault, A. Czarnecki, and M. Pospelov, *Electric dipole moments of leptons in the presence of majorana neutrinos*, Phys. Rev. D **70** (2004) 073006, arXiv: [hep-ph/0406089](https://arxiv.org/abs/hep-ph/0406089) (cit. on p. 16).
- [96] N. Craig, *Naturalness: past, present, and future*, Eur. Phys. J. C **83** (2023) 825, arXiv: [2205.05708 \[hep-ph\]](https://arxiv.org/abs/2205.05708) (cit. on p. 16).
- [97] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helvetica Physica Acta **6** (1933) 110 (cit. on p. 17).
- [98] V. C. Rubin and W. K. Ford Jr., *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, ApJ **159** (1970) 379 (cit. on p. 17).
- [99] Y. Akrami et al., *Planck 2018 results. VII. Isotropy and Statistics of the CMB*, Astron. Astrophys. **641** (2020) A7, arXiv: [1906.02552 \[astro-ph.CO\]](https://arxiv.org/abs/1906.02552) (cit. on p. 17).
- [100] F. Kahlhoefer, *Review of LHC Dark Matter Searches*, Int. J. Mod. Phys. A **32** (2017) 1730006, arXiv: [1702.02430 \[hep-ph\]](https://arxiv.org/abs/1702.02430) (cit. on p. 17).
- [101] A. Arbey and F. Mahmoudi, *Dark matter and the early Universe: a review*, Prog. Part. Nucl. Phys. **119** (2021) 103865, arXiv: [2104.11488 \[hep-ph\]](https://arxiv.org/abs/2104.11488) (cit. on p. 17).

[102] A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, *Annalen der Physik* **354** (1916) 769, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702> (cit. on p. 17).

[103] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, *Large  $N$  field theories, string theory and gravity*, *Phys. Rept.* **323** (2000) 183, arXiv: [hep-th/9905111](https://arxiv.org/abs/hep-th/9905111) (cit. on p. 17).

[104] F. Marchesano, G. Shiu, and T. Weigand, *The Standard Model from String Theory: What Have We Learned?* *Ann. Rev. Nucl. Part. Sci.* **74** (2024) 113, arXiv: [2401.01939 \[hep-th\]](https://arxiv.org/abs/2401.01939) (cit. on p. 17).

[105] N. S. Manton, *Topology in the Weinberg-Salam theory*, *Phys. Rev. D* **28** (8 1983) 2019, URL: <https://link.aps.org/doi/10.1103/PhysRevD.28.2019> (cit. on p. 17).

[106] F. R. Klinkhamer and N. S. Manton, *A saddle-point solution in the Weinberg-Salam theory*, *Phys. Rev. D* **30** (10 1984) 2212, URL: <https://link.aps.org/doi/10.1103/PhysRevD.30.2212> (cit. on p. 17).

[107] K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, *The Electroweak phase transition: A Nonperturbative analysis*, *Nucl. Phys. B* **466** (1996) 189, arXiv: [hep-lat/9510020](https://arxiv.org/abs/hep-lat/9510020) (cit. on p. 18).

[108] V. A. Rubakov and M. E. Shaposhnikov, *Electroweak baryon number nonconservation in the early universe and in high-energy collisions*, *Usp. Fiz. Nauk* **166** (1996) 493, arXiv: [hep-ph/9603208](https://arxiv.org/abs/hep-ph/9603208) (cit. on p. 18).

[109] G. Elor, R. Houtz, S. Ipek, and M. Ulloa, *The Standard Model CP Violation is Enough*, (2024), arXiv: [2408.12647 \[hep-ph\]](https://arxiv.org/abs/2408.12647) (cit. on p. 18).

[110] J. De Vries, M. Postma, and J. van de Vis, *The role of leptons in electroweak baryogenesis*, *JHEP* **04** (2019) 024, arXiv: [1811.11104 \[hep-ph\]](https://arxiv.org/abs/1811.11104) (cit. on p. 18).

[111] J. M. Cline and K. Kainulainen, *Electroweak baryogenesis at high bubble wall velocities*, *Phys. Rev. D* **101** (2020) 063525, arXiv: [2001.00568 \[hep-ph\]](https://arxiv.org/abs/2001.00568) (cit. on p. 18).

[112] J. M. Cline and B. Laurent, *Electroweak baryogenesis from light fermion sources: A critical study*, *Phys. Rev. D* **104** (2021) 083507, arXiv: [2108.04249 \[hep-ph\]](https://arxiv.org/abs/2108.04249) (cit. on p. 18).

[113] G. C. Dorsch, S. J. Huber, and T. Konstandin, *On the wall velocity dependence of electroweak baryogenesis*, *JCAP* **08** (2021) 020, arXiv: [2106.06547 \[hep-ph\]](https://arxiv.org/abs/2106.06547) (cit. on p. 18).

[114] D. E. Morrissey and M. J. Ramsey-Musolf, *Electroweak baryogenesis*, *New J. Phys.* **14** (2012) 125003, arXiv: [1206.2942 \[hep-ph\]](https://arxiv.org/abs/1206.2942) (cit. on p. 18).

[115] A. Hayrapetyan et al., *Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at  $s=13\text{TeV}$* , *Phys. Lett. B* **860** (2024) 139067, arXiv: [2405.18149 \[hep-ex\]](https://arxiv.org/abs/2405.18149) (cit. on p. 18).

[116] G. Aad et al., *Search for diphoton resonances in the 66 to 110 GeV mass range using  $pp$  collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector*, (2024), arXiv: [2407.07546 \[hep-ex\]](https://arxiv.org/abs/2407.07546) (cit. on p. 18).

[117] R. Aaij et al.,  
*Measurement of lepton universality parameters in  $B^+ \rightarrow K^+ \ell^+ \ell^-$  and  $B^0 \rightarrow K^{*0} \ell^+ \ell^-$  decays*, *Phys. Rev. D* **108** (2023) 032002, arXiv: [2212.09153 \[hep-ex\]](https://arxiv.org/abs/2212.09153) (cit. on p. 18).

[118] B. C. Toth et al., *Muon g-2: BMW calculation of the hadronic vacuum polarization contribution*, *PoS LATTICE2021* (2022) 005 (cit. on p. 18).

[119] G. Aad et al., *Measurement of the W-boson mass and width with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **84** (2024) 1309, arXiv: [2403.15085 \[hep-ex\]](https://arxiv.org/abs/2403.15085) (cit. on p. 18).

[120] V. Chekhovsky et al.,  
*High-precision measurement of the W boson mass with the CMS experiment at the LHC*, (2024), arXiv: [2412.13872 \[hep-ex\]](https://arxiv.org/abs/2412.13872) (cit. on p. 18).

[121] S. Weinberg, *Effective Gauge Theories*, *Phys. Lett. B* **91** (1980) 51 (cit. on p. 18).

[122] T. Appelquist and J. Carazzone, *Infrared Singularities and Massive Fields*, *Phys. Rev. D* **11** (1975) 2856 (cit. on p. 19).

[123] W. Buchmuller and D. Wyler,  
*Effective Lagrangian Analysis of New Interactions and Flavor Conservation*, *Nucl. Phys. B* **268** (1986) 621 (cit. on p. 19).

[124] F. Feruglio, *The Chiral approach to the electroweak interactions*, *Int. J. Mod. Phys. A* **8** (1993) 4937, arXiv: [hep-ph/9301281](https://arxiv.org/abs/hep-ph/9301281) (cit. on p. 19).

[125] R. Alonso, M. B. Gavela, L. Merlo, S. Rigolin, and J. Yepes,  
*The Effective Chiral Lagrangian for a Light Dynamical "Higgs Particle"*, *Phys. Lett. B* **722** (2013) 330, [Erratum: *Phys.Lett.B* 726, 926 (2013)], arXiv: [1212.3305 \[hep-ph\]](https://arxiv.org/abs/1212.3305) (cit. on p. 19).

[126] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1, arXiv: [1706.08945 \[hep-ph\]](https://arxiv.org/abs/1706.08945) (cit. on p. 19).

[127] M. Aker et al., *Direct neutrino-mass measurement with sub-electronvolt sensitivity*, *Nature Phys.* **18** (2022) 160, arXiv: [2105.08533 \[hep-ex\]](https://arxiv.org/abs/2105.08533) (cit. on p. 19).

[128] P. Artoisenet et al., *A framework for Higgs characterisation*, *JHEP* **11** (2013) 043, arXiv: [1306.6464 \[hep-ph\]](https://arxiv.org/abs/1306.6464) (cit. on pp. 20, 21, 127, 128).

[129] A. V. Gritsan, R. Röntsch, M. Schulze, and M. Xiao, *Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques*, *Phys. Rev. D* **94** (2016) 055023, arXiv: [1606.03107 \[hep-ph\]](https://arxiv.org/abs/1606.03107) (cit. on pp. 20, 54, 75, 77).

[130] A. Tumasyan et al., *Constraints on anomalous Higgs boson couplings to vector bosons and fermions from the production of Higgs bosons using the  $\tau\tau$  final state*, *Phys. Rev. D* **108** (2023) 032013, arXiv: [2205.05120 \[hep-ex\]](https://arxiv.org/abs/2205.05120) (cit. on pp. 22, 36, 55, 83).

[131] R. Barate et al., *Search for the standard model Higgs boson at LEP*, *Phys. Lett. B* **565** (2003) 61, arXiv: [hep-ex/0306033](https://arxiv.org/abs/hep-ex/0306033) (cit. on p. 23).

[132] *Precision Electroweak Measurements and Constraints on the Standard Model*, (2010), arXiv: [1012.2367 \[hep-ex\]](https://arxiv.org/abs/1012.2367) (cit. on p. 23).

- [133] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008) S08001, URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001> (cit. on p. 23).
- [134] ATLAS, *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008) S08003, URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003> (cit. on p. 24).
- [135] CMS, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08004, URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08004> (cit. on p. 24).
- [136] R. Reed, *The Upgrade of the ATLAS Tile Calorimeter Readout Electronics for Phase II*, *Journal of Physics Conference Series* 1742-6588 **623** (2014) (cit. on p. 24).
- [137] K. Bunkowski and J. Krolkowski, *Optimization, Synchronization, Calibration and Diagnostic of the RPC PAC Muon Trigger System for the CMS detector*, (2009) (cit. on p. 24).
- [138] J. L. Feng, I. Galon, F. Kling, and S. Trojanowski, *ForwArd Search ExpeRiment at the LHC*, *Phys. Rev. D* **97** (2018) 035001, arXiv: [1708.09389 \[hep-ph\]](https://arxiv.org/abs/1708.09389) (cit. on p. 25).
- [139] C. Ahdida et al., *SND@LHC*, (2020), arXiv: [2002.08722 \[physics.ins-det\]](https://arxiv.org/abs/2002.08722) (cit. on p. 25).
- [140] H. Abreu et al., *First Direct Observation of Collider Neutrinos with FASER at the LHC*, *Phys. Rev. Lett.* **131** (2023) 031801, arXiv: [2303.14185 \[hep-ex\]](https://arxiv.org/abs/2303.14185) (cit. on p. 25).
- [141] R. Albanese et al., *Observation of Collider Muon Neutrinos with the SND@LHC Experiment*, *Phys. Rev. Lett.* **131** (2023) 031802, arXiv: [2305.09383 \[hep-ex\]](https://arxiv.org/abs/2305.09383) (cit. on p. 25).
- [142] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, arXiv: [1106.0522 \[hep-ph\]](https://arxiv.org/abs/1106.0522) (cit. on p. 26).
- [143] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, arXiv: [1405.0301 \[hep-ph\]](https://arxiv.org/abs/1405.0301) (cit. on pp. 26, 57, 77, 104).
- [144] R. Frederix et al., *The automation of next-to-leading order electroweak calculations*, *JHEP* **07** (2018) 185, [Erratum: *JHEP* 11, 085 (2021)], arXiv: [1804.10017 \[hep-ph\]](https://arxiv.org/abs/1804.10017) (cit. on p. 26).
- [145] I. Brivio, Y. Jiang, and M. Trott, *The SMEFTsim package, theory and tools*, *JHEP* **12** (2017) 070, arXiv: [1709.06492 \[hep-ph\]](https://arxiv.org/abs/1709.06492) (cit. on p. 26).
- [146] I. Brivio, *SMEFTsim 3.0 — a practical guide*, *JHEP* **04** (2021) 073, arXiv: [2012.11343 \[hep-ph\]](https://arxiv.org/abs/2012.11343) (cit. on pp. 26, 127).
- [147] F. Dermartin and M. Zaro, *HC\_NLO\_X0\_UFO*, [https://github.com/lukasheinrich/higgs-mc-studies/tree/master/HC\\_NLO\\_X0\\_UFO](https://github.com/lukasheinrich/higgs-mc-studies/tree/master/HC_NLO_X0_UFO), Accessed: 20.01.2025, 2014 (cit. on p. 26).
- [148] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, *Nucl. Phys. B* **44** (1972) 189 (cit. on p. 26).
- [149] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, *JHEP* **04** (2007) 081, arXiv: [hep-ph/0702198](https://arxiv.org/abs/hep-ph/0702198) (cit. on p. 26).

[150] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015, arXiv: [1212.3460 \[hep-ph\]](https://arxiv.org/abs/1212.3460) (cit. on pp. 26, 110).

[151] H. M. Pilkuhn, *The interactions of hadrons*, North-Holland Pub. Co., 1967 (cit. on p. 26).

[152] D. A. Dicus, E. Sudarshan, and X. Tata, *Factorization theorem for decaying spinning particles*, *Physics Letters B* **154** (1985) 79, ISSN: 0370-2693, URL: <https://www.sciencedirect.com/science/article/pii/0370269385915710> (cit. on p. 26).

[153] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8, arXiv: [2203.11601 \[hep-ph\]](https://arxiv.org/abs/2203.11601) (cit. on pp. 27, 57, 104).

[154] S. Agostinelli et al., *GEANT4 - A Simulation Toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 (cit. on p. 28).

[155] J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, arXiv: [1307.6346 \[hep-ex\]](https://arxiv.org/abs/1307.6346) (cit. on pp. 28, 57, 104).

[156] A. M. Sirunyan et al., *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003, arXiv: [1706.04965 \[physics.ins-det\]](https://arxiv.org/abs/1706.04965) (cit. on p. 28).

[157] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, arXiv: [1111.6097 \[hep-ph\]](https://arxiv.org/abs/1111.6097) (cit. on pp. 28, 105).

[158] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063, arXiv: [0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189) (cit. on pp. 28, 57).

[159] A. M. Sirunyan et al., *Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu_\tau$  in  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, *JINST* **13** (2018) P10005, arXiv: [1809.02816 \[hep-ex\]](https://arxiv.org/abs/1809.02816) (cit. on p. 28).

[160] G. Aad et al., *Configuration and performance of the ATLAS  $b$ -jet triggers in Run 2*, *Eur. Phys. J. C* **81** (2021) 1087, arXiv: [2106.03584 \[hep-ex\]](https://arxiv.org/abs/2106.03584) (cit. on p. 28).

[161] A. Tumasyan et al., *Identification of hadronic tau lepton decays using a deep neural network*, *JINST* **17** (2022) P07023, arXiv: [2201.08458 \[hep-ex\]](https://arxiv.org/abs/2201.08458) (cit. on p. 28).

[162] G. Aad et al., *Fast  $b$ -tagging at the high-level trigger of the ATLAS experiment in LHC Run 3*, *JINST* **18** (2023) P11006, arXiv: [2306.09738 \[hep-ex\]](https://arxiv.org/abs/2306.09738) (cit. on p. 28).

[163] P. Gras et al., *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091, arXiv: [1704.03878 \[hep-ph\]](https://arxiv.org/abs/1704.03878) (cit. on p. 29).

[164] G. Aad et al., *ATLAS flavour-tagging algorithms for the LHC Run 2  $pp$  collision dataset*, *Eur. Phys. J. C* **83** (2023) 681, arXiv: [2211.16345 \[physics.data-an\]](https://arxiv.org/abs/2211.16345) (cit. on p. 29).

[165] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019, arXiv: [1902.08570 \[hep-ph\]](https://arxiv.org/abs/1902.08570) (cit. on p. 29).

[166] H. Qu, C. Li, and S. Qian, *Particle Transformer for Jet Tagging*, (2022), arXiv: [2202.03772 \[hep-ph\]](https://arxiv.org/abs/2202.03772) (cit. on p. 29).

[167] G. Aad et al., *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52, [Erratum: *Nature* 612, E24 (2022)], arXiv: [2207.00092 \[hep-ex\]](https://arxiv.org/abs/2207.00092) (cit. on pp. 29, 34).

[168] A. Tumasyan et al., *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) 60, [Erratum: *Nature* 623, (2023)], arXiv: [2207.00043 \[hep-ex\]](https://arxiv.org/abs/2207.00043) (cit. on pp. 29, 33, 38).

[169] C. Patrignani et al., *Review of Particle Physics*, *Chin. Phys. C* **40** (2016) 100001 (cit. on p. 29).

[170] J. D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*, *Phys. Rev.* **179** (1969) 1547 (cit. on p. 30).

[171] R. D. Ball et al., *The path to proton structure at 1% accuracy*, *Eur. Phys. J. C* **82** (2022) 428, arXiv: [2109.02653 \[hep-ph\]](https://arxiv.org/abs/2109.02653) (cit. on p. 30).

[172] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: [1610.07922 \[hep-ph\]](https://arxiv.org/abs/1610.07922) (cit. on pp. 30, 36, 55).

[173] C. Duhr, F. Dulat, V. Hirschi, and B. Mistlberger, *Higgs production in bottom quark fusion: matching the 4- and 5-flavour schemes to third order in the strong coupling*, *JHEP* **08** (2020) 017, arXiv: [2004.04752 \[hep-ph\]](https://arxiv.org/abs/2004.04752) (cit. on p. 31).

[174] H. Bahl et al., *Indirect  $\mathcal{CP}$  probes of the Higgs-top-quark interaction: current LHC constraints and future opportunities*, *JHEP* **11** (2020) 127, arXiv: [2007.08542 \[hep-ph\]](https://arxiv.org/abs/2007.08542) (cit. on pp. 31, 35, 54, 74, 76, 128).

[175] A. Tumasyan et al., *Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at  $s=13$  TeV*, *Phys. Lett. B* **842** (2023) 137531, arXiv: [2206.09401 \[hep-ex\]](https://arxiv.org/abs/2206.09401) (cit. on p. 31).

[176] G. Aad et al., *Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using  $pp$  collisions at  $s=13$  TeV*, *Phys. Lett. B* **843** (2023) 137745, arXiv: [2211.01216 \[hep-ex\]](https://arxiv.org/abs/2211.01216) (cit. on p. 31).

[177] ATLAS, *ATLAS closes in on rare Higgs decays*, <https://atlas.cern/Updates/Briefing/Rare-Higgs-Run3>, Accessed: 16/07/2025, 2025 (cit. on p. 33).

[178] C.-N. Yang, *Selection Rules for the Dematerialization of a Particle Into Two Photons*, *Phys. Rev.* **77** (1950) 242 (cit. on p. 33).

[179] J. Brehmer, F. Kling, T. Plehn, and T. M. P. Tait, *Better Higgs- $\mathcal{CP}$  Tests Through Information Geometry*, *Phys. Rev. D* **97** (2018) 095017, arXiv: [1712.02350 \[hep-ph\]](https://arxiv.org/abs/1712.02350) (cit. on pp. 34, 55, 102–104).

[180] S. Berge, W. Bernreuther, and S. Kirchner, *Determination of the Higgs  $\mathcal{CP}$ -mixing angle in the tau decay channels*, *Nucl. Part. Phys. Proc.* **273-275** (2016) 841, ed. by M. Aguilar-Benítez, J. Fuster, S. Martí-García, and A. Santamaría, arXiv: [1410.6362 \[hep-ph\]](https://arxiv.org/abs/1410.6362) (cit. on p. 34).

[181] S. Berge, W. Bernreuther, B. Niepelt, and H. Spiesberger, *How to pin down the  $CP$  quantum numbers of a Higgs boson in its tau decays at the LHC*, *Phys. Rev. D* **84** (2011) 116003, arXiv: [1108.0670 \[hep-ph\]](https://arxiv.org/abs/1108.0670) (cit. on p. 35).

[182] A. Tumasyan et al., *Analysis of the  $CP$  structure of the Yukawa coupling between the Higgs boson and  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **06** (2022) 012, arXiv: [2110.04836 \[hep-ex\]](https://arxiv.org/abs/2110.04836) (cit. on pp. 35, 74).

[183] G. Aad et al., *Measurement of the  $CP$  properties of Higgs boson interactions with  $\tau$ -leptons with the ATLAS detector*, *Eur. Phys. J. C* **83** (2023) 563, arXiv: [2212.05833 \[hep-ex\]](https://arxiv.org/abs/2212.05833) (cit. on p. 35).

[184] A. F. Falk and M. E. Peskin, *Production, decay, and polarization of excited heavy hadrons*, *Phys. Rev. D* **49** (1994) 3320, arXiv: [hep-ph/9308241](https://arxiv.org/abs/hep-ph/9308241) (cit. on p. 35).

[185] A. J. Bevan,  *$C$ ,  $P$ , and  $CP$  asymmetry observables based on triple product asymmetries*, (2014), arXiv: [1408.3813 \[hep-ph\]](https://arxiv.org/abs/1408.3813) (cit. on pp. 35, 103).

[186] N. Mileo, K. Kiers, A. Szynkman, D. Crane, and E. Gegner, *Pseudoscalar top-Higgs coupling: exploration of  $CP$ -odd observables to resolve the sign ambiguity*, *JHEP* **07** (2016) 056, arXiv: [1603.03632 \[hep-ph\]](https://arxiv.org/abs/1603.03632) (cit. on pp. 35, 54, 103).

[187] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi, *Disentangling Higgs-Top Couplings in Associated Production*, *JHEP* **04** (2014) 004, arXiv: [1312.5736 \[hep-ph\]](https://arxiv.org/abs/1312.5736) (cit. on pp. 35, 54, 77, 103).

[188] F. Boudjema, R. M. Godbole, D. Guadagnoli, and K. A. Mohan, *Lab-frame observables for probing the top-Higgs interaction*, *Phys. Rev. D* **92** (2015) 015019, arXiv: [1501.03157 \[hep-ph\]](https://arxiv.org/abs/1501.03157) (cit. on pp. 35, 54, 103).

[189] V. Miralles, Y. Peters, E. Vryonidou, and J. K. Winter, *Sensitivity to  $CP$ -violating effective couplings in the top-Higgs sector*, *JHEP* **06** (2025) 052, arXiv: [2412.10309 \[hep-ph\]](https://arxiv.org/abs/2412.10309) (cit. on pp. 35, 103).

[190] D. Azevedo, R. Capucha, A. Onofre, and R. Santos,  *$CP$ -violation, asymmetries and interferences in  $t\bar{t}\phi$* , *JHEP* **09** (2022) 246, arXiv: [2208.04271 \[hep-ph\]](https://arxiv.org/abs/2208.04271) (cit. on pp. 35, 54, 77, 110).

[191] G. Aad et al.,  *$CP$  Properties of Higgs Boson Interactions with Top Quarks in the  $t\bar{t}H$  and  $tH$  Processes Using  $H \rightarrow \gamma\gamma$  with the ATLAS Detector*, *Phys. Rev. Lett.* **125** (2020) 061802, arXiv: [2004.04545 \[hep-ex\]](https://arxiv.org/abs/2004.04545) (cit. on pp. 35, 42, 57, 77, 83, 85, 87, 90).

[192] A. Tumasyan et al., *Search for  $CP$  violation in  $t\bar{t}H$  and  $tH$  production in multilepton channels in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **07** (2023) 092, arXiv: [2208.02686 \[hep-ex\]](https://arxiv.org/abs/2208.02686) (cit. on pp. 35, 42, 77, 83, 85, 87, 90).

[193] D. A. Faroughy, J. F. Kamenik, N. Košnik, and A. Smolkovič, *Probing the  $CP$  nature of the top quark Yukawa at hadron colliders*, *JHEP* **02** (2020) 085, arXiv: [1909.00007 \[hep-ph\]](https://arxiv.org/abs/1909.00007) (cit. on pp. 35, 54, 103).

[194] Y. Chen, A. Falkowski, I. Low, and R. Vega-Morales, *New Observables for  $CP$  Violation in Higgs Decays*, *Phys. Rev. D* **90** (2014) 113006, arXiv: [1405.6723 \[hep-ph\]](https://arxiv.org/abs/1405.6723) (cit. on p. 36).

[195] J. R. Andersen et al.,  
“Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report,”  
*9th Les Houches Workshop on Physics at TeV Colliders*, 2016, arXiv: [1605.04692 \[hep-ph\]](https://arxiv.org/abs/1605.04692) (cit. on p. 36).

[196] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1*, (2019),  
arXiv: [1906.02754 \[hep-ph\]](https://arxiv.org/abs/1906.02754) (cit. on p. 37).

[197] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009  
(cit. on p. 38).

[198] D. Guest, K. Cranmer, and D. Whiteson, *Deep Learning and its Application to LHC Physics*,  
*Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161, arXiv: [1806.11484 \[hep-ex\]](https://arxiv.org/abs/1806.11484) (cit. on p. 38).

[199] A. Radovic et al., *Machine learning at the energy and intensity frontiers of particle physics*,  
*Nature* **560** (2018) 41 (cit. on p. 38).

[200] G. Carleo et al., *Machine learning and the physical sciences*, *Rev. Mod. Phys.* **91** (2019) 045002,  
arXiv: [1903.10563 \[physics.comp-ph\]](https://arxiv.org/abs/1903.10563) (cit. on p. 38).

[201] M. Feickert and B. Nachman, *A Living Review of Machine Learning for Particle Physics*, (2021),  
arXiv: [2102.02770 \[hep-ph\]](https://arxiv.org/abs/2102.02770) (cit. on pp. 38, 45).

[202] CERN, *Facts and figures about the LHC*,  
<https://home.cern/resources/faqs/facts-and-figures-about-lhc>,  
Accessed: 25/04/2025, 2025 (cit. on p. 38).

[203] T. R. S. A. of Sciences, *The Nobel Prize in Physics 2024 Press Release*,  
<https://www.nobelprize.org/prizes/physics/2024/press-release/>,  
Accessed: 25/04/2025, 2024 (cit. on p. 38).

[204] Y. Freund and R. E. Schapire, “Experiments with a New Boosting Algorithm,”  
*International Conference on Machine Learning*, 1996 148,  
URL: <https://api.semanticscholar.org/CorpusID:1836349> (cit. on p. 41).

[205] J. Friedman, T. Hastie, and R. Tibshirani,  
*Additive Logistic Regression: A Statistical View of Boosting*,  
*The Annals of Statistics* **28** (2000) 337 (cit. on p. 41).

[206] J. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*,  
*The Annals of Statistics* **29** (2000) (cit. on p. 42).

[207] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*,  
ACM, 2016 785, URL: <http://dx.doi.org/10.1145/2939672.2939785>  
(cit. on pp. 42, 108).

[208] B. Kegl et al., *Higgs Boson Machine Learning Challenge*,  
<https://kaggle.com/competitions/higgs-boson>, Kaggle, 2014 (cit. on p. 42).

[209] F. Rosenblatt,  
*The perceptron: A probabilistic model for information storage and organization in the brain*,  
*Psychological Review* **65** (1958) 386 (cit. on p. 43).

[210] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*, Cambridge, MA: MIT Press, 1969 (cit. on p. 43).

[211] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Neurocomputing: Foundations of Research*, Cambridge, MA, USA: MIT Press, 1988 696, ISBN: 0262010976 (cit. on p. 44).

[212] B. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics **4** (1964) 1, ISSN: 0041-5553, URL: <https://www.sciencedirect.com/science/article/pii/0041555364901375> (cit. on p. 45).

[213] J. Duchi, E. Hazan, and Y. Singer, *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*, J. Mach. Learn. Res. **12** (2011) 2121, ISSN: 1532-4435 (cit. on p. 45).

[214] M. D. Zeiler, *ADADELTA: An Adaptive Learning Rate Method*, 2012, arXiv: [1212.5701 \[cs.LG\]](https://arxiv.org/abs/1212.5701), URL: <https://arxiv.org/abs/1212.5701> (cit. on p. 45).

[215] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2017, arXiv: [1412.6980 \[cs.LG\]](https://arxiv.org/abs/1412.6980), URL: <https://arxiv.org/abs/1412.6980> (cit. on p. 45).

[216] G. Kasieczka et al., *The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics*, Reports on Progress in Physics **84** (2021) (cit. on p. 45).

[217] L. Shapley, “A Value for n-Person Games,” *Contributions to the Theory of Games II*, ed. by H. W. Kuhn and A. W. Tucker, Princeton: Princeton University Press, 1953 307 (cit. on pp. 46, 68).

[218] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems 30*, ed. by I. Guyon et al., Curran Associates, Inc., 2017 4765, URL: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf> (cit. on pp. 46, 68).

[219] M. Virgolin and S. P. Pissis, *Symbolic Regression is NP-hard*, 2022, arXiv: [2207.01018 \[cs.NE\]](https://arxiv.org/abs/2207.01018), URL: <https://arxiv.org/abs/2207.01018> (cit. on p. 47).

[220] P. Lemos, N. Jeffrey, M. Cranmer, S. Ho, and P. Battaglia, *Rediscovering orbital mechanics with machine learning*, 2022, arXiv: [2202.02306 \[astro-ph.EP\]](https://arxiv.org/abs/2202.02306), URL: <https://arxiv.org/abs/2202.02306> (cit. on p. 47).

[221] Z.-Y. Khoo, G. Rajiv, A. Yang, J. S. C. Low, and S. Bressan, “Celestial Machine Learning: Discovering the Planarity, Heliocentricity, and Orbital Equation of Mars with AI Feynman,” *Information Integration and Web Intelligence*, Springer Nature Switzerland, 2023 201, ISBN: 9783031483165, URL: [http://dx.doi.org/10.1007/978-3-031-48316-5\\_21](http://dx.doi.org/10.1007/978-3-031-48316-5_21) (cit. on p. 47).

[222] A. Butter, T. Plehn, N. Soybelman, and J. Brehmer, *Back to the formula - LHC edition*, SciPost Phys. **16** (2024) 037, arXiv: [2109.10414 \[hep-ph\]](https://arxiv.org/abs/2109.10414) (cit. on pp. 47, 49, 66, 70, 75, 104).

[223] Z. Dong, K. Kong, K. T. Matchev, and K. Matcheva, *Is the machine smarter than the theorist: Deriving formulas for particle kinematics with symbolic regression*, *Phys. Rev. D* **107** (2023) 055018, arXiv: [2211.08420 \[hep-ph\]](https://arxiv.org/abs/2211.08420) (cit. on p. 47).

[224] A. Alnuqaydan, S. Gleyzer, and H. Prosper, *SYMBA: symbolic computation of squared amplitudes in high energy physics with machine learning*, *Mach. Learn. Sci. Tech.* **4** (2023) 015007, arXiv: [2206.08901 \[hep-ph\]](https://arxiv.org/abs/2206.08901) (cit. on p. 47).

[225] H. F. Tsoi et al., *Symbolic Regression on FPGAs for Fast Machine Learning Inference*, *EPJ Web Conf.* **295** (2024) 09036, arXiv: [2305.04099 \[cs.LG\]](https://arxiv.org/abs/2305.04099) (cit. on p. 47).

[226] S. AbdusSalam, S. Abel, and M. Crispim Romão, *Symbolic regression for beyond the standard model physics*, *Phys. Rev. D* **111** (2025) 015022, arXiv: [2405.18471 \[hep-ph\]](https://arxiv.org/abs/2405.18471) (cit. on p. 47).

[227] M. Morales-Alvarado, D. Conde, J. Bendavid, V. Sanz, and M. Ubiali, “Symbolic regression for precision LHC physics,” *38th conference on Neural Information Processing Systems*, 2024, arXiv: [2412.07839 \[hep-ph\]](https://arxiv.org/abs/2412.07839) (cit. on p. 47).

[228] N. Sobelman, C. Schiavi, F. A. Di Bello, and E. Gross, *Accelerating graph-based tracking tasks with symbolic regression*, *Mach. Learn. Sci. Tech.* **5** (2024) 045042, arXiv: [2406.16752 \[hep-ex\]](https://arxiv.org/abs/2406.16752) (cit. on p. 47).

[229] M. Cranmer, *Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl*, 2023, arXiv: [2305.01582 \[astro-ph.IM\]](https://arxiv.org/abs/2305.01582), URL: <https://arxiv.org/abs/2305.01582> (cit. on pp. 47, 102, 103).

[230] H. Bahl, E. Fuchs, M. Menen, and T. Plehn, *CP-Analyses with Symbolic Regression*, (2025), arXiv: [2507.05858 \[hep-ph\]](https://arxiv.org/abs/2507.05858) (cit. on pp. 48, 50, 51, 102, 106–110, 112, 114–121, 146).

[231] H. F. Tsoi, V. Loncar, S. Dasu, and P. Harris, *SymbolNet: neural symbolic regression with adaptive dynamic pruning for compression*, *Mach. Learn. Sci. Tech.* **6** (2025) 015021, arXiv: [2401.09949 \[cs.LG\]](https://arxiv.org/abs/2401.09949) (cit. on pp. 49–51, 102, 103).

[232] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: <https://www.tensorflow.org/> (cit. on p. 49).

[233] H. Bahl, E. Fuchs, M. Hannig, and M. Menen, *Classifying the CP properties of the ggH coupling in H + 2j production*, *SciPost Phys. Core* **8** (2025) 006, arXiv: [2309.03146 \[hep-ph\]](https://arxiv.org/abs/2309.03146) (cit. on pp. 54, 56, 58, 59, 61–65, 67, 69–72, 74, 134, 135, 184).

[234] J. F. Gunion and X.-G. He, *Determining the CP nature of a neutral Higgs boson at the LHC*, *Phys. Rev. Lett.* **76** (1996) 4468, arXiv: [hep-ph/9602226](https://arxiv.org/abs/hep-ph/9602226) (cit. on pp. 54, 77, 79, 80).

[235] A. Freitas and P. Schwaller, *Higgs CP Properties From Early LHC Data*, *Phys. Rev. D* **87** (2013) 055014, arXiv: [1211.1980 \[hep-ph\]](https://arxiv.org/abs/1211.1980) (cit. on p. 54).

[236] P. Agrawal, S. Mitra, and A. Shivaji, *Effect of Anomalous Couplings on the Associated Production of a Single Top Quark and a Higgs Boson at the LHC*, *JHEP* **12** (2013) 077, arXiv: [1211.4362 \[hep-ph\]](https://arxiv.org/abs/1211.4362) (cit. on p. 54).

[237] A. Djouadi and G. Moreau, *The couplings of the Higgs boson and its CP properties from fits of the signal strengths and their ratios at the 7+8 TeV LHC*, *Eur. Phys. J. C* **73** (2013) 2512, arXiv: [1303.6591 \[hep-ph\]](#) (cit. on p. 54).

[238] J. Yue,  
*Enhanced  $t\bar{t}h$  signal at the LHC with  $h \rightarrow \gamma\gamma$  decay and CP-violating top-Higgs coupling*, *Phys. Lett. B* **744** (2015) 131, arXiv: [1410.2701 \[hep-ph\]](#) (cit. on pp. 54, 77).

[239] F. Demartin, F. Maltoni, K. Mawatari, B. Page, and M. Zaro,  
*Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction*, *Eur. Phys. J. C* **74** (2014) 3065, arXiv: [1407.5089 \[hep-ph\]](#) (cit. on pp. 54, 57, 60, 77, 83, 128).

[240] J. Chang, K. Cheung, J. S. Lee, and C.-T. Lu,  
*Probing the Top-Yukawa Coupling in Associated Higgs production with a Single Top Quark*, *JHEP* **05** (2014) 062, arXiv: [1403.2053 \[hep-ph\]](#) (cit. on p. 54).

[241] X.-G. He, G.-N. Li, and Y.-J. Zheng,  
*Probing Higgs boson CP Properties with  $t\bar{t}H$  at the LHC and the 100 TeV pp collider*, *Int. J. Mod. Phys. A* **30** (2015) 1550156, arXiv: [1501.00012 \[hep-ph\]](#) (cit. on pp. 54, 77).

[242] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro,  
*Higgs production in association with a single top quark at the LHC*, *Eur. Phys. J. C* **75** (2015) 267, arXiv: [1504.00611 \[hep-ph\]](#) (cit. on pp. 54, 77).

[243] M. R. Buckley and D. Goncalves,  
*Boosting the Direct CP Measurement of the Higgs-Top Coupling*, *Phys. Rev. Lett.* **116** (2016) 091801, arXiv: [1507.07926 \[hep-ph\]](#) (cit. on pp. 54, 77).

[244] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro,  
 *$tWH$  associated production at the LHC*, *Eur. Phys. J. C* **77** (2017) 34, arXiv: [1607.05862 \[hep-ph\]](#) (cit. on pp. 54, 77).

[245] A. Kobakhidze, N. Liu, L. Wu, and J. Yue,  
*Implications of CP-violating Top-Higgs Couplings at LHC and Higgs Factories*, *Phys. Rev. D* **95** (2017) 015016, arXiv: [1610.06676 \[hep-ph\]](#) (cit. on p. 54).

[246] D. Azevedo, A. Onofre, F. Filthaut, and R. Gonçalo,  
*CP tests of Higgs couplings in  $t\bar{t}h$  semileptonic events at the LHC*, *Phys. Rev. D* **98** (2018) 033004, arXiv: [1711.05292 \[hep-ph\]](#) (cit. on pp. 54, 77).

[247] D. Gonçalves, K. Kong, and J. H. Kim,  
*Probing the top-Higgs Yukawa CP structure in dileptonic  $t\bar{t}h$  with  $M_2$ -assisted reconstruction*, *JHEP* **06** (2018) 079, arXiv: [1804.05874 \[hep-ph\]](#) (cit. on pp. 54, 77, 80, 103, 113).

[248] W.-S. Hou, M. Kohda, and T. Modak,  
*Probing for extra top Yukawa couplings in light of  $t\bar{t}h(125)$  observation*, *Phys. Rev. D* **98** (2018) 075007, arXiv: [1806.06018 \[hep-ph\]](#) (cit. on p. 54).

[249] Q.-H. Cao, S.-L. Chen, Y. Liu, R. Zhang, and Y. Zhang,  
*Limiting top quark-Higgs boson interaction and Higgs-boson width from multitop productions*, *Phys. Rev. D* **99** (2019) 113003, arXiv: [1901.04567 \[hep-ph\]](#) (cit. on p. 54).

[250] J. Ren, L. Wu, and J. M. Yang,  
*Unveiling CP property of top-Higgs coupling with graph neural networks at the LHC*,  
*Phys. Lett. B* **802** (2020) 135198, arXiv: [1901.05627 \[hep-ph\]](#) (cit. on pp. 54, 77).

[251] M. Kraus, T. Martini, S. Peitzsch, and P. Uwer,  
*Exploring BSM Higgs couplings in single top-quark production*, (2019),  
arXiv: [1908.09100 \[hep-ph\]](#) (cit. on p. 54).

[252] B. Bortolato, J. F. Kamenik, N. Košnik, and A. Smolkovič,  
*Optimized probes of CP -odd effects in the  $t\bar{t}h$  process at hadron colliders*,  
*Nucl. Phys. B* **964** (2021) 115328, arXiv: [2006.13110 \[hep-ph\]](#) (cit. on pp. 54, 103).

[253] Q.-H. Cao, K.-P. Xie, H. Zhang, and R. Zhang,  
*A New Observable for Measuring CP Property of Top-Higgs Interaction*,  
*Chin. Phys. C* **45** (2021) 023117, arXiv: [2008.13442 \[hep-ph\]](#) (cit. on pp. 54, 77, 79, 80).

[254] T. Martini, R.-Q. Pan, M. Schulze, and M. Xiao, *Probing the CP structure of the top quark Yukawa coupling: Loop sensitivity versus on-shell sensitivity*, *Phys. Rev. D* **104** (2021) 055045, arXiv: [2104.04277 \[hep-ph\]](#) (cit. on pp. 54, 77).

[255] R. K. Barman, D. Gonçalves, and F. Kling,  
*Machine learning the Higgs boson-top quark CP phase*, *Phys. Rev. D* **105** (2022) 035023, arXiv: [2110.07635 \[hep-ph\]](#) (cit. on pp. 54, 77, 110, 113, 118).

[256] H. Bahl and S. Brass, *Constraining  $\mathcal{CP}$ -violation in the Higgs-top-quark interaction using machine-learning-based inference*, *JHEP* **03** (2022) 017, arXiv: [2110.10177 \[hep-ph\]](#) (cit. on pp. 54, 77).

[257] D. Gonçalves, J. H. Kim, K. Kong, and Y. Wu,  
*Direct Higgs-top CP-phase measurement with  $t\bar{t}h$  at the 14 TeV LHC and 100 TeV FCC*,  
*JHEP* **01** (2022) 158, arXiv: [2108.01083 \[hep-ph\]](#) (cit. on pp. 54, 103).

[258] A. Butter, T. Heimel, T. Martini, S. Peitzsch, and T. Plehn,  
*Two Invertible Networks for the Matrix Element Method*, (2022), arXiv: [2210.00019 \[hep-ph\]](#) (cit. on pp. 54, 102).

[259] J. Ackerschott, R. K. Barman, D. Gonçalves, T. Heimel, and T. Plehn,  
*Returning CP-Observables to The Frames They Belong*, (2023), arXiv: [2308.00027 \[hep-ph\]](#) (cit. on pp. 54, 77, 113).

[260] A. Bhardwaj, C. Englert, D. Gonçalves, and A. Navarro, *Non-linear top-Higgs CP violation*, (2023), arXiv: [2308.11722 \[hep-ph\]](#) (cit. on p. 54).

[261] Z. Yu, K. A. Mohan, and C. -. Yuan,  
*Determining the CP property of  $ht\bar{t}$  coupling via a novel jet substructure observable*,  
*Phys. Lett. B* **856** (2024) 138959, arXiv: [2211.00845 \[hep-ph\]](#) (cit. on p. 54).

[262] T. Plehn, D. L. Rainwater, and D. Zeppenfeld,  
*Determining the Structure of Higgs Couplings at the LHC*, *Phys. Rev. Lett.* **88** (2002) 051801, arXiv: [hep-ph/0105325](#) (cit. on pp. 54, 103, 104).

[263] V. Hankele, G. Klamke, and D. Zeppenfeld, “Higgs + 2 jets as a probe for CP properties,”  
*Meeting on CP Violation and Non-standard Higgs Physics*, 2006, arXiv: [hep-ph/0605117](#) (cit. on p. 54).

[264] S. Kraml et al., *CP Studies and Non-Standard Higgs Physics*, (2006), arXiv: [hep-ph/0608079](#) (cit. on p. 54).

[265] V. Del Duca et al., *Monte Carlo studies of the jet activity in Higgs + 2 jet events*, *JHEP* **10** (2006) 016, arXiv: [hep-ph/0608158](#) (cit. on p. 54).

[266] G. Klamke and D. Zeppenfeld, “Hjj production: Signals and CP measurements,” *42nd Rencontres de Moriond on QCD and High Energy Hadronic Interactions*, 2007 201, arXiv: [0705.2983](#) [hep-ph] (cit. on p. 54).

[267] G. Klamke and D. Zeppenfeld, *Higgs plus two jet production via gluon fusion as a signal at the CERN LHC*, *JHEP* **04** (2007) 052, arXiv: [hep-ph/0703202](#) (cit. on p. 54).

[268] K. Hagiwara, Q. Li, and K. Mawatari, *Jet angular correlation in vector-boson fusion processes at hadron colliders*, *JHEP* **07** (2009) 101, arXiv: [0905.4314](#) [hep-ph] (cit. on pp. 54, 58).

[269] C. Englert, D. Goncalves-Netto, K. Mawatari, and T. Plehn, *Higgs Quantum Numbers in Weak Boson Fusion*, *JHEP* **01** (2013) 148, arXiv: [1212.0843](#) [hep-ph] (cit. on p. 54).

[270] C. Englert, M. Spannowsky, and M. Takeuchi, *Measuring Higgs CP and couplings with hadronic event shapes*, *JHEP* **06** (2012) 108, arXiv: [1203.5788](#) [hep-ph] (cit. on p. 54).

[271] I. Anderson et al., *Constraining Anomalous HVV Interactions at Proton and Lepton Colliders*, *Phys. Rev. D* **89** (2014) 035007, ed. by N. A. Graf, M. E. Peskin, and J. L. Rosner, arXiv: [1309.4819](#) [hep-ph] (cit. on p. 54).

[272] M. J. Dolan, P. Harris, M. Jankowiak, and M. Spannowsky, *Constraining CP-violating Higgs Sectors at the LHC using gluon fusion*, *Phys. Rev. D* **90** (2014) 073008, arXiv: [1406.3322](#) [hep-ph] (cit. on p. 54).

[273] F. U. Bernlochner et al., *Angles on CP-violation in Higgs boson interactions*, *Phys. Lett. B* **790** (2019) 372, arXiv: [1808.06577](#) [hep-ph] (cit. on p. 54).

[274] C. Englert, P. Galler, A. Pilkington, and M. Spannowsky, *Approaching robust EFT limits for CP-violation in the Higgs sector*, *Phys. Rev. D* **99** (2019) 095007, arXiv: [1901.05982](#) [hep-ph] (cit. on p. 54).

[275] ATLAS Collaboration, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-014/>, 2024 (cit. on p. 54).

[276] CMS Collaboration, [https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEX013TeV#Long\\_lived\\_particle\\_summary\\_plot](https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEX013TeV#Long_lived_particle_summary_plot), 2025 (cit. on p. 54).

[277] U. Haisch and A. Hala, *Bounds on CP-violating Higgs-gluon interactions: the case of vanishing light-quark Yukawa couplings*, *JHEP* **11** (2019) 117, arXiv: [1909.09373](#) [hep-ph] (cit. on p. 55).

[278] J. R. Andersen, T. Binoth, G. Heinrich, and J. M. Smillie, *Loop induced interference effects in Higgs Boson plus two jet production at the LHC*, *JHEP* **02** (2008) 057, arXiv: [0709.3513](#) [hep-ph] (cit. on p. 56).

[279] A. Bredenstein, K. Hagiwara, and B. Jager,  
*Mixed QCD-electroweak contributions to Higgs-plus-dijet production at the LHC*,  
*Phys. Rev. D* **77** (2008) 073004, arXiv: [0801.4231 \[hep-ph\]](#) (cit. on p. 56).

[280] A. Dedes, W. Materkowska, M. Paraskevas, J. Rosiek, and K. Suxho,  
*Feynman rules for the Standard Model Effective Field Theory in  $R_\xi$ -gauges*, *JHEP* **06** () 143, arXiv: [1704.03888 \[hep-ph\]](#) (cit. on pp. 56, 104).

[281] J. Davis et al., *Constraining anomalous Higgs boson couplings to virtual photons*,  
*Phys. Rev. D* **105** (2022) 096027, arXiv: [2109.13363 \[hep-ph\]](#) (cit. on pp. 56, 104).

[282] M. Cacciari, F. A. Dreyer, A. Karlberg, G. P. Salam, and G. Zanderighi,  
*Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order*,  
*Phys. Rev. Lett.* **115** (2015) 082002, [Erratum: *Phys. Rev. Lett.* 120, 139901 (2018)], arXiv: [1506.02660 \[hep-ph\]](#) (cit. on p. 57).

[283] P. Obul, S. Dulat, T.-J. Hou, A. Tursun, and N. Yalkun, *Next-to-leading order QCD and electroweak corrections to Higgs-strahlung processes at the LHC*,  
*Chin. Phys. C* **42** (2018) 093105, arXiv: [1801.06851 \[hep-ph\]](#) (cit. on p. 57).

[284] CMS Collaboration, *Measurements of  $t\bar{t}H$  Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel*,  
*Phys. Rev. Lett.* **125** (2020) 061801, arXiv: [2003.10866 \[hep-ex\]](#) (cit. on pp. 57, 77, 83).

[285] V. Del Duca, W. Kilgore, C. Oleari, C. Schmidt, and D. Zeppenfeld,  
*Higgs + 2 jets via gluon fusion*, *Phys. Rev. Lett.* **87** (2001) 122001, arXiv: [hep-ph/0105129](#) (cit. on p. 57).

[286] V. Del Duca, W. Kilgore, C. Oleari, C. Schmidt, and D. Zeppenfeld,  
*Gluon fusion contributions to  $H + 2$  jet production*, *Nucl. Phys. B* **616** (2001) 367, arXiv: [hep-ph/0108030](#) (cit. on p. 57).

[287] M. Buschmann, C. Englert, D. Goncalves, T. Plehn, and M. Spannowsky,  
*Resolving the Higgs-Gluon Coupling with Jets*, *Phys. Rev. D* **90** (2014) 013010, arXiv: [1405.7651 \[hep-ph\]](#) (cit. on p. 57).

[288] F. Maltoni, E. Vryonidou, and M. Zaro,  
*Top-quark mass effects in double and triple Higgs production in gluon-gluon fusion at NLO*, *JHEP* **11** (2014) 079, arXiv: [1408.6542 \[hep-ph\]](#) (cit. on p. 57).

[289] J. Ansel et al., “PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation,” *ASPLOS ’24*, La Jolla, CA, USA: Association for Computing Machinery, 2024 929, URL: <https://doi.org/10.1145/3620665.3640366> (cit. on p. 59).

[290] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825 (cit. on p. 62).

[291] L. Buitinck et al.,  
“API design for machine learning software: experiences from the scikit-learn project,” *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013 108 (cit. on p. 62).

[292] A. Bhardwaj, C. Englert, R. Hankache, and A. D. Pilkington, *Machine-enhanced CP-asymmetries in the Higgs sector*, *Phys. Lett. B* **832** (2022) 137246, arXiv: [2112.05052 \[hep-ph\]](#) (cit. on pp. 63, 102).

[293] A. V. Gritsan et al., *New features in the JHU generator framework: constraining Higgs boson properties from on-shell and off-shell production*, *Phys. Rev. D* **102** (2020) 056022, arXiv: [2002.09888 \[hep-ph\]](#) (cit. on p. 63).

[294] N. Castro et al., *LHC EFT WG Report: Experimental Measurements and Observables*, (2022), arXiv: [2211.08353 \[hep-ph\]](#) (cit. on pp. 63, 102).

[295] P. Bechtle, S. Heinemeyer, O. Stål, T. Stefaniak, and G. Weiglein, *HiggsSignals: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC*, *Eur. Phys. J. C* **74** (2014) 2711, arXiv: [1305.1933 \[hep-ph\]](#) (cit. on p. 74).

[296] P. Bechtle et al., *HiggsSignals-2: Probing new physics with precision Higgs measurements in the LHC 13 TeV era*, *Eur. Phys. J. C* **81** (2021) 145, arXiv: [2012.09197 \[hep-ph\]](#) (cit. on p. 74).

[297] H. Bahl et al., *HiggsTools: BSM scalar phenomenology with new versions of HiggsBounds and HiggsSignals*, *Comput. Phys. Commun.* **291** (2023) 108803, arXiv: [2210.09332 \[hep-ph\]](#) (cit. on p. 74).

[298] G. Aad et al., *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **06** (2022) 097, arXiv: [2111.06712 \[hep-ex\]](#) (cit. on pp. 74, 85).

[299] V. M. Abazov et al., *A precision measurement of the mass of the top quark*, *Nature* **429** (2004) 638, arXiv: [hep-ex/0406031](#) (cit. on p. 75).

[300] Y. Gao et al., *Spin Determination of Single-Produced Resonances at Hadron Colliders*, *Phys. Rev. D* **81** (2010) 075022, arXiv: [1001.3396 \[hep-ph\]](#) (cit. on p. 75).

[301] J. Alwall, A. Freitas, and O. Mattelaer, *The Matrix Element Method and QCD Radiation*, *Phys. Rev. D* **83** (2011) 074010, arXiv: [1010.2263 \[hep-ph\]](#) (cit. on p. 75).

[302] S. Bolognesi et al., *On the Spin and Parity of a Single-Produced Resonance at the LHC*, *Phys. Rev. D* **86** (2012) 095031, arXiv: [1208.4018 \[hep-ph\]](#) (cit. on p. 75).

[303] P. Avery et al., *Precision studies of the Higgs boson decay channel  $H \rightarrow ZZ \rightarrow 4\ell$  with MEKD*, *Phys. Rev. D* **87** (2013) 055006, arXiv: [1210.0896 \[hep-ph\]](#) (cit. on p. 75).

[304] J. R. Andersen, C. Englert, and M. Spannowsky, *Extracting precise Higgs couplings by using the matrix element method*, *Phys. Rev. D* **87** (2013) 015019, arXiv: [1211.3011 \[hep-ph\]](#) (cit. on p. 75).

[305] P. Artoisenet, P. de Aquino, F. Maltoni, and O. Mattelaer, *Unravelling  $t\bar{t}h$  via the Matrix Element Method*, *Phys. Rev. Lett.* **111** (2013) 091802, arXiv: [1304.6414 \[hep-ph\]](#) (cit. on p. 75).

[306] J. M. Campbell, R. K. Ellis, W. T. Giele, and C. Williams, *Finding the Higgs Boson in Decays to  $Z\gamma$  using the Matrix Element Method at Next-to-Leading Order*, *Phys. Rev. D* **87** (2013) 073005, arXiv: [1301.7086 \[hep-ph\]](#) (cit. on p. 75).

[307] T. Martini and P. Uwer, *The Matrix Element Method at next-to-leading order QCD for hadronic collisions: Single top-quark production at the LHC as an example application*, *JHEP* **05** (2018) 141, arXiv: [1712.04527 \[hep-ph\]](https://arxiv.org/abs/1712.04527) (cit. on p. 75).

[308] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *Mining gold from implicit models to improve likelihood-free inference*, *Proc. Nat. Acad. Sci.* **117** (2020) 5242, arXiv: [1805.12244 \[stat.ML\]](https://arxiv.org/abs/1805.12244) (cit. on p. 75).

[309] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *Constraining Effective Field Theories with Machine Learning*, *Phys. Rev. Lett.* **121** (2018) 111801, arXiv: [1805.00013 \[hep-ph\]](https://arxiv.org/abs/1805.00013) (cit. on p. 75).

[310] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *A Guide to Constraining Effective Field Theories with Machine Learning*, *Phys. Rev. D* **98** (2018) 052004, arXiv: [1805.00020 \[hep-ph\]](https://arxiv.org/abs/1805.00020) (cit. on p. 75).

[311] M. Stoye, J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *Likelihood-free inference with an improved cross-entropy estimator*, (2018), arXiv: [1808.00973 \[stat.ML\]](https://arxiv.org/abs/1808.00973) (cit. on p. 75).

[312] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, *MadMiner: Machine learning-based inference for particle physics*, *Comput. Softw. Big Sci.* **4** (2020) 3, arXiv: [1907.10621 \[hep-ph\]](https://arxiv.org/abs/1907.10621) (cit. on p. 75).

[313] H. Bahl et al.,  *$C\bar{P}$ -sensitive simplified template cross-sections for  $t\bar{t}H$* , *JHEP* **10** (2024) 214, arXiv: [2406.03950 \[hep-ph\]](https://arxiv.org/abs/2406.03950) (cit. on pp. 76, 79–84, 88, 91, 92, 94–96, 98, 100, 101, 136–138, 142–145, 184).

[314] W. Esmail, A. Hammad, A. Jueid, and S. Moretti, *Boosting probes of  $CP$  violation in the top Yukawa coupling with Deep Learning*, (2024), arXiv: [2405.16499 \[hep-ph\]](https://arxiv.org/abs/2405.16499) (cit. on p. 77).

[315] G. Aad et al., *Probing the  $CP$  nature of the top–Higgs Yukawa coupling in  $t\bar{t}H$  and  $tH$  events with  $H \rightarrow b\bar{b}$  decays using the ATLAS detector at the LHC*, *Phys. Lett. B* **849** (2024) 138469, arXiv: [2303.05974 \[hep-ex\]](https://arxiv.org/abs/2303.05974) (cit. on pp. 77, 83, 85, 87, 97).

[316] J. C. Collins and D. E. Soper, *Angular Distribution of Dileptons in High-Energy Hadron Collisions*, *Phys. Rev. D* **16** (1977) 2219 (cit. on pp. 79, 80, 113).

[317] A. Ferroglio, M. C. N. Fiolhais, E. Gouveia, and A. Onofre, *Role of the  $t\bar{t}h$  rest frame in direct top-quark Yukawa coupling measurements*, *Phys. Rev. D* **100** (2019) 075034, arXiv: [1909.00490 \[hep-ph\]](https://arxiv.org/abs/1909.00490) (cit. on p. 79).

[318] G. Aad et al., *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J. C* **74** (2014) 3071, arXiv: [1407.5063 \[hep-ex\]](https://arxiv.org/abs/1407.5063) (cit. on p. 85).

[319] G. Aad et al., *Measurement of the charge asymmetry in top quark pair production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **72** (2012) 2039, arXiv: [1203.4211 \[hep-ex\]](https://arxiv.org/abs/1203.4211) (cit. on p. 85).

[320] G. Aad et al., *Measurements of top-quark pair differential and double-differential cross-sections in the  $\ell+jets$  channel with  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 1028, [Erratum: Eur.Phys.J.C 80, 1092 (2020)], arXiv: [1908.07305 \[hep-ex\]](https://arxiv.org/abs/1908.07305) (cit. on p. 85).

[321] *Formulae for Estimating Significance*, (2020) (cit. on pp. 85, 86, 131).

[322] S. Raschka, *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*, 2020, arXiv: [1811.12808 \[cs.LG\]](https://arxiv.org/abs/1811.12808), URL: <https://arxiv.org/abs/1811.12808> (cit. on p. 93).

[323] LHCTopWG, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/ParticleLevelTopDefinitions>, Accessed: 12/06/2025 (cit. on p. 100).

[324] T. Heimel, N. Huetsch, R. Winterhalder, T. Plehn, and A. Butter, *Precision-machine learning for the matrix element method*, *SciPost Phys.* **17** (2024) 129, arXiv: [2310.07752 \[hep-ph\]](https://arxiv.org/abs/2310.07752) (cit. on p. 102).

[325] M. Schmidt and H. Lipson, *Distilling Free-Form Natural Laws from Experimental Data*, *Science* **324** (2009) 81, eprint: <https://www.science.org/doi/pdf/10.1126/science.1165893> (cit. on p. 102).

[326] B. Burlacu, G. Kronberger, and M. Kommenda, “Operon C++: An efficient genetic programming framework for symbolic regression,” English, *GECCO 2020 Companion - Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, Association for Computing Machinery, Inc, 2020 1562 (cit. on p. 102).

[327] M. Virgolin, T. Alderliesten, C. Witteveen, and P. A. N. Bosman, *Improving Model-Based Genetic Programming for Symbolic Regression of Small Expressions*, *Evolutionary Computation* **29** (2021) 211, ISSN: 1063-6560, eprint: [https://direct.mit.edu/evco/article-pdf/29/2/211/1921067/evco\\_a\\_00278.pdf](https://direct.mit.edu/evco/article-pdf/29/2/211/1921067/evco_a_00278.pdf), URL: [https://doi.org/10.1162/evco%5C\\_a%5C\\_00278](https://doi.org/10.1162/evco%5C_a%5C_00278) (cit. on p. 102).

[328] T. Stephens, <https://gplearn.readthedocs.io/en/stable/>, Accessed: 27/06/2025 (cit. on p. 102).

[329] S.-M. Udrescu and M. Tegmark, *AI Feynman: a Physics-Inspired Method for Symbolic Regression*, *Sci. Adv.* **6** (2020) eaay2631, arXiv: [1905.11481 \[physics.comp-ph\]](https://arxiv.org/abs/1905.11481) (cit. on p. 102).

[330] B. K. Petersen et al., *Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients*, 2021, arXiv: [1912.04871 \[cs.LG\]](https://arxiv.org/abs/1912.04871), URL: <https://arxiv.org/abs/1912.04871> (cit. on p. 102).

[331] S. Kim et al., *Integration of Neural Network-Based Symbolic Regression in Deep Learning for Scientific Discovery*, *IEEE Transactions on Neural Networks and Learning Systems* **32** (2021) 4166, ISSN: 2162-2388, URL: [http://dx.doi.org/10.1109/TNNLS.2020.3017010](https://dx.doi.org/10.1109/TNNLS.2020.3017010) (cit. on p. 102).

[332] J. Kubalík, E. Derner, and R. Babuška, *Toward Physically Plausible Data-Driven Models: A Novel Neural Network Approach to Symbolic Regression*, *IEEE Access* **11** (2023) 61481 (cit. on p. 102).

[333] T. Han and Y. Li, *Genuine CP-odd Observables at the LHC*, *Phys. Lett. B* **683** (2010) 278, arXiv: [0911.2933 \[hep-ph\]](#) (cit. on p. 103).

[334] I. I. Y. Bigi, Y. L. Dokshitzer, V. A. Khoze, J. H. Kuhn, and P. M. Zerwas, *Production and Decay Properties of Ultraheavy Quarks*, *Phys. Lett. B* **181** (1986) 157 (cit. on p. 103).

[335] A. Czarnecki, M. Jezabek, and J. H. Kuhn, *Lepton Spectra From Decays of Polarized Top Quarks*, *Nucl. Phys. B* **351** (1991) 70 (cit. on p. 103).

[336] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Phil. Trans. Roy. Soc. Lond. A* **231** (1933) 289 (cit. on pp. 104, 130).

[337] M. Davier, L. Duflot, F. Le Diberder, and A. Rouge, *The Optimal method for the measurement of tau polarization*, *Phys. Lett. B* **306** (1993) 411 (cit. on p. 104).

[338] V. Hankele, G. Klamke, D. Zeppenfeld, and T. Figy, *Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC*, *Phys. Rev. D* **74** (2006) 095001, arXiv: [hep-ph/0609075](#) (cit. on p. 104).

[339] A. V. Gritsan et al., *Snowmass White Paper: Prospects of CP-violation measurements with the Higgs boson at future experiments*, (2022), arXiv: [2205.07715 \[hep-ex\]](#) (cit. on p. 123).

[340] F. Renga, *Status of the muEDM experiment at PSI*, *PoS ICHEP2024* (2025) 465, arXiv: [2409.20050 \[hep-ex\]](#) (cit. on p. 123).

[341] J. Alexander et al., *Status of the Proton EDM Experiment (pEDM)*, (2025), arXiv: [2504.12797 \[hep-ex\]](#) (cit. on p. 123).

[342] R. Bause et al., *Prospects for measuring the electron's electric dipole moment with polyatomic molecules in an optical lattice*, *Physical Review A* **111** (2025), ISSN: 2469-9934, URL: <http://dx.doi.org/10.1103/81tl-7wsb> (cit. on p. 123).

[343] T. Higuchi et al., “Neutron EDM Experiment with an Advanced Ultracold Neutron Source at TRIUMF,” 2025, arXiv: [2507.05278 \[physics.ins-det\]](#) (cit. on p. 123).

[344] P. Ferreira, I. P. Ivanov, E. Jiménez, R. Pasechnik, and H. Serôdio, *CP4 miracle: shaping Yukawa sector with CP symmetry of order four*, *Journal of High Energy Physics* **2018** (2018), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP01\(2018\)065](http://dx.doi.org/10.1007/JHEP01(2018)065) (cit. on p. 124).

[345] H. Georgi, *Lie Algebras In Particle Physics: from Isospin To Unified Theories*, 2018 (cit. on p. 124).

[346] C. Degrande et al., *UFO - The Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012) 1201, arXiv: [1108.2040 \[hep-ph\]](#) (cit. on p. 127).

## Bibliography

---

[347] G. Cowan, K. Cranmer, E. Gross, and O. Vitells,  
*Asymptotic formulae for likelihood-based tests of new physics*,  
*Eur. Phys. J. C* **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)],  
arXiv: [1007.1727 \[physics.data-an\]](https://arxiv.org/abs/1007.1727) (cit. on p. 130).

[348] S. S. Wilks,  
*The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*,  
*The Annals of Mathematical Statistics* **9** (1938) 60,  
URL: <https://doi.org/10.1214/aoms/1177732360> (cit. on p. 131).

---

# List of Figures

---

2.1	SM particle content	4
2.2	Mexican hat potential	12
3.1	ATLAS and CMS detectors	24
3.2	Event display from <code>Pythia8</code>	27
3.3	Higgs production channel cross sections	29
3.4	Higgs production channel Feynman diagrams	30
3.5	Higgs decay channel Feynman diagrams	31
3.6	CMS cross section measurements	33
3.7	ATLAS coupling strength measurements	34
3.8	STXS Stage 1.2	37
4.1	Parameter space of a decision tree	40
4.2	Adaboost algorithm	41
4.3	Perceptron structure	44
4.4	PySR function tree	48
4.5	PySR crossover operation	48
4.6	SymbolNet structure	50
4.7	SymbolNet schematic	51
5.1	ggF2j Feynman diagrams	56
5.2	ggF2j background Feynman diagrams	56
5.3	Flowchart for the ggF2j analysis	59
5.4	Exemplary features for the signal-background separation	60
5.5	Metrics for the ggF2j-classifier training	61
5.6	Metrics for the VBF-classifier training	61
5.7	Scores of the signal-background classifiers	61
5.8	$\Delta\phi_{jj}$ in the ggF2j-SR and VBF-SR	62
5.9	$\mathcal{CP}$ classifiers in the ggF2j-SR	63
5.10	$\mathcal{CP}$ classifiers in the VBF-SR	64
5.11	Exclusion limits for the Higgs-gluon coupling in the ggF2j-SR	65
5.12	Exclusion limits for the Higgs-gluon coupling in the VBF-SR	67
5.13	Exclusion limits when both signal regions are combined	69
5.14	Projected limits for $\mathcal{L} = 300 \text{ fb}^{-1}$ and $\mathcal{L} = 3000 \text{ fb}^{-1}$	70

5.15	SHAP values for the $\mathcal{CP}$ -odd classifiers	71
5.16	SHAP values for the $\mathcal{CP}$ -even classifiers	71
5.17	Exclusion limits with added $\mathcal{CP}$ violation in the $HVV$ coupling	72
5.18	Exclusion limits for the $\mathcal{CP}$ -mixing angle	73
5.19	Limits obtained from <code>HiggsTools</code>	74
6.1	Boosting procedure	79
6.2	Parton-level distributions of basic kinematic variables	81
6.3	Parton-level distributions of the $b_i$ variables	82
6.4	Parton-level distributions of $\phi_C$ and $ \cos \theta^* $	83
6.5	Selected significances at $\alpha_t = 35^\circ$	88
6.6	Selected significances at $\alpha_t = 45^\circ$	89
6.7	Distributions with optimized binning in the lab frame	91
6.8	Distributions with optimized binning in the $t\bar{t}$ frame	92
6.9	Significances at $\alpha_t = 35^\circ$ with optimized binning	92
6.10	Distributions of the BDT score	93
6.11	Exclusion limits for the top-Yukawa coupling	95
6.12	Exclusion limits for the top-Yukawa coupling from BDT	96
6.13	Non-Higgs background distributions	98
6.14	Combined exclusion limits	100
6.15	Proposal of $t\bar{t}H$ STXS extension	101
7.1	Asymmetry of the $\Delta\phi_{jj}$ variable in VBF	106
7.2	Asymmetry of the PySR variable in VBF	107
7.3	Asymmetry of the <code>SymbolNet</code> variable in VBF	107
7.4	Classifier significances for various $c_{\phi\widetilde{W}}$	109
7.5	Classifier significances for different amounts of training data	110
7.6	Schematic of <code>SymbolNet</code> to incorporate TPs	111
7.7	$\epsilon$ -tensors for $t\bar{t}H$	112
7.8	<code>SymbolNet</code> prediction for a $\mathcal{CP}$ -odd $t\bar{t}H$ variable	112
7.9	Reconstruction of $\cos \theta^*$ with PySR and <code>SymbolNet</code>	117
7.10	MSE losses in scenarios 1 to 6	118
7.11	Predictions of PySR and <code>SymbolNet</code> in scenario 6	119
7.12	$\mathcal{CP}$ sensitivity in scenarios 1 to 6	120
7.13	$\mathcal{CP}$ sensitivity for various $\alpha_t$	121
D.1	ggF2j distributions for the signal-background classifiers	134
D.2	ggF2j distributions for the $\mathcal{CP}$ classifiers	135
D.3	$t\bar{t}H$ parton-level distributions in the Higgs rest frame	136
D.4	$t\bar{t}H$ parton-level distributions in the $t\bar{t}$ rest frame	137
D.5	$t\bar{t}H$ parton-level distributions in the $t\bar{t}H$ rest frame	138
E.1	$\mathcal{CP}$ classifier uncertainties in the ggF2j-SR	139
E.2	$\mathcal{CP}$ classifier uncertainties in the VBF-SR	140
F.1	Full significance tables for $t\bar{t}H(\rightarrow \gamma\gamma)$	142

## List of Figures

---

F.2	Full significance tables for $t\bar{t}H$ (multilep.)	143
F.3	Full significance tables for $t\bar{t}H(\rightarrow b\bar{b})$	144
F.4	Full significance tables for the combination of channels	145

---

# List of Tables

---

3.1	LHC runs and luminosities	24
4.1	SymbolNet parameter definitions	50
5.1	Cutflow table ggF2j	58
6.1	$t\bar{t}H$ observable definitions	80
6.2	$t\bar{t}H$ yields per channel	84
6.3	Uncertainties per channel	87
6.4	$t\bar{t}H$ optimized variable binning	90
6.5	Comparison of $t\bar{t}H$ significances	94
7.1	VBF $\mathcal{CP}$ -odd classifier significances	108
7.2	Benchmark scenarios for SR in $t\bar{t}H$	114
7.3	SymbolNet hyperparameters for $t\bar{t}H$	114
7.4	SymbolNet function sets for $t\bar{t}H$	115
B.1	Syntax for ggF2j, VBF, and $VH$ event generation	128
B.2	Syntax for $t\bar{t}H$ event generation	129

---

## Acknowledgements

---

This thesis marks the end of a roughly 3.5 year-long journey in which I had a lot of fun, learned tons of new stuff, and had the privilege of meeting and working with many amazing people. Therefore, I would like to take a few words to express my gratitude.

First, I want to say a big thanks to my supervisor, Elina Fuchs, whom I had the pleasure of meeting while working on my master's thesis and who coincidentally started her own group right when I finished my master's. I am very grateful for the opportunity to work on this thesis under her guidance and for her open ear when it comes to discussions about ongoing projects, new projects, or my academic career. She is arguably the person with the most positive attitude I have met in physics, and I hope that she will keep this positivity in the future. It made working in her group even more enjoyable.

Next, I would like to thank Tilman Plehn, who endured me when I first bothered him when applying for a PhD position and then again for becoming my second supervisor. He made Heidelberg what feels like the international capital of machine learning in HEP and invited me to this beautiful city several times. I enjoyed every stay, and I look forward to discussions in the future.

A special thanks also goes out to Philip Bechtle, who already supervised my master's thesis and now agreed to work himself through another thesis of mine. He was my first contact person when I got in touch with particle physics and set the first stones of this path. I wish him and his many projects, both in and outside of particle physics, all the best.

I would like to say a huge thank you to Henning Bahl, who shaped my scientific career to date like no other. He was always available when I had questions or needed an opinion, and he proofread this entire thesis. I could not have asked for a better unofficial supervisor, and I wish him all the best for the future, wherever it may lead him.

Then, I want to thank Christian Ospelkaus for taking the time to act as the head of my PhD committee.

Even if research is my passion, it alone does not bring food to the table. Therefore, I would like to acknowledge the funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC-2123 "QuantumFrontiers" — 390837967. Furthermore, I also acknowledge the "indirect funding" by being allowed to use the resources of the LUIS HPC cluster. It is funded by the DFG — Projektnummern INST 187/742-1 FUGG and INST 187/592-1 FUGG.

Although I joined Elina's group as her first PhD student, the group quickly grew, and I am happy that many amazing people joined me on my journey in Hannover. You all made my time here very special. First, I want to thank Jan Richter and Agnese Mariotti for being my two PhD partners in crime and making sure I do not get lost in the CERN basement. Thank you also to our two Post-Docs, Fiona Kirk

## Acknowledgements

---

and Nathaniel Sherrill, who I am sure will be excellent future group leaders. Thank you to Marc Hannig and Vincent Riechers for being some of the most independent young researchers I know and for making supervising their theses a breeze. I also enjoyed the time with all the other master students of the group, so thank you to Brian, Althea, Apostolos, Finn, and Prithvi, who I am sure have great careers ahead of them. Finally, a thank you goes to our two "adopted" group members, Jan and Alex.

I was very lucky to join a collaboration with the people from CEA Saclay. First, I want to thank Frédéric Déliot, who greatly supported my Post-Doc applications, and Matthias Saimpert, who guided me through the rocks of Fontainebleau for the first time, for inviting me and starting this collaboration. A special thanks goes out to Alberto Carnelli and Anastasia Kotsokechagia, who made me feel at home in their group in such a short period of time. Furthermore, I would like to thank Vera Maiboroda and Tristan Daret for making sure I survive the French Alps, Marianna Liberatore for her awesome drawing skills, and Peter Fitzhugh, without whom I would not be able to pitch a tent. Thank you also to everyone else for adopting me into your group, particularly Xiang, Tanguy, and Ema.

I am very grateful to Klaus Desch for always providing me with an office space in Bonn, even after the end of my master's thesis. Thank you also to Lena, Jan, and Murillo for always making me want to come back to Bonn.

A huge thanks goes out to my girlfriend Lara, who has been my emotional support over all of this time and who agreed to move with me, despite not having the safety of a signed contract that I did. She is the kindest person I know, and I am very happy to have her by my side. Without her, this thesis would not have been possible.

Another emotional support has been my dog Mara, who became a regular in the office and always brought joy to everyone. She has attended so many physics meetings that she deserves a PhD of her own and only the best treats.

I would like to thank my entire family and my friends who always made me look forward to visiting Köln and providing exactly the distraction from work that I needed. One day, they will understand that my thesis is not about hiccups (in German, the sound associated with it is "Higgs"). I am very grateful to my mum for being the most fun person at parties and to my dad for showing me the best hobby out there. Thank you also to my grandparents for always spoiling me, and to my sister for all the fun little competitions. I owe you all much more than I could ever pay back.

In memory of Peer Fischer, who taught me everything I know about the guitar and who left us too soon.

# MARCO MENEN | CV

Contact | [marco.menen@itp.uni-hannover.de](mailto:marco.menen@itp.uni-hannover.de)

## Education

<b>PhD Student</b> - PTB Braunschweig / University of Hanover	<a href="#">From 02/2022</a>
· Examine $\mathcal{CP}$ violation in the Higgs couplings to the top quark and vector bosons	
· Implement machine learning techniques to improve current experimental analyses	
· PhD thesis: Advancing methods for probing $\mathcal{CP}$ violation in Higgs couplings	
<b>Physics, M.Sc. (1.0, with honors)</b> - University of Cologne / Bonn	<a href="#">10/2019 - 12/2021</a>
· Master physics courses with special focus on particle, nuclear, and astrophysics	
· Master thesis: Constraining possible $\mathcal{CP}$ -admixtures in modified Higgs sectors using HiggsSignals	
<b>Physics, B.Sc. (1.2)</b> - University of Cologne	<a href="#">10/2016 - 09/2019</a>
· Bachelor thesis: Determination of projectile number with energy spectra of elastically scattered protons between 2 and 15 MeV	

## Awards and scholarships

<b>Higgs 2024: Price for best parallel session talk</b>	<a href="#">2024</a>
<b>Travel grant by PTB</b>	<a href="#">2024</a>
<b>Travel grant by the Graduate Academy of LUH</b>	<a href="#">2023</a>
<b>LHCP 2023: ECR grant</b>	<a href="#">2023</a>
<b>Deutschlandstipendium</b>	<a href="#">10/2019 - 09/2021</a>
<b>Deutschlandstipendium</b>	<a href="#">04/2017 - 09/2017</a>

## Teaching experience

<b>Python introductory course</b> - University of Hanover	<a href="#">08/2024</a>
· 1-week long introduction to Python for undergraduate students	
<b>Research assistant</b> - University of Bonn	<a href="#">09/2021 - 11/2021</a>
· Student supervisor in the preparatory physics course	
<b>Research assistant</b> - University of Cologne	<a href="#">11/2019 - 09/2020</a>
· Student supervisor in experimental physics and physics for medicine students	
<b>Student assistant</b> - University of Cologne	<a href="#">04/2018 - 09/2018</a>
· Student supervisor in experimental physics	

## Talks

<b>DPG spring meeting 2025</b> - Göttingen, Germany	<a href="#">04/2025</a>
<b>Higgs 2024</b> - Uppsala, Sweden	<a href="#">11/2024</a>
<b>DESY Theory Workshop</b> - Hamburg, Germany	<a href="#">09/2024</a>
<b>DPG spring meeting 2024</b> - Karlsruhe, Germany	<a href="#">03/2024</a>
<b>ITP Seminar (invited)</b> - Heidelberg, Germany	<a href="#">11/2023</a>
<b>CERN Higgs cross section working group 2&amp;3</b> - online	<a href="#">09/2023</a>
<b>LHCP 2023</b> - Belgrade, Serbia	<a href="#">05/2023</a>
<b>Physics at the Terascale</b> - Hamburg, Germany	<a href="#">11/2022</a>
<b>Maria Laach autumn school for high energy physics</b> - Glees, Germany	<a href="#">09/2022</a>
<b>DPG spring meeting 2022</b> - online	<a href="#">03/2022</a>

## List of publications

- H. Bahl et al.,  *$\mathcal{CP}$ -Analyses with Symbolic Regression*, SciPost Phys. (submitted) (2025), arXiv: [2507.05858 \[hep-ph\]](#)
- H. Bahl et al.,  *$\mathcal{CP}$ -sensitive simplified template cross-sections for  $t\bar{t}H$* , JHEP **10** (2024) 214, arXiv: [2406.03950 \[hep-ph\]](#)
- H. Bahl et al., *Classifying the  $CP$  properties of the  $ggH$  coupling in  $H + 2j$  production*, SciPost Phys. Core **8** (2025) 006, arXiv: [2309.03146 \[hep-ph\]](#)
- H. Bahl et al., *Constraining the  $CP$  structure of Higgs-fermion couplings with a global LHC fit, the electron EDM and baryogenesis*, Eur. Phys. J. C **82**.7 (2022) 604, arXiv: [2202.11753 \[hep-ph\]](#)