Searches for CP Violation in the Higgs Sector with the $H \to ZZ^* \to 4l$ Decay Channel in the ATLAS Experiment at the Large Hadron Collider

by

YANG, Tianyi

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in Physics

August 2023, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Yang Tianyi

YANG, Tianyi

4 August 2023

Searches for CP Violation in the Higgs Sector with the $H \to ZZ^* \to 4l$ Decay Channel in the ATLAS Experiment at the Large Hadron Collider

by

YANG, Tianyi

This is to certify that I have examined the above PhD thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.

Prof. Kirill Prokofiev, Thesis Supervisor

Prof. Jiannong Wang, Department Head

Department of Physics
4 August 2023

Acknowledgement

After many years of working on this project towards the end, I would like to express my gratitude to all the people who have helped me along the way.

First of all, I would like to thank my supervisor, Professor Kirill Prokofiev. He is a very nice and patient supervisor. He has been with me throughout the entire project. During my years of study, he has always provided me with very useful comments and guided me to view the project in a more global picture. He has also spent a lot of time helping me with my thesis work and presenting my work in the meetings. His accurate and practical views have inspired me a lot from time to time. Without his great help, I could not finish this thesis smoothly. I feel very fortunate to have worked with Kirill on my Ph.D. project. He has also been very helpful with administrative issues and has given me the opportunity to work onsite at CERN.

I would like to thank our colleagues in the Hong Kong ATLAS group. Professor Ki Lie has helped me countless times in solving my confusion in the work, providing me with useful knowledge about high-energy physics analysis, sharing his enlightening opinions and helping me a lot with the paperwork. He also picked me up at the Geneva airport and helped me a lot in my CERN life. Professor Luis Roberto Flores-Castillo always holds meetings for the team and keeps up with our progress. Mr. Jiawei Wang has cooperated with me in this analysis and previous projects. He is always glad to share his excellent and firm frontline work experience with me when I encounter difficulties. Doctor Jianhuan Xiang spends a lot of time discussing my confusion in physics, programming, and anxiety when the research does not progress smoothly. His passion for machine learning techniques also inspires me a lot.

I would like to thank the Hong Kong University of Science and Technology for giving me the opportunity to do my Ph.D. project in high-energy physics. The University provides me with useful lectures to gain the knowledge that helped me in my research, the opportunity to join the IAS conference every year, and the chance to travel to CERN to work onsite closely with the physicists there. The teaching experience has also given me more ideas to convey my knowledge and work. The fellowship that the university provides has allowed me to concentrate on my scientific research.

I would like to thank the ATLAS HZZ CP group, especially Professor R.D. Schaffer and

Doctor Maximilian Emanuel Goblirsch-Kolb. R.D. guides the route of the whole project. He always integrates our work in different aspects. He has also spent his private time helping me with additional checks on the analysis and helping me through the difficulties in generating samples. Max has excellent programming ability. He has upgraded our simple scripts into a very powerful and fast analysis program. He also helped me a lot in coding on my working part. The people in our group, including Antoine, Chiara, Jiawei, Varena and Yi-Lin have all helped me a lot in my research and explained their studies to me very patiently.

In addition, I would also like to show my special gratitude to my friends for encouraging and helping me through years of study. Doctor Yuchen Yan is a graduate of computer science at the Hong Kong University of Science and Technology. He is my programming savior. He has spent a lot of time introducing new features of C++ to me, leading me to the machine learning techniques, recommending me the Windows subsystem of Linux and helping me understand a lot of external packages. Doctor Yucheng Qiu is a graduate of high-energy physics at the Hong Kong University of Science and Technology. He has solid knowledge and a deep understanding of theoretical physics. He has solved many of my doubts about the theoretical background of my research.

TABLE OF CONTENTS

摘要

自从在位于CERN的Large Hadron Collider实验中发现Higgs玻色子，对Higgs玻色子性质的测量不断开展。这些测量的结果均符合标准模型的预测。在Higgs玻色子的性质中目前仍未发现新物理的迹象。这些测量受限于数据的量。对于持续增长的数据集的继续测量可以进一步压缩这些性质上的不确定性。这将有助于实验上更加精准地探测对标准模型的偏移，并指明寻找新物理的方向。

这项研究目的在于通过对于ATLAS实验在2015-2018年间收集的数据的分析，寻找Higgs玻色子的耦合中潜在的被称为CP破坏的对电荷共轭与宇称反演的对称性的违反。在Higgs玻色子的不同衰变通道中，尽管其支链比较为保守，通过Z玻色子到达四费米子的衰变模式是一个Large Hadron Collider中非常“干净”的通道。这个通道中的事件与大量的多喷注背景可以较好区分。而其末态粒子的电子和μ子可被以很高的效率完整重建。这些特性让这个通道非常适宜对于Higgs玻色子的玻色子耦合测量。除此之外，在Large Hadron Collider中的第二高的Higgs玻色子产生渠道的矢量玻色子融合中具有对于CP破坏高灵敏度的喷注。为优化测量的灵敏度，它们的信息也会被和四费米子衰变的信息共同使用。

这项研究中搜寻的Higgs玻色子和Z玻色子耦合中潜在的CP破坏通过对于超越标准模型的潜在CP-奇性的耦合的测量来进行。本研究中，这些要测量的CP-奇性的耦合通过标准模型有效场论的框架来参数化。这项测量的方法基于对归一化的轻子和喷注的动力学观测量的分布的似然性拟合。

这项研究的结果在两倍标准差的偏离以内和标准模型的预期符合。数据的统计涨落是这项分析中最主要的不确定性来源。该分析结果对于CP-奇性耦合的约束超越了以往基于产额的统计测量。这个测量将更有助于深化对于Higgs玻色子耦合结构的理解。

Abstract

Since the discovery of the Higgs boson at the Large Hadron Collider at CERN by the ATLAS and CMS experiments, measurements of its properties have been carried out. These measurements have yielded results in good agreement with the Standard Model expectation. Signs of new physics have not been discovered in the properties of the Higgs boson. The precision of these measurements is limited by the amount of data. Continuous measurements on the dataset of increasing size would be able to constrain the properties with shrinking uncertainties. This would enable experiments to detect potential deviations from the Standard Model more precisely and indicate the direction of future new physics searches.

This study aims to search for the potential violation in simultaneous charge conjugation and parity inversion, known as CP violation, in the Higgs boson couplings with the 2015-2018 ATLAS data. Among different Higgs boson decay modes, the four-lepton decay mode through Z bosons is a "clean" channel at the Large Hadron Collider, despite its moderate branching ratio. Events in this channel are well-separated from heavy multi-jet backgrounds. The electrons and muons in this final state are fully reconstructable with high reconstruction efficiency. These features make it a suitable channel to study the bosonic couplings of the Higgs boson. In addition, vector boson fusion, as the subleading Higgs boson production channel at the Large Hadron Collider, produces jets that are sensitive to CP violation. Their information is used together with the four-lepton decay information in this study to optimize sensitivity.

In this study, the potential CP violation in the Higgs boson coupling to Z bosons is searched through the measurement of the CP-odd couplings beyond the Standard Model. The CP-odd couplings measured are parametrized using the effective field theory framework. The measurement is based on the likelihood fit method of the normalized kinematic observable distributions of the leptons and jets in the final state.

The results of this study are consistent with the Standard Model expectation within two standard deviations. The statistical fluctuations of data are the dominant source of uncertainties in this analysis. More strict constraints than the previous yield-based analysis have been set in the CP-odd effective couplings. This measurement inspires a deeper understanding of the coupling structure of the Higgs boson.

Introduction

In 2012, a new particle was discovered at the Large Hadron Collider at CERN by the ATLAS and CMS experiments [1, 2]. The properties of this new particle are consistent with the predictions of the Higgs boson in the Standard Model.

The Standard Model is the minimal theory of particle physics describing the electromagnetic, strong and weak interactions well. It is built on quantum field theory and the principle of gauge symmetries [3]. In the Standard Model, the Higgs mechanism is proposed to solve the difficulty in the weak interaction boson mass generation [4]. The Higgs mechanism introduces additional scalar fields that break the electroweak gauge symmetry spontaneously. The weak interaction bosons $W^{\pm}$ and Z acquire their mass from the vacuum expectation values of the scalar fields in this process. The mass of Standard Model fermions is simultaneously generated through the Yukawa couplings to the introduced scalar fields [5]. The Higgs mechanism in the Standard Model predicts the existence of a single scalar particle. The scalar particle named the Higgs boson had been searched for a long time.

In continuing analyses, the properties of the Higgs boson were gradually revealed. In the spin and parity test in the vector boson decay channels, the scalar particle hypothesis of the Higgs boson is favoured against the pseudoscalar, vector or tensor particle hypotheses [6–8]. The cross-sections and couplings of the Higgs boson to different particles through multiple final states currently measured do not have significant deviations from the Standard Model predictions [9–14].

Despite the discovery of the Higgs boson at the Large Hadron Collider has made the Standard Model a great success, there are still puzzles left with the Standard Model. From cosmic observations, there is a sign of more matter than antimatter in the Universe [15]. This is known as matter-antimatter asymmetry. The production mechanism of the extra baryons named baryogenesis is a potential solution to this puzzle. The simultaneous violation of the charge conjugation and parity inversion symmetry, known as CP violation, is necessary for the baryogenesis [16]. The measured CP violation within the Standard Model processes is not strong enough. This makes it difficult to match the matter-antimatter asymmetry from this observation. The sources of CP violation beyond the Standard Model could help with the explanation [17].

In this study, CP violation is searched in the Higgs boson couplings to Z bosons. The extension of the Higgs sector beyond the Standard Model introduces potential CP violation to the couplings of the Higgs boson to other elementary particles in the Standard Model. This study uses the effective field theory approach to make a model-independent measurement [18]. In this approach, physics beyond the Standard Model is parametrized to all possible forms of operators obeying the Standard Model gauge with the least assumptions. With the CP-even coupling in the Standard Model, if non-zero CP-odd couplings were measured, CP violation would be achieved. This study is based on the newly submitted ATLAS paper in which I contributed to the work as a member of the ATLAS Higgs CP group [19].

The Higgs boson decay to four leptons through Z bosons, labeled as $H \to ZZ^* \to 4l$, is a good channel to study the Higgs boson coupling to vector bosons. In this analysis, the leptons in the final state only refer to electrons, muons and their antiparticles unless otherwise specified. The properties of τ lepton are significantly different from these particles. The events with τ lepton in the final state were not used in this study. The correlated final state leptons in this channel are well-isolated from the heavy multi-jet backgrounds at the Large Hadron Collider. This makes the channel “clean" to study. There are no neutrinos in the final state, which cannot be observed by the ATLAS detector and thus leave missing momentum and energy. This makes the final state fully reconstructable.

In this study, the CP violation measurement focuses specifically on the vector boson fusion channel. Among the main production channels of the Higgs boson at the Large Hadron Collider, the merging of two virtual weak bosons emitted from quarks into the Higgs boson is known as vector boson fusion. The vector boson fusion channel involves the Higgs boson to Z bosons couplings in the production vector. This gives this production channel additional information in effective couplings over the other channels. In this channel, there are at least two jets produced together with the Higgs boson. The kinematic observables of these jets are of high CP sensitivity.

The non-Higgs boson events with four leptons in the final states are the backgrounds in this analysis. The events with four leptons originating from the primary vertices are named irreducible background events. These leptons are named prompt leptons. The events with non-prompt leptons originating from the decay of particles traveling in the detector or misidentified fake leptons are named reducible background events. The dominant background in this study is the events with non-resonant Z bosons or photons from gluons or quarks collisions, labeled as ZZ^*. The other irreducible backgrounds include the triple vector boson events and vector boson produced together with top quarks events. The reducible backgrounds include Z boson together with jets production events and top quark pair production events. These contributions are very small compared to the ZZ^* backgrounds.

The measurement of the potential CP-odd couplings is based on the likelihood scan. Distributions of the observables are compared between the data and the hypotheses with varied CP-odd couplings. The differences in the likelihood values scanned through the full phase space of the couplings of interest are presented. The models of the hypotheses are predicted by interpolations from Monte Carlo simulation for a few optimized coupling points. The main uncertainty in this measurement is the statistical fluctuations in the ATLAS data. The systematical uncertainties from the theoretical modeling, experimental measurement, and statistical fluctuations in the Monte Carlo model building are estimated. The impact of systematical uncertainties is negligible.

This study is a group work done by the HZZ analysis group in the ATLAS collaboration. As a complicated research, the measurement contains different jobs including the preparation of the Monte Carlo dataset, determination of the measured parameters, optimizations of analysis strategies, and estimation of uncertainties. In this analysis, I proposed the production and decay combined measurement and joined the observable binning optimization for it. Meanwhile, I joined the estimation of backgrounds and systematical uncertainties. For the closure test, I participated in the examination for higher-order effects with additional toy samples generated as well, which is not included in the result of the measurement.

In this thesis, the first chapter introduces the Standard Model as the theoretical foundation and reviews previous experimental results on Higgs physics. The second chapter explains the role of CP violation in this analysis. The third chapter describes the Standard Model Effective Field Theory as the framework of this analysis. The fourth chapter introduces the structure and performance of the experimental setup. The fifth chapter describes the data and the Monte Carlo simulations. The sixth chapter analyses the methodology, optimizations and uncertainties in this analysis. The seventh chapter shows the measurement results and comparisons with previous and parallel studies. The eighth chapter discusses the prospects of this analysis. The ninth chapter is the summary of this thesis.

Chapter 1

The Standard Model

Particle physics is a relatively new field in investigating the microscopic nature of the world. In 1897, the electron was discovered by Thomson in the study of the cathode rays and its mass-to-charge ratio was measured [20]. The electron was the first fundamental particle discovered. In 1901, Planck's law was published to describe black-body radiation with the assumption of the discrete energy element [21]. This is the beginning of quantum mechanics. In 1905, a quantum theory of light was proposed by Einstein to explain the photoelectric effect [22]. This theory extended the quantum nature to the electromagnetic field. Based on these and other studies, quantum mechanics was formulated and became the fundamental language for particle physics in the 1920s [23–25].

The contemporary theory of particle physics is the Standard Model (SM) [3]. The SM is a theory based on the quantum field theory (QFT), describing three fundamental interactions: the strong, the electromagnetic and the weak interactions. The modern quantum field theory uses quantum fields to describe both matter and interactions. In the QFT, the elementary particles are excitations of the quantum fields. The matter is composed of fermions, which are the excitations of the spinor fields. The interactions correspond to bosons, which are the excitations of the vector fields.

The interactions in the SM are described by the gauge theories. The gauge is the local symmetry on the internal structure of the fields. Different gauges characterize different interactions. The electromagnetic and the weak interactions are described by the unified electroweak theory (EW) [5, 26, 27]. The strong interaction is described by the quantum chromodynamics (QCD) [28].

The fermions of the SM are quarks and leptons. Quarks participate in strong interactions and leptons do not. Vector bosons are the force carriers. The strong interactions are realized by exchanging gluons between quarks. The electromagnetic interactions correspond to the photon exchange between electrically charged particles. In weak interactions, unstable heavy elemen-

tary particles decay to stable elementary particles mediated by $W^{\pm}$ bosons and Z boson.

In the SM, the Higgs mechanism introduces a scalar field to generate the mass of the $W^{\pm}$ and Z bosons and the fermions [4]. In its simplest formulation, the Higgs mechanism predicts the existence of a scalar elementary particle named the Higgs boson.

The SM elementary particles and their quantum numbers including mass, electric charge and spin are summarized in Figure 1.1 [29]. There are three generations of elementary fermions with the same charges and different masses. The up quarks, charm quarks and top quarks are with $2/3$ elementary electric charge, known as up-type quarks. The down quarks, strange quarks and bottom quarks are with $-1/3$ elementary electric charge, known as down-type quarks. The electrons, muons and tau leptons are negative electrically charged leptons. The corresponding electrically chargeless leptons are the neutrinos only involved in weak interactions.

1.1 Gauge field theories

In the QFT, the fermions are described by the excitation of the spinor field denoted by ψ. The free fermion field theory does not realize interactions and gauge invariances in Maxwell's equations. In the formalism of quantum mechanics, the electromagnetic gauge transformation adds a phase on the states [30]. As a generalization, fermions assigned to a gauge are involved in the corresponding fundamental interaction. The theories of this type are named gauge theories.

In a gauge theory, the Lagrangian of the fermion field needs to be invariant under a group of gauge transformations. A gauge transformation on a fermion field is a local unitary transformation with a general form $\psi \rightarrow \psi'(x) = U(x)\psi(x)$, where $U(x)$ is the operator of a gauge transformation. The locality in these transformations requires the gauge invariance to be an intrinsic structure of the field different from the symmetries in the field spacetime configuration. These gauge symmetries are achieved by the introduction of vector fields named gauge fields. The mass terms of the gauge fields are not gauge-invariant and therefore forbidden in the Lagrangian. This means gauge bosons as excitations of these fields must be massless without additional mechanisms introduced.

Quantum electrodynamics (QED) is the simplest gauge theory describing the electromagnetic interactions with the local $U(1)$ gauge group [25]. To make the Lagrangian $U(1)$ gauge-invariant, a single gauge field is required. Photons are the excitations of this field. The conserved charge for the $U(1)$ gauge transformation is the electric charge of the fermion. This theory demonstrates the need for photons and the electric charges from the $U(1)$ gauge invariance.

In the SM, there are multiple gauge structures associated with the fermions. The EW theory follows the $SU(2)_L \times U(1)_Y$ gauge group. The QCD theory follows the $SU(3)$ gauge

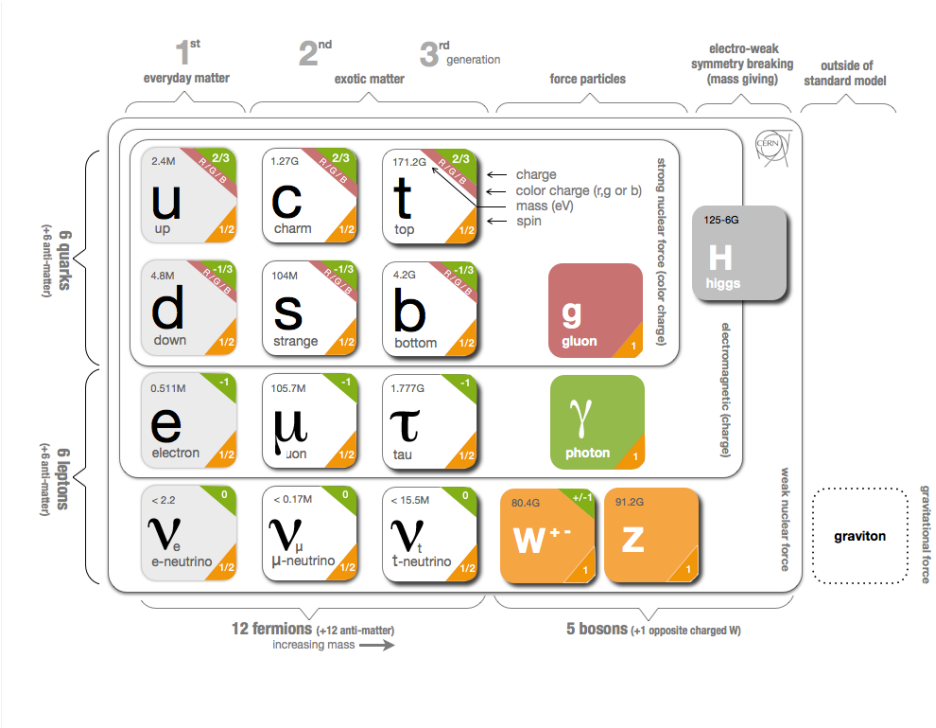


Figure 1.1: The SM elementary particles and quantum numbers.

group. These non-Abelian gauges are described by the Yang-Mills theory [31]. The QCD theory describes strong interactions. The fermions subject to this gauge are the quarks with $SU(3)$ charges named colors. The gauge bosons for this theory are the gluons. The EW theory describes the unified electromagnetic and weak interactions. This theory accumulates contributions from many physicists known as Glashow-Salam-Weinberg (GSW) theory [5, 26, 27]. The subscript L means only left-handed fermions (and right-handed anti-fermions) observe the $SU(2)$ part of the gauge. The subscript Y denotes the $U(1)$ charge known as hypercharge in this theory, distinguished from the electric charge. The electric charge is produced by the combination of $SU(2)_L$ charge named weak isospin and hypercharge.

In Fermi weak interaction, the current is electrically charged [32]. This requires a mixed description of the electromagnetic and weak interactions. Parity violation in the weak interaction was proposed by Lee and Yang [33] and confirmed by the Wu experiment [34]. The fermions described by the spinor fields have the components invariant under the Lorentz transformations in the right-hand and left-hand frames respectively. This property is known as chirality. The chirality of a fermion is inversed under parity transformation. Therefore, parity violation in the weak interaction requires its theory should be chiral, which means left-hand particles and right-hand particles behave differently under weak interaction. The neutral current was predicted by the GSW theory and experimentally confirmed later [35]. The mass production of weak bosons

and chiral fermions needs additional mechanisms to manifest the gauge invariance. The weak boson mass generation problem is achieved by the Higgs mechanism in the GSW theory [4]. This mechanism provides a simple solution to the fermion mass generation as well.

1.2 The Higgs mechanism

The Higgs mechanism generates mass for vector bosons by introducing scalar fields ϕ to break the gauge symmetry [4]. In this mechanism, there are some requirements on the potential of ϕ. The potential distribution over the ϕ configuration conserves the gauge symmetry. The potential attains the minimum value at multiple ϕ points that are not at $\phi = 0$. Since the system always tends to achieve the minimum potential state, there has to be a choice of vacuum at one of the minimum potential points. The Lagrangian of ϕ is re-expanded around the chosen vacuum. The vacuum and the re-expanded form of Lagrangian do not observe the original symmetry anymore. This process is named spontaneous symmetry breaking (SSB) [36]. In the maximum breaking of gauge symmetries, the ϕ fields are gauge transformed to make the expansion with the vacuum expectation value v and a single scalar field defined around the vacuum point. The gauge bosons acquire mass from the vacuum configuration. This new scalar field from the expansion predicts the existence of the Higgs boson.

In the Standard Model, the Higgs mechanism is applied minimally with a doublet of complex scalar fields conserving the EW gauge $SU(2) \times U(1)_Y$. The scalar fields are transformed as:

$$\phi \rightarrow e^{i\alpha^a \tau^a} e^{i\theta/2} \phi, \tag{1.2.1}$$

where τ^a are the generators for the $SU(2)$ group. The symbols α^a and θ are the real numbers indicating the amount of transformation. The SSB process breaks the EW gauge to $U(1)_{EM}$ which is the electromagnetic gauge. A general doublet of complex scalar fields has four degrees of freedom. Three of them are eliminated by the choice of gauge and only one real scalar field around the vacuum point leaves. These eliminated degrees of freedom generate mass for $W^{\pm}$ and Z bosons through the vacuum expectation value of the scalar field. Photons are not coupled to the vacuum expectation value and remain massless upon the gauge choice. The left real scalar field predicts the Higgs boson [4].

The mass generation of chiral fermions is achieved simultaneously by introducing their Yukawa couplings to the ϕ field [5]. The expansion around the vacuum in the SSB applies to the Yukawa coupling terms as well. In this process, the left chiral and right chiral fermions are coupled together by the vacuum expectation value. These terms provide mass to fermions. These mass terms are accompanied by the interaction terms with the Higgs boson.

With the Higgs mechanism included, the Lagrangian of the SM is summarized in the fol-

lowing compact expression:

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}D_\mu\gamma^\mu\psi + \text{h.c.} \\ &+ y_{ij}\bar{\psi}^i\psi^j\phi + \text{h.c.} + |D_\mu\phi|^2 - V(\phi).\end{aligned} \tag{1.2.2}$$

The first term is the kinetic term for the gauge bosons, where $F_{\mu\nu}$ is the field strength tensor. The second term and its Hermitian conjugation are the kinetic terms for the SM fermions, where $\bar{\psi} = \psi^\dagger\gamma^0$ denotes the conjugation of the fermion fields, γ^μ denotes the gamma matrices and D_μ is the covariant derivative. The third term and its Hermitian conjugation are the Yukawa interactions of the fermions and the Higgs field, where y_{ij} is the Yukawa coupling where the indices i and j refer to the fermions. The last two terms are the Lagrangian for the scalar field introduced by the Higgs mechanism, where $V(\phi)$ is the potential term. Through the Higgs mechanism, the mass of chiral theory fermions and weak gauge bosons are explained well. An additional scalar particle is predicted as the Higgs boson. The discovery of the Higgs boson at CERN in 2012 found this last missing piece in the SM, making it a great success [1, 2].

1.3 Higgs boson production and decay at the Large Hadron Collider

This analysis is carried out at the Large Hadron Collider (LHC) at CERN [37]. This is the world's largest proton collider in which the Higgs boson was discovered. The designed centre-of-mass energy for the protons in collisions ranges from 7 TeV to 14 TeV. The production channels of the Higgs boson are determined by the properties of proton collisions.

A proton is not an elementary particle. It is a bound state composed of quarks and gluons through the strong interaction [38]. The quantum numbers of a proton are determined by two up quarks and one down quark. These are the valence quarks. From the quantum fluctuation inside a proton, many gluons and other quarks are created and annihilated. The proton collision events are described by the interactions between these quarks and gluons as a "part" of a proton. This is the parton model [39]. The probability density distribution of different flavour partons involved in the collision is described by the parton distribution function (PDF). The PDF of interacting protons at the LHC is dominated by gluons with a low momentum share among the total proton momentum [40]. The deep inelastic collisions at the LHC are intrinsically gluon-gluon collisions and quark-quark collisions, where the gluon-gluon collisions are dominant.

The dominant production channel of the Higgs boson at the LHC is gluon fusion (ggF). The other important channels include vector bosons fusion (VBF), associated production with W/Z boson (VH), and associated production with heavy quarks [41]. An example of the associ-

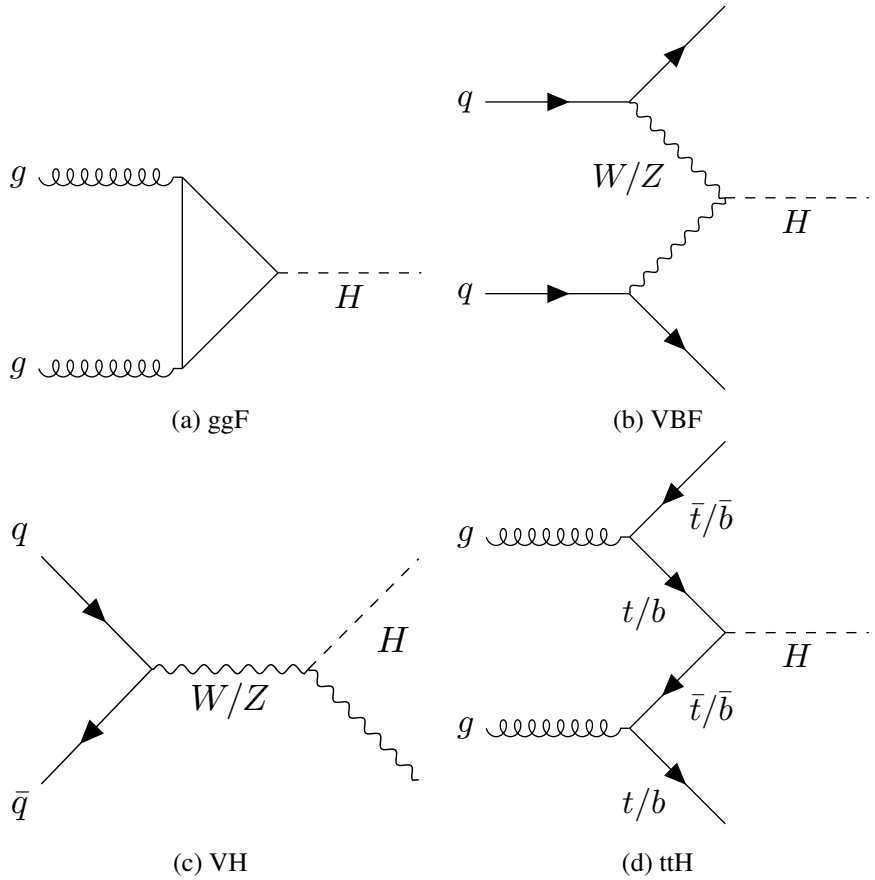
ated production with heavy quarks is top quark pair fusion (ttH). These processes are visualized by the Feynman diagrams [42, 43]. The leading order contributions of ggF, VBF, VH and ttH processes are illustrated in Figure 1.2.

The theoretical predictions of the cross-section of these channels are computed by the LHC Higgs cross-section working group [44]. These cross-section values with the mass of the Higgs boson at 125 GeV assumption are shown in Figure 1.3 as a function of the centre-of-mass energy of protons in the collision [44]. The cross-section computation for the ggF is with QCD accuracy up to next-to-next-to-next-to-leading order (N^3LO) and EW accuracy up to next-to-leading order (NLO). The cross-section computation for the VBF is with QCD accuracy up to next-to-next-to-leading order (NNLO) and EW accuracy up to NLO. The cross-section computation for the VH is with QCD accuracy up to NNLO and EW accuracy up to NLO. The cross-section computation for the ttH is with both QCD and EW accuracy up to NLO. The dataset used in this analysis has the centre-of-mass energy $\sqrt{s} = 13$ TeV., during the second stable operation period known as Run 2. At $\sqrt{s} = 13$ TeV, the ggF, labeled by $pp \to H$ in the plot, has the largest cross-section of these channels. The value is of an order larger than the second dominant VBF channel, which is labeled as $pp \to qqH$. The VH production channel, labeled as $pp \to WH$ and $pp \to ZH$ depending on vector boson type, has a cross-section of about half of the VBF value.

In the VBF and VH channels, the Higgs boson has couplings to $W^{\pm}$ or Z bosons as shown in the Feynman diagrams. The VBF channel has at least two non-resonant jets produced together with the Higgs boson. Kinematic observables of these two jets are sensitive to the Higgs boson beyond the Standard Model (BSM) couplings to vector bosons couplings. This feature makes the VBF channel a suitable channel for measuring the couplings using the information in the production process.

The properties of the Higgs boson are studied with its decay products in the collider. The theoretical predictions of the branching ratio of the Higgs boson decay modes are computed by the LHC Higgs cross-section working group up to NLO EW accuracy [44]. The branching ratio values are shown in Figure 1.4 as a function of the Higgs boson mass [44]. The b quarks decay mode, with about 58% branching ratio, dominates the decay channels around 125 GeV Higgs boson mass. The bosonic decay modes of the Higgs boson have relatively smaller branching ratios. The branching ratio of the $W^{\pm}$ bosons decay mode is about 22%. The value for the decay into Z bosons mode is about 2.6%.

Among decay products of the Higgs boson, the Z boson has leptonic decay modes into a pair of electron and positron denoted by e^- and e^+ and a pair of muon and anti-muon denoted by μ^- and μ^+. This results in a final state with four leptons from the full leptonic decay of the Z bosons decayed from the Higgs boson denoted as $H \to ZZ^* \to 4l$. In the ATLAS



(a) ggF (b) VBF (c) VH (d) ttH

Figure 1.2: Leading order Feynman diagrams for the main Higgs boson production channels at the LHC: (a) ggF (b) VBF (c) VH (d) associated production with heavy quarks.

and CMS experiments, this decay channel has played an important role in the discovery of the Higgs boson [1, 2]. This final state is neutrino-free. This makes it fully reconstructable. The lepton-only final state is well separated from the heavy multi-jet backgrounds at the LHC. The $e^{\pm}$ and $\mu^{\pm}$ are identified with high efficiency at the LHC [45, 46]. These features make this decay channel suitable for precise measurements of the Higgs boson couplings. In this study, the measurement of the Higgs boson couplings focuses on the $H \to ZZ^* \to 4l$ decay channel.

1.4 Experimental results

As the last missing piece of the Standard Model, searching for the Higgs boson has a long history since the prediction from the Higgs mechanism [47]. In the analyses at multiple accelerator experiments, the exclusion limits on the Higgs boson mass have been set and the significance of the potential Higgs boson signal has been studied [48–51]. In 2012, the Higgs boson has been finally discovered by the ATLAS and CMS experiments at the LHC independently based on the

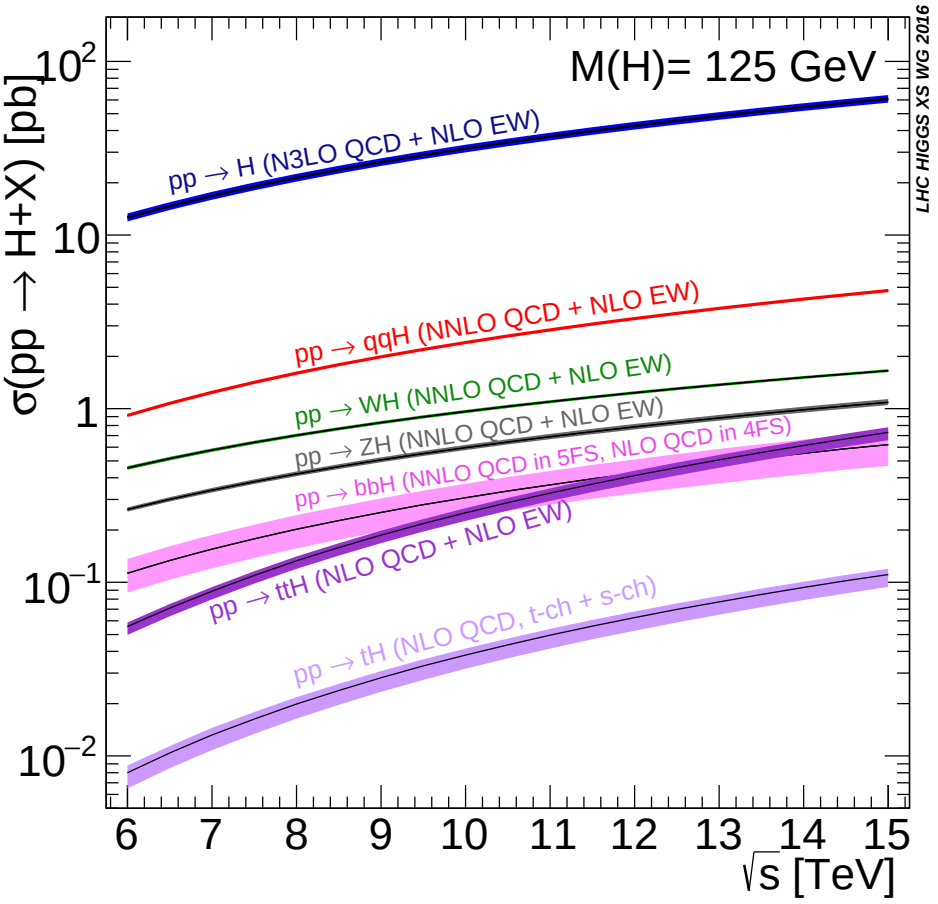


Figure 1.3: Cross-sections of Higgs boson production channels at the LHC as a function of the centre-of-mass energy in pp collisions.

analysis of the 2011-2012 data [1, 2]. Measurements of the properties of the Higgs boson have been carried out since its discovery.

In high-energy physics experiments, the discoveries and measurements are based on statistical analyses from enormous data produced in accelerators. The number of collision events available is crucial to analyses of an experiment. In high-energy physics, the concept of luminosity is introduced to quantify the ability of an accelerator to deliver collision events [52].

The instantaneous luminosity is defined as the proportional factor between the cross-section and the event rate shown as follows:

$$\frac{dR}{dt} = \mathcal{L}\sigma_p, \tag{1.4.1}$$

where dR/dt is the event rate, σ_p is the cross-section of the process of interest and $\mathcal{L}$ is the instantaneous luminosity. This quantity depends only on the properties of the machine. By increasing the beam intensity, more particles can pass through the area that two beams cross in

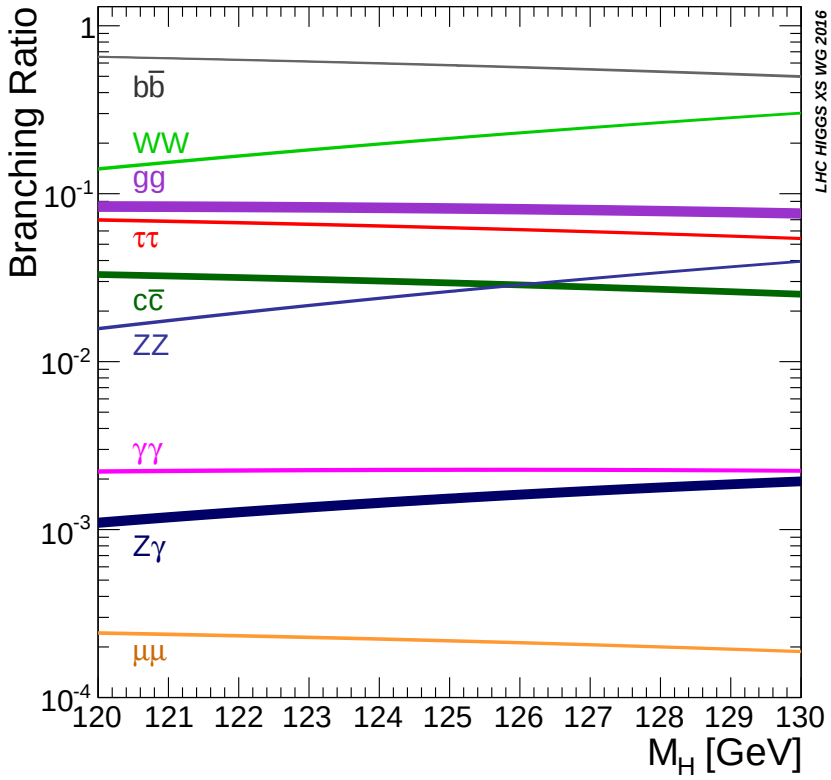


Figure 1.4: Higgs boson decays branching ratios as a function of the Higgs boson mass.

the same period. By squeezing the transverse size of two beams, particles in one beam are more likely to collide with particles in the opposite beam. Both methods can increase luminosity. The expression of the instantaneous luminosity for a head-to-head collision of Gaussian distributed beams in a circular collider is shown as follows:

$$\mathcal{L} = \frac{N_1 N_2 f n_b}{4\pi\sigma_x\sigma_y}, \tag{1.4.2}$$

where N_1 and N_2 are the numbers of protons in each bunch of collision, f is the revolution frequency of each beam in the circular collider rings, n_b is the number of bunches, σ_x and σ_y are the bunch sizes orthogonal to the beam traveling [52]. The integrated luminosity is the integrated value of instantaneous luminosity over time. This quantity is a measure of the total amount of data available in an accelerator.

In 2012, the ATLAS and the CMS experiments at the LHC announced the discovery of a new particle [1, 2] The ATLAS experiment combined the 2011 data at central-mass energy $\sqrt{s} = 7$ TeV of integrated luminosity 4.8 fb^{-1} and 2012 data at $\sqrt{s} = 8$ TeV of 5.8 fb^{-1} received in $H \to ZZ^* \to 4l$, $H \to \gamma\gamma$ and $H \to WW^* \to e\nu\mu\nu$ channels. The obvious excesses with

high mass resolution were observed $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels. The analysis in the $H \to WW^* \to e\nu\mu\nu$ channel found the excess in a broad mass region. The CMS experiment combined the 2011 data at $\sqrt{s} = 7$ TeV of integrated luminosity 5.1 fb^{-1} and 2012 data at $\sqrt{s} = 8$ TeV of 5.3 fb^{-1} received in $H \to ZZ^* \to 4l$, $H \to \gamma\gamma$, $H \to WW^* \to e\nu\mu\nu$, $H \to b\bar{b}$ and $H \to \tau^+\tau-$ channels. The $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channel analyses found obvious excesses with high mass resolution. The $H \to WW^* \to e\nu\mu\nu$ channel analysis found the low mass resolution excess. The analyses in the $H \to b\bar{b}$ and $H \to \tau^+\tau-$ channels did not find clear sign of signal. The result of the ATLAS and the CMS experiment agrees with each other. The observed particle is compatible with the SM Higgs boson of mass 125 GeV hypothesis. The properties of the new particle have been studied in the analyses following.

The mass of the Higgs boson was measured jointly by the ATLAS experiment and the CMS experiment on the analyses using 2011-2012 data in the $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels. The result is $125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})$ GeV [53], where the two errors are statistical uncertainties and systematic uncertainties respectively. The latest Higgs boson mass measurement result is 125.25 ± 0.17 GeV combined the ATLAS and CMS experiments analyses on 2015-2018 data in these channels [54–56].

The spin and parity of the Higgs boson were measured by both the ATLAS and CMS experiments. The behavior of the interactions between the Higgs boson and the other elementary particles under the simultaneous charge conjugation and parity inversion (CP) was tested meanwhile in these studies. In the SM, the Higgs boson is a scalar particle with spin-0 and CP-even couplings. This means the physics is invariant under the CP. The $H \to \gamma\gamma$ process disfavoured the spin-1 hypothesis based on the selection rule [57, 58]. The ATLAS analysis using 2011-2012 data in the $H \to ZZ^* \to 4l$ and $H \to WW^* \to e\nu\mu\nu$ channels excluded the spin-1 hypothesis at more than 99% confidence level [6]. The succeeding study on the $H \to ZZ^* \to 4l$, $H \to WW^* \to e\nu\mu\nu$ and $H \to \gamma\gamma$ channels using 2011-2012 data strongly favoured the spin-0 with CP-even couplings hypothesis against the spin-0 with CP-odd couplings hypothesis and spin-2 hypotheses tested at more than 99.9% confidence level [7]. The CMS analysis using Run 1 data in the $H \to 4l$, $H \to WW^* \to e\nu\mu\nu$ and $H \to \gamma\gamma$ channels had results consistent with the spin-0 with CP-even couplings hypothesis against the spin-1 and spin-2 hypotheses at more than 99.999% and 99.87% confidence level respectively [8]. The CP-odd couplings spin-0 hypothesis was constrained well. The measured spin and parity properties of the Higgs boson in the ATLAS and CMS experiments on Run 1 data are highly consistent with the SM expectations.

There are updated CP measurements of BSM Higgs boson coupling structures in different channels in the ATLAS and CMS experiments. The ATLAS experiment has measured the Higgs boson production cross-section and interpreted them in terms of effective couplings to gauge

bosons using 139 fb^{-1} of data in the $H \to ZZ^* \to 4l$ channel [9]. The CP invariance has been tested in the Higgs boson coupling to τ leptons on the 139 fb^{-1} of data in the $H \to \tau\tau$ channel [10]. The test in the Higgs boson coupling to top quarks has been done using 139 fb^{-1} of data in the $H \to \gamma\gamma$ channel in the ATLAS experiment [11]. The test in the Higgs boson coupling to EW gauge bosons has been done using 139 fb^{-1} of data in the VBF $H \to \gamma\gamma$ channel in the ATLAS experiment [12]. The CMS experiment has measured the CP structure of Higgs boson coupling to top quarks using 138 fb^{-1} of data in the ttH and top quark associated production with $H \to WW^*$ and $H \to \tau\tau$ decay channels [13]. The CP structure in the Higgs boson coupling to τ leptons was measured using 137 fb^{-1} of data in the $H \to \tau\tau$ decay channels in the CMS experiment [14]. The results of these measurements are all consistent with the SM predictions with available data samples. Higher sensitivity is required to further investigate the CP structure in the Higgs couplings.

The analysis described in this thesis is specially designed to investigate the CP structure of the Higgs boson couplings to Z bosons using 139 fb^{-1} of data. The shape distributions of the kinematic observables in the VBF and $H \to ZZ^* \to 4l$ events are used to constrain the effective CP-odd Higgs boson to Z bosons couplings. The use of the shape information is expected to produce results with higher sensitivity from the effective coupling measurement succeeding the ATLAS 139 fb^{-1} cross-section measurement in the $H \to ZZ^* \to 4l$ channel.

Chapter 2

CP violation and baryon asymmetry

2.1 CP violation in the Standard Model

Charge conjugation is the transformation that reverses the charge of particles. If a theory is invariant under the charge conjugation, it has the C symmetry. Parity inversion is the transformation that flips the spatial coordinates of particles. If a theory is invariant under the parity inversion, it has the P symmetry. A theory has the CP symmetry if it is invariant under the simultaneous charge conjugation and parity inversion.

In the SM, the strong interactions and electromagnetic interactions preserve the C symmetry, P symmetry and CP symmetry. The weak interaction in the SM violates the C and P symmetry significantly [5, 26, 27]. In the weak interaction, only left chiral fermions and right chiral anti-fermions involve in the interaction with the weak bosons. The parity violation in the weak interaction was first raised by Lee and Yang in the analysis of the kaon decay [33]. The violation was observed in the Wu experiment through β-decay of cobalt-60 atoms [34]. The cobalt-60 nucleus is radioactive. It decays into an excited nickel nucleus and the nickel nucleus radiates γ-rays to fall into the ground state. If the parity were conserved in this process, the electrons should be emitted along both the polarization direction of the cobalt atom and the reverse for the same probability. In this experiment, a magnetic field was applied to align the polarization of the atoms. The experiment started at a low starting at a low temperature and gradually warmed up. The alignment of the atom polarization became looser during this process. Two γ-ray counters at the equator and around the polar direction were used to monitor the polarization situation of the cobalt atoms. The β-decay electrons were detected along the polar direction. These detections were done again with the magnetic field direction reversed. The result showed for the more aligned atom polarization indicated by the γ-ray counting, the larger share of the electrons was emitted along the magnetic field direction both before and after the magnetic field reversion. This indicated the existence of the parity violation in the β-decay.

As a step forward, the CP symmetry is violated as well in the weak interactions with the existence of fermions of multiple generations in the Standard Model. Both CP-even and CP-odd mesons are constructed from a pair of quarks with each quark from a different generation. For example, the neutral kaon constructed by a pair of down quark (d) and strange quark (s) and their anti-particles has the CP-even state $(d\bar{s}+s\bar{d})/\sqrt{2}$ and the CP-odd state $(d\bar{s}-s\bar{d})/\sqrt{2}$ with different lifetimes. The bar above the quark denotes the anti-particle. The fermion masses are generated from their Yukawa couplings to the scalar field in the Higgs mechanism [5]. The mass eigenstates of the quarks are mixed from their weak interaction eigenstates. Therefore, the weak interaction between massive quarks crosses generations. The weak interactions between three-generation massive quarks in the SM are characterized by the CKM matrix [59]. The small off-diagonal entries are measured for the couplings between quarks from different generations. These CP-even and CP-odd mesons oscillate between each other mediated by the weak bosons following the CKM matrix [60]. CP violation is generated from this process.

In the Brookhaven National Laboratory, the "indirect" CP violation was first observed in an experiment studying the decay of the long-lifetime kaons [61]. If the CP were conserved, the final state of CP-odd neutral kaon hadronic decay should consist of three pions. The one for CP-even neutral kaon should consist of two pions. This leads to a longer lifetime for the CP-odd neutral kaon. There was a branching ratio of about two in a thousand for the long-lifetime kaon decaying into two pions observed in this experiment. This indicated only the long-lifetime neutral kaon should not be pure CP-odd but with a CP-even component. The CP violation observed could be due to the state mixing but not the decay coupling and therefore was named indirect.

The succeeding experiment studies have been carried out to search for the so-called "direct" CP violation in the weak interaction vertex. The KTeV experiments at Fermilab and the NA48 experiments at CERN first observed the evidence for the direct CP violation in the decay of kaons [62, 63]. The theoretical parametrization indicated the observable for direct CP violation searching as follows:

$$\Re(\epsilon'/\epsilon) = \frac{1}{6}[\frac{\Gamma(K_L \to \pi^+\pi^-)/\Gamma(K_S \to \pi^+\pi^-)}{\Gamma(K_L \to \pi^0\pi^0)/\Gamma(K_S \to \pi^0\pi^0)} - 1], \tag{2.1.1}$$

where ϵ' is the direct CP violation contribution to the neutral kaon decay into two pions and ϵ is the total CP violation contribution to the processes [64]. The Γ factors are the decay rate of the corresponding processes where K_L indicates the long-lifetime kaon and K_S indicates the short-lifetime kaon. The KTeV experiment measured the result $\Re(\epsilon'/\epsilon) = (28.0 \pm 3.0(\text{stat}) \pm 2.8(\text{syst})) \times 10^{-4}$ [62]. The NA48 experiment measured the result $\Re(\epsilon'/\epsilon) = (18.5 \pm 4.5(\text{stat}) \pm 5.8(\text{syst})) \times 10^{-4}$ [63]. Both experiments confirmed the CP symmetry is directly violated in the

kaon decay processes.

The CP violation has been found continuously in the weak interaction for different generation-mixed mesons in the SM. The CP violation in B^0 mesons which are constructed by a pair of down quark and bottom quark has been observed in the Belle experiment at the KEKB accelerator [65]. The CP violation in D^0 mesons which are constructed by a pair of up quark and charm quarks has been observed in the LHCb experiment at the LHC [66].

In a more formal description, the quantity Jarlskog invariant is derived from the commutation of the mass matrices for up-type quarks and down-type quarks in their weak interaction eigenstates [67]:

$$J = \Im(V_{ij}V_{kl}V_{kj}^{*}V_{il}^{*}), \tag{2.1.2}$$

where V_{ij} denotes the CKM matrix element with the indices ranging from 1 to 3. This quantity is independent of the parametrization in the CKM matrices. Given the CKM matrix is the only source of the CP violation in the SM, any CP violation effect in the SM is proportional to this quantity. The latest fit value for the Jarlskog invariant is $3.08^{+0.15}_{-0.13} \times 10^{-5}$ [54].

2.2 Baryogenesis and the need of beyond the Standard Model CP violation

The observable Universe is dominated by matter. If there were regions of matter and antimatter in the Universe, the antimatter and matter should annihilate to emit γ-rays at the boundaries. Studies based on cosmological observations have excluded this scenario [68]. This fact suggests the Universe is purely matter-dominant. The phenomenon is known as matter-antimatter asymmetry.

To produce the matter-dominant Universe, there have to be some mechanisms producing more matter than anti-matter. The nuclei of atoms consist of protons and neutrons, which are baryons. Ordinary matter formed from atoms is dominant by the baryons and is therefore known as baryonic matter. Baryogenesis is the process to produce extra baryons in the early Universe. It is one of the possible mechanisms. Sakharov raised three requirements known as Sakharov's conditions for the baryogenesis happening [16]. The conditions include:

- The baryon number is violated.
- Both C and CP are violated.
- The Universe is not always evolving under thermal equilibrium.

The baryon number is a quantum number in the strong interaction. The baryon number is conserved for all perturbative interactions in the SM. Each quark is with a $+1/3$ baryon number

and the anti-quark with $-1/3$. A baryon consisting of three quarks has a $+1$ baryon number. An anti-baryon consisting of three antiquarks has a -1 baryon number. A meson consisting of one quark and one antiquark has a 0 baryon number. The baryon number changes in generating extra baryons. Therefore this mechanism must allow the baryon number violation.

For a process of generating baryons, the charge-conjugated process generates antibaryons at the same rate if the C symmetry is preserved. In this scenario, there would be no extra number of baryons generated. Furthermore, for a process generating left-chiral (right-chiral) baryons, the CP-inverted process generates right-chiral (left-chiral) antibaryons at the same rate if the CP symmetry is preserved. There would be no extra number of baryons generated as well. For more baryons produced in baryogenesis, the C and CP symmetries must be violated.

If the Universe evolves always under the thermal equilibrium, the rates of producing and vanishing baryons are the same. No net baryon number increasing should be kept given the thermal equilibrium. To achieve the current matter-dominant Universe, there has to be an exit mechanism for the thermal equilibrium.

In the cosmological observation results of the Wilkinson Microwave Anisotropy Probe experiment, the baryon-to-photon ratio is $6.5^{+0.4}_{-0.3} \times 10^{-10}$ [15]. As the Universe cooled down, the baryons and antibaryons in the early Universe annihilated into photons. This observation suggests there is about one extra baryon left for a billion baryons and antibaryons. Together with the baryon number violation and non-equilibrium, CP violation is one of the necessary conditions for baryogenesis. The only source of this violation in the SM is described by the Jarlskog invariant, which would be difficult to match this observed value. In a theoretical estimation with certain baryon number violation and non-equilibrium conditions, a dimensionless quantity constructed from the Jarlskog determinant is of a value around the order 10^{-20} [69]. This value is too small compared to the observed baryon-to-photon ratio. The 10^{10} order amplified Jarlskog invariant value would possibly account for the matter-antimatter asymmetry naturally. For additional sources of CP violation, physics beyond the SM should be considered. The potential CP violation contribution in the precision measurements of the Higgs boson couplings to SM particles might be one of the sources beyond the SM. This study tests the BSM CP-odd contributions in the Higgs boson couplings to Z bosons. With the CP-even couplings in the SM, additional CP violation would be produced if non-zero CP-odd coupling could be measured. This emphasizes the importance of measuring the properties of the Higgs boson, as it potentially opens a window into physics beyond the Standard Model.

Chapter 3

Effective field theory

An enormous number of BSM theories and models exist. These models are constructed based on different assumptions and have different predictions. It is impossible to design a specific experiment and do the analysis for each of them one by one. To be able to test different theories efficiently, experimental studies should be designed in a model-independent approach. This means that experimental analysis should rely on the assumptions of a specific new physics model at the minimum level. The results from the model-independent experiments can be interpreted by parameters predicted by different theories. Theories with different mechanisms and assumptions can test their predictions to these interpretations.

The processes studied at the LHC reach the electroweak scale around $v = 246$ GeV. Potential new physics above this scale is of interest. The effective field theory (EFT) offers a model-independent approach for the searches of new physics at higher-energy scales. In this approach, all possible operators with required symmetries in the SM are added to the effective Lagrangian. The potential high-energy theories of new physics may contribute to processes at lower-energy scales. These contributions are folded as the high-order operators at the electroweak scale [70]. If non-zero coupling values for these operators are observed, this would indicate some new physics existing above the electroweak scale. These EFT measurements only assume the symmetry requirements for the operators to measure. This minimization of the assumptions makes this approach highly model-independent. On the other side, a non-zero coupling measured does not indicate what new physics models lead to the effect at the electroweak scale. Further investigations are required to identify the new physics models that lead to the observations.

3.1 The Standard Model Effective Field Theory in the Higgs sector

The strong, electromagnetic and weak interactions are described well using the gauges in the Standard Model. The up-to-date experimental results are in good agreement with the SM predictions as described in Section 1.4. Therefore the gauge structures of the SM should be preserved in probing new physics beyond the Standard Model. In this model-independent analysis, the Standard Model Effective Field Theory (SMEFT) as a direct extension of the Standard Model is used [71].

The SMEFT is an EFT model that includes all the SM effects and parametrizes the BSM effects up to a scale of interest for theorists. The requirements for building the effective Lagrangian for the SMEFT are as follows:

1. The SM gauge symmetry $SU(3) \times SU(2)_L \times U(1)_Y$ is preserved;
2. The final theory should contain all SM degrees of freedom;
3. No low energy limit BSM theory should be present in the Lagrangian.

The first two criteria make sure that the full physics in the SM are satisfied. The third criterion makes sure that the SMEFT only parametrizes the new physics above the scales the current experiments can reach. If there exists new physics at the low-energy limit, it would predict deviations from the SM expectations, which is not consistent with the current observations. The low energy limit new physics would have predictions deviated from the SM expectations, which is not preferred by the observations consistent with the SM expectations. With these criteria, the SMEFT degenerates to the SM at the low-energy limit. Any new physics of interest above the electroweak scale and below the cutoff is parametrized as higher-order corrections to the SM.

The Lagrangian of the SMEFT takes the form:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum \frac{c_i}{\Lambda}.O_i^{(5)} + \sum \frac{c_i}{\Lambda^2}.O_i^{(6)} + \sum \frac{c_i}{\Lambda^3}.O_i^{(7)} + \dots . \tag{3.1.1}$$

The term $\mathcal{L}_{\text{SM}}$ is the SM Lagrangian. Each $O_i^{(N)}$ is an effective operator of dimension-N indexed by i. The coefficient c_i is the corresponding dimensionless coupling of the i-th operator. The dimension-N operator terms are suppressed by the cutoff Λ to $(N-4)$ orders, keeping each term of dimension 4. In this study, no specific high-energy physics or corresponding cutoff scale is assumed. The interpretations with potential new physics hypotheses are subjective to

theorists. Before the electroweak SSB, the SM Lagrangian is written as follows:

$$\mathcal{L}_{SM} = -\frac{1}{4}G^{A}_{\mu\nu}G^{A\mu\nu} - \frac{1}{4}W_{I\mu\nu}W^{I\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) + m^2\phi^\dagger\phi - \frac{1}{2}(\phi^\dagger\phi)^2 + i(\bar{l}\not{D}l + \bar{e}\not{D}e + \bar{q}\not{D}q + \bar{u}\not{D}u + \bar{d}\not{D}d) - (\bar{l}\Gamma_e e\phi + \bar{q}\Gamma_u u\tilde{\phi} + \bar{q}\Gamma_d d\phi + h.c.). \tag{3.1.2}$$

The first three terms are the kinetic terms for the SM gauge bosons. In these terms, $G_{\mu\nu}$, $W_{\mu\nu}$ and $B_{\mu\nu}$ are the strength tensor for gluons, EW $SU(2)_L$ gauge bosons, and EW $U(1)_Y$ gauge bosons respectively. The next three terms are the Lagrangian for the scalar field in the Higgs mechanism with the simplest potential. The symbol D_μ denotes the covariant derivative for the EW gauge. The following terms in the bracket are the SM fermions Lagrangian. The slash is a convention denoting the contraction with γ matrices. Symbols l and q represent the left-chiral lepton doublet and quark doublets respectively. Symbols e, u and d represent the right-chiral electron, up-type quarks, and down-type quarks respectively. The last terms in the bracket are the Yukawa terms for the SM fermions and scalar field from the Higgs mechanism. The symbol Γ is the Yukawa couplings. The Hermitian conjugation of these terms is included as $h.c.$ to keep the Lagrangian Hermitian.

The operators above dimension-8 are suppressed by the Λ scale and not considered in the study. The dimension-5 operators violate the lepton number conservation [72]. The dimension-7 operators violate the $(B-L)$ symmetry, where B is the baryon number and L is the lepton number [73]. These violations are constrained by previous experimental observations [74]. Therefore these terms are not considered in this study. In this study, only dimension-6 operators are considered, illustrated by the effective Lagrangian as follows:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum \frac{c_i}{\Lambda^2}.O_i^{(6)}. \tag{3.1.3}$$

The dimension-6 operators in the SMEFT are built by the strength tensors of gauge bosons denoted by X, the scalar fields of the Higgs mechanism denoted by ϕ, fermion fields in the SM denoted by ψ and the covariant derivatives denoted by D. All possible combinations of these forms are not linearly independent. There are different bases for parametrizing the Higgs boson couplings. Different bases are defined upon this redundancy. The Warsaw basis as a standard basis in the theoretical study is selected in this analysis [71]. With this basis, dimension-6 operators in the SMEFT are categorized by their tensor forms on this basis as follows: X^3, ϕ^6, ϕ^4D^2, $\psi^2\phi^3$, $X^2\phi^2$, $\psi^2X\phi$, $\psi^2\phi^2D$, ψ^4 [75].

This study aims at searching the BSM CP violation in the single Higgs boson coupling to Z bosons. Couplings for the operators of form $X^2\phi^2$ are of interest. Through the spontaneous symmetry breaking, the Z boson is mixed from the $SU(2)_L \times U(1)_Y$ gauge electroweak bosons W^i and B and have no relation with the gluon fields. Therefore, $W^{i\mu\nu}$ and $B^{\mu\nu}$ are the only X

field strength tensors considered in this study. Since the SM coupling is CP-even, only CP-odd BSM couplings are measured for CP-violating effects. In the Warsaw basis, such couplings and the corresponding operators are as follows:

$$\begin{aligned} c_{H\widetilde{W}}: &\quad \phi^\dagger\phi\tilde{W}_{i\mu\nu}W_i^{\mu\nu}, \\ c_{H\widetilde{B}}: &\quad \phi^\dagger\phi\tilde{B}_{\mu\nu}B^{\mu\nu}, \\ c_{H\widetilde{W}B}: &\quad \phi^\dagger\tau_i\phi\tilde{W}_{i\mu\nu}B^{\mu\nu}, \end{aligned} \tag{3.1.4}$$

The $\tilde{}$ sign makes operators CP-odd with antisymmetric tensors in the form $\tilde{W}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}W^{\rho\sigma}$ and $\tilde{B}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma}B^{\rho\sigma}$. The $\epsilon_{\mu\nu\rho\sigma}$ is an antisymmetric pseudotensor generating additional (-1) factor under the CP transformation.

3.2 The Higgs basis

Different choices of bases for the parametrization of SMEFT operators may have experimental utility. The Warsaw basis operators are constructed with the $SU(2)_L \times U(1)_Y$ bosons and ϕ field before the EW SSB. These particles are not directly observable in the experiments. The current Universe consists of the massive $W^\pm$ and Z bosons and the Higgs boson after the EW SSB. It would be convenient if the experiment results are interpreted on a basis consisting of these observed bosons. The Higgs basis is proposed to meet these needs [44]. This basis is constructed by rewriting the SMEFT Lagrangian using the mass eigenstates of physical particles and eliminating the dependent operators based on the criteria of the SMEFT. The effective couplings and operators in the Higgs basis are linear combinations of the couplings and operators in the Warsaw basis.

In the measurement of an effective coupling, the analysis usually only extracted some aspects of the information from the raw data. In addition, the precision of the experiments is limited by the finite number of events in the collected dataset. The precision of momentum and energy measurements is not infinitely high in a real detector. The efficiency of event reconstruction and identification is not 100%. Background events are present in the collider. Therefore, different bases could provide different perspectives on the experimental measurements.

In the Higgs basis, the couplings and corresponding dimension-6 CP-odd operators for the

Higgs boson and weak gauge bosons interactions are as follows:

$$\begin{aligned}
\widetilde{c}_{zz}: &\quad \frac{g^2+g'^2}{4} Z_{\mu\nu}\tilde{Z}^{\mu\nu}, \\
\widetilde{c}_{z\gamma}: &\quad \frac{e\sqrt{g^2+g'^2}}{2} Z_{\mu\nu}\tilde{A}^{\mu\nu}, \\
\widetilde{c}_{\gamma\gamma}: &\quad \frac{e^2}{4} A_{\mu\nu}\tilde{A}^{\mu\nu},
\end{aligned} \tag{3.2.1}$$

where the symbols g and g' are the $SU(2)_L$ and $U(1)_Y$ couplings respectively. The symbol $e = gg'/\sqrt{g^2+g'^2}$ is the electromagnetic coupling. The symbols $Z_{\mu\nu}$ and $A_{\mu\nu}$ are the strength tensors for the Z boson and photon fields. The transformation between the Warsaw and Higgs bases is defined in the report of the LHC Higgs Cross Section Working Group [44] and is given below:

$$\begin{aligned}
\widetilde{c}_{zz} &= \frac{v^2}{\Lambda^2}\frac{4g^4 c_{H\widetilde{W}} + 4g^2 g'^2 c_{H\widetilde{W}B} + 4g'^4 c_{H\widetilde{B}}}{(g^2+g'^2)^2}, \\
\widetilde{c}_{z\gamma} &= \frac{v^2}{\Lambda^2}\frac{4c_{H\widetilde{W}} - 2\frac{g^2-g'^2}{gg'} c_{H\widetilde{W}B} - 4c_{H\widetilde{B}}}{g^2+g'^2}, \\
\widetilde{c}_{\gamma\gamma} &= \frac{v^2}{\Lambda^2}\frac{4g'^2 c_{H\widetilde{W}} - 4gg' c_{H\widetilde{W}B} + 4g^2 c_{H\widetilde{B}}}{g^2 g'^2}.
\end{aligned} \tag{3.2.2}$$

3.3 Parameters measured in the effective field theory

In this analysis, the test of the CP invariance is done through the measurement of the SMEFT couplings parametrized in previous sections. The measurement is carried out with both the Warsaw basis and the Higgs basis. The parameters measured include the effective couplings $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ in the Warsaw basis and $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$, $\widetilde{c}_{\gamma\gamma}$ in the Higgs basis.

To allow comparison and combination of this analysis with the measurements in other channels, an additional parameter $\tilde{d}$ is measured as well. This is the parameter measured in the Higgs decay to $\tau\tau$ analysis [76]. In the EFT framework, this parameter is constructed with constrained Warsaw basis couplings:

$$\tilde{d} = \frac{\nu^2}{\Lambda^2} \times c_{H\widetilde{W}}, c_{H\widetilde{W}} = c_{H\widetilde{B}}, c_{H\widetilde{W}B} = 0. \tag{3.3.1}$$

This parameter facilitates a cross-channel validation and potential combination of the measurement on different decay modes for VBF events.

Chapter 4

Experimental setup

Particle physics experiments aim to measure the properties of elementary particles and their interactions. In particle accelerators, high-energy and high-intensity beams of particles are produced. Various types of electromagnetic fields are employed to accelerate and guide the beams to collide. Detectors are constructed to collect information from high-energy collisions of particles within the accelerator. This study analyzed the datasets collected by the ATLAS detector during 2015-2018 at the Large Hadron Collider.

4.1 The Large Hadron Collider

The Large Hadron Collider is the world's largest proton-proton circular collider and synchrotron, located at CERN [37]. The circumference of the LHC tunnel is about 27 km. The tunnel is located underground at a depth of around 100 m, crossing the border of France and Switzerland. The tunnel was previously occupied by the Large Electron Positron collider (LEP) [77]. After the end of the LEP operation, the old machine was removed for the installation of the LHC. Proton beams are collided at the LHC. The designed centre-of-mass energy of these proton beams at the LHC can reach up to 14 TeV [37]. This is much higher than the value for LEP: up to 209 GeV [78].

The layout of the LHC is illustrated in Figure 4.1 [79]. The separated LHC beams travel in opposite directions, within the two rings of the accelerator [80]. The two beams consist of protons for the same charge. Magnetic fields in the opposite direction to guide them traveling in the opposite directions. Each LHC main ring is composed of eight arcs and eight insertion regions. The arcs are the curved parts of the rings. Each contains dipole magnets for bending the beams and quadrupole magnets for squeezing the beams. The insertion regions are the long straight regions between the arcs. In each insertion region, the two beams cross at the interaction point. These regions serve diverse functions in the collider, including accommodating

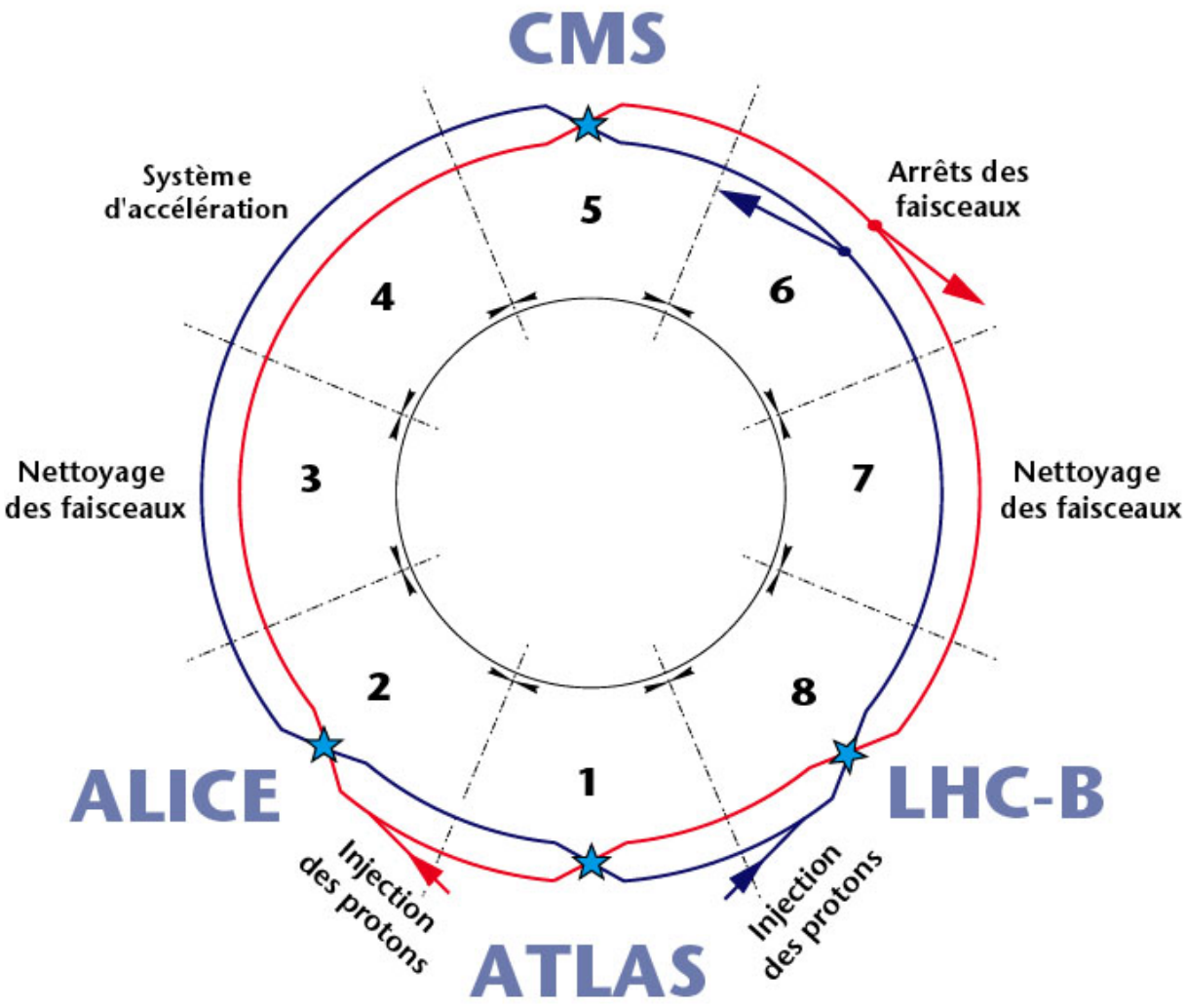


Figure 4.1: The LHC layout.

experiments, collimating and accelerating beams, and absorbing beams.

There are four main experiments at four interaction points of the rings: ATLAS [81], CMS [82], LHCb [83] and ALICE [84]. The ATLAS and the CMS are general-purpose detectors, designed to study the Higgs physics and potential new physics at the energy frontier. The LHCb experiment is dedicated to the studies of hadron physics, focusing on detecting CP-violating processes and rare decays of bottom and charm hadrons. The ALICE is a general-purpose heavy-ion detector focusing on the QCD measurements within the SM.

The Run 2 CERN accelerator complex is illustrated in Figure 4.2 [85]. The injection scheme of the LHC Run 2 beams using the accelerator complex is described in [86]. At the beginning of the LHC Run 2 injection chain, protons are prepared from the ionization of hydrogen by an electric field. The linear collider LINAC 2 accelerates the protons to 50 MeV. Then the proton

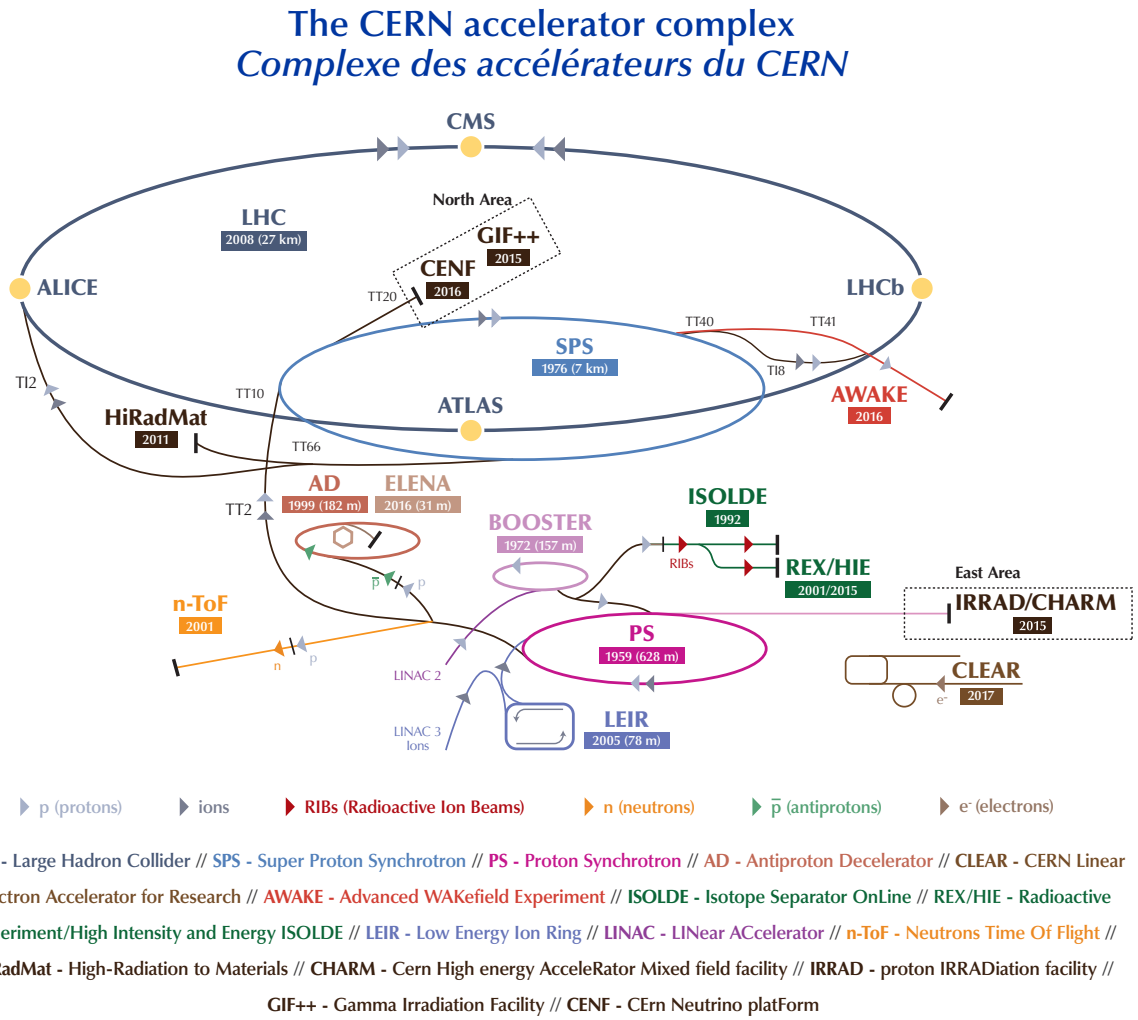


Figure 4.2: The CERN accelerator Complex in 2018.

beam is injected into the Proton Synchrotron Booster (PSB). The PSB is a circular accelerator with four stacked rings. Each ring takes the injection from the LINAC 2. The particles in the four rings are accelerated together in the PSB to 1.4 GeV. Accelerated particles in each PSB ring are then injected into the Proton Synchrotron (PS) as a beam bunch. For the nominal LHC beams, the PS injects two batches of six beam bunches during each acceleration period. The injected beams in the PS are accelerated to 25 GeV and split into 72 bunches separated by a 25 ns time interval. The beams are then injected into the Super Proton Synchrotron (SPS) and accelerated to 450 GeV. The accelerated beams in the SPS are separated into 2 directions and injected into the LHC rings at interaction points 2 and 8 counter-clockwise and clockwise respectively. The proton beams inside the LHC are accelerated and bunched further to the energy for collisions.

4.2 The ATLAS detector

The ATLAS detector is the world's largest general-purpose particle detector with 46 m length, 25 m height and 25 m width [87]. The ATLAS experiment employs a right-handed coordinate system with the origin at the nominal interaction point at the centre of the detector and the z-axis is aligned with the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ as

$$\eta = -\ln \tan(\theta/2), \tag{4.2.1}$$

and the rapidity is defined in terms of energy E and momentum p as $y = 1/2 \ln[(E + p_z)/(E - p_z)]$. Angular distance is measured in units of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The layout of the ATLAS detector is shown in Figure 4.3 [87]. The ATLAS detector consists of three main components as sub-detectors. The Inner Detector (ID) is the innermost part of the detector around the beam pipe. This part measures the trajectories and the momentum of charged particles. It is known as the tracking system of the detector. The calorimeters are located in the middle parts surrounding the ID. The calorimeters consist of electromagnetic calorimeters (ECAL) and hadronic calorimeters (HCAL). The calorimeters are used to absorb scattered particles and measure their energies. The outermost part is the muon spectrometer. It detects the muons and reconstructs their trajectories together with the inner detectors.

In the ATLAS detector, the measurement of the momentum of charged particles is supported by the magnet system. A charged particle traveling in the magnetic field has its trajectory bent by the Lorentz force as follows:

$$qv_\perp B = m\frac{v_\perp^2}{R}, \tag{4.2.2}$$

where $v_\perp$ is the velocity perpendicular to the magnetic field. The perpendicular momentum of the charged particle is determined by the radius of curvature of its trajectory.

For the transverse momentum measurement of charged particles, the entire ID is immersed in a 2 T magnetic field that is parallel to the beam axis [88]. This field is provided by a solenoid magnet with 5.8 m length and 2.46 m inner diameter surrounding the ID. The trajectories of the charged particles are deflected by the solenoidal field in the transverse plane. For the muon momentum measurement, the magnetic field is provided in the muon spectrometer in the azimuthal direction of the ATLAS coordinate. The field is approximately 0.5 T for the barrel parts, where the pseudorapidity range is $0 < |\eta| < 1.4$, and approximately 1.0 T for the endcap parts, where the pseudorapidity range is $1.6 < |\eta| < 2.7$. The transition region is affected by both fields. The magnetic field for the barrel region of the muon spectrometer is provided by a toroid consisting of eight coils located within the muon spectrometer in the transverse plane. The magnetic

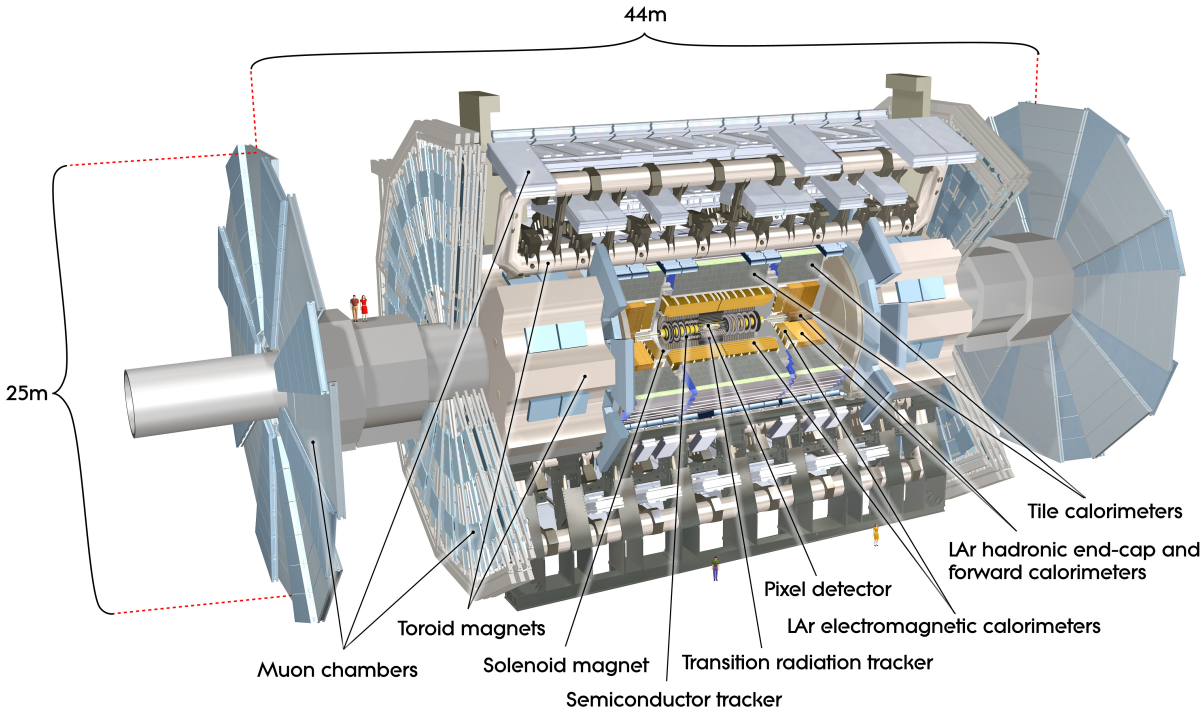


Figure 4.3: The ATLAS detector layout.

field for the endcap region of the muon spectrometer is provided by two toroids each consisting of eight coils surrounded by the barrel muon spectrometer. Muons passing through the muon spectrometer are deflected by it in the $r - z$ plane.

ATLAS tracking system

The layout of the ATLAS tracking system is shown in Figure 4.4 [87]. From the inside out, it consists of the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). Each component contains the barrel and the endcap parts, which cover different but overlapping pseudorapidity ranges. The positioning and size of the components are demonstrated in Figure 4.5 [89]. The ID is designed for the high-resolution measurement of the trajectories and momentum of charged particles within the range of $|\eta| \leq 2.5$. The position of the proton-proton collision is reconstructed from the trajectory of outgoing particles as well. These points are named the primary vertices.

The Pixel detector and the SCT both use semiconductor sensors to measure particle trajectories. The Pixel detector comprises four layers of pixel sensors in the barrel part and three disks of pixel sensors in the endcap part. The barrel layers are installed at radii of 33.5 mm, 50.5 mm, 88.5 mm and 122.5 mm respectively. The endcap disks are installed at $z = 495.0$ mm,

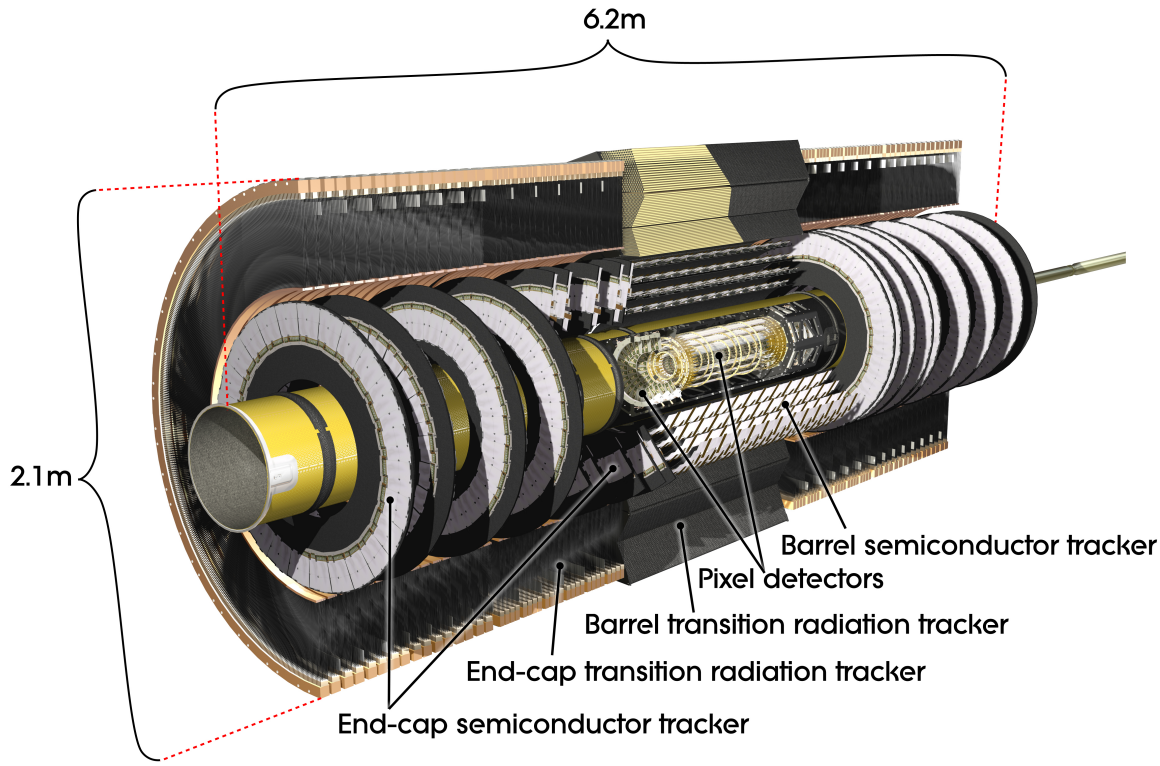


Figure 4.4: The ATLAS tracking system layout.

580.0 mm and 650.0 mm respectively. The SCT comprises four layers of micro-strip sensors in the barrel part and nine disks of micro-strip sensors in the endcap part. The barrel layers are placed at radii of 299.0 mm, 371.0 mm, 443.0 mm and 514.0 mm. The endcap disks are placed at $z = 853.8$ mm, 934.0 mm, 1091.5 mm, 1299.9 mm, 1399.7 mm, 1771.4 mm, 2115.2 mm, 2505.0 mm and 2720.2 mm. The TRT uses polyimide drift tubes named straws to measure particle trajectories and identify particles. The tubes are placed parallel to the beam pipe in the barrel part and radial from the beam pipe in the endcap part.

The silicon pixel sensors of the Pixel detector are each composed of thousands of tiny pixels. Most of the pixels are of a very small nominal pixel size 50×400 μm^2. With 10^7 pixels in total in the Pixel detector, the accuracy of these pixel sensors is extremely high. The position resolution is up to 10 μm in the azimuthal direction and 115 μm in the longitudinal or radial directions. The SCT employs microstrip sensors to achieve a balance between performance and cost. With the higher radius, the precision required for the SCT is lower. The SCT has a resolution up to 17 μm in the azimuthal direction and 580 μm in the longitudinal or radial directions.

The straws used in the TRT are long tubes filled with a gas mixture dominated by Xe and CO_2 with an anode wire in the centre. In the TRT, the straws are stacked as thick walls in

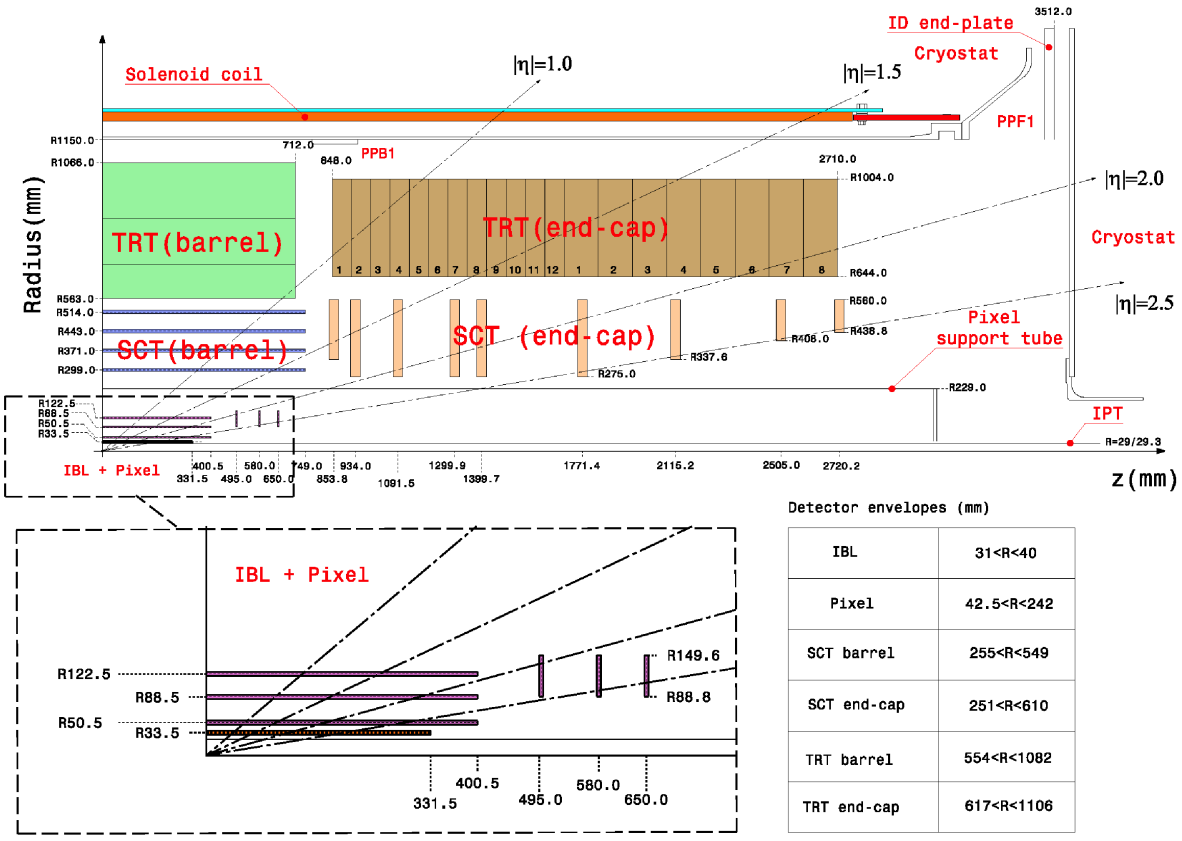


Figure 4.5: Quarter-section Plan view ATLAS ID layout.

the barrel and endcap detectors. The straws in the barrel part are placed parallel to the beam direction covering a radius from 563.0 mm to 1066.0 mm. The straws in the endcap are placed radially with a z coverage from 848.0 mm to 2710.0 mm. A passing through charged particles can be detected in tens of straws. The azimuthal resolution of these straws reaches 130 μm.

ATLAS calorimetry system

The layout of the ATLAS calorimetry system is shown in Figure 4.6 [87]. The calorimeters are located outside the ID. The ECAL is the inner part and the HCAL is the outer part. The ECAL contains the barrel and endcap parts. The HCAL contains the barrel, extended barrel, endcap parts and the forward calorimeter (FCal) [87].

The ATLAS calorimeters contain the absorbers and the active materials in each module. When the particles from the proton-proton collisions enter a calorimeter module, they have interactions with the metal layers named absorber. A lot of low-energy particles known as showers are emitted from the interactions. The energy of these showers is transferred to the active material between the layers. Measurement of the current from the absorbed energy of the

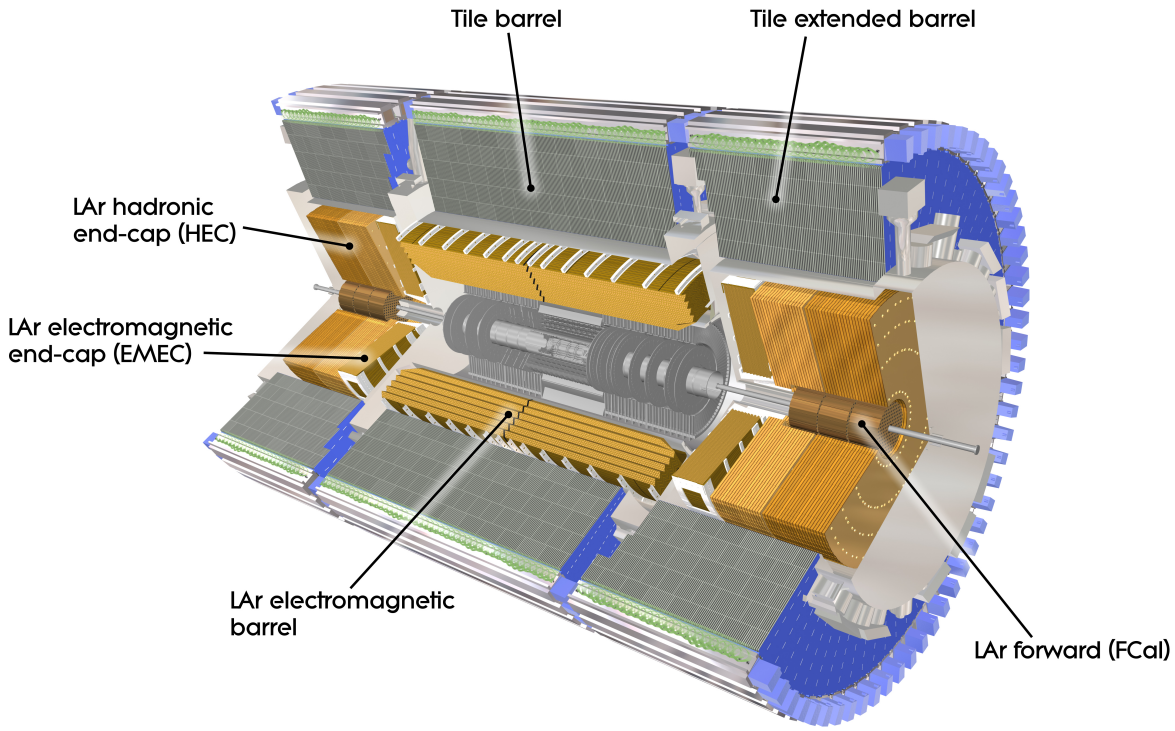


Figure 4.6: The ATLAS calorimeter layout.

shower in the detector module determines the energy deposit of the particle passing through. Most of the measured particles are supposed to be fully absorbed by the calorimeter and their energy is reconstructed mostly from the energy deposit in the calorimeter.

The ECAL, the HCAL endcap and the HCAL FCal are the liquid argon (LAr) calorimeters. In LAr calorimeters of the ATLAS, the absorber is made of copper, lead or tungsten. LAr is used as active material. The particles in the shower ionize LAr atoms and the current is produced for measuring. The barrel and extended barrel parts of the HCAL are the tile calorimeters. In this type of calorimeter, the absorber is made of steel. The scintillators are used as active materials. Scintillators emit light when hadrons in the shower pass through. The light signal is collected and transferred by wavelength-shifting fibres and measured by the photomultiplier blocks at the outer edge of the module.

In the ECAL, the electrons and photons produced from the proton collisions in the pseudorapidity range $|\eta| \leq 3.2$ are absorbed. The radiation length is defined as the mean travel distance for the particle to lose its energy by $1/e$ fraction. The thickness of the ECAL is designed greater than 22 times of electron radiation length. Most of the electrons and photons can be stopped in the ECAL. The energy deposits of their showers are measured to high granularities. The reso-

lution of the cell for deposits varies in its layer and $|\eta|$ value. The resolution is up to $0.025/8$ in η and 0.025 in ϕ. The cells achieve the maximum η resolution within the ranges $|\eta| < 1.40$ and $1.5 < |\eta| < 1.8$ in the first layer. The cells achieve the maximum ϕ resolution within the range $1.40 < |\eta| < 1.475$ in the first layer and within $|\eta| < 2.5$ in the second and third layers. The energy resolution of the ATLAS detector is parametrized by

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{4.2.3}$$

where a, b and c are the coefficients for the source of the stochastic term, noise term and the constant resolution [90]. The energy resolution for the LAr calorimeter is about 0.5% during the ATLAS Run 2 [91].

The hadrons produced from the proton collisions cannot be stopped by the ECAL and travel into HCAL. With the FCal, the HCAL absorbs and measures the energy of hadrons in the pseudorapidity range $|\eta| \leq 4.9$. The interaction length is defined as the mean free path between two inelastic interactions. The thickness of the HCAL is more than 9.7 times of hadron interaction length. The resolution of HCAL varies in position. The resolution of the tile calorimeter is 0.1 in η and 0.1 in ϕ except for the last layer. The FCal achieves its finest resolution of 0.75 cm$\times 0.65$ cm in $\Delta x \times \Delta y$ within the range $3.10 < |\eta| < 3.15$.

ATLAS muon spectrometer

The layout of the ATLAS muon spectrometer is shown in Figure 4.4 [87]. Muons from the proton collision usually punch through the calorimeters. The calorimeters cannot determine the full energy of not absorbed muons. The muon spectrometer is installed outside the calorimeters as the outmost part of the detector. This provides additional measurements for the muon trajectories within a pseudorapidity range $|\eta| \leq 2.7$. The large radius of the muon spectrometer is suitable for measuring the momentum of muons reconstructed. The momentum resolution contains three main contributions:

$$\frac{\sigma(p)}{p} = \frac{p_0^{MS}}{p_T} \oplus p_2^{MS} \oplus p_2^{MS} \cdot p_T, \tag{4.2.4}$$

where the coefficients p_0^{MS}, p_1^{MS} and p_2^{MS} model the uncertainties from the energy loss in calorimeters, the multiple scattering with detector materials during the muon propagation and the intrinsic resolution of the muon spectrometers [92]. The designed muon momentum resolution in the muon spectrometer for the $H \to ZZ^*$ decay events is around 3% [93].

The muon spectrometer uses four types of chambers: monitored drift tubes (MDT), cathode strip chambers (CSC), resistive plate chambers (RPC), and thin gap chambers (TGC) [94].

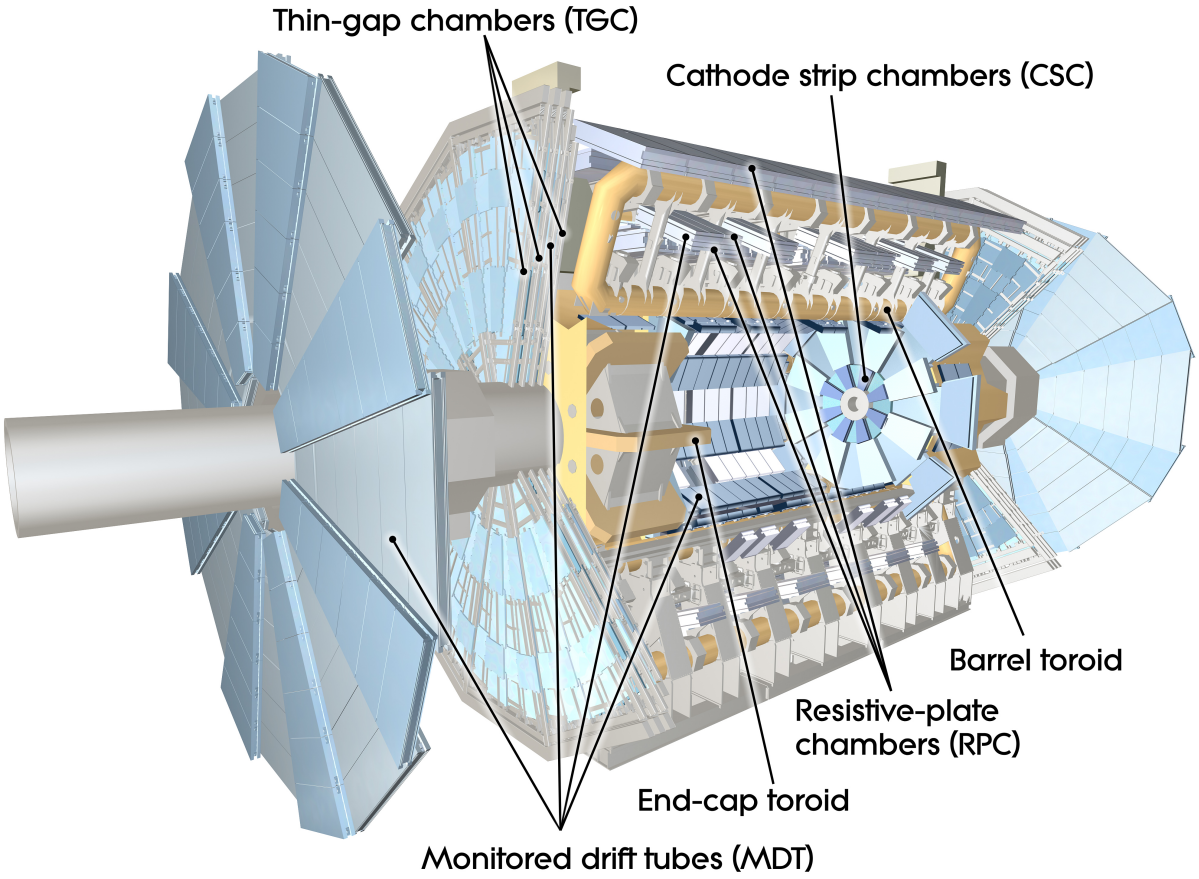


Figure 4.7: The ATLAS Muon Spectrometer layout.

The RPCs and TGCs serve as the trigger chambers in the barrel and endcap parts respectively. These chambers also record the azimuthal angle of incident muons. The MDTs are the primary chambers used for precise muon measurement up to the range of $|\eta| < 2.7$. Each chamber contains several drift tubes in $3 - 4$ layers. These tubes are filled with Ar/CO_2 gas and each possesses an anode wire in the centre [95]. When a muon passed through a tube, electrons are ionized from the gas molecules along the muon trajectory and drift to the anode wire. The electrons ionized have different distances to the wire ranging from the distance between the trajectory to the wire to the radius of the tube. The distance of the trajectory to the wire is determined by the time duration of electron drift signals. The resolution of a single space point is around 80 μm and for the chambers is around 35 μm in z direction [87, 96].

The CSCs are the chambers used in the high pseudorapidity range $2.0 \leq |\eta| \leq 2.7$. These chambers contain anode wires along radial directions between two cathodes consisting of stripes one along the radial direction and one in the perpendicular direction. Since the distances between each cathode and a wire are uniform, the designed drift duration of the electrons is shorter

than that of the MDT. Therefore, these chambers are with better time resolution overcoming the heavier event rates and background conditions in the high η region. The chamber precision is around 40 μm in the radial direction [87].

ATLAS trigger and data acquisition system

In the LHC, the designed separation of the proton bunches is 25 ns. During the ATLAS data collection between 2015-2018, the average number of proton collisions for each bunch crossing is about 40. About 10^9 interactions per second occur in the ATLAS experiment [97]. The amount of data produced in these collisions is huge. It is impossible to record all these collisions. In contrast, interactions of great physical interests are relatively rare events. There are in total about 8×10^6 Higgs bosons produced in the whole ATLAS Run 2. The ATLAS detector has the trigger system to select the events of interest [98]. Only these events are saved for analysis.

The ATLAS trigger system consists of different levels of trigger: the first level (L1), the second level (L2) and the Event Filter (EF) [99]. The L1 trigger is based on hardware devices that can analyze the events at the collision rate of 40 MHz. In the muon spectrometer, the trigger chambers belong to the L1 trigger. These chambers analyze the pattern of muons to apply the momentum threshold. The L1 trigger for the calorimeter (L1Calo) is an electronics system outside the detector in a cavern named USA15 [100]. The input of the L1Calo is from the trigger towers throughout the ECAL and HCAL. The information from the muon system and calorimetry system is also analyzed together by processors. The event rate passing the L1 trigger is greatly reduced to 90 kHz for further triggering.

The L2 and EF triggers are based on software processing. They are combined to form the Higher-Level Trigger (HLT). The events accepted by the L1 are sent to HLT for further selection. The L2 trigger is an online software trigger that processes events only in the regions of interest where the L1 trigger has observed muon tracks or calorimeter clusters. The EF uses offline algorithms for further event selections after the L2 trigger. The event rate output from the HLT is greatly reduced to about 1.2 kHz. These interesting events are stored permanently in the data centres of CERN for offline reconstruction and analysis.

Chapter 5

Data preparation and event selection

5.1 The LHC Run 2

The LHC Run 2 refers to the stable LHC operation period from 2015 to 2018. The progress of data collection as a function of time is shown in Figure 5.1 [101]. The LHC delivered the integrated luminosity of 156 fb^{-1} in the entire Run 2 period, among which 139 fb^{-1} is suitable for physical analysis in the ATLAS detector. This analysis is based on this dataset.

The beam condition varies over time during the Run 2 period. For each bunch crossing, there are usually multiple proton-proton interactions happen. The events of interest are with large momentum transfer at the EW scale, known as the hard-scatter interactions. The hard-scatter events are very rare. Therefore among the interactions in each bunch crossing, most of the events are from QCD processes with low momentum transferred. These are pile-up events. The average number of pile-up interactions for each bunch crossing is described by the parameter μ. The distribution of the μ parameter each year during Run 2 is shown in Figure 5.2 [101]. During the Run 2 period, each year the peak instantaneous luminosity increases [97]. The value of the pile-up μ factor increases as well. The pile-up effect would affect the reconstructions of physical objects from the detector information and further influence the analysis [102]. In order to obtain a consistent description of the data samples, Monte Carlo simulations need to be performed separately for the dataset each year.

5.2 Reconstruction of objects

This study aims to measure the Higgs boson couplings on the Higgs boson-enriched dataset with the VBF production mode and the $H \rightarrow ZZ^* \rightarrow 4l$ decay channel. The VBF events have at least two jets produced with the Higgs boson. The $H \rightarrow ZZ^* \rightarrow 4l$ decay final state

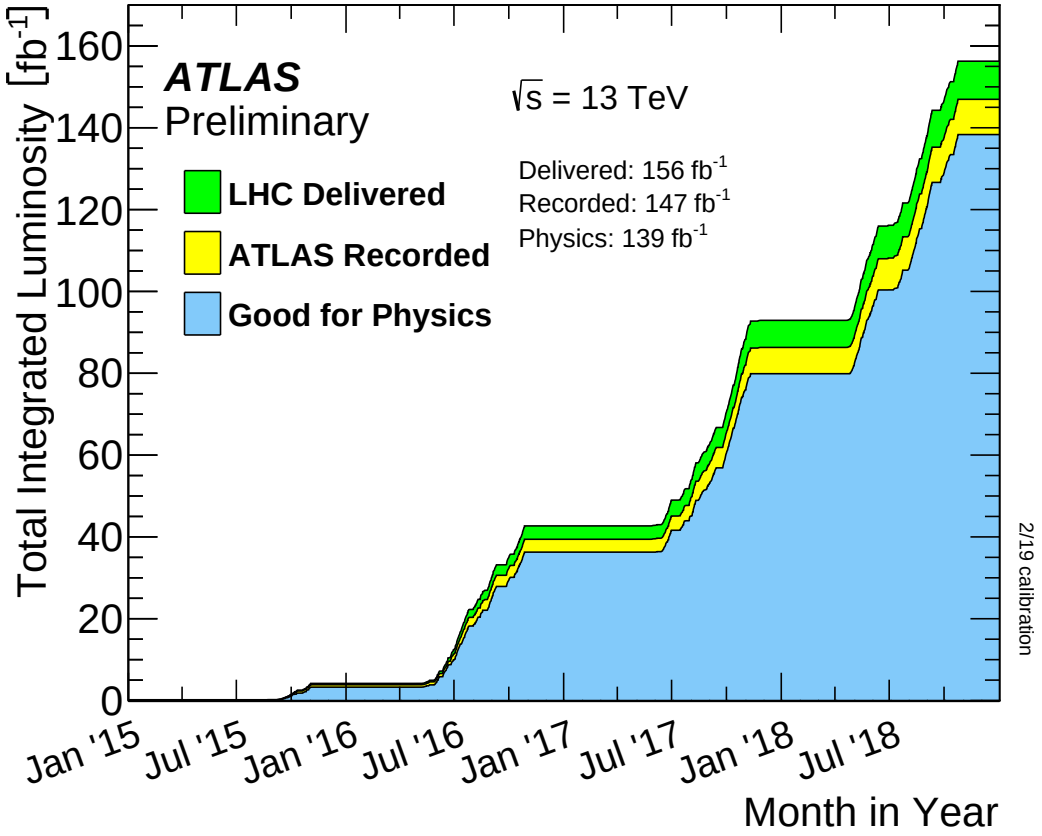


Figure 5.1: The total integrated luminosity during the LHC Run 2.

has four leptons. The final state leptons in this channel are electrons or muons. τ leptons are not included due to their unique properties. The selection criteria for events are based on the properties of jets, electrons, and muons. The reconstruction and identification of these objects form the foundation of this study.

Electron reconstruction and identification

The reconstruction of the electrons at the ATLAS experiments combines the information from ID trajectories and ECAL energy deposits [45]. When an electron passes through the Inner Detector, it leaves information recorded by silicon detector layers and TRT. The electron then propagates into the calorimeter and deposits its energy in the calorimeter. Most electrons deposit all their energy and are absorbed in the ECAL.

The cluster of an electron energy deposit in the calorimeter is reconstructed by the topological cluster approach [103]. This approach starts by forming the proto-clusters with a "$4-2-0$" method. The cells with energy deposits of at least four times the noise level are the center of the proto-clusters. Then iteratively the neighbour cells of included cells with energy deposits of at

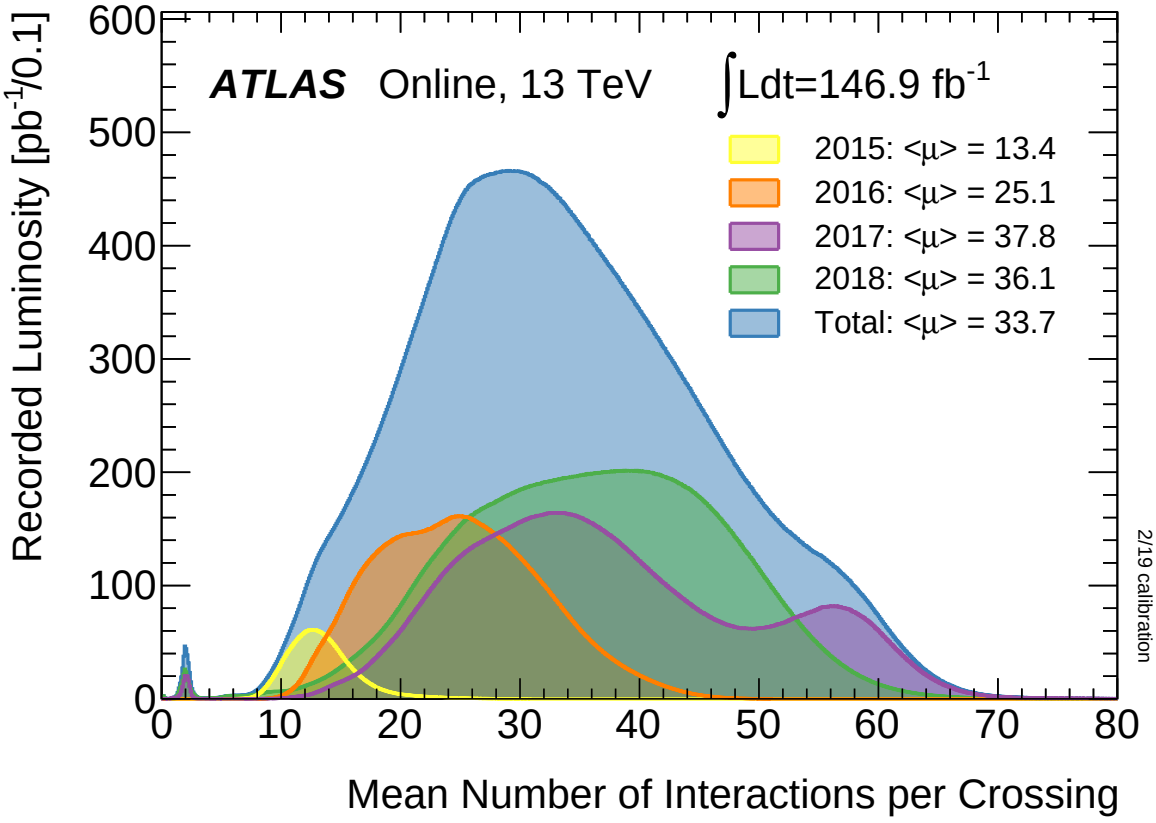


Figure 5.2: The number of interactions in each bench-crossing in Run 2 for each year.

least twice the noise level are included. Finally, the surrounding neighbour cells are added independent of their energy deposits. The formed proto-clusters sharing the same cells with energy deposits at least twice the noise are merged. The proto-clusters with multiple local maxima are split.

The reconstruction of electron trajectories is based on the Inner Detector information and matching with the ECAL clusters [104]. The trajectory candidates are reconstructed by extension from the seeds consisting of three space points recorded by separate silicon detector layers [105]. Among these candidates, the silicon detector space points might be shared by different trajectories. A global χ^2 fit on the candidates is applied to solve the ambiguity and determine the trajectory parameters including the position and momentum information [106]. The selected trajectories are extended to the TRT volume where the TRT information improves the fit results [107].

The electrons might have strong bremsstrahlung radiation when interacting with the material in the silicon detectors. This effect might cause a large change in the curvature and trajectories. An optimized Gaussian Sum Filter is introduced to contain this effect for the trajectory

candidates matching the ECAL clusters with a small angular separation and having at least four silicon detector space points [108]. In this algorithm, the bremsstrahlung of electrons is modeled to improve the determination of trajectory parameters and to increase the reconstruction efficiency of electrons.

The reconstructed trajectory and reconstructed ECAL clusters are matched using their momentum and energy deposits, and their angular separation [104]. Then superclusters are formed to include the electromagnetic showers from the electron [45]. The seeds of superclusters are clusters with at least 1 GeV transverse energy and associated with a trajectory with at least four silicon detector space points. The clusters around with a small angular separation and the clusters with a looser separation requirement and associated with the same trajectory are included to form the superclusters. The electron candidates are reconstructed with the match again with the trajectories and the superclusters. The reconstruction efficiencies of the electrons are shown in Figure 5.3 from the ATLAS simulation [45].

The quality of reconstructed electron candidates is further tested in a process named "identification" [45, 104]. The electrons coming from the collision point are considered signal-like named "prompt" electrons. Some reconstructed candidates are with energy deposits of light-flavour hadrons, come from photon conversion, or come from the decay of heavy-flavour hadrons. The identification algorithm distinguishes the prompt electrons and background-like electrons. The properties of reconstructed candidates from the Inner Detector, ECAL and leakage into the HCAL are taken into account. A likelihood discriminant is built from the probability density distributions of the prompt and non-prompt electrons of these properties. The identification has different working points including *Loose*, *Medium* and *Tight*. The candidates passing the identification with tighter working points must pass the looser identification. The efficiency on $Z \to ee$ data is shown in Figure 5.4 [45]. In this analysis, the *Loose* working point is used.

Muon reconstruction and identification

In the ATLAS experiments, muons produced in the proton collisions are usually not stopped by the calorimeters. They leave information in the chambers of the muon spectrometer outside the calorimeters. The reconstruction of muons mostly uses the trajectory information collected in the Inner Detector and muon chambers [46]. In each station of muon chambers, a straight-line named "segment" is reconstructed from the space points [109]. The trajectories of muons in the muon spectrometer are reconstructed by combining these segments and performing a global χ^2 fit [46]. The calorimeter information is considered in the reconstruction.

Based on different information and strategy used, there are different types of muon reconstruction in the ATLAS experiment. The muon reconstruction types are described as follows:

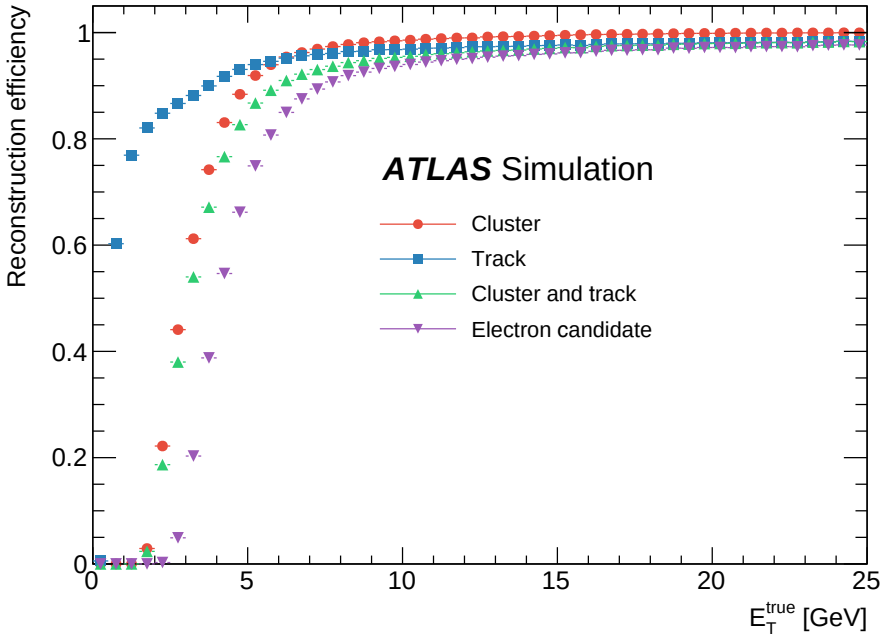


Figure 5.3: Reconstruction efficiency of electrons from Monte Carlo simulation as a function of the transverse energy at the generator level.

- Combined muons: Most of the muons are of this type. These muons require the matching of ID trajectories and muon spectrometer trajectories. They are reconstructed by a combined fit on the space points of matched subdetector trajectories with energy loss in the calorimeter considered. The combined fit of muon spectrometer trajectory and ID trajectory segments reconstructed from silicon detector space points produce the silicon-associated forward muons.

- Inside-out combined muons: These muons require the extrapolation of ID trajectories to match with at least three space points in the muon spectrometer. They are reconstructed by a combined fit of these space points of the ID trajectories and in the matching.

- Muon spectrometer extrapolated muons: These muons are reconstructed by extrapolating muon spectrometer trajectories inwards to the beamline.

- Segment-tagged muons: These muons are reconstructed by extrapolating ID trajectories to match the segments reconstructed in the muon spectrometer. The trajectory parameters are determined by the ID information.

- Calorimeter-tagged muons: These muons are reconstructed by extrapolating ID trajectories to match the energy deposits of a minimum-ionizing particle in the calorimeter. The

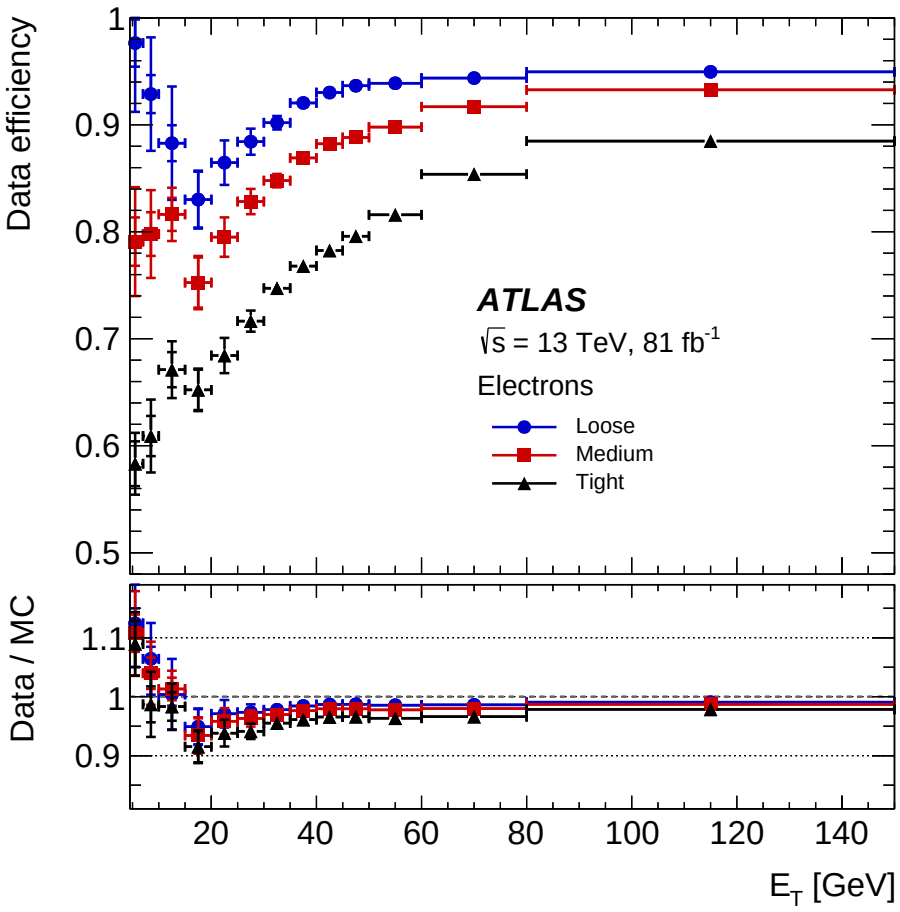


Figure 5.4: Identification efficiency of electrons from $Z \to ee$ data on *Loose*, *Medium* and *Tight* working points as a function of the transverse energy. The lower panel shows the ratio of the data efficiencies to the simulated efficiencies.

trajectory parameters are determined by the ID information.

For the muons with low transverse momentum, the space points left in the muon spectrometer might not be enough to reconstruct full trajectories in the muon spectrometer. The inside-out combined muons are reconstructed to recover the efficiency of these low transverse momentum muons. The pseudorapidity ranges $2.5 < |\eta| < 2.7$ are beyond the coverage of the Inner Detector. Therefore, the muon spectrometer extrapolated muons are reconstructed to recover this part of efficiency. There is a gap in the centre part of the muon spectrometer of $|\eta| < 0.1$. The segment-tagged and calorimeter-tagged muons are reconstructed to recover this part of efficiency.

Similar to the electrons, there is a muon identification procedure to distinguish prompt muon candidates from the backgrounds. The dominant backgrounds are muon candidates from the in-flight decay of light-flavour hadrons. The non-prompt muons originating from the light-flavour

hadrons usually have changes in the trajectory direction. Therefore, the matching of trajectories of these muon candidates reconstructed in the ID and the muon spectrometer is not usually poor. For muons with combined trajectory information, the matching quality for the trajectories reconstructed in the ID and muon spectrometers are characterized by the q/p compatibility

$$\frac{|q/p_{\text{ID}} - q/p_{\text{MS}}|}{\sqrt{\sigma^2(q/p_{\text{ID}}) + \sigma^2(q/p_{\text{MS}})}}, \tag{5.2.1}$$

which describes the significance of the charge-momentum-ratio difference, and the variable

$$\rho' = \frac{|p_{\text{T,ID}} - p_{\text{T,MS}}|}{p_{\text{T,CB}}} \tag{5.2.2}$$

which describes the relative difference in the transverse momentum from different components used in the trajectory fit. In these variables, the subscript ID, MS and CB denote the trajectory reconstructed in the ID, in the muon spectrometer and from combined fit. The symbol q is the charge and σ denotes the uncertainty.

There are different working points for muon identification as well. The standard working points include *Loose*, *Medium* and *Tight*. Similar to electron identification, candidates passing tighter working points identification must pass the looser identification for these standard working points. Only for the *Loose* working point, the calorimeter-tagged and segment-tagged muons in the central region of $|\eta| < 0.1$, and inside-out combined muons with transverse momentum p_{T} below 7 GeV within $|\eta| < 1.3$ are considered. The efficiency of the muon identification for these working points is shown in Figure 5.5 [46]. In this study, the *Loose* working point is used in muon identification.

Jet reconstruction

The quarks and gluons produced in the collider cannot propagate freely. They must form hadrons to travel due to very strong QCD coupling on long-distance scales. With radiation in QCD processes, the resulting physical object is a spray of collimated particles in a cone. In the ATLAS experiment, the hadronic products are described as complex objects named jets constructed using jet algorithms [102]. The major jet reconstruction method for the ATLAS analysis is the anti-k_t clustering algorithm [110]. The rough description of the algorithm is as follows. In this algorithm, the distance of two clusters in the calorimeter is angular separation in the $\eta - \phi$ plane inverse-squared rescaled by a configurable constant radius parameter R and

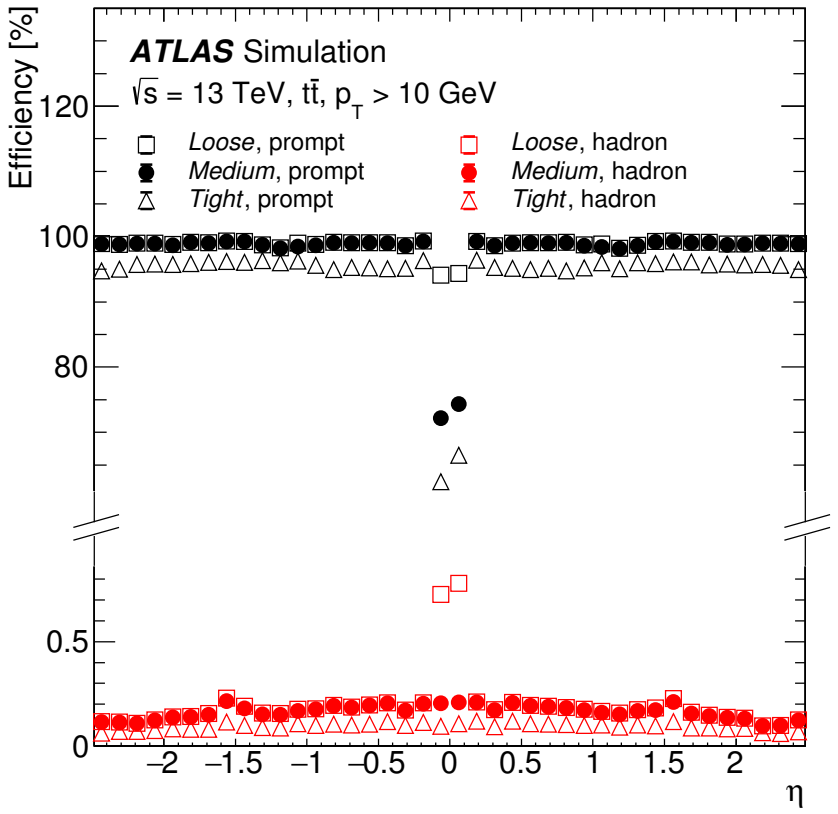


Figure 5.5: Muon identification efficiencies for $t\bar{t}$ events as a function of pseudorapidity η for *Loose*, *Medium* and *Tight* working points with $p_{\mathrm{T}} > 10$ GeV.

the transverse momentum. The distance for clusters i and j is:

$$d_{ij} = \min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right)\frac{\Delta_{ij}^2}{R^2}, \tag{5.2.3}$$

where k_{ti} is the transverse momentum of cluster i, Δ_{ij} is the angular separation and R is the radius parameter. This algorithm tends to cluster hadrons to the hadron with high transverse momentum. The hadrons from interesting events usually have high transverse momentum. In the ATLAS Run 2 data, the anti-k_t algorithm is implemented in the FastJet 3.2.2 software package with the radius parameter $R = 0.4$ [111, 112].

The clusters used in the jet reconstruction are formed from the calorimeter cells topologically connected. The energy deposit is corrected at the electromagnetic energy scale. The jets reconstructed from only calorimeter information are the *EMTopo* jets [103]. The jet reconstruction is improved by matching with the trajectories and clusters of energy deposits. The energy

deposit of the cluster can be corrected with the momentum of the trajectories. This is the particle flow algorithm and these jets are the *PFlow* jets [113].

The jet energy scale and resolution are determined by comparing the reconstructed jets and particle-level truth jets from Monte Carlo simulations. The jet energy scale receives a chain of calibrations from the pile-up and Monte Carlo events. The jet energy resolution is estimated from the electronic noise, random fluctuations and fluctuations of a constant transverse momentum fraction [102]. The energy resolution result ranges between 24% and 6% for the jet energy of 20 GeV to 300 GeV.

5.3 Event selection

The Higgs boson event candidates are selected based on the reconstructed leptons and jets information. The selection requirements of the objects are described as follows. For the reconstructed leptons and jets in selected events, there are requirements on the transverse energy E_{T}, transverse momentum p_{T} and pseudorapidity η. The electrons are required to satisfy $E_{\mathrm{T}} > 7$ GeV and $|\eta| < 2.47$. The muons are required to satisfy $p_{\mathrm{T}} > 5$ GeV and $|\eta| < 2.7$. The calorimeter-tagged muons need more restricted $p_{\mathrm{T}} > 15$ GeV. The jets are required to satisfy $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.5$.

The selected events are required to contain a proper quadruplet of leptons. Each quadruplet is composed of four leptons that form two pairs with the same flavour and the opposite charge. The lepton pair with the invariant mass closest to the Z boson mass is the leading lepton pair. Its invariant mass is labeled as m_{12}. The pair with the invariant mass second closest to the Z boson mass is the subleading lepton pair. Its invariant mass is labeled as m_{34}. The final states are classified by the flavour of the leading and subleading lepton pairs into 4μ, $2e2\mu$, $2\mu2e$ and $4e$ in decreasing order of priority.

There are selection criteria for the quadruplet. The three leptons of highest p_{T} have minimum values of p_{T} of 20, 15 and 10 GeV respectively. There can be no more than one calorimeter-tagged or segment-tagged muon. The invariant mass of the leading lepton pair is required to fall within the range of $50 < m_{12} < 106$ GeV. The invariant mass of the subleading lepton pair is required to fall within the range of $m_{min} < m_{34} < 105$ GeV. The lower bound m_{min} is 12 GeV for the four leptons invariant mass of $m_{4l} \leq 140$ GeV and increases linearly to 50 GeV until m_{4l} reach 190 GeV. The threshold keeps the same for higher m_{4l}. The other choices of the same-flavour opposite-charge lepton pairs are required to have an invariant mass of $m_{ll} > 5$ GeV. In all the lepton pairs, the angular separations $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ between two leptons need to be larger than 0.10, where $\Delta\phi$ is the difference in the azimuthal angles of the two leptons and $\Delta\eta$ is the difference in their pseudorapidity.

The lepton isolation requirements are applied consisting of isolation of the trajectories and isolation of the calorimeter records. The impact parameter d_0 is the distance of the trajectory from the interaction point. The leptons are required to have the impact parameter significance $|d_0|/\sigma(d_0) < 5$ and $|d_0|/\sigma(d_0) < 3$ for electrons and muons respectively, where $\sigma(d_0)$ is the resolution of the impact parameter. The leptons are required to meet the isolation requirement. To suppress the reducible background, there is a requirement on the common vertex. The four leptons need to originate from the same vertex. A χ^2 fit is performed using the four lepton trajectories to the common vertex with the VKalVrt algorithm [114]. The requirement on the fit quality is χ^2/N_{dof} must be no greater than 6 for the four muons events and must be no greater than 9 for the other flavours, where N_{dof} is the degrees of freedom. The four-momenta of the four leptons consist of 16 parameters. The constraints on the mass of leptons and the Higgs boson four-momentum reduce the degrees of freedom to 8. If there is an additional lepton with $p_{\text{T}} > 12$ TeV satisfying the selections of isolation, impact parameter and angular separation, the quadruplet with the highest amplitude is selected.

In the analysis, the Higgs boson candidates are studied within the four-lepton mass window $115 < m_{4l} < 130$ GeV around the Higgs boson mass. This mass window is the Higgs bosons signal region of the study. The sideband $105 < m_{4l} < 115$ GeV and $130 < m_{4l} < 160$ GeV serve to confine the four-lepton backgrounds. These are the ZZ^* control regions.

5.4 Monte Carlo simulation

Signal samples for the Standard Model

In this study, Monte Carlo samples are prepared based on hypothetical modeling for comparisons with the ATLAS data. In the preparation, event generators are used to compute the amplitude for events, randomly generate events based on the expected amplitude, and compute the shower and hadronization of their particles. The detector simulation and reconstruction processes are processed on these Monte Carlo events for an analogue to the data events retrieving.

The Monte Carlo Standard Model Higgs boson samples with different production channels ggF, VBF, VH, tH, ttH and bbH are produced separately. The SM ggF produced Higgs boson sample is generated at the NLO accuracy for the strong coupling α_s using POWHEG-BOX generator [115–119]. The accuracy is raised to NNLO through the reweighting using the NNLOPS method [120–123]. Therefore, the NNLO PDF set PDF4LHC15NNLO is used for the ggF generation [124].

The samples for other Higgs boson production modes are modeled with strong coupling accuracy at the NLO. The VBF, VH and ttH channel samples are generated using the POWHEG-BOX generator separately [115–117, 125, 126]. The PDF set for producing these samples is

PDF4LHC15NLO [124]. The Higgs boson associated production with a top quark sample is generated using the MADGRAPH5_AMC@NLO generator [127, 128]. The PDF set for this channel using the NNPDF23 [129].

For the SM Higgs boson samples in all these channels, the decay of charm and bottom quarks are modeled using EVTGEN generator [130]. The $H \rightarrow ZZ^* \rightarrow 4l$ decay and showering of the signal samples are modeled using the PYTHIA 8 generator [131, 132].

Background samples for the Standard Model

The irreducible backgrounds are the non-signal events with four prompt leptons. In this study, the irreducible backgrounds simulated include the $(Z^{(*)}/\gamma^*)(Z^{(*)}/\gamma^*)$ continuum referred to as ZZ^*, three $W^{\pm}$ and Z bosons events with at least four leptons in the final state, and $t\bar{t} + Z$ events with both top quarks semi-leptonic decay and Z boson decay to two leptons.

The main irreducible backgrounds ZZ^* are produced from the annihilation of the quark-antiquark pair denoted as qqZZ and from the gluon pair through a quark box loop denoted as ggZZ. In the quantum field theory, the transition amplitudes between the initial and final states form the S-matrix. The cross-section of a transition process is proportional to the absolute square of the corresponding matrix element. The matrix elements of both of the processes are modeled by SHERPA generator [133, 134]. The NNPDF3.0 NNLO PDF set is used [129, 135]. The QCD accuracy is NLO for 0 and 1 jet final state and LO for 2 and 3 jets final state. Higher jet number processes and the LO ggZZ process with a loop in the Feynman diagram are with higher QCD accuracy. The software COMIX [136] and OPENLOOP [137] are used for modeling. The parton shower is simulated by SHERPA [138] and merged by MEPS@NLO [139]. The EW accuracy is to NLO applied as a function of ZZ^* invariant mass [140, 141]. The three weak boson and $t\bar{t} + Z$ background samples are modeled by SHERPA generator as well

The reducible backgrounds are the events without four prompt leptons. The final state contains non-prompt or misidentified electrons or muons. In this study, the reducible backgrounds simulated include WZ events, Z+jets events and $t\bar{t}$ events.

The WZ samples are modeled using POWHEG-BOX v2 interface [115]. The parton shower and hadronization are simulated by PYTHIA 8 [132]. The decays of the bottom and charm hadrons are simulated by EVTGEN generator [130]. The CT10nlo and CTEQ6L1 PDF set [142] is used for the matrix element and parton shower respectively. The AZNLO tune [143] is used.

The dominant reducible background Z+jets samples are modeled by SHERPA generator. The QCD accuracy is to NLO for up to two jets and LO for more jets. The matrix element modeling uses COMIX and OPENLOOP. The parton shower is simulated by SHERPA and merged by MEPS@NLO.

The $t\bar{t}$ samples are modeled by POWHEG-BOX v2 and the showering is also done with the PYTHIA 8. The A14 tune [144] is used for hadronization. The heavy-flavour hadron decay simulation is with EVTGEN generator.

EFT samples beyond the Standard Model

In this analysis, the potential CP violation is searched through the measurement of effective couplings. The Monte Carlo samples beyond the Standard Model are parametrized following the SMEFT framework as described in Section 3.1. The VBF and VH-produced Higgs boson events have the Higgs boson to Z bosons couplings in both the production and decay vertices. In the VBF production measurement, the BSM effect is estimated for the VBF and VH channel events. The other Higgs boson production modes have these couplings only in the decay vertex. In the $H \rightarrow ZZ^* \rightarrow 4l$ decay measurement, the BSM effect in the non-VBF produced Higgs boson events is estimated in the dominated ggF events. Therefore, the ggF samples are generated independently and the VBF and VH samples are generated together as the VBF/VH samples with the BSM couplings.

The EFT samples are generated using the MADGRAPH5_AMC@NLO generator with the SMEFTsim_A_U35_MwScheme_UFO_v2.1 model [127, 145]. The jets in the final state of ggF events are from the strong interaction. The CKKW-L jet merging process is applied to reduce the double counting of QCD jet modeling in the ggF event generation and the parton shower [146]. The SMEFTsim model in the current analysis only supports the leading order modeling of the VBF events. Therefore, the BSM EFT samples in this study are only at the leading order QCD accuracy.

Chapter 6

Analysis

6.1 Optimal observables

In this analysis of effective coupling measurements, the amplitudes of the scattering processes are supposed to be with contributions from both SM and effective couplings in BSM hypotheses. The squared absolute values of S-matrix elements characterizing the amplitudes are expanded as:

$$|\mathcal{M}_{\text{Mix}}(c)|^2 = |\mathcal{M}_{\text{SM}} + \mathcal{M}_{\text{BSM}}(c)|^2 = |\mathcal{M}_{\text{SM}}|^2 + 2\Re(\mathcal{M}_{\text{SM}}\mathcal{M}^*_{\text{BSM}}(c)) + |\mathcal{M}_{\text{BSM}}(c)|^2, \quad (6.1.1)$$

where $\mathcal{M}_{\text{Mix}}$ represents the total matrix element, $\mathcal{M}_{\text{SM}}$ is the matrix element of the pure SM, and $\mathcal{M}_{\text{BSM}}$ is the matrix element of the BSM contribution only. The label c stands for the BSM couplings.

Based on these expanded terms, the optimal observables (OO) of a process are constructed for the measurement [147–149]. The optimal observables are defined as follows:

$$OO(c) = \frac{|\mathcal{M}_{\text{mix}}(c)|^2 - |\mathcal{M}_{\text{BSM}}(c)|^2 - |\mathcal{M}_{\text{SM}}(c)|^2}{|\mathcal{M}_{\text{SM}}(c)|^2}. \quad (6.1.2)$$

The definition of these observables aims to minimize the statistical uncertainties and maximize the significance. Formally, these observables are the relatively increased amplitude contribution from additional BSM couplings of their interference with the SM. From the definition, these observables are defined under a certain BSM hypothesis with couplings to measure. Their values vanish under the pure SM assumption. In this analysis, the Higgs boson couplings to Z bosons appear in the production vertices in the VBF channel and in the decay vertices in all $H \to ZZ^* \to 4l$ events. Therefore, OO can be constructed for both production and decay vertices. For the production vertices, the amplitudes are computed from the four-momenta of

two jets and the observables are denoted as OO_{jj}. For the decay vertices, the amplitudes are computed from the four-momenta of four leptons and the observables are denoted as $OO_{4\ell}$. The OO_{jj} distributions based on $\widetilde{c}_{zz} = 1$ hypothesis simulated with the SM and BSM expectations are shown in Figure 6.1 as an example [19]. The SM expected distribution has a symmetric shape. The $\widetilde{c}_{zz} = 5$ positive coupling is expected to shift the distribution to the right. The $\widetilde{c}_{zz}$ =-5 negative coupling is expected to shift the distribution to the left inversely. These simulations indicate the sensitivity of the optimal observables to the CP-odd couplings.

To construct the optimal observables in both production and decay processes, a BSM hypothesis needs to be specified. This includes the choice of the coupling type and the value of this coupling. In this study, the OO for measuring a certain coupling is constructed based on the hypothesis with the corresponding coupling. The coupling values are all fixed to 1 since this does not affect the measurement. The comparison of the SM expected $OO_{jj}^{\widetilde{c}_{zz}}$ distributions based on the hypotheses with the different $\widetilde{c}_{zz}$ values is shown in Figure 6.2. These distributions are from the Monte Carlo sample without parton shower modeling. The lower panel of the second plot shows the contents in the same indexed bins for $\widetilde{c}_{zz} = 1$ and $\widetilde{c}_{zz} = 2.65$ hypotheses are almost the same. This indicates the choice of the coupling value of the hypothesis constructing the OO does not distort but only squeezes or stretches the shape. In the coupling measurement, the total likelihood is computed by the product of the likelihood for each bin content. With properly optimized binning schemes, the observables based on different coupling value hypotheses are equivalent. In this example, the bin edges are multiplied by 2.65 with the coupling value in the hypothesis enhanced by a 2.65 factor to align the bins. Therefore the hypothetical coupling values are chosen to be 1 for convenience.

In this analysis, the measurement is based on the shape of OO. The binning scheme is essential to the measurement significance. Intuitively, increasing the bin number captures the BSM impact on the finer characteristic in OO distributions. On the other side, too many bins cause failure in predicting bin contents for low-concentration regions, making the measurement unstable. It is ideal to get as fine binning as possible while keeping the bin contents well-simulated. To achieve maximum significance, the bin width is allowed to vary. The optimization scheme aiming at getting equal-populated events in each bin is used. The effect of BSM couplings on distribution is taken into account. In this study, a half contribution from the expected SM distribution and a half contribution from the BSM distributions is used in the optimization. The BSM contribution used is the OO distribution expectation with the coupling value at the expected 1σ significance for each coupling. In the VBF production measurement, each OO_{jj} is separated into 12 bins in the optimization. In the $H \to ZZ^* \to 4l$ decay measurement, the Higgs boson events are used inclusively in the measurement. The event population is larger. Each $OO_{4\ell}$ is separated into 48 bins in the optimization. The optimization results for these observables are

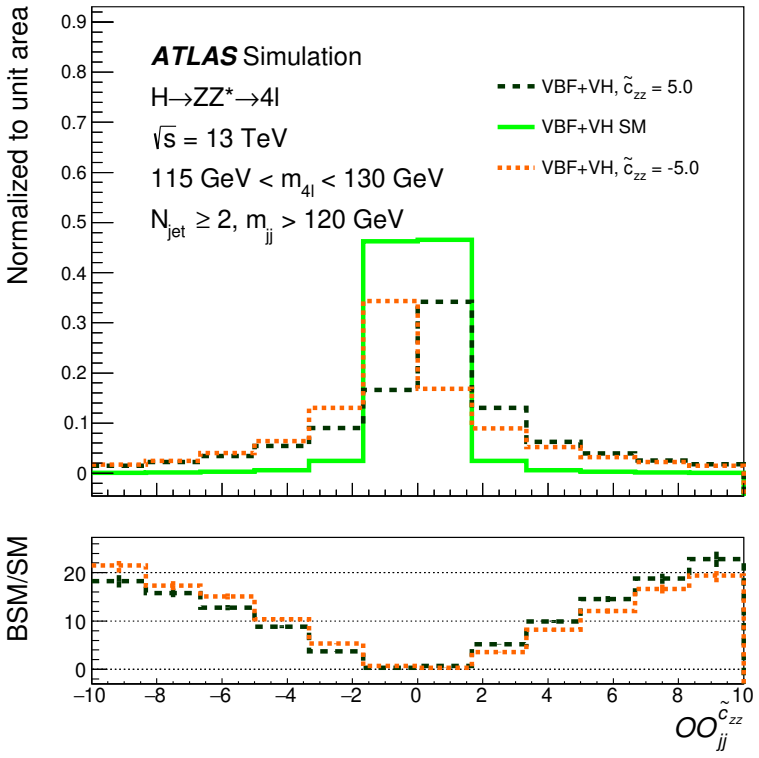


Figure 6.1: OO_{jj} distributions for the SM expectation and $\widetilde{c}_{zz} = \pm 5$ expectations VBF/VH produced Higgs boson candidates based on $\widetilde{c}_{zz} = 1$ hypotheses in the four-lepton mass window 115 GeV$< m_{4l} <$ 130 GeV with two-jet mass $m_{jj} > 120$ GeV.

summarized in Appendix A.

6.2 Events categorization

Production side scheme

Among the main Higgs boson production channels at the LHC, VBF/VH production channels involve the Higgs boson to Z bosons couplings. Both VBF and VH production channels are characterized by the two jets produced together with the Higgs boson. These channels have the same type of particles in the final state. In the VBF production channel, the Higgs boson is emitted from the vector boson exchanging between the two jets. The distributions of kinematic observables for the two jets reflect the couplings of the Higgs boson to the vector bosons. The VH production channel has one of its final states with the same particles. The two jets in a VH event come from the hadronic decay of the vector boson produced together with the Higgs boson. Their invariant mass peaks around $W^{\pm}$ and Z bosons mass. The kinematic observable

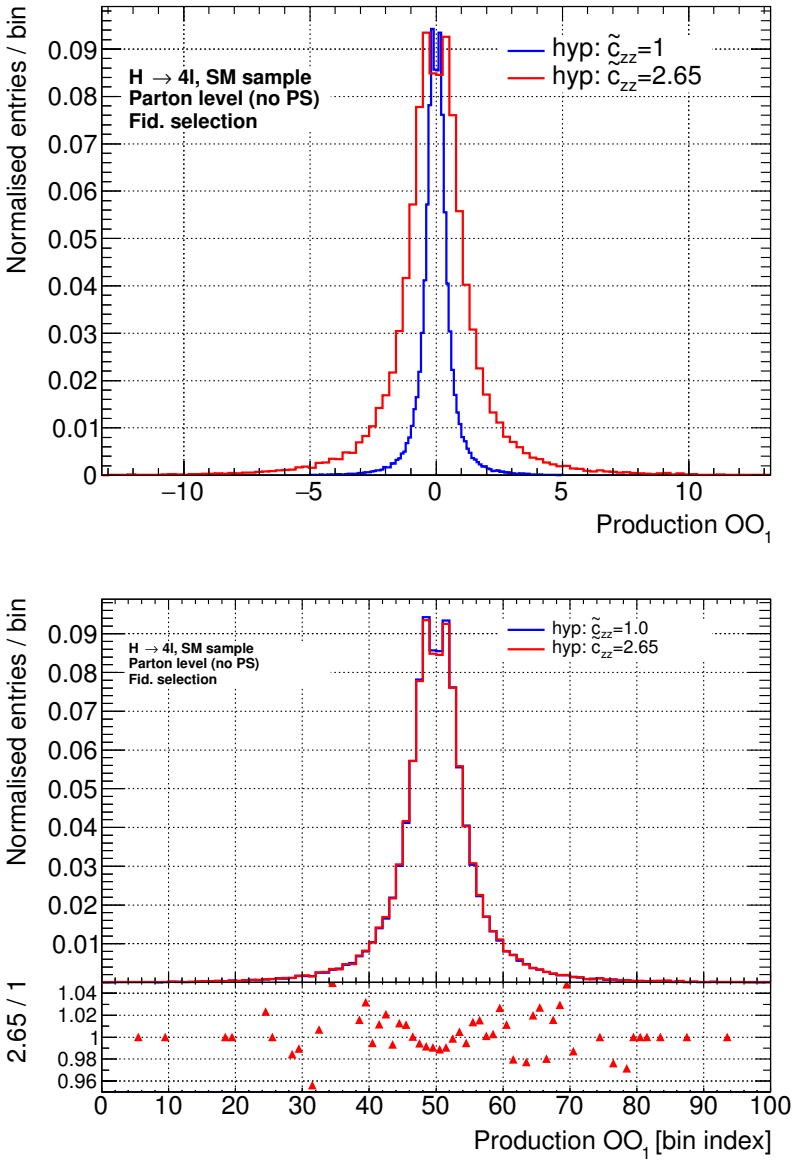


Figure 6.2: The comparison of OO_{jj} distribution based on $\widetilde{c}_{zz} = 1$ (blue curve) and $\widetilde{c}_{zz} = 2.65$ (red curve) hypotheses on a parton-level SM-like sample. The upper plot is the overlay 100 bins plots ranging $[-13.25, 13.25]$ for hypothesis $\widetilde{c}_{zz} = 2.65$ and $[-5, 5]$ for hypothesis $\widetilde{c}_{zz} = 1$. The lower plot is a bin-by-bin comparison of the two distributions. The lower panel of this plot is the ratio of the $\widetilde{c}_{zz} = 2.65$ curve to the $\widetilde{c}_{zz} = 1$ curve in each bin.

distributions in this channel are not sensitive to the Higgs boson to Z bosons couplings. Therefore, the VBF Higgs boson events are the only signal events in the production measurement.

The irreducible backgrounds for the production measurement contain the Higgs boson events from the other production channels and the non-Higgs boson four-lepton events. The Higgs

boson backgrounds include the events from ggF, VH, ttH and bbH channels. The dominant background events are from ggF. In these background events, there are hadrons emitted from quarks or gluons in these channels through strong interaction. The associated productions have additional jets from the decay of weak bosons or heavy flavor quarks as well. The non-Higgs boson irreducible backgrounds include the ZZ^* continuum events not from the Higgs boson decay, events with three $W^{\pm}$ and Z bosons and $t\bar{t} + Z$ boson events. These backgrounds are dominated by the ZZ^* continuum events. In these background events, the four leptons come from the decay of $W^{\pm}$ and Z bosons and top quarks. The jets come from the strong interactions and decay of $W^{\pm}$ bosons or top quarks in these channels. There are small contributions from the reducible backgrounds without four prompt leptons. These events are from the $t\bar{t}$ and Z+jets processes.

The event categorization of the production measurement is determined based on this principle. The signal regions (SR) are the region abundant with signal events and suppressing the backgrounds. In the production measurement, the VBF events are the only signals. These events are with at least two jets. The invariant mass for the two leading jets m_{jj} is required to be larger than 120 GeV to reject most of the VH background. The contribution of the ggF Higgs boson backgrounds in this signal region is not small. Therefore, the machine learning technique is introduced to the production SR for further categorization to improve sensitivity [150].

In this study, a neural network (NN) score is used as the discriminant between the VBF signal events and background events to enhance the sensitivity of the measurement. This discrimination follows the NN used in the ATLAS Run 2 $H \to ZZ^* \to 4l$ simplified template cross-section and EFT measurement [9]. In this NN, the transverse momentum and pseudorapidity of the leading jets and leptons in the final state are input to recurrent neural networks. This information together with the enhancement from kinematic observables is fed into a multilayer perceptron for deep learning. These observables include the total transverse momentum for four leptons and two jets p_{T}^{4ljj} and the Zeppenfeld variable defined as the pseudorapidity difference between four leptons and two jets $\eta_{ZZ}^{\mathrm{Zepp}} = |\eta_{4l} - (\eta_{j1} + \eta_{j2})/2|$ [151]. The trained VBF NN score ranges from 0 to 1. The larger value means the event is more VBF-like. It is used to separate the signal region into four sub-categories VBF SR1, VBF SR2, VBF SR3 and VBF SR4 with NN score value inside the intervals $(0.0, 0.2)$, $(0.2, 0.6)$, $(0.6, 0.9)$ and $(0.9, 1.0)$ respectively. The discriminating power of this NN score is shown by the SM expected event composition in each category in Figure 6.3 [19]. In the SM expectation, the VBF and VH-generated events take increasing shares in the total events from VBF SR1 to VBF SR4. The events are partitioned into these regions in the VBF measurements.

The background events in the signal regions take non-negligible shares from the simulations based on the SM expectation. Their kinematic observable distributions do not have strong

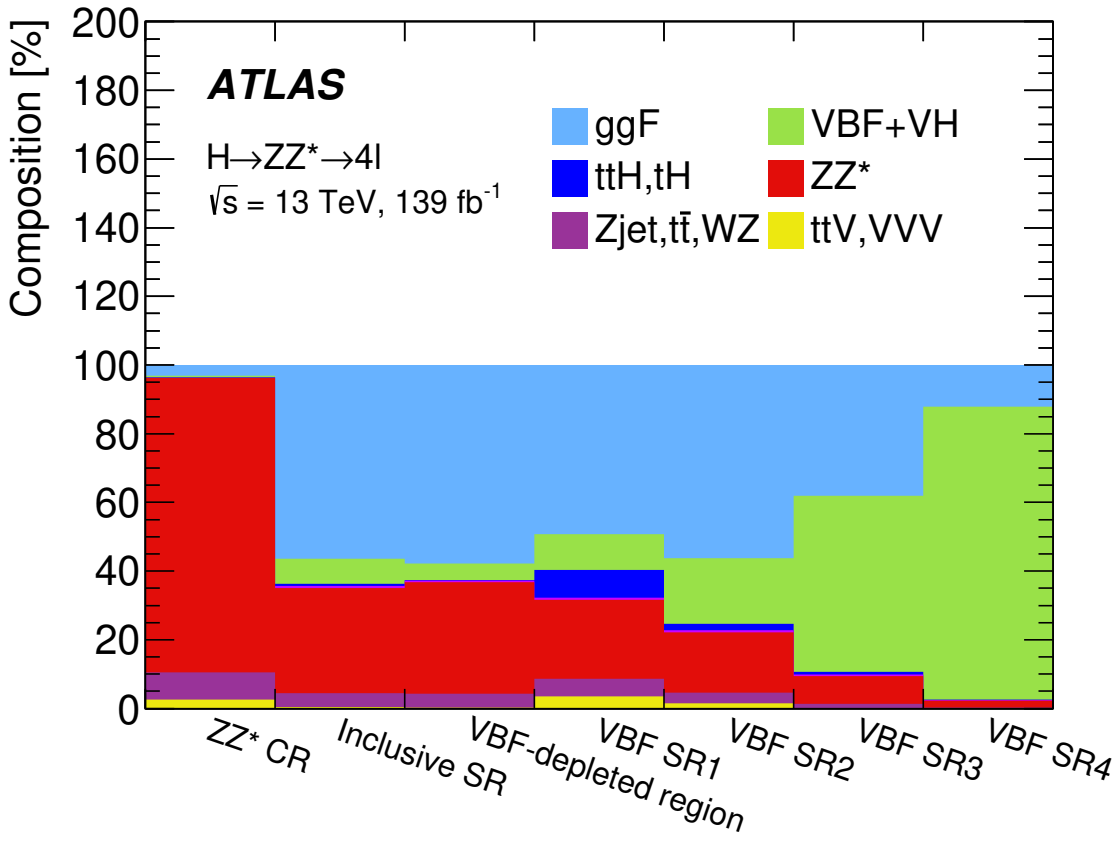


Figure 6.3: Expected percentage composition of the different channel events in each signal region or control region.

sensitivity as the distributions for the signal events. Therefore, their yields should be estimated properly. The control regions (CR) are introduced where the backgrounds are dominant. These regions are used to constrain the contribution of background events. In the production measurement, the region where the events pass the Higgs boson selection but fail the VBF signal selection named the VBF-depleted region is the control region for constraining the Higgs boson backgrounds. The sidebands with four-lepton invariant mass m_{4l} in $[105, 115]$ GeV or $[130, 160]$ GeV are with suppressed Higgs boson events and dominated by the ZZ^* continuum events. This range is used for constraining the non-Higgs boson irreducible backgrounds denoted by ZZ^* CR. The signal regions and the control regions of the production side study are summarized in Table 6.1.

Decay side scheme

In decay side analysis, all the Higgs boson events with $H \rightarrow ZZ^* \rightarrow 4l$ decay are signal events. Since the final state four-lepton kinematic observables do not depend on the production

Table 6.1: Signal regions and control regions of the VBF production measurement

Category	Selection
Signal Regions	
VBF SR1	($m_{jj} > 120$ GeV AND $n_{jets} > 1$) AND ($115 < m_{4l} < 130$ GeV) AND ($0.0 < NN_{VBF} < 0.2$)
VBF SR2	($m_{jj} > 120$ GeV AND $n_{jets} > 1$) AND ($115 < m_{4l} < 130$ GeV) AND ($0.2 < NN_{VBF} < 0.6$)
VBF SR3	($m_{jj} > 120$ GeV AND $n_{jets} > 1$) AND ($115 < m_{4l} < 130$ GeV) AND ($0.6 < NN_{VBF} < 0.9$)
VBF SR4	($m_{jj} > 120$ GeV AND $n_{jets} > 1$) AND ($115 < m_{4l} < 130$ GeV) AND ($0.9 < NN_{VBF} < 1.0$)
Control Regions	
VBF-depleted region	($n_{jets} < 2$ OR $m_{jj} < 120$ GeV) AND ($115 < m_{4l} < 130$ GeV)
ZZ^* CR	($105 < m_{4l} < 115$ GeV) OR ($130 < m_{4l} < 160$ GeV)

channels, there is no need to split the inclusive Higgs boson signal region in this measurement. The irreducible backgrounds of this measurement contain the same non-Higgs boson $4l$ backgrounds as in the production measurement. The yields of these backgrounds are constrained in the same ZZ^* CR. The SR and the CR of the decay side study are summarized in Table 6.2. The events are used inclusively in the $H \to ZZ^* \to 4l$ measurement.

Combined measurement scheme

In the investigation of the categorization and measurement scheme, both production and decay side analyses show relatively high sensitivities in the measurement of $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$. A combined measurement of production and decay vertices is proposed for these couplings. In the combined measurement, the signal includes the Higgs boson events with all production channels and $H \to ZZ^* \to 4l$ decay. Among these signals, the VBF events have kinematic information in leading jets sensitive to CP-odd couplings. Therefore, the combined measurement adapts the same categorization definition as the production measurement. For VBF-like events, information from their production vertex is used, while for the other Higgs boson events, their decay information is used. The VBF-depleted region is considered an SR instead of a CR in the production measurement. The observables for the VBF-like regions are the OO_{jj}. The

observables used in the VBF-depleted region are the $OO_{4\ell}$.

6.3 Measurement method

In particle physics experiments, the results are usually reported as a statistical inference of some hypotheses. The hypotheses at least include a null hypothesis H_0 and an alternative hypothesis H_1. The alternative hypothesis H_1 needs to have predictions different from the H_0. If H_0 is excluded in favour of H_1 to a certain satisfied confidence level by the statistical evidence observed from the experiment, there is a discovery. In many analyses, there are a series of alternative hypotheses characterized by some parameters θ. In this study, the null hypothesis is the SM and the alternative hypotheses include hypotheses characterized by additional BSM CP-odd couplings.

The statistical inference method used in this study is the maximum likelihood estimation method. A likelihood is a quantity measuring how reasonable a hypothesis is based on an observation outcome. It is defined as the probability of the observation given the hypothesis. In this shape-based measurement, the likelihood for a hypothesis with BSM coupling c_{BSM} has the form:

$$\begin{aligned} L(c_{\text{BSM}}, \theta) =& \Pi_i^{N_{\text{categories}}} \text{Poisson}(n^{\text{obs}} | \mu \cdot n_{s,i}(c_{\text{BSM}}) + n_{b,i}) \times \\ & P(x | \mu \cdot n_{s,i}(c_{\text{BSM}}) + n_{b,i}) \times \Pi_m^{N_{\text{nuisance}}} C_m(\theta). \end{aligned} \tag{6.3.1}$$

The high-energy physics experiments are essentially the counting experiments in each phase space. Therefore, Poisson distributions are used to describe the probability density of event numbers in each category:

$$\text{Poisson}(n|\bar{n}) = \frac{\mu^n e^{-\bar{n}}}{n!}, \tag{6.3.2}$$

where $\bar{n}$ is the expected count. The symbol n_i^{obs} is the observed event number in the category labeled by the index i, $n_{s,i}$ is the expected signal event number and $n_{b,i}$ is the expected background event number. As a shape-only study, the μ factors are introduced as normalization factors determined by data. The shape distribution of observable x is described by the probability density function P. The systematical uncertainties in this study are estimated by nuisance parameters θ and the constraints C_m are set correspondingly.

The likelihood of a hypothesis might be misunderstood as the probability of the hypothesis being true. It should not be interpreted as a probability as it is not normalized to 1. There are different choices of observables as the statistics of a likelihood analysis. The likelihood value of a hypothesis from a certain observation depends on this choice. Therefore the interpretation should be based on the comparison of likelihood values with different hypotheses instead of the values themselves. For the maximum likelihood estimation (MLE), the estimated values for the

parameters are the ones that maximize the likelihood function:

$$L(\hat{c}_{\text{BSM}}, \hat{\theta}) = \max L(c_{\text{BSM}}, \theta), \tag{6.3.3}$$

where $\hat{c}$ denotes the estimated values of the couplings. In this analysis, the effective couplings are estimated given an observation with the MLE.

The likelihood is further interpreted in the frequentist view in this analysis. In this view, the confidence interval and the *p*-value or significance computed from the likelihood values are the results of measurement [152]. The confidence interval is a range of hypothesis parameters that are considered to contain the true parameter with a certain confidence level. In this analysis, the 68% and 95% confidence intervals are estimated. This means with all the EFT coupling values in this range, the probability of observing a more "extreme" distribution over the real observation is less than 16% and 2.5% respectively. The *p*-value of an observation based on the null hypothesis is defined as the probability of observing a more "extreme" result than the actual observation. In this analysis, the *p*-value describes the probability to reject the SM hypothesis with the observation.

In this analysis, the log-likelihood ratio used as the test statistic [153]:

$$q = -2\ln(L(c_{\text{BSM}}, \hat{\hat{\theta}})/L(\hat{c}_{\text{BSM}}, \hat{\theta}), \tag{6.3.4}$$

where the double-hat $\hat{\hat{\theta}}$ is the local nuisance parameter maximizing the likelihood depending on the floating c_{BSM}. For the nearly Gaussian-distributed event counts in each bin, the statistic value at the boundary confidence interval corresponds to the square of the significance. Therefore the limits for 68% confidence interval are with about $q = 1$ and the limits for 95% confidence interval are with about $q = 4$.

There are one-dimensional and two-dimensional measurements in this analysis. In the one-dimensional measurement, each time only one coupling is changing along the tested hypotheses. The other couplings are always set to 0. A confidence interval of that coupling is received for each measurement. The two-dimensional measurement each time simultaneously floats two couplings in the measurement with the third coupling set to 0. In this approach, the results are presented as contours at the 68% and 95% confidence levels. The correlation of the couplings is seen as well from the shape of the contours.

6.4 Morphing technique

In this study, a morphing technique is used for building the expected distributions of the models with different couplings [154]. The likelihood fit compares the observable distributions of data

and the expectations of the hypotheses with different coupling values. The observable distributions for all the sampling points in the coupling value range of interest are needed. In principle, Monte Carlo simulation would provide expected observable distributions given any set of coupling parameters. In reality, it is inefficient and impractical to perform dedicated simulations for every required coupling value. The method of morphing is to interpolate the observable distribution expectations of any points, based on a relatively smaller size of Monte Carlo simulated samples. This technique is applied to alleviate the burden of generating and simulating heavy amounts of datasets.

In an observable distribution histogram, the content in each bin is the count of observed or expected events within the bin coverage. The yield of events in a channel and a kinematic region is squared proportional to the absolute value of the transition amplitude:

$$N_{\text{event}} \propto |\mathcal{M}|^2, \tag{6.4.1}$$

where N_{event} is the expected number of events and $\mathcal{M}$ is the transition amplitude as a matrix element of the S-matrix. The matrix element receives contributions from both the SM and BSM couplings. In the simplest case, the matrix element of a process with one BSM coupling is expanded with the coupling extracted out as:

$$|\mathcal{M}_{\text{mix}}|^2 = \kappa_{\text{SM}}^2|\mathcal{M}_{\text{SM}}|^2 + g_{\text{BSM}}^2|\mathcal{M}_{\text{BSM}}|^2 + 2\kappa_{\text{SM}}g_{\text{BSM}}\Re(\mathcal{M}_{\text{SM}}^*\mathcal{M}_{\text{BSM}}), \tag{6.4.2}$$

where κ_{SM} is the factor for scaling the SM coupling, g_{BSM} is the BSM coupling, $\mathcal{M}_{\text{SM}}$ is the matrix element in the SM and $\mathcal{M}_{\text{BSM}}$ is the matrix element of a theory with only unit BSM coupling. From this expansion, the yield can be viewed as a linear superposition of the pure SM term $|\mathcal{M}_{\text{SM}}|^2$, pure BSM term $|\mathcal{M}_{\text{BSM}}|^2$ and interference term $\Re(\mathcal{M}_{\text{SM}}^*\mathcal{M}_{\text{BSM}})$ with the coefficients as functions of coupling values. The total amplitude will be predicted if these three terms are known. Each expansion with a specific coupling value is a linear equation of three unknown variables. Three samples with different values of the BSM coupling are needed to solve for the three terms.

In the more general scenario, there are multiple BSM couplings appearing in production and decay processes in a model. In this study, the couplings of interest are the CP-odd Higgs boson effective couplings to Z bosons. In the VBF and VH production channels, these couplings affect both production and decay of the Higgs boson. In the ggF-dominated other production channels, these couplings affect only the decay process. The computation of the amplitude using the Feynman rule includes the contribution from interaction vertices, incoming and outgoing particle legs, and intermediate particle propagators [42]. The change in the couplings only affects the interaction vertices. The legs and propagators are independent of the couplings. Therefore their

Table 6.2: Signal region and control region of the $H \to ZZ^* \to 4l$ decay measurement

Category	Selection
Signal Regions	
inclusive SR	$115 < m_{4l} < 130$ GeV
Control Regions	
ZZ^* CR	$(105 < m_{4l} < 115)$ GeV OR $(130 < m_{4l} < 160)$ GeV

contribution to the amplitude is eliminated in the interpolation. The expansion of the matrix element based on the Feynman rule focuses on the vertex contributions. The contributions from legs and propagators independent of the couplings are absorbed into a proportionality factor. The general matrix element expansion is:

$$\begin{aligned} |\mathcal{M}_{\text{mix}}|^2 \propto |(\kappa_{\text{SM}}\mathcal{O}_{\text{SM}} + \sum_{i\in\text{prod}} g_{\text{BSM},i}\mathcal{O}_{\text{BSM},i} + \sum_{k\in\text{shared}} g_{\text{BSM},k}\mathcal{O}_{\text{BSM},k}) \\ (\kappa_{\text{SM}}\mathcal{O}_{\text{SM}} + \sum_{j\in\text{decay}} g_{\text{BSM},j}\mathcal{O}_{\text{BSM},j} + \sum_{k\in\text{shared}} g_{\text{BSM},k}\mathcal{O}_{\text{BSM},k})|^2, \end{aligned} \tag{6.4.3}$$

where the subscript i represents all the couplings involved in the production process, j represents all the couplings involved in the decay process and k represents all the couplings shared by both the production and decay process. The symbol $\mathcal{O}$ represents the contribution to the amplitude of the SM coupling or a unit BSM coupling to the vertex [154].

The yield of events with a series of couplings is:

$$\text{yield} = C\sum_{i}^{N} f_i(\kappa_{\text{SM}}, g_{\text{BSM1}}, g_{\text{BSM2}}, \ldots)\mathcal{O}_i, \tag{6.4.4}$$

where C is the proportionality factor, independent of the vertex contribution and couplings. The symbol $\mathcal{O}_i$ is a product of the vertex contributions from a coupling in the production and a coupling in the decay, and the symbol f_i is the corresponding sum and product of the couplings. In the example of the VBF production process with $c_{H\widetilde{W}}$ coupling shared by the production and decay of the Higgs boson, the yield is expanded as:

$$\begin{aligned} \text{yield} &\propto |(\kappa_{\text{SM}}\mathcal{O}_{\text{SM}} + c_{H\widetilde{W}}\mathcal{O}_{c_{H\widetilde{W}}})(\kappa_{\text{SM}}\mathcal{O}_{\text{SM}} + c_{H\widetilde{W}}\mathcal{O}_{c_{H\widetilde{W}}})|^2 \\ &\propto |\kappa_{\text{SM}}|^2|\mathcal{O}_{\text{SM}}|^2 + |2\kappa_{\text{SM}}c_{H\widetilde{W}}||\mathcal{O}_{\text{SM}}\mathcal{O}_{c_{H\widetilde{W}}}| + |c_{H\widetilde{W}}|^2|\mathcal{O}_{c_{H\widetilde{W}}}|^2 \end{aligned} \tag{6.4.5}$$

where $|\kappa_{\text{SM}}|^2$, $|2\kappa_{\text{SM}}c_{H\widetilde{W}}|$ and $|c_{H\widetilde{W}}|^2$ are the functions f_i and $|\mathcal{O}_{\text{SM}}|^2$, $|\mathcal{O}_{\text{SM}}\mathcal{O}_{c_{H\widetilde{W}}}|$ and $|\mathcal{O}_{c_{H\widetilde{W}}}|$

are the operators $\mathcal{O}_i$. Counting based on the binomial expansion gives the number of terms:

$$\begin{aligned} N =& \frac{1}{24}s(s+1)(s+2)[(s+3)+4(p+d)] \\ &+ \frac{1}{4}[s(s+1)p(p+1)+s(s+1)d(d+1)+p(p+1)d(d+1)] \\ &+ \frac{1}{2}pds(p+d+s+3), \end{aligned} \tag{6.4.6}$$

where N is the number of terms, p is the number of independent couplings only in the production vertex, d is the number of decay-only ones, and s is the number of shared by production and decay ones [154].

To predict the yield with any coupling values for which Equation 6.4.4 contain N terms, the couplings and yields for N different samples are required. The morphing technique is to predicts the yield of events for given coupling configurations from the Monte Carlo samples input. Yields of both MC sample events and predicted coupling events follow Eqaution 6.4.4. If the equations of a set of enough number samples form an independent linear system, each $C \times \mathcal{O}_i$ can be solved and the yield of any coupling configurations can be predicted. The number of input samples needed is determined by the number of independent terms predicted by Equation 6.4.6. The set of input samples is named "morphing basis". The yield of events for any coupling configuration is predictable by the linear superposition of the yields of morphing basis samples. This can be written as follows:

$$\text{yield}_{\text{out}} = \sum_{i}^{N} c_i(g_{\text{out}}, g_{\text{in}})\text{yield}_{\text{in},\,i}, \tag{6.4.7}$$

where $c_i(g_{\text{out}}, g_{\text{in}})$ are coefficient depends only on the coupling of predicted couplings and the couplings of the i-th input sample. These coefficients are named morphing functions.

For a given binning scheme, the shape distribution of an observable is a series of yields in each bin. Each bin defines a kinematic region that does not intersect. These yields are independent of each other. For every bin, the morphing functions are the same. The shape of a determined coupling point is predicted by:

$$T_{\text{out}} = \sum_{i}^{N} c_i(g_{\text{out}}, g_{\text{in}})T_{\text{in},\,i}, \tag{6.4.8}$$

where T_{out} is the vector of bin contents with the predicted couplings, $T_{\text{in},\,i}$ is the vector of bin contents of the i-th input samples.

For the ggF process, the couplings of the Higgs boson to Z bosons appear only in the decay

process. From the above formula, the n_s is always 1 for the SM coupling scaling κ_{SM} and n_d is the number of BSM HZZ couplings. For the VBF and VH processes, both production and decay of the Higgs boson involve the couplings of the Higgs boson to Z bosons. The n_s for the processes is the number of the BSM HZZ couplings $+1$ for the SM coupling. The number of samples needed is shown in Table 6.3.

Morphing basis optimization

In the morphing procedure, the input MC samples have events obeying the Poisson distribution. The yield in each bin has uncertainty. These uncertainties make the prediction from the morphing technique imperfect. The uncertainties in the shape distribution predicted from Equation 6.4.8 are propagated from the uncertainties in the input MC samples:

$$\Delta T_{\text{out}} = \sqrt{(c_i(g_{\text{out}}, g_{\text{in}})\Delta T_{\text{in}})^2}. \tag{6.4.9}$$

The morphing functions c_i can be positive or negative. In the prediction of the shape distribution, the contributions from different samples with opposite sign morphing functions cancel each other. In the uncertainty estimation, the uncertainties in the bin content of each input sample accumulate independently of the sign of the morphing functions. If the morphing functions have very large absolute values with different signs, the uncertainties accumulated in the morphing results are very large. The uncertainties in the morphing results are larger than those of directly generated datasets with the same yield. This means the morphing results get decreased statistical power. The "effective statistics" of morphing results are defined as the yields in a directly generated dataset with the same uncertainties.

The morphing method predicts the distribution with acceptable uncertainties only within a certain range of parameter space where the morphing functions are not too large. The morphing functions depend on the couplings of the input MC samples. For an interesting coupling range, the morphing basis should be optimized for minimizing the uncertainties throughout the range. The morphing basis needs to be validated for a morphing application.

Morphing basis determination is highly non-intuitive for a multi-dimensional coupling space. In this study, the morphing basis is chosen by an iteration algorithm demonstrated in Figure 6.4. The algorithm begins with an arbitrary basis. In iterative executions, the basis is randomly modified by changing one of the coupling values in a random input morphing basis. A set of test points is selected within the interested coupling range for evaluating the modification. The modifications that increase the total effective statistics at these test points are retained. A detailed description and validation of the optimization algorithm are reported in Appendix B.

Table 6.3: The number of samples needed and couplings involved in the ggF and VBF/VH channels.

Channel	HVV vertices	BSM couplings	n_p	n_d	n_s	N
ggF	1	1	0	1	1	3
ggF	1	2	0	2	1	6
ggF	1	3	0	3	1	10
VBF/VH	2	1	0	0	2	5
VBF/VH	2	2	0	0	3	15
VBF/VH	2	3	0	0	4	35

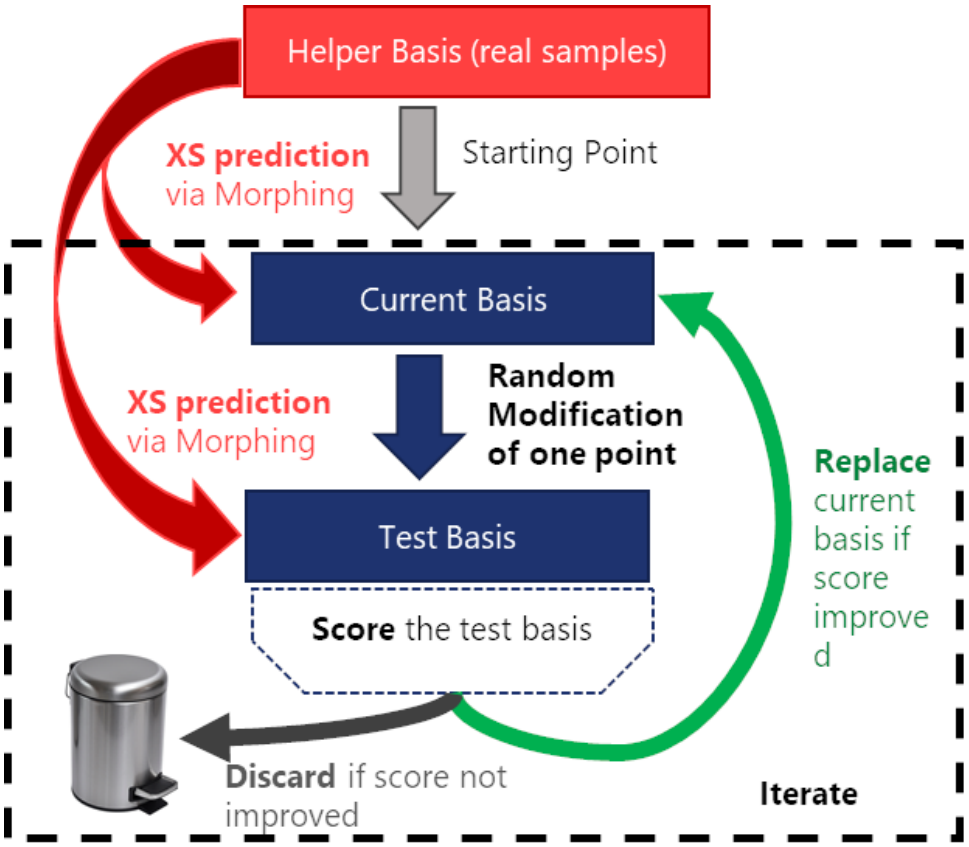


Figure 6.4: The illustration of the morphing basis optimization algorithm.

Morphing Basis

In this analysis, the expected observable distributions with any coupling values are estimated using the morphing method with three couplings simultaneously, namely $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$ and $c_{H\widetilde{W}B}$, or $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$ and $\widetilde{c}_{\gamma\gamma}$. The histograms for VBF and VH channels have these couplings in both production and decay vertices. The histograms for ggF channels have these couplings in only

decay vertices. The number of corresponding couplings is recorded in Table 6.3. 35 samples are needed to form a morphing basis for VBF and VH channels. 10 samples are needed for the ggF channel. The couplings for the bases in this study are optimized by the algorithm described above. The coupling values of the samples used in the morphing bases and their validation are reported in Appendix B.

The contributions from the other Higgs boson production channels, including bbH and ttH, are small in the overall observable distributions. These contributions are not modeled using the morphing method. The SM samples from the MC simulation are used for these channels. The potential BSM contribution to the non-resonant ZZ^* backgrounds is not estimated in this study. The MC-simulated SM samples are used as well.

6.5 Sample normalization

This analysis is a "shape-only" measurement inspired by the asymmetric OO distributions with CP-odd couplings. In the likelihood computation, the comparisons between the observation and the hypothetical predictions use the shape of OO distribution only, without considering the overall normalization. With this criterion, the precision requirement in yield modeling of the signal and background samples is relaxed. The expected OO distribution with certain effective couplings is constructed by adding up the distributions in each channel obtained from the MC simulation or morphing method. In this construction, the relative share of each channel affects the final OO distribution. The yield values of each channel in each SR and CR are retrieved from the SM samples with the highest-order corrections in this study. The presence of BSM couplings will increase the yield expectation. To minimize the impact of this effect, normalization factors relative to the SM expectation values for each channel are introduced. These normalization factors are shared in all the categories in the fit. They are allowed to float as nuisance parameters in the fit. In this approach, the relative shares of each component in the construction of hypothetical distribution are fit to data.

In the $H \to ZZ^* \to 4l$ measurement, the OO shape studied contains properties of four leptons from the Higgs boson decay. Different production channels of the Higgs boson should not make differences in their decay observable distributions. The "shape-only" criterion only needs to profile the total yield of the Higgs boson signals inclusively for different production channels. Therefore the normalization factor μ_{Signal} is introduced for all the Higgs boson events. The hypothetical change in the Higgs boson event yield due to the BSM couplings is absorbed in this factor.

The other component of the decay-side fit model is the non-resonant ZZ^* background events. For these backgrounds the normalization factor μ_{qqZZ} is introduced. The name fol-

lows the domination of the qqZZ events in ZZ^* backgrounds. This μ_{qqZZ} factor is universal in the inclusive SR and the ZZ^* CR. In the profiled likelihood fit, the OO distribution shapes in the SR are used in conjunction with the total yield in the ZZ^* CR. Since the ZZ^* events dominate in the ZZ^* CR, the μ_{qqZZ} factor is mainly controlled by data in the ZZ^* CR. With constraint on this factor, the impact of the relative share of non-Higgs boson backgrounds in the SR OO distribution is eliminated.

In the VBF production measurement, the signal only comprises VBF events. The irreducible backgrounds consist of the Higgs boson events with other production channels and the same non-Higgs boson backgrounds as the $H \to ZZ^* \to 4l$ measurement. To eliminate the impact of yield modeling on the measurement, normalization factors μ_{VBF}, μ_{ggF} and μ_{qqZZ} are introduced for VBF signal, ggF background and non-Higgs boson background respectively. The yield modeling for the Higgs boson background is determined by the normalization profile in ggF events due to their domination. In the profiled likelihood fit, the OO distribution shapes in four VBF signal regions and the total yields in the VBF-depleted region and ZZ^* CR are used simultaneously. The constraint of the μ_{qqZZ} factor is set in the same way as $H \to ZZ^* \to 4l$ measurement. The VBF-depleted region is dominated by the ggF events and ZZ^* events. With ZZ^* yield already constrained in the ZZ^* CR, the ggF normalization factor μ_{ggF} is mainly determined by the yield in the VBF-depleted region.

In this measurement, the SR separation is based on the NN score trained with the SR simulated events. The BSM couplings might affect the categorization of the VBF events and change the yields in four VBF regions. Therefore, additional normalization factors are introduced to float the yields in these regions while not affecting the total normalization in the VBF SR. These normalization factors are floated to change the relative share of each VBF SR but not influence the total yield of the four regions.

The event categorization in the production and decay combined measurement is the same as the VBF production measurement. Therefore, the same normalization factors as the production measurement are introduced in this measurement. A more detailed description and the validation of these normalization constraining schemes are in Appendix C.

6.6 Reducible backgrounds estimation

This analysis tests the CP of the Higgs boson coupling in its four-lepton decay channel. The backgrounds considered for this channel should contain four leptons in the final state as the signal events have. The measurement and constraints on the signal and the four-lepton backgrounds dominated by non-resonant ZZ^* based on their kinematic properties have been discussed in the previous sections. In addition to the four-lepton backgrounds, there are events

without four leptons in the final state contributing to the backgrounds. These events have misidentified leptons or non-prompt leptons faking the final state with four prompt leptons. They are the reducible backgrounds.

In this study, the reducible backgrounds are events with only fake or non-prompt lepton candidates in the subleading pair. The contributions from the $Z + \mu\mu$ ($2e2\mu$ and 4μ) and $Z + ee$ ($2\mu2e$ and $4e$) channels are estimated separately. The $Z + \mu\mu$ reducible backgrounds are dominated by Z boson accompanied by jets denoted as $Z+$ Jets, top quark production denoted as $t\bar{t}$ and diboson WZ process events. The non-prompt muons in jets come from the semileptonic decays of heavy-flavour jets or from π/K decays from light-flavour jets. The $Z + ee$ reducible backgrounds are dominated by events with light-flavour jets faking the electrons denoted as f, non-prompt electrons from semileptonic decay of heavy-flavour quarks denoted as q or photon electrons from photon conversions or final-state radiations denoted as γ.

The estimation of reducible backgrounds based on the MC simulation is subject to parametrization. These are not certain channel events but events contain misidentification and fake leptons. The estimation is highly affected by the modeling of selection efficiency. Therefore, the estimation of the reducible backgrounds uses a data-driven method following the STXS measurement [9]. In this approach, the possible sources are modeled by data fitting in the regions with abundant events first. These regions are control regions required to be orthogonal to the signal region of the analysis. Then the different sources are extrapolated to a common control region. The fit in this region validates the modeling. The reducible backgrounds in the analysis signal region are estimated by the extrapolation from the common control region. The process is described by the formulas:

$$\mathrm{N}_{SR} = \mathrm{TF} \times \mathrm{N}_{CR}, \tag{6.6.1}$$

where N_{CR} is the yield of reducible backgrounds in the control region, and N_{SR} is the estimation results in the signal region. The transfer factor for the extrapolation denoted as TF is estimated as the ratio of MC simulated yields in the signal region and control region.

In this study, the data-driven approach is extended to shape estimation. Each bin in a OO shape distribution is viewed as an independent yield count with the corresponding observable range requirement. In principle, the reducible background shapes should be estimated through bin-by-bin data-driven. In the real analysis, the shape analysis has many bings and the number of reducible background events is limited. To reduce the statistical uncertainties, the shape estimation scheme has the simplifications as follows. The TF estimation with MC simulation is done using a reduced number of bins to decrease the statistical uncertainties in the MC simulation. The TF value is shared by adjacent bins in the same reduced bin. In the OO_{jj} shape estimation for reducible backgrounds, the extrapolation is done into the inclusive SR without the separation into four VBF SRs to reduce the relative statistical uncertainties in the data for

each bin. The total yield in each VBF signal region is estimated independently.

$Z + \mu\mu$ shape estimation

For the $Z + \mu\mu$ reducible backgrounds, non-prompt or fake muons are allowed in the sub-leading lepton pair, which has an invariant mass father from the Z boson mass. The $Z+$ Jets and $t\bar{t}$ reducible backgrounds are estimated using the data-driven approach. These reducible backgrounds are fitted to data using the leading lepton pair mass m_{12} in four control regions simultaneously. There is also a small background from WZ events estimated using MC simulation.

The common control region for validation is the relaxIsoD0 control region defined by relaxing the isolation and impact parameter requirements for lepton candidates and the vertex requirement in the four-lepton selections. The four-lepton selected region is excluded from this region to suppress the Higgs boson events. In this region, both $Z+$ Jets and $t\bar{t}$ reducible backgrounds have large contributions. The source of the extrapolation is the remaining events after subtracting the MC-estimated Higgs boson events and other background events from the data. The transfer factors are estimated by averaging those of each component from the MC simulations with the MC yield used as weights. The systematic uncertainties of the reducible backgrounds are estimated from the data fitting uncertainties and transfer factors computation. The $Z + \mu\mu$ backgrounds take only 1%-2% share of the events in the region studied in this analysis from the estimation. The contribution of these reducible backgrounds to the likelihood fit is very small. The shape distribution estimation and details in the data-driven computation are reported in Appendix D.

$Z + ee$ shape estimation

For the $Z + ee$ reducible backgrounds, the leading lepton pair and the lepton with the larger transverse momentum p_{T} are required to be prompt leptons. The last electron in the background event is the fake or non-prompt electron denoted by "X". The reducible backgrounds contain events, where "X" is f, q or γ from the $Z+$ Jets and $t\bar{t}$ events. The f and γ backgrounds are estimated in the "$3l + X$" region, where the selection and identification criteria for X are relaxed. These backgrounds are fit to data with the number of detections on pixel sensors in the Inner Detector n_{InnerPix}. The sPlot method is used to extract the probability of an event is f or γ as p_{T}, the probability with a certain number of jets N_{jets}, and the probability distribution of NN score [155]. The q background is estimated by MC simulation instead.

The estimation of the reducible backgrounds in the regions used for the CP-odd observable measurements follows the same strategy as the $Z + \mu\mu$ distribution estimation. In the "$3l +$

X" region, the source of the extrapolation is the remaining events after subtracting the MC-estimated Higgs boson events and other background events from the data. The transfer factors are estimated by averaging those of each component from the MC simulations with the summing probability distribution from the sPlot method as yield. The $Z + ee$ backgrounds account for less than 1% share of the events in the region studied in this analysis from the estimation. The contribution to the CP-odd coupling measurements is negligible. The shape distribution estimation and details in the data-driven computation are reported in Appendix D.

6.7 Systematic uncertainties

In this analysis, the uncertainties of the measurement are dominated by the statistical fluctuations of the ATLAS Run 2 data. The systematic uncertainties from the model building affect the measurement very slightly in comparison. In this study, systematic uncertainties enter the likelihood measurement as nuisance parameters representing the upward and downward fluctuations from each type of uncertainty. The impacts of these uncertainties are represented as the change in the confidence interval limits. The main sources of systematic uncertainties contain theoretical modeling, experimental measurement, and statistical uncertainties in the Monte-Carlo simulations.

Theoretical uncertainties

Theoretical uncertainties refer to the systematic uncertainties in modeling the signal and background expectations from a theoretical perspective. The main concerns of the theoretical uncertainties include the missing higher-order corrections, the choice of the PDF set, the estimation of the strong coupling constant and the parametrization in parton showers. The first three sources affect the computation of the matrix element. These uncertainties were estimated independently in different channels. The estimation strategy of the theoretical uncertainties contribution inherits the ATLAS Run 2 simplified template cross-section measurement [156].

The dominant theoretical uncertainties in this study are related to the jet number categorization and NN categorization in the ggF channel. The ggF Higgs boson production process includes the strong interaction in its heavy quark loop and the radiation of additional jets. Therefore the uncertainties in the QCD perturbative computation were considered in this study. In this study, the BSM samples with effective couplings were generated with the QCD accuracy at leading order only. The impact on the measurement from the missing higher-order QCD effect was evaluated by varying the factorization scale and renormalization scale from the half of Higgs boson mass to twice the Higgs boson mass [157]. The expected impact of this theoretical uncertainty on the limits of confidence intervals in the coupling measurement is about

5%. In the parton shower modeling, the uncertainty is estimations by comparing the result from PYTHIA 8 [131, 132] and HERWIG 7 [158]. The shower parameters were set with the AZNLO tune, which is a set of parameters used in the ATLAS experiment [143]. The impact of this theoretical uncertainty on the limits is $1-5\%$. For the other Higgs boson production channels including VBF, VH and ttH, the uncertainties in missing higher order QCD parametrization were estimated following the STXS measurement [156, 159]. Their impact is negligible in the coupling measurements.

The uncertainties on the PDF choice were estimated from the variations of the PDF set based on the PDF4LHC recommendations [124]. These PDFs were estimated from the fit to data. The uncertainties in the strong coupling constant, the functional form selected and the uncertainties in the dataset used in the estimation affect the PDF modeling. Each choice of variation was treated as an independent source of uncertainty in the model for each Higgs boson production channel. The impact of these uncertainties is negligible in the coupling measurement.

For the ZZ^* background samples, the effect of missing higher-order QCD in the migration between different regions was estimated by varying the factorization scale and the renormalization scale. The uncertainties in the PDF choice were estimated from the variations in the NNPDF3.0 PDF set [135]. The uncertainties in the parton shower modeling of the ZZ^* were estimated as well. The impact of all theoretical uncertainties in the ZZ^* background on the fit limits is within 2%.

The VBF production measurement is with four signal regions categorization based on the NN trained with the SM expectations. The dominant theoretical uncertainty is the migration effects in the ggF channel from the QCD scales. The migration effect in the ZZ^* backgrounds takes place in the production measurements as well. In the $H \to ZZ^* \to 4l$ decay measurement, the signal region for the Higgs boson events is inclusive without the migration issue. Therefore theoretical uncertainties take a minor role.

The effect of missing higher-order QCD in the VBF measurement has been tested through direct simulation at the generator level with another model named the Higgs characterisation model [160]. This model is not a dedicated EFT model but allows next-to-leading order computation. The test shows the impact of missing higher-order QCD on CP-odd coupling measurements in this analysis is small. The details are reported in Appendix G.

Experimental uncertainties

The dominant experimental uncertainties in the $H \to ZZ^* \to 4l$ measurement are related to the energy scale and resolution of leptons. The impact of these uncertainties on the fit limits of couplings is expected to be $1-2\%$. The uncertainties in the pile-up evaluation, reconstruction, identification and isolation of the leptons, track-to-vertex association and flavour-tagging

were estimated as well. The impact from these uncertainties is within 2%. The dominant experimental uncertainties in the VBF production measurement are related to the energy scale and resolution of jets. These uncertainties are expected to have $1-3\%$ impacts on the CP measurement.

The uncertainty in the ATLAS Run 2 luminosity is 1.7% [97]. This uncertainty was not evaluated in this study since the yields in this study are determined by fitting to data. The luminosity fluctuations are absorbed in the normalization factors. In the reducible backgrounds estimation, the statistical uncertainty and systematic uncertainties in the fake estimations were included following the STXS study [156]. The effect on the coupling measurements is negligible.

Monte-Carlo uncertainties

In this shape-based analysis, the MC simulated events were split by the bins of optimal observables. The number of events in each bin is smaller than the total events generated. Therefore the statistical uncertainties from the MC simulation are expected to have a larger impact than the yield-based measurement. The impact of this uncertainty was estimated using a bootstrapping method [161]. In this approach, an "MC toy" was generated by randomly assigning an integer bootstrapping weight following the Poisson distribution with the mean value at 1 to each event in the MC samples. This MC toy is considered statistically independent from the original MC dataset. An ensemble of such MC toys was generated and fitted with a pseudo-dataset independently. The multiple likelihood fit curves generated form a band of likelihood ratio variations.

The root-mean-squared values of these variations on the best-fit value and limits from these MC toys fit are the MC statistical uncertainties. In this estimation, the MC statistical uncertainties were estimated with different CP-odd coupling measurements with different OO. The largest impacts among these measurements from the root-mean-squared variation of 100 MC toys fit are shown in Table 6.4. The impact from the MC statistical uncertainties is negligible compared to the statistical uncertainties in the Run 2 data. The impact is of a comparable order with the experimental and theoretical uncertainties.

Table 6.4: The maximum impact of MC statistical uncertainties on the likelihood fit among the CP-odd coupling measurements with different OO. The impact in each case was estimated using the root-mean-square variations of 100 bootstrapping MC toys.

Channel	Best fit Value	$+1\sigma$ limit	-1σ limit	$+2\sigma$ limit	-2σ limit
VBF production	0.004	2.5%	1.8%	2.2%	2.0%
$H \to ZZ^* \to 4l$ decay	0.003	2.4%	2.9%	1.8%	2.5%

Chapter 7

Results and discussion

7.1 Observable distributions in data

The event yields of the ATLAS Run 2 data and the SM expectation in the signal and control regions are presented in Table 7.1. For most of the categories, the observed yield is consistent with the SM expectation within the counting error. In the VBF SR4, there is an upward fluctuation of data compared to the SM expectation. The four VBF signal regions partitioned by the NN scores all contain small numbers of events. This leads to large statistical uncertainties in data. In the control region for the ZZ^*-dominant irreducible background, the data also have a slight upward fluctuation from the SM expectation.

Production OO distributions

The mean values of the production optimal observables in data are shown in Table 7.2. The mean values are consistent with the SM expectation. The values are close to zero within the statistical uncertainties. The large mean value and statistical uncertainties are due to the limited number of events. The high NN categories are highly VBF enriched and with lower background contributions. The mean values for high NN categories are more closed to zero. The $\widetilde{d}$ coupling is proportional to the $c_{H\widetilde{W}}$ coupling based on its definition. The equal value constrained $c_{H\widetilde{B}}$ is not sensitive to the VBF measurement. The $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4jj}^{\widetilde{d}}$ values of an event are almost the same. The consistency in mean values for $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4jj}^{\widetilde{d}}$ agrees with the expectation.

The distributions of $OO_{jj}^{\widetilde{c}_{zz}}$ in each VBF-enriched category are shown in Figure 7.1 [19]. The number of events in each bin is consistent with the SM expectation within a range of two standard deviations. There is an excess in the VBF SR1 for the last bin. In each category, there are some empty bins and fluctuations in the distribution, indicating the statistical limits. The distributions of OO corresponding to other sensitive couplings $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$, $\widetilde{d}$ have similar features.

Table 7.1: Unblinding data yield counting and SM expectation yield values for signal and backgrounds in signal regions and control regions

Category	data	total SM	ggF	bbH	ttH/tH	VBF/VH	ZZ^*	ttV/VVV	reducible
VBF SR1	14	15^{+3}_{-3}	$7.5^{+3.0}_{-2.4}$	$0.1^{+0.0}_{-0.0}$	$1.2^{+0.0}_{-0.0}$	$1.60^{+0.09}_{-0.08}$	$3.5^{+1.3}_{-1.7}$	$0.53^{+0.04}_{-0.04}$	$0.8^{+0.9}_{-0.9}$
VBF SR2	9	$9.9^{+2.0}_{-1.7}$	$5.6^{+1.8}_{-1.5}$	$0.1^{+0.0}_{-0.0}$	$0.2^{+0.0}_{-0.0}$	$1.89^{+0.11}_{-0.11}$	$1.7^{+0.6}_{-0.6}$	$0.150^{+0.020}_{-0.020}$	$0.3^{+0.6}_{-0.6}$
VBF SR3	4	$5.9^{+0.7}_{-0.7}$	$2.2^{+0.6}_{-0.6}$	$0.0^{+0.0}_{-0.0}$	$0.1^{+0.0}_{-0.0}$	$3.01^{+0.18}_{-0.18}$	$0.48^{+0.16}_{-0.15}$	$0.0^{+0.0}_{-0.0}$	$0.07^{+0.26}_{-0.26}$
VBF SR4	8	$4.1^{+0.5}_{-0.5}$	$0.49^{+0.17}_{-0.17}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$	$3.5^{+0.4}_{-0.4}$	$0.086^{+0.025}_{-0.028}$	$0.0^{+0.0}_{-0.0}$	$0.01^{+0.09}_{-0.09}$
inclusive SR	311	321^{+14}_{-15}	181^{+12}_{-12}	$1.853^{+0.021}_{-0.021}$	$1.89^{+0.01}_{-0.01}$	$23.8^{+0.7}_{-0.7}$	98^{+6}_{-9}	$1.31^{+0.06}_{-0.06}$	13^{+4}_{-4}
VBF-depleted SR	276	286^{+14}_{-15}	165^{+12}_{-12}	$1.648^{+0.021}_{-0.021}$	$0.4^{+0.0}_{-0.0}$	$13.8^{+0.6}_{-0.6}$	92^{+6}_{-8}	$0.62^{+0.04}_{-0.04}$	12^{+3}_{-3}
ZZ^* CR	294	267^{+21}_{-26}	$8.2^{+1.3}_{-1.3}$	$0.1^{+0.1}_{-0.1}$	$0.2^{+0.0}_{-0.0}$	$1.39^{+0.16}_{-0.16}$	229^{+20}_{-25}	$6.79^{+0.13}_{-0.13}$	21^{+5}_{-5}

Table 7.2: OO_{jj} mean values for the ATLAS Run 2 data in VBF signal regions

Category	$OO_{jj}^{\widetilde{c}_{zz}}$	$OO_{jj}^{\widetilde{c}_{za}}$	$OO_{jj}^{c_{H\widetilde{W}}}$	$OO_{4jj}^{\widetilde{d}}$
VBF SR1	0.4 ± 0.7	0.15 ± 0.29	0.2 ± 0.4	0.2 ± 0.4
VBF SR2	0.4 ± 0.4	0.25 ± 0.20	0.28 ± 0.24	0.28 ± 0.24
VBF SR3	0.06 ± 0.27	-0.00 ± 0.14	0.02 ± 0.16	0.02 ± 0.06
VBF SR4	0.00 ± 0.16	0.00 ± 0.08	0.00 ± 0.10	0.00 ± 0.10

The $OO_{jj}^{\widetilde{c}_{za}}$, $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4jj}^{\widetilde{d}}$ distributions in the VBF SRs are presented in Appendix H.

Decay OO distributions

In the $H \to ZZ^* \to 4l$ decay side measurement, the decay optimal observables are used in the study. The decay measurement uses the whole Higgs boson production events as the signal in the Higgs boson mass window defined in Section 5.3. The decay optimal observables are used in the production-decay combined measurement as well. In this scenario, the production optimal observables are used for events in the VBF-enriched regions. The decay optimal observables are used for non-VBF-like events in the VBF-depleted region. The mean values of the decay optimal observables of data are shown in Table 7.3. In both regions, the mean values are close to zero within the statistical uncertainties. There is a good agreement between the data and the SM expectation.

The distributions of $OO_{4\ell}^{c_{H\widetilde{W}}}$ in the inclusive signal region are shown in Figure 7.2 as an example [19]. In these distribution plots, the data is consistent with the SM expectation within two standard deviations. The statistical fluctuations in data are large in many bins. The $OO_{4\ell}$ plots for $\widetilde{c}_{z\gamma}$, $\widetilde{c}_{\gamma\gamma}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ and $\widetilde{d}$ are presented in Appendix H.

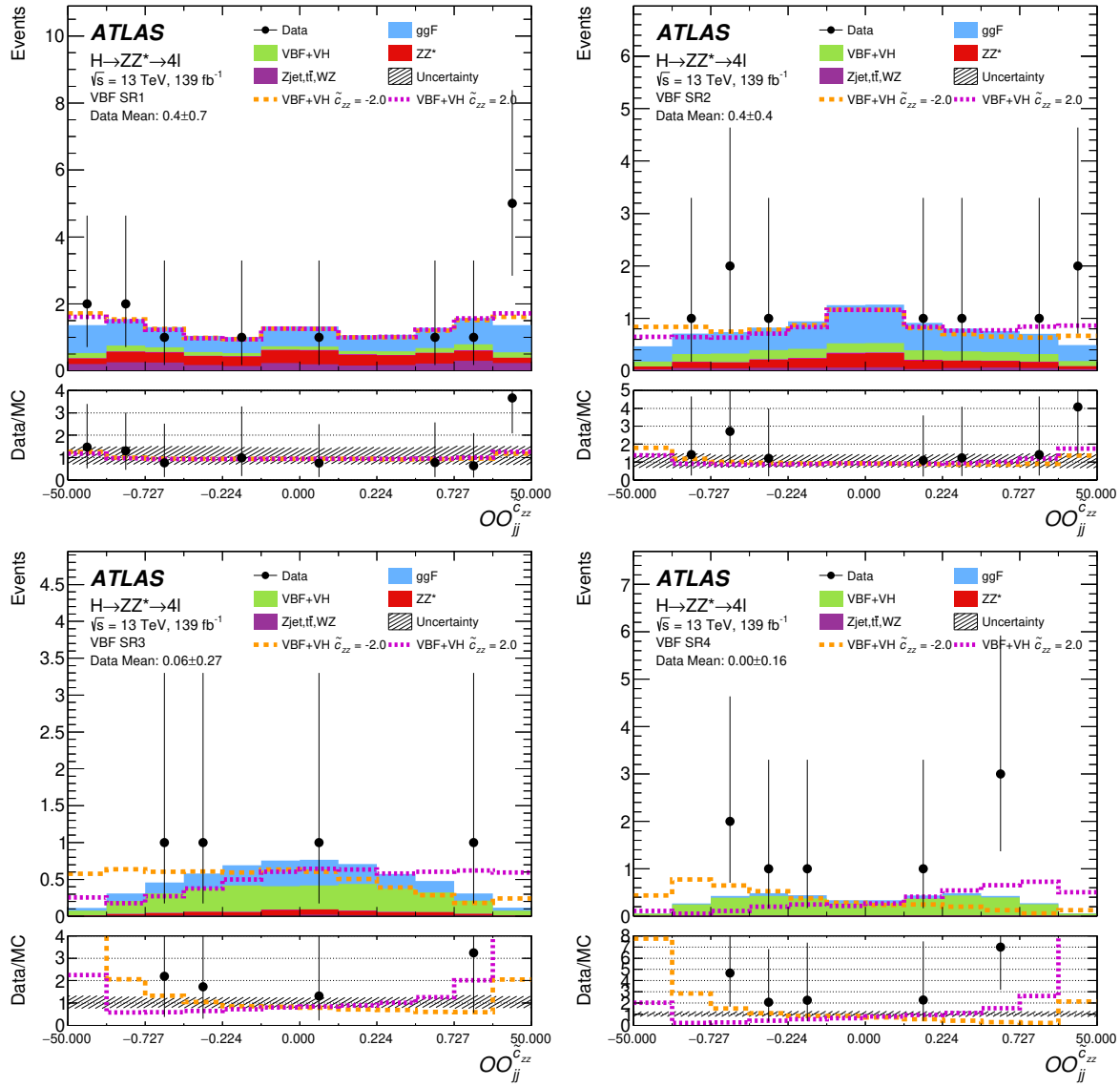


Figure 7.1: $OO_{jj}^{\tilde{c}_{zz}}$ distributions for the VBF-like NN signal regions of production side measurement. The dot represents unblinding data with statistical uncertainties. The green block represents the VBF/VH signal MC distributions. The other stacks represent different MC background components. The dashed area is the total uncertainties for the MC simulation.

Table 7.3: $OO_{4\ell}$ mean values for the ATLAS Run 2 data in inclusive and VBF-depleted signal regions

Category	$OO_{4\ell}^{\tilde{c}_{za}}$	$OO_{4\ell}^{\tilde{c}_{aa}}$	$OO_{4\ell}^{c_{H\tilde{W}}}$	$OO_{4\ell}^{c_{H\tilde{B}}}$	$OO_{4\ell}^{c_{H\tilde{W}B}}$	$OO_{4\ell}^{\tilde{d}}$
inclusive SR	0.001 ± 0.011	-0.03 ± 0.05	-0.017 ± 0.030	0.06 ± 0.11	0.03 ± 0.06	-0.08 ± 0.14
VBF-depleted SR	0.003 ± 0.012	-0.03 ± 0.06	-0.018 ± 0.034	-0.07 ± 0.12	0.03 ± 0.06	-0.09 ± 0.16

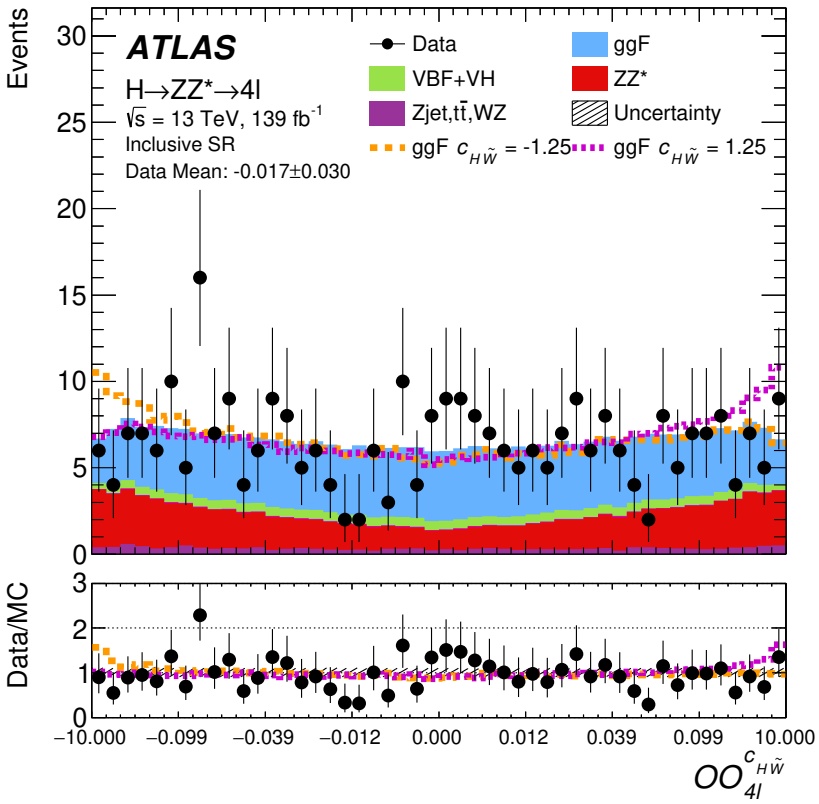


Figure 7.2: $OO_{4\ell}$ distributions for the inclusive signal regions for decay side measurements. The dots represent unblinding data with statistical uncertainties. The light blue and green blocks represent the dominant Higgs boson production channels ggF and VBF/VH MC distributions. The red block represents the dominant backgrounds ZZ^* MC distributions.

7.2 Likelihood fit results

One-dimensional coupling measurement

In the one-dimensional coupling measurement of the Run 2 data, each coupling is measured through likelihood fit assuming no presence of other couplings. The couplings measured include $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$ and $c_{H\widetilde{W}B}$ in the Warsaw basis and $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$ and $\widetilde{c}_{\gamma\gamma}$ in the Higgs basis. An additional parameter $\widetilde{d}$ is measured for comparison to the studies in other channels as well. The couplings that are sensitive to the di-jet optimal observables are measured in the VBF production-only analysis. The couplings that are sensitive to the four-lepton optimal observables are measured in the $H \rightarrow ZZ^* \rightarrow 4l$ decay-only analysis. The couplings with sensitivities on both sides are measured in the production-decay combined analysis. For each coupling, the measurement with the best significance among these schemes is presented as the result.

In VBF production-only analysis, the couplings $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$ measurements show good sensitivities to the shape-only fit on the Run 2 data. The likelihood scan curves for the

$\widetilde{c}_{zz}$ within $[-3, 3]$ range on the Run 2 data and SM simulation are shown as an example in Figure 7.3 [19]. The best-fit value in the measurement has a shift to the right. This is likely due to the upward fluctuation in the right boundary bin of the VBF SR1 histogram in Figure 7.1. This is the best result ever achieved on the $\widetilde{c}_{zz}$ measurement in the $H \to ZZ^* \to 4l$ channel. The measurements on the other couplings have similar behaviors. The likelihood scan curves of the other couplings are presented in Appendix I.

In the $H \to ZZ^* \to 4l$ decay-only analysis, the couplings $\widetilde{c}_{z\gamma}$, $\widetilde{c}_{\gamma\gamma}$, $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ and $\widetilde{d}$ measurements show good sensitivities. The likelihood scan curves for the $c_{H\widetilde{W}}$ within $[-2, 2]$ range on the Run 2 data and SM simulation are shown as an example in Figure 7.4 [19]. The fit results on the Run 2 data are in good agreement with the SM expectation. The data fit has narrower limits compared to the SM expectation due to the statistical fluctuations. The measurements on the other couplings have results consistent with the SM expectations as well. The likelihood scan curves of the other couplings are presented in Appendix I.

The production and decay combined measurements have been done with the couplings $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$. These are sensitive to both VBF production measurement and $H \to ZZ^* \to 4l$ decay analyses. The likelihood scan curves for the $c_{H\widetilde{W}}$ within $[-2, 2]$ range on the Run 2 data and SM simulation are shown as an example in Figure 7.5 [19]. The data fit results are consistent with the SM expectation. The best-fit value in the observation has a shift to the right similar to the VBF production measurement. This is due to the usage of the OO_{jj} in the VBF signal regions same as the production-only measurement. The shift is not as large as the production measurement with the dilution of the non-biased VBF-depleted decay information.

In the measurement with data, the selection of the measurement channels is based on their performance with the MC samples of the SM expectation. The SM expected limits are summarized in Table 7.4 [19]. The combined measurement in the $c_{H\widetilde{W}}$ coupling is expected to have the advantage over the single vertex measurement. The $\widetilde{c}_{zz}$ coupling is sensitive to the production-only measurement. The other couplings $\widetilde{c}_{z\gamma}$, $\widetilde{c}_{\gamma\gamma}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ and $\widetilde{d}$ are most sensitive to the decay-only measurement. The measurement results on the corresponding vertex for the measurement in the ATLAS Run 2 data are summarized in Table 7.5 [19]. These results are consistent with the SM expectation.

Two-dimensional coupling measurement

In this analysis, the constraints of the BSM CP-odd couplings with the presence of other couplings are studied through two-dimensional likelihood fits. In the decay-only measurement, each pair of couplings among $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$ and $c_{H\widetilde{W}B}$ is allowed to float simultaneously, while the third coupling is fixed to 0. The two-dimensional likelihood scan contours for the three pairs of couplings on the Run 2 data and SM simulation are shown in Figure 7.6 [19]. All of these mea-

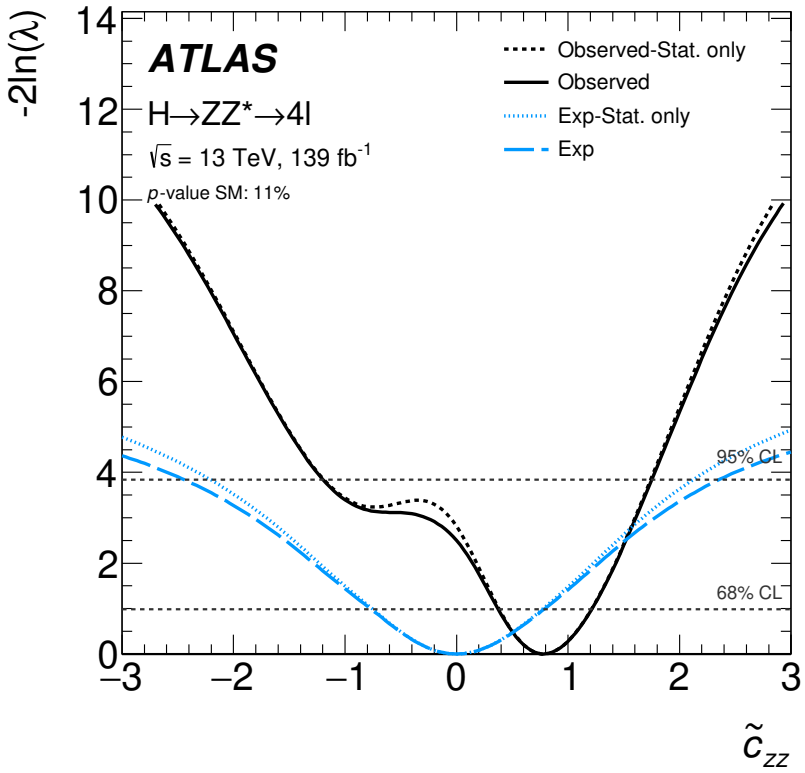


Figure 7.3: Likelihood scans along $\widetilde{c}_{zz}$ axis on the SM expectation and ATLAS Run 2 data with $OO_{jj}^{\widetilde{c}_{zz}}$. The black curves represent measures on ATLAS Run 2 data. The blue curves represent measures on the SM expectation. The solid curves include full systematical uncertainties in the scans. The dashed curves are the statistical uncertainties-only scans.

Table 7.4: The SM expected limits of 95% confident level in the VBF production-only, $H \to ZZ^* \to 4l$ decay-only and production and decay combined likelihood scan for BSM couplings.

	68% production-only	decay-only	combined
$c_{H\widetilde{B}}$	-	± 0.38	-
$c_{H\widetilde{W}B}$	-	± 0.75	-
$c_{H\widetilde{W}}$	± 4.8	± 1.4	± 1.3
$\widetilde{d}$	± 0.63	± 0.018	± 0.019
$\widetilde{c}_{\gamma\gamma}$	± 2.4	-	-
$\widetilde{c}_{z\gamma}$	± 6.6	± 0.76	± 0.80
$\widetilde{c}_{zz}$	-	± 0.80	-

surements utilize the two-dimensional observable $OO_{4\ell}^{c_{H\widetilde{B}}}$ vs $OO_{4\ell}^{c_{H\widetilde{W}}}$, which is built from the direct product of $OO_{4\ell}^{c_{H\widetilde{B}}}$ and $OO_{4\ell}^{c_{H\widetilde{W}}}$. The performance of this observable is the best among all the OO tested on the SM expectation. The appearance of the second coupling in the fit

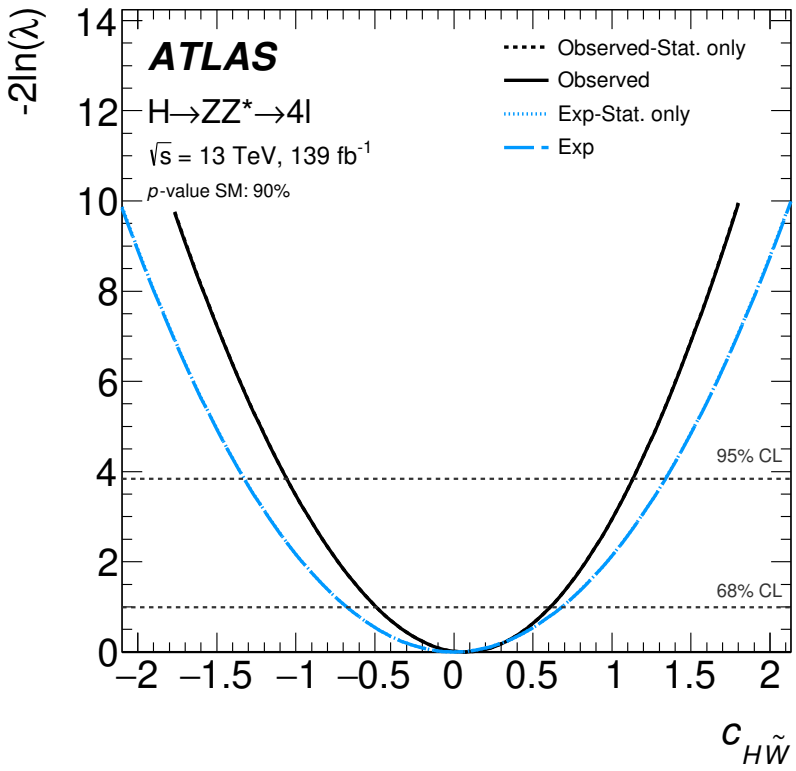


Figure 7.4: Likelihood scan along $c_{H\widetilde{W}}$ axis on the SM expectation and ATLAS Run 2 data with $OO_{4\ell}^{c_{H\widetilde{W}}}$. The black curves represent measures on ATLAS Run 2 data. The blue curves represent measures on the SM expectation. The solid curves include full systematical uncertainties in the scan. The dashed curves are the statistical uncertainties-only scan.

Table 7.5: The SM expected and Run 2 data observed results of 68% and 95% confident level limits and best-fit values for BSM couplings. The measured vertex of the best-expected scheme was selected.

	expected		observed		Best-fit	Fit type
	68% CL	95% CL	68% CL	95% CL	value	
$c_{H\widetilde{B}}$	$[-0.18, 0.19]$	$[-0.37, 0.37]$	$[-0.42, 0.31]$	$[-0.61, 0.54]$	-0.08	decay
$c_{H\widetilde{W}B}$	$[-0.36, 0.36]$	$[-0.72, 0.72]$	$[-0.56, 0.53]$	$[-0.97, 0.98]$	-0.02	decay
$c_{H\widetilde{W}}$	$[-0.63, 0.63]$	$[-1.26, 1.28]$	$[-0.07, 1.09]$	$[-0.81, 1.54]$	0.60	combined
$\widetilde{d}$	$[-0.009, 0.009]$	$[-0.018, 0.018]$	$[-0.017, 0.014]$	$[-0.026, 0.025]$	-0.003	decay
$\widetilde{c}_{zz}$	$[-0.77, 0.79]$	$[-2.44, 2.36]$	$[0.37, 1.21]$	$[-1.20, 1.75]$	0.78	production
$\widetilde{c}_{z\gamma}$	$[-0.47, 0.47]$	$[-0.76, 0.76]$	$[-0.54, 0.54]$	$[-0.84, 0.83]$	0.08	decay
$\widetilde{c}_{\gamma\gamma}$	$[-0.38, 0.38]$	$[-0.76, 0.77]$	$[-0.52, 0.48]$	$[-0.99, 0.93]$	-0.01	decay

makes the constraint looser in the first coupling. The observed results in these two-dimensional measurements are in good agreement with the SM expectations.

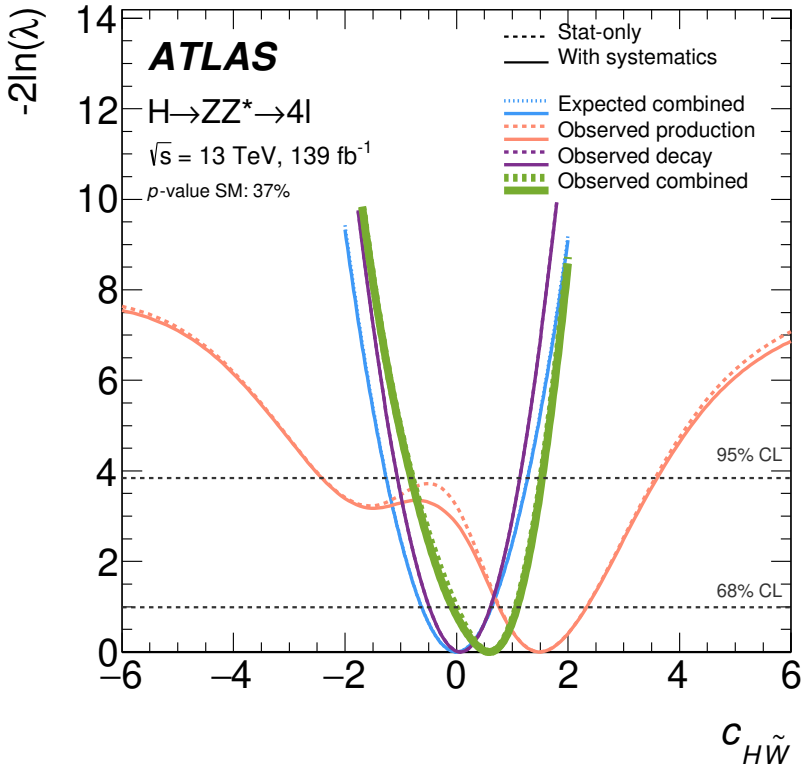


Figure 7.5: Likelihood scan along $c_{H\widetilde{W}}$ axis on the SM expectation and ATLAS Run 2 data. The blue curves represent measures on the SM expectation simultaneously with $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4\ell}^{c_{H\widetilde{W}}}$ in VBF and VBF-depleted signal regions respectively. The orange, violet and green curves represent measures on the ATLAS Run 2 data with only $OO_{jj}^{c_{H\widetilde{W}}}$, only $OO_{4\ell}^{c_{H\widetilde{W}}}$ and simultaneously with $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4\ell}^{c_{H\widetilde{W}}}$ respectively. The solid curves include full systematical uncertainties in the scan. The dashed curves are the statistical uncertainties-only scan.

In addition, the information from both the production and decay vertices was combined in a two-dimensional measurement. The couplings $\widetilde{c}_{zz}$ and $\widetilde{c}_{\gamma\gamma}$ are measured simultaneously without the presence of other BSM couplings in this measurement. The $\widetilde{c}_{zz}$ coupling is only sensitive to the production vertex measurement. The $\widetilde{c}_{\gamma\gamma}$ is only sensitive to the decay vertex measurement. The two-dimensional likelihood scan contours for $\widetilde{c}_{zz}$ and $\widetilde{c}_{\gamma\gamma}$ simultaneously on the Run 2 data and SM simulation are shown in Figure 7.7 [19]. In this measurement, the production observable $OO_{jj}^{\widetilde{c}_{zz}}$ is used for the VBF signal regions and the decay observable $OO_{4\ell}^{\widetilde{c}_{aa}}$ is used for the VBF-depleted region. The constraints on the two couplings are loosened slightly compared to the one-dimensional scan. The result obtained is consistent with the SM expectation.

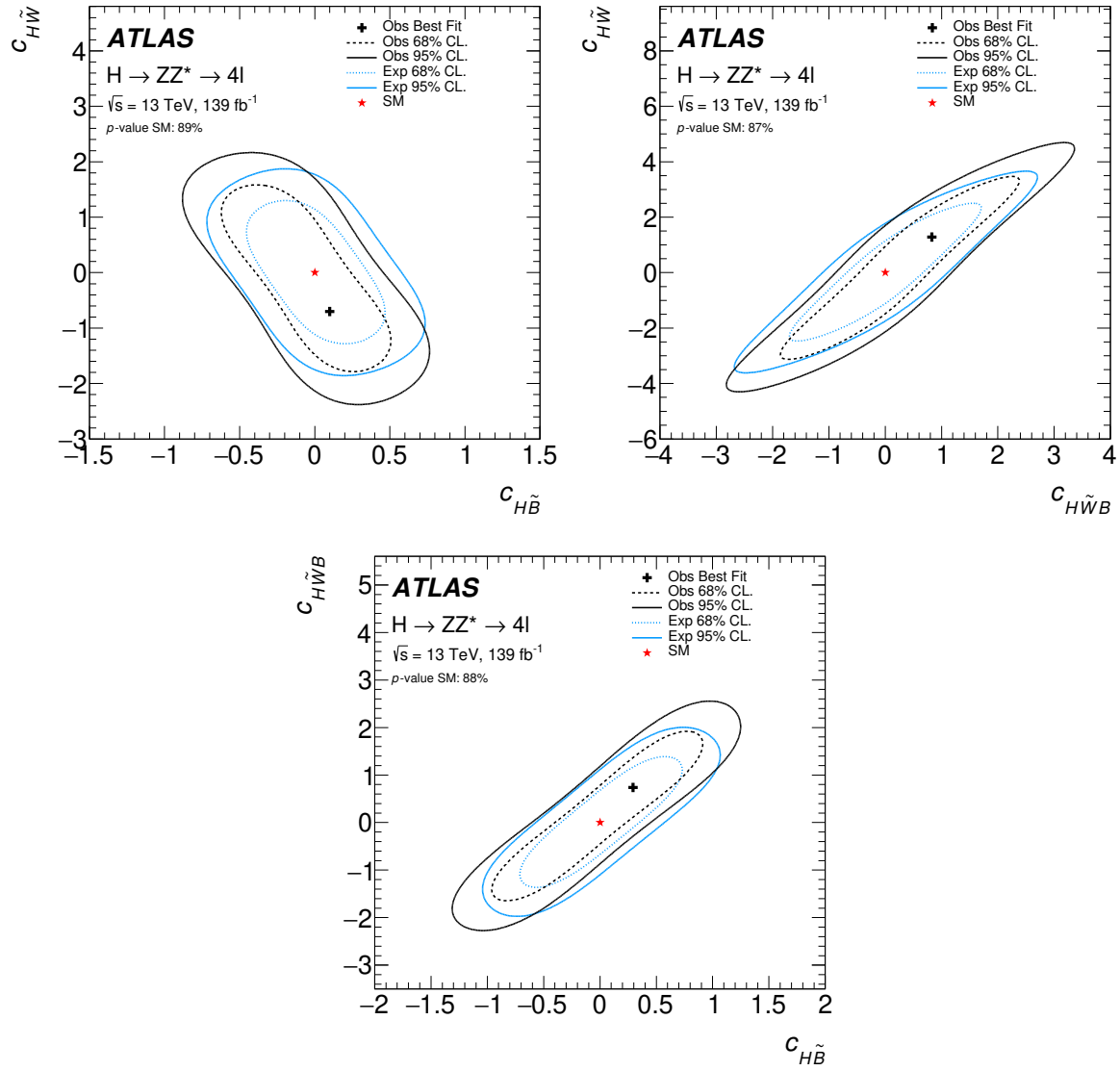


Figure 7.6: Simultaneous likelihood fit along combinations of $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$ and $c_{H\widetilde{W}B}$ axes on the SM expectation and ATLAS Run 2 data. The black contours represent measures on Run 2 data. The blue contours represent measures on the SM expectation. The solid contours are results at 95% confidence level. The dashed contours are results at 68% confidence level. All the curves in the plot include full systematical uncertainties in the fit.

7.3 Comparison with previous results and parallel studies

Comparison with the simplified template cross-section measurement

In a previous ATLAS study on the 139 fb^{-1} Run 2 $H \to ZZ^* \to 4l$ data, cross-sections and EFT couplings have been measured with the simplified template cross-section (STXS) framework [9]. In that measurement, the EFT coupling interpretation was determined using cross-

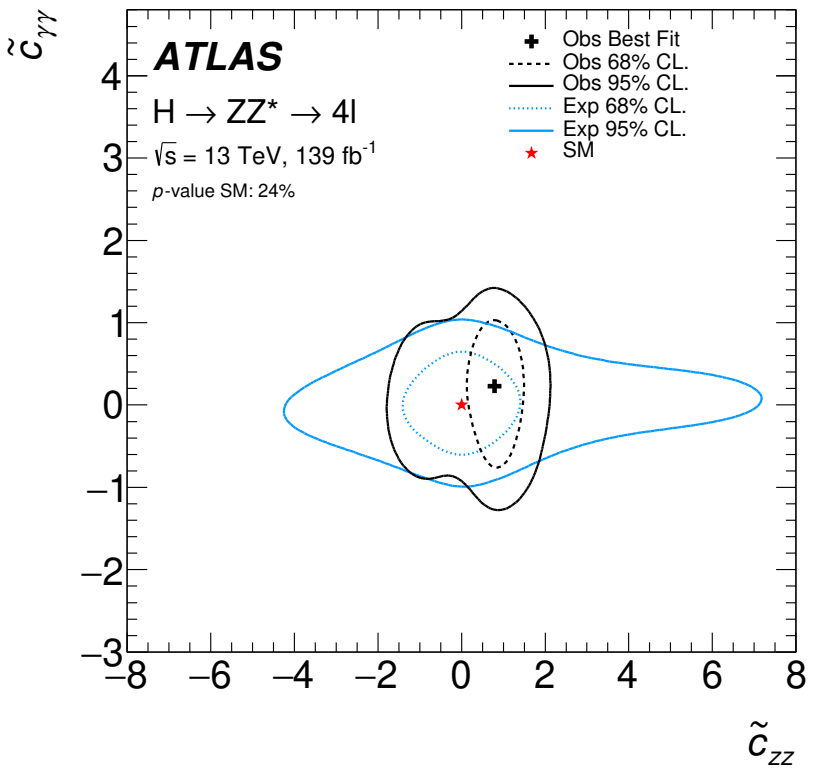


Figure 7.7: Simultaneous likelihood fit along $\widetilde{c}_{zz}$ and $\widetilde{c}_{\gamma\gamma}$ axes on the SM expectation and ATLAS Run 2 data. The black contours represent measures on Run 2 data. The blue contours represent measures on the SM expectation. The solid contours are results at 95% confidence level. The dashed contours are results at 68% confidence level. All the curves in the plot include full systematical uncertainties in the fit.

section information only. The results obtained are in good agreement with the Standard Model expectation.

In comparison, this study is optimized for measuring CP properties in the Higgs boson couplings to Z bosons. This study relies on the shape information with the optimal observables. It is expected that this study would improve upon the previous STXS study. The comparison of the SM expected and data observed 95% confidence level results in the previous STXS study and this study is shown in Table 7.6. The comparison shows an improvement from the shape information with optimal observables in the $c_{H\widetilde{W}}$ and $c_{H\widetilde{B}}$ measurements.

Comparison with the $H \to \tau\tau$ and $H \to \gamma\gamma$ channels measurement

In the previous study of the CP properties in the VBF production channel through different decay channels, the parameter $\widetilde{d}$ was raised as a single parameter to measure the strength of

Table 7.6: The 95% confidence level limits of the couplings on the Standard Model expectation and the data observation in the STXS measurement and this study

couplings	STXS study		this study	
	SM expected	Observed	SM expected	Observed
$c_{H\widetilde{W}}$	$[-2.1, 2.1]$	$[-2.4, 2.4]$	$[-1.3, 1.4]$	$[-1.1, 1.1]$
$c_{H\widetilde{B}}$	$[-0.57, 0.57]$	$[-0.56, 0.56]$	$[-0.37, 0.38]$	$[-0.54, 0.61]$
$c_{H\widetilde{W}B}$	$[-1.05, 1.05]$	$[-1.0, 1.0]$	$[-0.73, 0.73]$	$[-0.96, 1.0]$

CP violation [162]. This parameter is measurable in other Higgs boson decay channel analyses using the EFT framework. This enables studies in different channels to compare and combine their result through it. In the ATLAS Run 2 analyses, this parameter was measured in the VBF $H \to \tau\tau$ channel on 36.1 fb^{-1} and the VBF $H \to \gamma\gamma$ channel on 139 fb^{-1} data [12, 76]. The $H \to \gamma\gamma$ analysis measured the coupling $c_{H\widetilde{W}}$ as well. The results of these studies in the SM expected and data observed 95% confidence level limits and the comparison with this analysis are shown in Table 7.7. The $c_{H\widetilde{W}}$ measurement in the $H \to \gamma\gamma$ channel is more sensitive than in this study. The first observation is the larger VBF event numbers available in the $H \to \gamma\gamma$ channel compared to this analysis. This may explain the difference in sensitivity between these two studies. Further investigations are needed to provide more concrete explanations. On the other side, this study has a higher sensitivity in the $\widetilde{d}$ measurement. This might be explained by the insensitivity of $c_{H\widetilde{B}}$ of the VBF channel measurement. In this analysis, the $\widetilde{d}$ sensitive comes from the decay measurement. These previous studies did not have the Higgs boson to $W^{\pm}$ and Z bosons coupling in the decay processes. Combining the results of this study with those from other channel measurements in the future may yield a more significant result.

Comparison with the CMS measurements

The CMS is the other general-purpose detector at the LHC [2]. The ATLAS and CMS experiments have different detector designs and comparable beam conditions during the LHC operation. Analyses in these two experiments serve as cross-validations of each other. Therefore, the results obtained from the CMS experiments are crucial to this study.

Recently, an anomalous couplings study was conducted in $H \to ZZ^* \to 4l$ decay channel on the 137 fb^{-1} CMS Run 2 data [163]. The measurement was interpreted using the SMEFT framework as well. The SMEFT couplings were constrained on the Warsaw Basis. The comparison of the SM expected 68% confidence level results in the CMS measurement and this study in the Warsaw Basis is listed in Table 7.8. The results obtained in this study appear to be more sensitive than those obtained in the CMS measurement. Further investigation is needed to

Table 7.7: The 95% confidence level limits of the couplings on the Standard Model expectation and the data observation in the $H \to \gamma\gamma$ measurement and this study. The "$\widetilde{d}$ combined" row shows the $\widetilde{d}$ limits with the 36.1 fb^{-1} $H \to \tau\tau$ measurement result combined.

couplings	$H \to \gamma\gamma$ study		this study	
	SM expected	Observed	SM expected	Observed
$c_{H\widetilde{W}}$	$[-0.95, 0.95]$	$[-0.55, 1.07]$	$[-1.3, 1.4]$	$[-1.1, 1.1]$
$\widetilde{d}$	$[-0.061, 0.060]$	$[-0.034, 0.071]$	$[-0.018, 0.018]$	$[-0.023, 0.028]$
$\widetilde{d}$ combined	$[-0.046, 0.045]$	$[-0.034, 0.057]$	–	–

Table 7.8: The 68% confidence level limits on the Standard Model expected and data observed 95% confidence level limits of the CMS Run 2 $H \to ZZ^* \to 4l$ measurement and this study.

couplings	the CMS measurement		this study	
	SM expected	Observed	SM expected	Observed
$c_{H\widetilde{W}}$	$[-1.11, 1.11]$	$[-0.75, 0.28]$	$[-0.68, 0.68]$	$[-0.56, 0.54]$
$c_{H\widetilde{B}}$	$[-0.33, 0.33]$	$[-0.22, 0.09]$	$[-0.19, 0.19]$	$[-0.34, 0.39]$
$c_{H\widetilde{W}B}$	$[-1.21, 1.21]$	$[-0.82, 0.31]$	$[-0.37, 0.37]$	$[-0.54, 0.55]$

understand the difference in the sensitivity of these studies.

Chapter 8

Future directions of work

In this analysis, the CP conservation has been tested using the observable shape information in the VBF and $H \rightarrow ZZ^* \rightarrow 4l$ channel. The informativity of the OO distribution shape is seen in the more restricted constraints compared to the cross-section-based measurement. On the other side, the sensitivity of this measurement is limited by the statistics in the data. The statistical uncertainties become larger compared to the cross-section-based analyses on the same dataset since the events are split into different bins. In the VBF measurement, the VBF-like selection and the categorization with the NN score further restrict the statistics. For the 139 fb^{-1} ATLAS Run 2 data, most of the bins in the VBF SR in this study have less than four events. Some of the bins are empty. The event numbers increase in the future will help the shape-based CP measurement achieve more stable and significant results.

Study on the future Run at the LHC and High Luminosity LHC

The currently undergoing operation of the LHC is Run 3. This number is moderately larger than the Run 2 luminosity. More strict constraints on the CP-odd couplings are expected to be set by combining the future Run 3 analysis. The Run 3 shape-based analysis itself is very likely to suffer the low statistics condition. In the LHC future upgrade plan, there will be a major upgrade during the long shutdown period after Run 3. The upgraded machine is named high luminosity LHC (HL-LHC). The total integrated luminosity of the HL-LHC is expected to be 3000 fb^{-1} [164].

There is a prospective study on the CP measurement using the $H \rightarrow ZZ^* \rightarrow 4l$ channel [165]. This prospective study did not parametrize the couplings in the EFT framework. The parametrization on the scattering amplitude of $H \rightarrow ZZ^*$ processes used in this study is:

$$A(H \rightarrow VV) \propto g_1 m_V^2 \epsilon_1^{*\mu} \epsilon_2^{*\nu} + g_2 f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + g_3 f^{*(1),\mu\nu} f_{\mu\alpha}^{*(2)} \frac{q_\nu q^\alpha}{\Lambda^2} + g_4 f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu}, \quad (8.0.1)$$

Table 8.1: The SM expected upper limits of 95% confident level in the yield-based fit of the CP-odd contribution in the future 300 fb^{-1} and 3000 fb^{-1} HL-LHC simulation.

Luminosity	f_{g_4}
300 fb^{-1}	0.15
3000 fb^{-1}	0.037

where $A(H \rightarrow VV)$ is the amplitude for the Higgs boson decay into vector bosons. The symbols g_1 to g_4 are the couplings characterizing different contributions in the amplitude. The symbol m_V is the mass, q and ϵ are the four-momentum and polarization vector, and $f_{\mu\nu}$ is the field strength tensor of the vector bosons. The symbol $\tilde{f}$ denotes the conjugate field strength tensor, which makes corresponding terms CP-odd [166]. In this decomposition, the g_1 term characterizes the tree-level SM contribution. The g_2 and g_3 terms represent the loop-induced CP-even contributions. The g_4 term represents the CP-odd contribution. The tree-level SM amplitude has $g_1 = 1$ and $g_2 = g_3 = g_4 = 0$. The fractional contributions are defined as:

$$f_{g_i} = \frac{|g_i|^2\sigma_i}{|g_1|^2\sigma_1 + |g_2|^2\sigma_2 + |g_4|^2\sigma_4}, \tag{8.0.2}$$

is the relative contribution of each type of amplitude component in the total cross-section. The σ_i symbol denotes the cross-section given only the $g_i = 1$ and others zero. The g_3 term is assumed very small and ignored in this study.

The fit with matrix element-based observables on 300 fb^{-1} and 3000 fb^{-1} MC dataset simulated for HL-LHC has an expected 95% confidence level exclusion limit for f_{g_4}, as shown in Table 8.1. The exclusion power for the 3000 fb^{-1} measurement is expected to be much stronger than the 300 fb^{-1} measurement. This prospective study illustrates the future potential of the CP-odd couplings measurement at the HL-LHC with the future dataset.

Studies on the future Circular Electron Positron Collider

There are other future machines planned with higher luminosity and enhanced precision measurement ability. The Circular Electron Positron Collider (CEPC) planned in China is a future electron and positron collider with one of its goals to serve as a high luminosity Higgs boson factory [167, 168]. As the name indicates, collision events in the CEPC are produced by the crossing of a bunch of electrons and a bunch of positrons. Compared to the LHC as a hadron collider, the CEPC is not affected by pile-up effects. On the other side, the more severe synchrotron radiation of low-mass leptons limits the bunch spacing at the Higgs factory due to the power limitation. The Higgs boson production at the CEPC will be dominated by the $e^+e^- \rightarrow ZH$

Table 8.2: The SM expected $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ ranges at 68% confident level in the yield-based fit for 3000 fb^{-1} HL-LHC simulation and 5600 fb^{-1} and 20000 fb^{-1} CEPC simulations.

Luminosity	pp 3000 fb^{-1}	e^+e^- 5600 fb^{-1}	e^+e^- 20000 fb^{-1}
$\widetilde{c}_{z\gamma}$	$[-0.22, 0.22]$	$[-0.30, 0.27]$	$[-0.16, 0.14]$
$\widetilde{c}_{zz}$	$[-0.33, 0.33]$	$[-0.06, 0.06]$	$[-0.03, 0.03]$

process with electron and positron beam energy at 120 GeV. This channel includes the decay of a virtual Z boson into a Higgs boson and a Z boson with the full kinematic information of initial state particles. Therefore, these events will be suitable for the HZZ coupling measurement. The high precision in these couplings measurement is expected from the abundant events.

There is a prospective study on the CP-odd EFT coupling measurement at the CEPC [169]. The performance of yield-based $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ measurement has been tested with 5600 fb^{-1} integrated luminosity simulated dataset for about ten years of the CEPC operation and some prospective larger luminosity 20000 fb^{-1}. From the simulation in this paper, the optimal observables corresponding to $\widetilde{c}_{zz}$ is more sensitive to the $\widetilde{c}_{zz}$ coupling than the case for $\widetilde{c}_{z\gamma}$. The expected results at 68% confidence level limit for $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ measurements and the comparison to the HL-LHC expectation are shown in Table 8.2. The performance of the $\widetilde{c}_{zz}$ measurement at the CEPC is expected to be increased significantly and $\widetilde{c}_{z\gamma}$ is at the same level of performance as the HL-LHC.

Summary

This study aims to test the potential CP invariance in the Higgs to Z bosons couplings on the 139 fb^{-1} of the ATLAS Run 2 data. The potential CP violation in this channel could provide hints toward new physics beyond the Standard Model. This study is carried out by measuring the BSM CP-odd Higgs boson to Z bosons couplings parametrized using the SMEFT model. This measurement succeeds the yield-based STXS measurement on the 139 fb^{-1} of ATLAS Run 2 data. The strategy of shape-only likelihood fit with the optimal observable distributions in this analysis has been optimized for the CP-odd measurement. This measurement focuses on the VBF production events and $H \to ZZ^* \to 4l$ decay events.

In the VBF production channel measurement, the VBF-like events selected within a mass window of $[115, 130]$ GeV, with at least two jets and $m_{jj} > 120$ GeV are the signal events. The dominant backgrounds are the ggF production Higgs boson events and non-Higgs boson ZZ^* continuum events. In the $H \to ZZ^* \to 4l$ decay measurement, all the Higgs boson events within the $[115, 130]$ GeV mass windows are taken as the signal events. The ZZ^* continuum events dominate in the backgrounds. The $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$, $\widetilde{c}_{\gamma\gamma}$, $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ and $\widetilde{d}$ couplings are measured through production-only, decay-only and production and decay combined measurements based on the SM expected performance.

In these measurements, the shapes of the optimal observables in the model are predicted by the morphing technique. The shapes by this technique are interpolated from a basis of MC-simulated samples with optimized couplings. The yield of each component is constrained by the simultaneous fit in the background-dominated regions to the data. The neural network technique for the VBF and ggF discriminant is included in this study to enhance the VBF production measurement. The dedicated experimental and theoretical systematical uncertainties dominated by the energy scales of leptons and jets and QCD modeling are taken into consideration. The reducible backgrounds dominated by Z+jets and $t\bar{t}$ are estimated in the model.

In the VBF production likelihood fit in the $\widetilde{c}_{zz}$, an around 1σ shift to the right has been observed. This phenomenon is consistent with the SM predictions within the statistical uncertainties. This shift to the right is most likely due to the upward fluctuations of backgrounds in the right bins of the VBF SR1. In the $H \to ZZ^* \to 4l$ decay likelihood, no obvious shift

in the best-fit value. The results are in good agreement with the SM expectation. The boundary bins have a large contribution to the fit sensitivity. As a result, the couplings with upward fluctuations in the boundary bins of the corresponding observables get higher sensitivity compared to the SM expectation. The couplings with downward fluctuations in the backgrounds bins of the corresponding observables in contrast get lower sensitivity compared to the SM expectation. The production and decay combined measurement for $c_{H\widetilde{W}}$ exhibits characteristics of both production-only and decay-only fits. The sensitivity depends on the boundary bins of the decay observables and the shift is caused by the VBF SR1 fluctuation in the production observables. These measurements set more strict constraints on the CP-odd HZZ couplings compared to the previous STXS measurement based on the cross-section on the ATLAS Run 2 data. In this analysis, the parameter $\tilde{d}$ measurable in different channels has been measured as well. The result has an improved sensitivity compared to the $H \to \gamma\gamma$ measurement on the 139 fb^{-1} ATLAS Run 2 data combined with the $H \to \tau\tau$ measurement on the 36.1 fb^{-1} early ATLAS Run 2 data. This would possibly come from the inclusion of non-VBF channel events in the decay measurement in this study.

The expected VBF and $H \to ZZ^* \to 4l$ event numbers in the ATLAS Run 2 dataset are statistically limited for the shape-based CP measurement. Future studies on higher luminosity datasets would boost the measurement sensitivity and robustness. The future high luminosity machines including HL-LHC and CEPC shade lights on the future more sensitive CP measurement. In future analysis, there is also the possibility to combine the yield information and the shape information simultaneously to increase the sensitivity with the future development of the software.

Appendix A

Optimized variable binnings

The optimized binning scheme for the optimal observables in this study is to keep the same weight of events in each bin. The shape is influenced by the potential BSM couplings. To maintain the stability for BSM scenarios, the optimization for each OO variable includes a 50% contribution from the SM distribution expectation and a 50% contribution from the BSM distribution expectation.

In the VBF measurement, the expected total number of events in the VBF signal regions varies from 4 to 15. Each OO_{jj} is optimized into 12 bins of varied sizes. The detailed binning scheme is listed as follows:

$OO_{jj}^{\widetilde{c}_{zz}}$ { -50.0000, -1.6678, -0.7274, -0.3977, -0.2237, -0.1040, 0.0000, 0.1040, 0.2237, 0.3977, 0.7274, 1.6678, 50.0000 }

$OO_{jj}^{\widetilde{c}_{za}}$ { -30.0000, -0.7622, -0.3588, -0.2112, -0.1280, -0.0645, 0.0000, 0.0645, 0.1280, 0.2112, 0.3588, 0.7622, 30.0000 }

$OO_{jj}^{\widetilde{c}_{aa}}$ { -4.0000, -0.1353, -0.0619, -0.0351, -0.0224, -0.0121, 0.0000, 0.0121, 0.0224, 0.0351, 0.0619, 0.1353, 4.0000 }

$OO_{jj}^{c_{H\tilde{B}}}$ { -1.0000, -0.0140, -0.0055, -0.0027, -0.0013, -0.0005, 0.0000, 0.0005, 0.0013, 0.0027, 0.0055, 0.0140, 1.0000 }

$OO_{jj}^{c_{H\widetilde{W}B}}$ { -5.0000, -0.0603, -0.0237, -0.0143, -0.0082, -0.0034, 0.0000, 0.0034, 0.0082, 0.0143, 0.0237, 0.0603, 5.0000 }

$OO_{jj}^{c_{H\widetilde{W}}}$ { -30.0000, -1.0188, -0.4501, -0.2543, -0.1477, -0.0734, 0.0000, 0.0734, 0.1477, 0.2543, 0.4501, 1.0188, 30.0000 }

In the $H \to ZZ^* \to 4l$ decay measurement, the expected total number of events in the inclusive signal region is around 300. A finer binning scheme is allowed. Each $OO_{4\ell}$ is optimized

into 48 bins of varied sizes. Finer binning will get diminishing improvement in sensitivity. The detailed binning scheme is listed as follows:

$OO_{4\ell}^{c_{H\tilde{B}}}$ { -40.0000, -2.6422, -1.1817, -0.7242, -0.4920, -0.3610, -0.2778, -0.2197, -0.1778, -0.1460, -0.1208, -0.0994, -0.0819, -0.0674, -0.0549, -0.0446, -0.0356, -0.0282, -0.0219, -0.0166, -0.0122, -0.0085, -0.0054, -0.0026, 0.0000, 0.0026, 0.0054, 0.0085, 0.0122, 0.0166, 0.0219, 0.0282, 0.0356, 0.0446, 0.0549, 0.0674, 0.0819, 0.0994, 0.1208, 0.1460, 0.1778, 0.2197, 0.2778, 0.3610, 0.4920, 0.7242, 1.1817, 2.6422, 40.0000 }

$OO_{4\ell}^{c_{H\widetilde{W}B}}$ { -20.0000, -1.1164, -0.5182, -0.3200, -0.2228, -0.1672, -0.1309, -0.1056, -0.0867, -0.0715, -0.0591, -0.0486, -0.0400, -0.0326, -0.0264, -0.0210, -0.0166, -0.0129, -0.0099, -0.0074, -0.0054, -0.0038, -0.0024, -0.0011, 0.0000, 0.0011, 0.0024, 0.0038, 0.0054, 0.0074, 0.0099, 0.0129, 0.0166, 0.0210, 0.0264, 0.0326, 0.0400, 0.0486, 0.0591, 0.0715, 0.0867, 0.1056, 0.1309, 0.1672, 0.2228, 0.3200, 0.5182, 1.1164, 20.0000 }

$OO_{4\ell}^{c_{H\widetilde{W}}}$ { -10.0000, -0.3579, -0.2241, -0.1699, -0.1386, -0.1160, -0.0986, -0.0844, -0.0725, -0.0622, -0.0532, -0.0454, -0.0386, -0.0326, -0.0274, -0.0229, -0.0188, -0.0153, -0.0123, -0.0095, -0.0072, -0.0051, -0.0032, -0.0015, 0.0000, 0.0015, 0.0032, 0.0051, 0.0072, 0.0095, 0.0123, 0.0153, 0.0188, 0.0229, 0.0274, 0.0326, 0.0386, 0.0454, 0.0532, 0.0622, 0.0725, 0.0844, 0.0986, 0.1160, 0.1386, 0.1699, 0.2241, 0.3579, 10.0000 }

$OO_{4\ell}^{\widetilde{d}}$ { -50.0000, -3.4316, -1.4883, -0.8917, -0.5918, -0.4186, -0.3075, -0.2303, -0.1753, -0.1349, -0.1047, -0.0821, -0.0648, -0.0511, -0.0402, -0.0316, -0.0244, -0.0187, -0.0140, -0.0103, -0.0073, -0.0048, -0.0028, -0.0012, 0.0000, 0.0012, 0.0028, 0.0048, 0.0073, 0.0103, 0.0140, 0.0187, 0.0244, 0.0316, 0.0402, 0.0511, 0.0648, 0.0821, 0.1047, 0.1349, 0.1753, 0.2303, 0.3075, 0.4186, 0.5918, 0.8917, 1.4883, 3.4316, 50.0000 }

$OO_{4\ell}^{\widetilde{c}_{zz}}$ { -0.2000, -0.0330, -0.0286, -0.0256, -0.0233, -0.0212, -0.0194, -0.0178, -0.0163, -0.0149, -0.0135, -0.0122, -0.0110, -0.0098, -0.0087, -0.0076, -0.0066, -0.0056, -0.0046, -0.0037, -0.0028, -0.0020, -0.0013, -0.0006, 0.0000, 0.0006, 0.0013, 0.0020, 0.0028, 0.0037, 0.0046, 0.0056, 0.0066, 0.0076, 0.0087, 0.0098, 0.0110, 0.0122, 0.0135, 0.0149, 0.0163, 0.0178, 0.0194, 0.0212, 0.0233, 0.0256, 0.0286, 0.0330, 0.2000 }

$OO_{4\ell}^{\widetilde{c}_{za}}$ { -1.5000, -0.3871, -0.2903, -0.2318, -0.1910, -0.1615, -0.1376, -0.1181, -0.1019, -0.0883, -0.0760, -0.0657, -0.0564, -0.0484, -0.0413, -0.0351, -0.0294, -0.0243, -0.0196, -0.0155, -0.0117, -0.0083, -0.0053, -0.0024, 0.0000, 0.0024, 0.0053, 0.0083, 0.0117, 0.0155, 0.0196, 0.0243, 0.0294, 0.0351, 0.0413, 0.0484, 0.0564, 0.0657, 0.0760, 0.0883, 0.1019, 0.1181, 0.1376, 0.1615, 0.1910, 0.2318, 0.2903, 0.3871, 1.5000 }

$OO_{4\ell}^{\widetilde{c}_{aa}}$ { -15.0000, -0.9138, -0.4505, -0.2855, -0.1969, -0.1415, -0.1048, -0.0796, -0.0612, -0.0475, -0.0374, -0.0293, -0.0231, -0.0182, -0.0143, -0.0112, -0.0087, -0.0067, -0.0050, -0.0037, -0.0026, -0.0017, -0.0010, -0.0004, 0.0000, 0.0004, 0.0010, 0.0017, 0.0026, 0.0037, 0.0050, 0.0067, 0.0087, 0.0112, 0.0143, 0.0182, 0.0231, 0.0293, 0.0374, 0.0475, 0.0612, 0.0796, 0.1048, 0.1415, 0.1969, 0.2855, 0.4505, 0.9138, 15.0000 }

In the combined measurement, the 2D observables constructed from the production of two OO are used to enhance the sensitivity. The binning of the 2D observables is the direct product of each component with a reduced binning. For decay 2D OO, the reduced binning scheme for the components is listed as follows:

$OO_{4\ell}^{c_{H\tilde{B}}}$ { -40.0000, -0.1747, -0.0351, 0.0000, 0.0351, 0.1747, 40.0000 }

$OO_{4\ell}^{c_{H\widetilde{W}B}}$ { -20.0000, -0.0853, -0.0163, 0.0000, 0.0163, 0.0853, 20.0000 }

$OO_{4\ell}^{c_{H\widetilde{W}}}$ { -10.0000, -0.0721, -0.0187, 0.0000, 0.0187, 0.0721, 10.0000 }

Appendix B

Morphing basis and validations

Morphing validation

In this study, the performance of the morphing basis optimization algorithm shown in Figure 6.4 has been tested. The test is demonstrated with the VBF 2D morphing on the $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$. The test points in this validation are shown in Figure B.1. The morphing basis score is defined as follows:

$$\text{Score} = \sum_i N_{\text{eff}}, \tag{B.0.1}$$

where the index i iterates for the test points, the N_{eff} represents the effective statistics on the test points.

The summing effective statistics scores at the beginning of the algorithm and after 5000 iterations execution are shown in Figure B.2. Through 5000 iterations of execution, the effective statistics in the test points become uniform. Most areas of the plane have more than 50 thousand effective events. There is a direct comparison of the statistical uncertainties in the bin contents as well. The largest errors among bins at the beginning of the algorithm and after 5000 iterations execution are shown in Figure B.3. The largest bin-wise statistical errors in most areas of the plane are within 1 %. This iterative algorithm has performed well in achieving the morphing basis with reliable statistical predictions in the measured area.

Morphing basis

The ggF morphing basis needs 10 samples with the optimized couplings. The couplings in the morphing basis are optimized by the iterative algorithm described above. The couplings of the morphing basis samples are summarized in Table B.1. The optimization of these couplings is done on the Warsaw Basis. The Higgs Basis couplings are from direct basis transformation. These two bases are equivalent in this analysis.

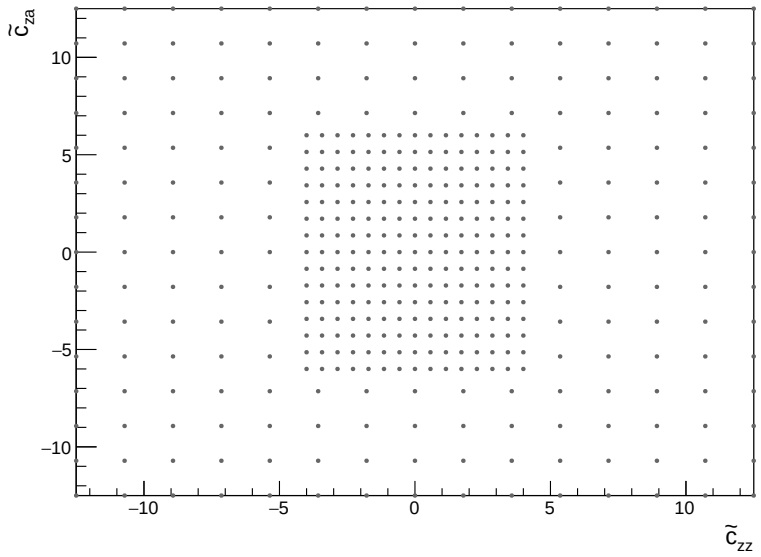


Figure B.1: The test points in the $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ plain for scoring the morphing basis optimization.

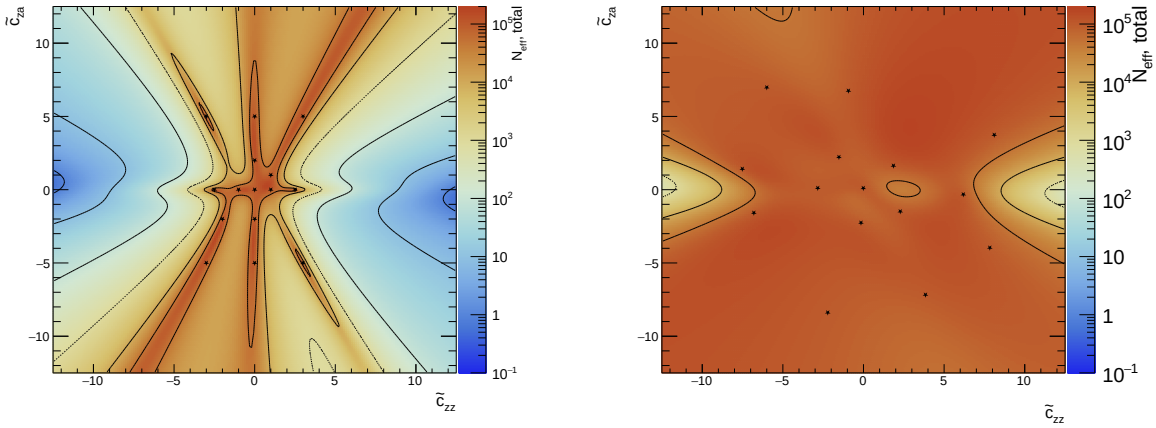


Figure B.2: The effective statistics in the $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ coupling plane from morphing at the beginning (left) and the result after 5000 iterations (right).

Table B.2: VBFVH samples and optimized couplings in the morphing basis.

Index	DSID	$\widetilde{c}_{zz}$	$\widetilde{c}_{z\gamma}$	$\widetilde{c}_{\gamma\gamma}$	$c_{H\widetilde{W}}$	$c_{H\widetilde{B}}$	$c_{H\widetilde{W}B}$
1	510315	9.03	-5.88	-8.33	10.53525	6.629140	25.86299
2	510316	-1.38	-0.76	0.56	-2.972843	0.07018251	-1.989089
3	510317	6.03	0.20	-9.71	9.908042	-0.07205297	14.31847
4	510318	-11.97	0.96	5.09	-19.84440	-5.241837	-25.20945

5	510319	11.59	1.37	9.32	22.26313	7.612600	17.35926
6	510320	-5.33	7.43	2.69	-3.308793	-7.695564	-18.67030
7	510321	2.72	-0.86	11.99	5.155475	5.698939	2.107367
8	510322	5.74	-11.45	-9.55	0.2777997	8.964800	25.87624
9	510323	8.45	-0.55	1.57	14.56032	5.172999	15.97772
10	510324	-8.81	-3.80	6.29	-17.91668	0.4502818	-14.68216
11	510325	1.48	1.15	11.99	4.551658	3.497300	-2.325611
12	510326	-4.07	3.63	10.31	-3.407096	-1.759242	-14.81891
13	510327	-3.27	0.57	0.06	-5.296182	-2.078554	-6.773654
14	510328	-0.60	7.11	2.20	4.712353	-5.207353	-9.267141
15	510329	-0.13	0.90	0.14	0.4893773	-0.7286126	-1.229667
16	510330	-11.68	2.65	-3.07	-18.72294	-8.906188	-23.76335
17	510331	-1.50	-0.67	-12.00	-4.211336	-3.886792	1.790346
18	510332	-3.24	-6.36	-7.13	-11.30568	1.180079	2.860678
19	510333	1.99	0.48	2.17	4.064306	1.288715	2.539213
20	510334	7.56	-9.75	5.96	6.166136	13.27463	22.46828
21	510335	2.04	0.10	-0.92	3.584098	0.6706558	4.038428
22	510336	1.02	-3.81	9.49	-0.3645483	6.392683	2.799427
23	510337	1.04	-0.53	11.47	2.415476	4.434140	-1.231666
24	510338	-3.73	0.13	-11.73	-7.480600	-5.556974	-3.332921
25	510339	-3.90	6.95	-8.38	-2.139185	-9.969687	-11.86286
26	510340	0.04	-0.20	0.47	-0.04540648	0.3201620	0.1306335
27	510341	-1.60	1.60	-11.77	-2.587174	-5.647032	-0.8417182
28	510342	0.48	5.40	-8.91	4.298789	-6.705847	-1.822452
29	510343	1.98	-11.09	7.03	-4.600080	11.83577	13.00690
30	510344	-8.00	-0.63	-1.16	-14.65862	-3.895854	-14.03265
31	510345	-0.74	0.79	-5.01	-1.119453	-2.518871	-0.5829953
32	510346	11.34	2.43	-0.45	21.80059	3.679336	18.97107
33	510347	0.16	0.08	4.37	0.7260562	1.349135	-1.208399
34	510348	-0.59	-0.53	-2.75	-1.693152	-0.7198006	0.3392728

35	510349	1.21	-3.23	-0.96	-0.4894199	2.850647	5.962811

The VBF/VH morphing basis needs 35 samples with the optimized couplings. The couplings are optimized by the same algorithm. The couplings of the morphing basis samples are summarized in Table B.2. The couplings for the VBF/VH samples are optimized on the Higgs Basis. The Warsaw Basis couplings are from direct basis transformation as well.

These morphing bases are validated in this study. Additional samples with different couplings are generated from the MC simulation. The morphing predictions and the MC predictions are compared to check the difference. In Figure B.4, the morphing results from the optimized basis and the direct MC simulation are compared for the ggF channels. The agreement between morphing and direct simulation is good. In most of the bins, the difference is within several percentages.

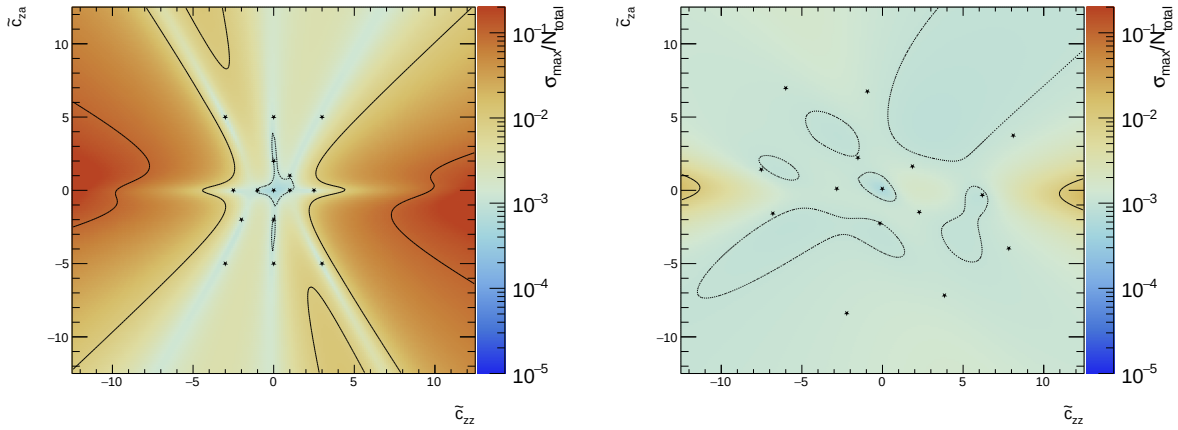


Figure B.3: The maximum single bin statistical error over the total events in the $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ coupling plane from morphing at the beginning (left) and the result after 5000 iterations (right).

Table B.1: ggF samples and optimized couplings in the morphing basis.

Index	DSID	$c_{H\widetilde{W}}$	$c_{H\widetilde{B}}$	$c_{H\widetilde{W}B}$	$\widetilde{c}_{zz}$	$\widetilde{c}_{z\gamma}$	$\widetilde{c}_{\gamma\gamma}$
1	507439	9.06	2.49	10.95	5.37155	-0.316557	-1.53794
2	507440	-9.71	1.99	-0.06	-3.14954	-5.15371	-1.51588
3	507441	-9.82	-6.83	-7.3	-5.38879	0.825412	-11.3716
4	507442	4.08	1.4	-0.02	1.53549	1.19043	5.11688
5	507443	2.99	-1.72	1.78	1.18480	1.55829	-3.59805
6	507444	2.58	3.65	4.89	2.14552	-1.91112	3.50945
7	507445	-7.14	1.75	-10.12	-4.14152	-0.952982	10.1522
8	507446	-1.46	-2.98	-6.71	-2.02970	2.64529	0.388058
9	507447	1.65	-0.73	10.49	2.42473	-2.03323	-11.6466
10	507448	-2.42	-0.52	-1.31	-1.12337	-0.454512	-1.01627

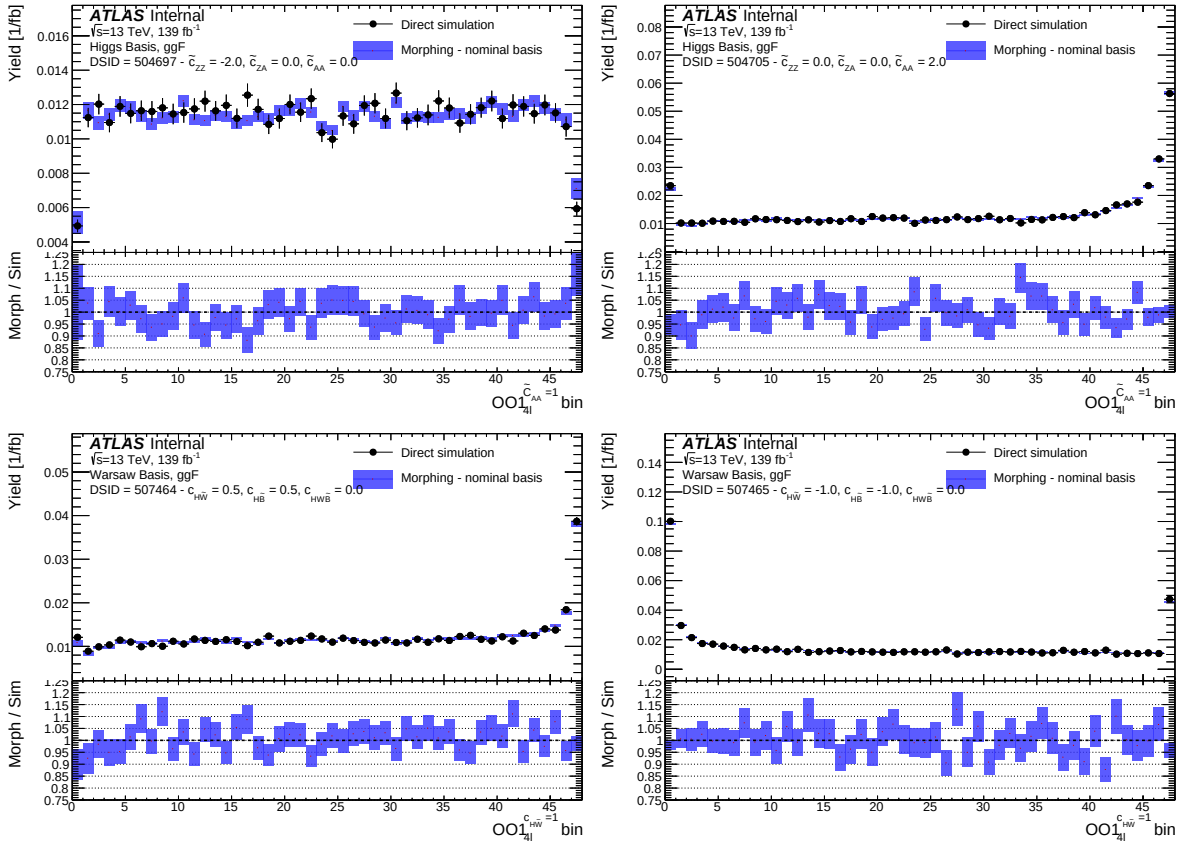


Figure B.4: The ggF morphing validation in the inclusive Higgs boson mass region. The black dot shows the result from direct MC simulation. The blue line is the result of morphing with MC uncertainties.

Appendix C

Normalization floating validation

Normalization Schemes

As introduced in Section 6.5, the yields of events in this study are fit to data. In the $H \rightarrow ZZ^* \rightarrow 4l$ measurement, the normalization factors μ_{Signal} and μ_{qqZZ} are introduced to constrain float the yield of inclusive Higgs boson events and the non-resonant ZZ^* backgrounds. The categorization and normalization scheme for the $H \rightarrow ZZ^* \rightarrow 4l$ measurement is shown in Figure C.1.

In the VBF production measurement, the normalization factors μ_{VBF} for the VBF events, μ_{ggF} for the other Higgs boson production channels and μ_{qqZZ} for the ZZ^* backgrounds are introduced. For the NN sub-categories in the VBF region, there are additional factors $\mu_{\text{NN1,2}}$ and $f(\mu_{\text{NN1,2}})$ introduced. The normalization factor $\mu_{\text{NN1,2}}$ is defined for the first two NN categories simultaneously in the VBF region. The normalization modifier $f(\mu_{\text{NN1,2}})$ is introduced as a universal factor for all four VBF signal regions to keep the inclusive yield of events in the VBF inclusive SR invariant. The modifier $f(\mu_{\text{NN1,2}})$ is defined as follows:

$$f_{\mu_{\text{NN1,2}}} = \frac{N^{\text{MC}}_{\text{NN1}} + N^{\text{MC}}_{\text{NN2}} + N^{\text{MC}}_{\text{NN3}} + N^{\text{MC}}_{\text{NN4}}}{\mu^2_{\text{NN1,2}}(N^{\text{MC}}_{\text{NN1}} + N^{\text{MC}}_{\text{NN2}}) + N^{\text{MC}}_{\text{NN3}} + N^{\text{MC}}_{\text{NN4}}}, \tag{C.0.1}$$

where each $N^{\text{MC}}_{\text{NN}n}$ is the MC predicted yield of signal events in the n-th NN category. The categorization and normalization scheme for the VBF production measurement is shown in Figure C.2.

In the production and decay combined measurement, the categorization is the same as the production-only measurement. The same normalization factors for the production-only measurement are introduced in the combined measurement normalization. The categorization and normalization scheme for the production and decay combined measurement is shown in Figure C.3.

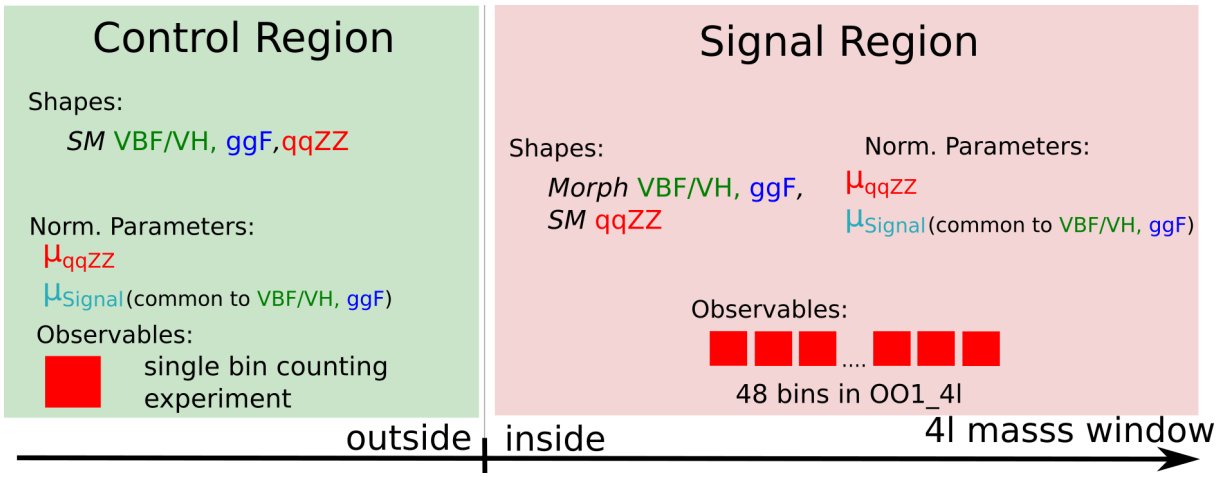


Figure C.1: The categorization and normalization floating scheme of the four leptons decay measurement

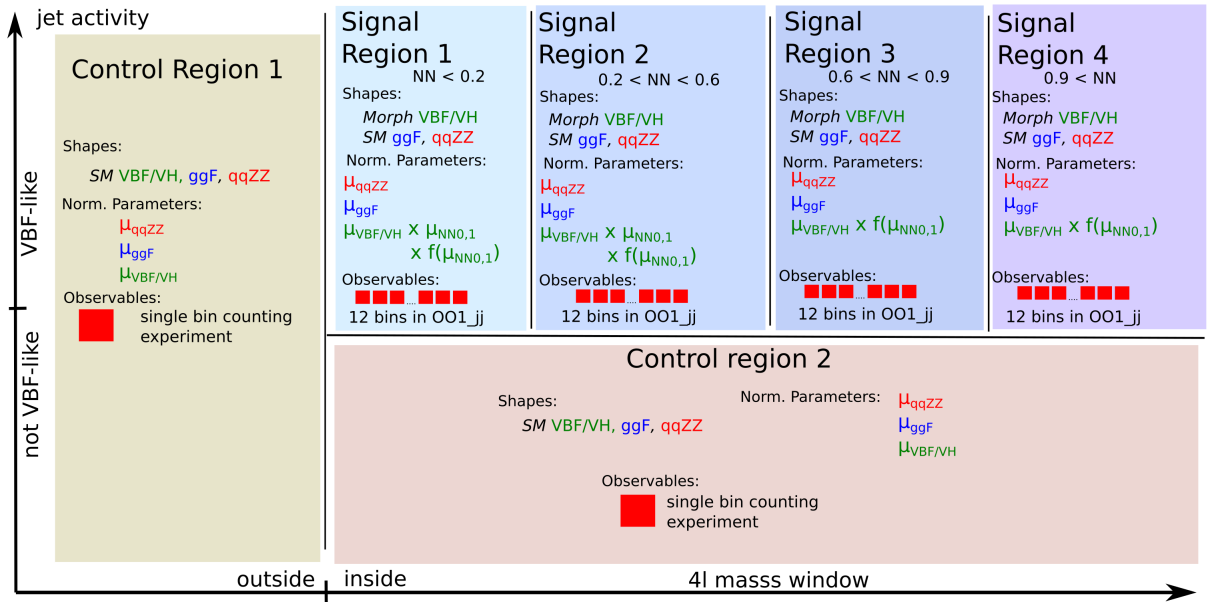


Figure C.2: The categorization and normalization floating scheme of the VBF production measurement

Validation on the Normalization

The validation of floating normalization has been validated with a 3 NN categories scheme. This normalization scheme was tested in the VBF production $c_{H\widetilde{W}}$ measurement on the SM expectation model and the model with VBF/VH yield 3.5 times amplified from the SM expectation. The likelihood fit with and without μ_{VBFVH} floating on these models are shown in Figure C.4. The μ_{ggF} and μ_{VBFVH} best-fit values in these scans are shown in the Figure C.5. For the fits on

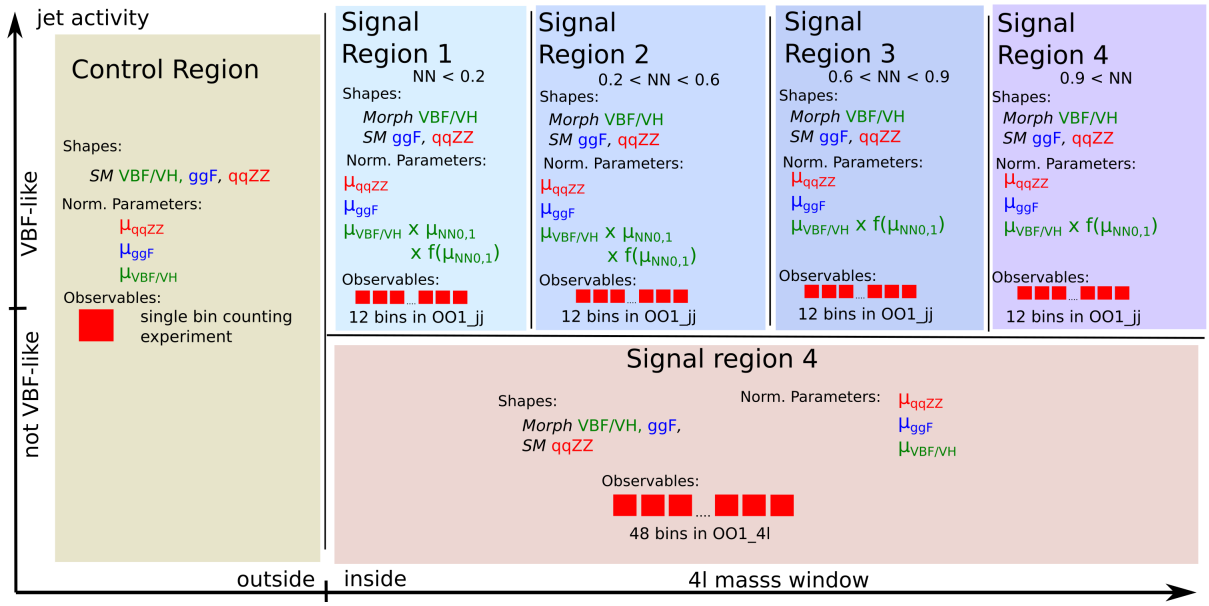


Figure C.3: The categorization and normalization floating scheme of the four leptons decay measurement

the SM expectation model, the VBF/VH yield floating makes only a very small difference in the fit. Under the SM hypothesis, the fit gets $\mu_{\text{VBFVH}} = 1$. For the BSM hypotheses, the μ_{VBFVH} factor floats to lower values. The μ_{ggF} factor compensates the difference. For the fits on the model with 3.5 times amplified VBF/VH yield from the SM expectation, μ_{VBFVH} for the SM hypothesis captures the change in the VBF/VH yield well. Without this floating, the ggF yield is wrongly amplified in the green curve.

The normalization floating test for the BSM expectation fit has been carried out with the coupling $c_{H\widetilde{W}} = 3.0$. The fit results with and without μ_{VBFVH} floating on the $c_{H\widetilde{W}} = 3.0$ BSM expectation model and the model with 3.5 times amplified VBF/VH yield are shown in Figure C.7. The μ_{ggF} and μ_{VBFVH} best-fit values in these scans are shown in the Figure C.7. For the fit on the $c_{H\widetilde{W}} = 3.0$ model, μ_{VBFVH} floating cause a very small impact on the fit. For the fit on the amplified VBF/VH yield model, the floating μ_{VBFVH} captures the yield amplification correctly. Without the floating, the ggF backgrounds get a wrongly amplified yield. The green curves showed a large bias in the fit without correct VBF/VH yield estimation. Floating of the μ_{VBFVH} factor is needed to reduce the bias from VBF/VH yield estimation uncertainties for the BSM scenarios.

The background yield floating has been tested with the production-only $\widetilde{c}_{zz}$ measurement. The $\widetilde{c}_{zz}$ fits on the SM expected model without the floating of μ_{VBFVH}, μ_{qqZZ} and μ_{ggF} factors, and with floating factors constrained in different control regions are shown in Figure C.8. The corresponding best-fit values of the normalization factors are shown in Figure C.9. The μ_{qqZZ}

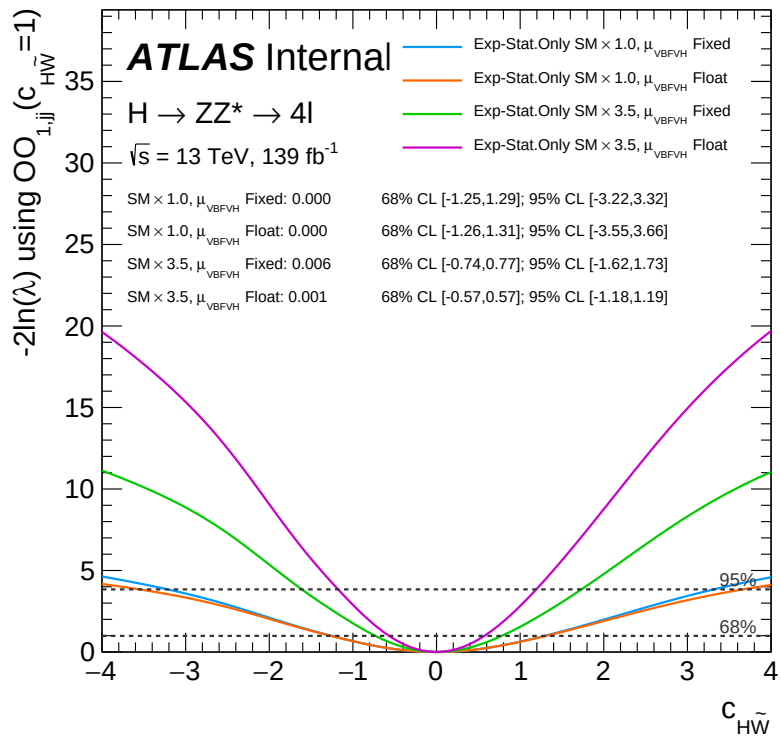


Figure C.4: The $c_{H\widetilde{W}}$ fit with and without signal normalization floating on the SM expectation and increased signal yield expectation. The blue and orange curves show the fits on the SM expectation model without and with μ_{VBFVH} floating. The violet and green curves show the fits on the SM-like model with 3.5 times amplified VBF/VH yield without and with μ_{VBFVH} floating.

factor is mainly constrained in the ZZ^* CR. The absence of the ggF CR affects μ_{qqZZ} constraint slightly. In comparison, the ggF-dominated VBF-depleted region is essential for the μ_{ggF} constraining. Using only ZZ^* CR for the background constraining leads to a large overestimation of the ggF event yield.

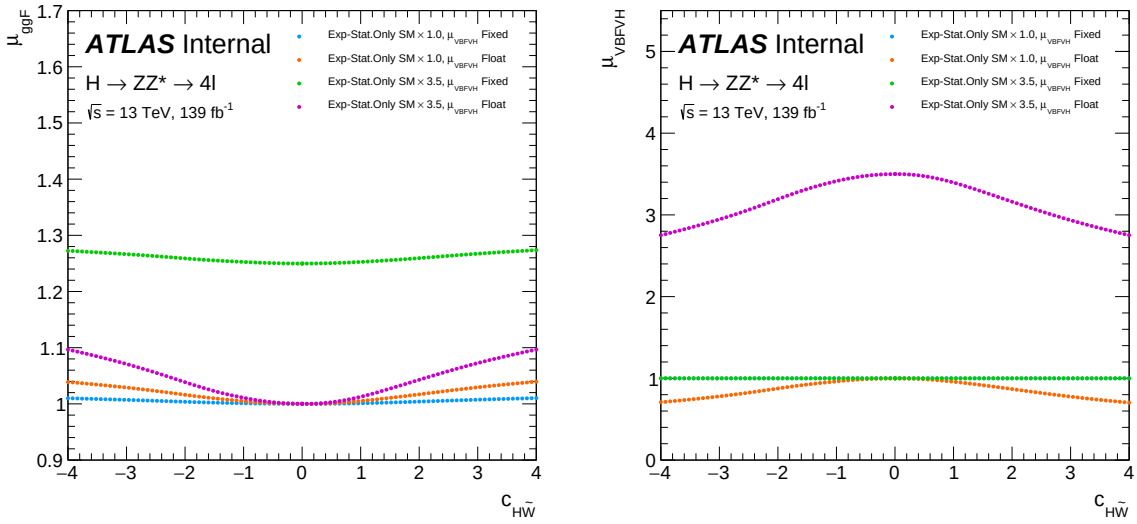


Figure C.5: The μ_{ggF}, μ_{VBFVH} best-fit values along $c_{H\widetilde{W}}$ scan with and without signal normalization floating on the SM expectation and increased signal yield expectation. The blue and orange curves show the fits on the SM expectation model without and with μ_{VBFVH} floating. The violet and green curves show the fits on the SM-like model with 3.5 times amplified VBF/VH yield without and with μ_{VBFVH} floating.

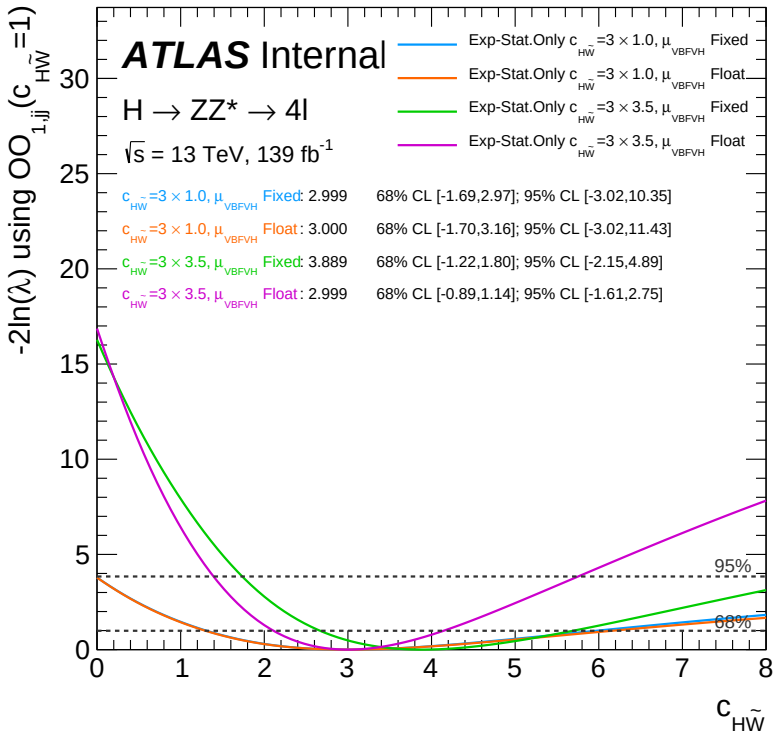


Figure C.6: The $c_{H\widetilde{W}}$ fit results with and without signal normalization floating from BSM expectation and increased signal yield expectation at $c_{H\widetilde{W}} = 3.0$.

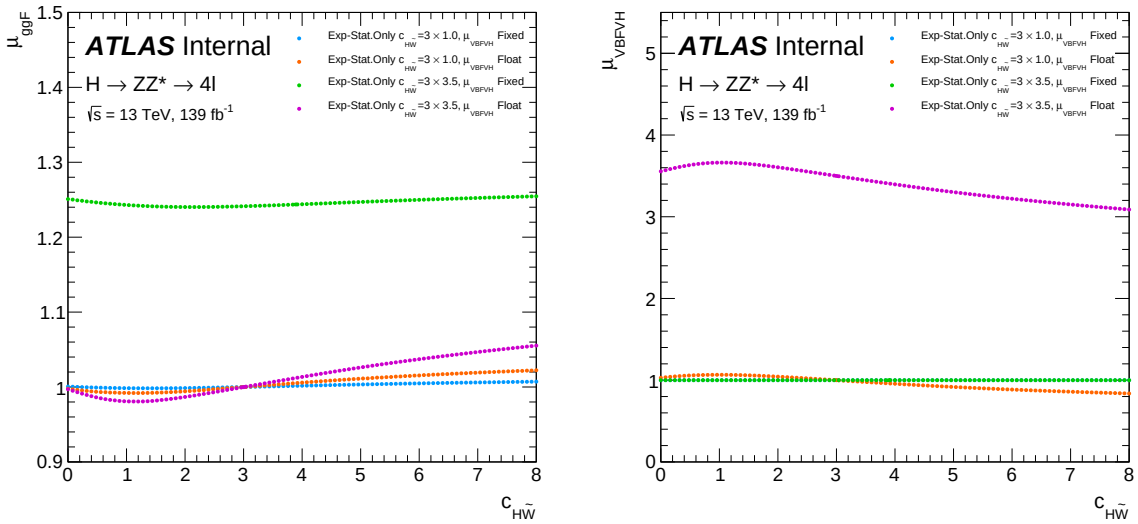


Figure C.7: The μ_{ggF}, μ_{VBFVH} profiled results along $c_{H\widetilde{W}}$ scan with and without signal normalization floating from BSM expectation and increased signal yield expectation at $c_{H\widetilde{W}} = 3.0$.

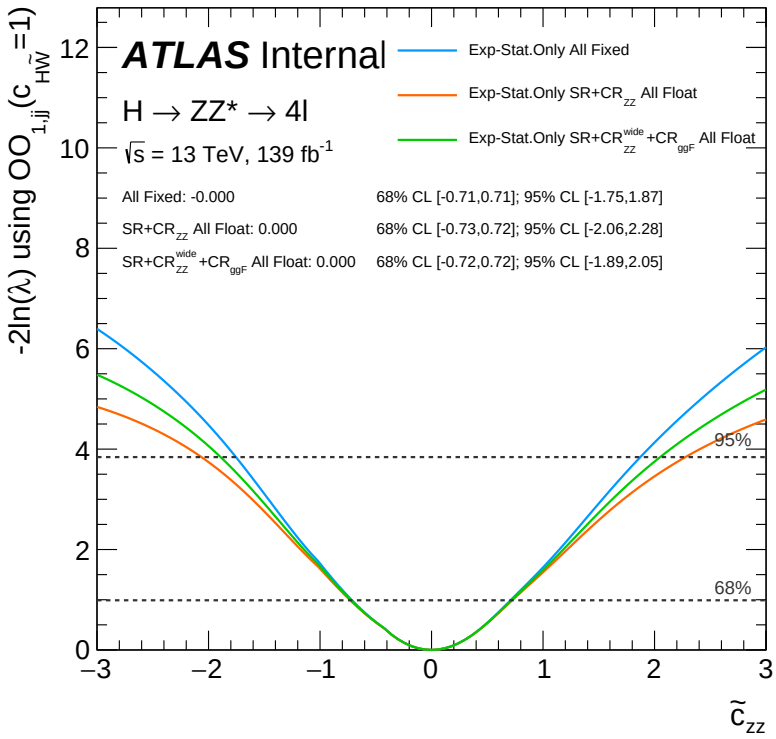


Figure C.8: The SM expectation $\widetilde{c}_{zz}$ fit results from fixing normalization and floating the signal and backgrounds normalization floating constrained by control regions. The blue curve is the $\widetilde{c}_{zz}$ fit with all normalization factors fixed. The orange curve is the fit allowing μ_{VBFVH}, μ_{qqZZ} and μ_{ggF} floating in the SR and ZZ^* CR. The green curve is the fit allowing μ_{VBFVH}, μ_{qqZZ} and μ_{ggF} floating in all of the SR, ggF CR and ZZ^* CR.

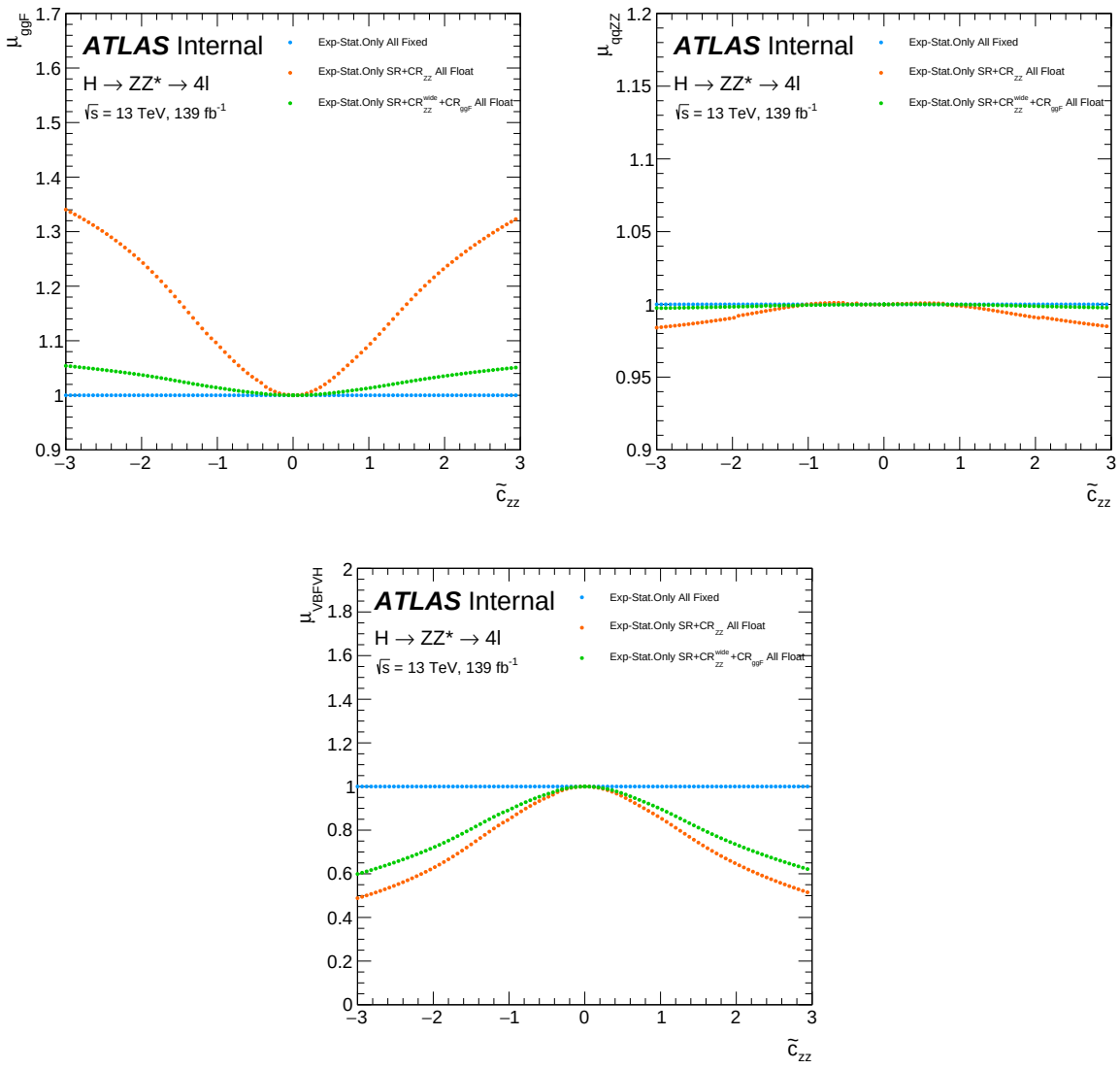


Figure C.9: The corresponding best-fit μ_{VBFVH}, μ_{qqZZ} and μ_{ggF} within the floating of these normalization factors turnning on and off. The blue curve is the $\widetilde{c}_{zz}$ fit with all normalization factors fixed. The orange curve is the fit allowing μ_{VBFVH}, μ_{qqZZ} and μ_{ggF} floating in the SR and ZZ^* CR. The green curve is the fit allowing μ_{VBFVH}, μ_{qqZZ} and μ_{ggF} floating in all of the SR, ggF CR and ZZ^* CR.

Appendix D

Reducible background distributions

Reducible backgrounds data-driven formula

As introduced in Section 6.6, the estimation of $Z + \mu\mu$ and $Z + ee$ are done separately in this analysis. The $Z + \mu\mu$ reducible backgrounds are fit to data with leading lepton pair mass m_{12} in the orthogonal control regions described as follows:

- Inverted d_0 CR is the four-lepton selection region without the impact parameter d_0 and isolation selections on the subleading lepton pair. At least one lepton in the subleading pair needs to fail the d_0 selection to make sure the orthogonality to the four-lepton selection region. This region is abundant with heavy flavour $Z+$ Jets and $t\bar{t}$ events.
- $e\mu+\mu\mu$ CR is the four-lepton selection region with the requirement of the opposite-charge different-flavour subleading lepton pair. This region is abundant with $t\bar{t}$ events.
- Inverted isolation CR is the four-lepton selection region without the isolation selection on the subleading lepton pair. At least one lepton in the subleading pair needs to fail the isolation selection. This region is abundant with light flavour Z+Jets events.
- Same-sign CR is the four-lepton selection region without the d_0 and isolation selections on the subleading lepton pair. There is the requirement of the same-charge subleading lepton pair applied. This region is abundant with all types of reducible backgrounds.

The reducible backgrounds are fit to data in these orthogonal control regions simultaneously. The fit results are validated and used for the data-driven estimation in the relaxIsoD0 region. From the fit, the scaling factors between the MC simulation and data are achieved.

The resulting data-driven $Z + \mu\mu$ prediction for the signal region follows:

$$N_{\text{SR},j} = N_{\text{data,relaxIsoD0},j} \frac{\sum_i N_{\text{MC,relaxIsoD0},i,j} * \text{SF}_i * \text{TF}_i}{\sum_i N_{\text{MC,relaxIsoD0},i,j} * \text{SF}_i}, \tag{D.0.1}$$

where the symbol j is the bin index for the OO distribution. The symbol i is the index for the type of reducible backgrounds, in the $Z+\mu\mu$ backgrounds including $Z+$ Jets and $t\bar{t}$. The symbol $N_{\text{SR},j}$ is the j-th bin content of the reducible backgrounds in the regions used in this analysis. The symbol $N_{\text{data,relaxIsoD0},j}$ is the j-th bin content of the Run 2 data in the relaxIsoD0 region. The symbol $N_{\text{MC,relaxIsoD0},i,j}$ is the j-th bin content of the MC simulated type i backgrounds in the relaxIsoD0 region. The SF and TF are the scaling factors between the MC simulation and data and the transfer factors between the signal region and relaxIsoD0 region estimated from the data fit and MC simulation.

The $Z + ee$ reducible backgrounds are fit to data in the $3l + X$ control region to data with the number of detections on the Inner Detector pixel sensors n_{InnerPix}. Modeling of $Z + ee$ background components is done with the sPlot method. Different from the direct fit with analytical function distributions in the $Z + \mu\mu$ backgrounds, the sPlot method assigns a weight to each event standing for the probability to be f or γ. The yield, selection efficiency and fake rate distributions of the f and γ backgrounds in the X p_T and categorization in this analysis are accumulated from each background event weight.

The resulting data-driven $Z + ee$ prediction for the signal region follows:

$$\begin{aligned}
\text{SF}_i =& \frac{\sum_k N_{\text{sPlot},i,k} - N_{\text{MC,Higgs}+ZZ^*,3l+X,i}}{N_{\text{MC,fake},3l+X,i}}, \\
\text{TF}_i =& \frac{\sum_k \text{Eff}_{i,k} * \text{S}_k * N_{\text{sPlot},i,k}}{\sum_k N_{\text{sPlot},i,k}}, \\
N_{\text{SR},j} =& (N_{\text{data},3l+X,j} - N_{q,3l+X,j} - N_{\text{MC,Higgs}+ZZ^*,3l+X,j}) \\
& * \frac{\sum_i N_{\text{MC},3l+X,i,j} * \text{SF}_i * \text{TF}_i}{\sum_i N_{\text{MC},3l+X,i,j} * \text{SF}_i},
\end{aligned} \tag{D.0.2}$$

where the symbol k is the index for the categorization in this study. The symbol j is the bin index for the OO distribution. The symbol i is the index for the type of reducible backgrounds, in the $Z + ee$ backgrounds estimation including f and γ. The scaling factor SF_i for i-type background is the data-to-MC yield ratio. The symbol $\sum_k N_{\text{sPlot},i,k}$ is the yield of i-type backgrounds in $3l + X$ region estimated by the sPlot method on data fit. The symbol $N_{\text{MC,Higgs}+ZZ^*,3l+X\text{ fullmass},i}$ is the yield of events identified as i-type background estimated from the MC simulation on Higgs boson and ZZ^* channels. The symbol $N_{\text{MC,fake},3l+X\text{ fullmass},i}$ is the yield of events identified as i-type background estimated from the MC simulation on the $Z + ee$ backgrounds. The transfer factor TF_i for i-type background is the ratio of the signal region events yield and control region events yield. The symbol $\text{Eff}_{i,k}$ is the efficiency of the i-type events in the k category passing four-lepton signal region selections. The symbol S_k is the efficiency scale factor defined as the ratio of efficiency in the signal region and $3l + X$ region.

$Z + \mu\mu$ shape estimation

The estimation results of the $Z + \mu\mu$ OO_{jj} distributions in the inclusive VBF signal region are shown in Figure D.1. The estimation results of the $Z + \mu\mu$ Warsaw Basis $OO_{4\ell}$ distributions in the inclusive Higgs boson signal region are shown in Figure D.2. The estimation results of the $Z + \mu\mu$ Higgs Basis $OO_{4\ell}$ distributions in the inclusive Higgs boson signal region are shown in Figure D.3. The estimated yields of reducible backgrounds in the regions studied in this analysis are summarized in Table D.1 for $Z + \mu\mu$. The nominal values are estimated without uncertainties. The statistical uncertainties are based on the statistical uncertainties of the Run 2 data in the relaxIsoD0 region. The systematical uncertainties are estimated from the fluctuations in the transfer factors and scaling factors estimation.

$Z + ee$ shape estimation

The estimation results of the $Z + ee$ OO_{jj} distributions in the inclusive VBF signal region are shown in Figure D.4. The estimation results of the $Z + ee$ Warsaw Basis $OO_{4\ell}$ distributions in the inclusive Higgs boson signal region are shown in Figure D.5. The estimation results of the $Z + ee$ Higgs Basis $OO_{4\ell}$ distributions in the inclusive Higgs boson signal region are shown in Figure D.6. The estimated yields of reducible backgrounds in the regions studied in this analysis are summarized in Table D.2 for $Z + ee$. The nominal values are estimated without uncertainties. The statistical uncertainties are based on the statistical uncertainties of the Run 2 data in the $3l + X$ region. The systematical uncertainties are estimated from the fluctuations in the transfer factors and scaling factors estimation.

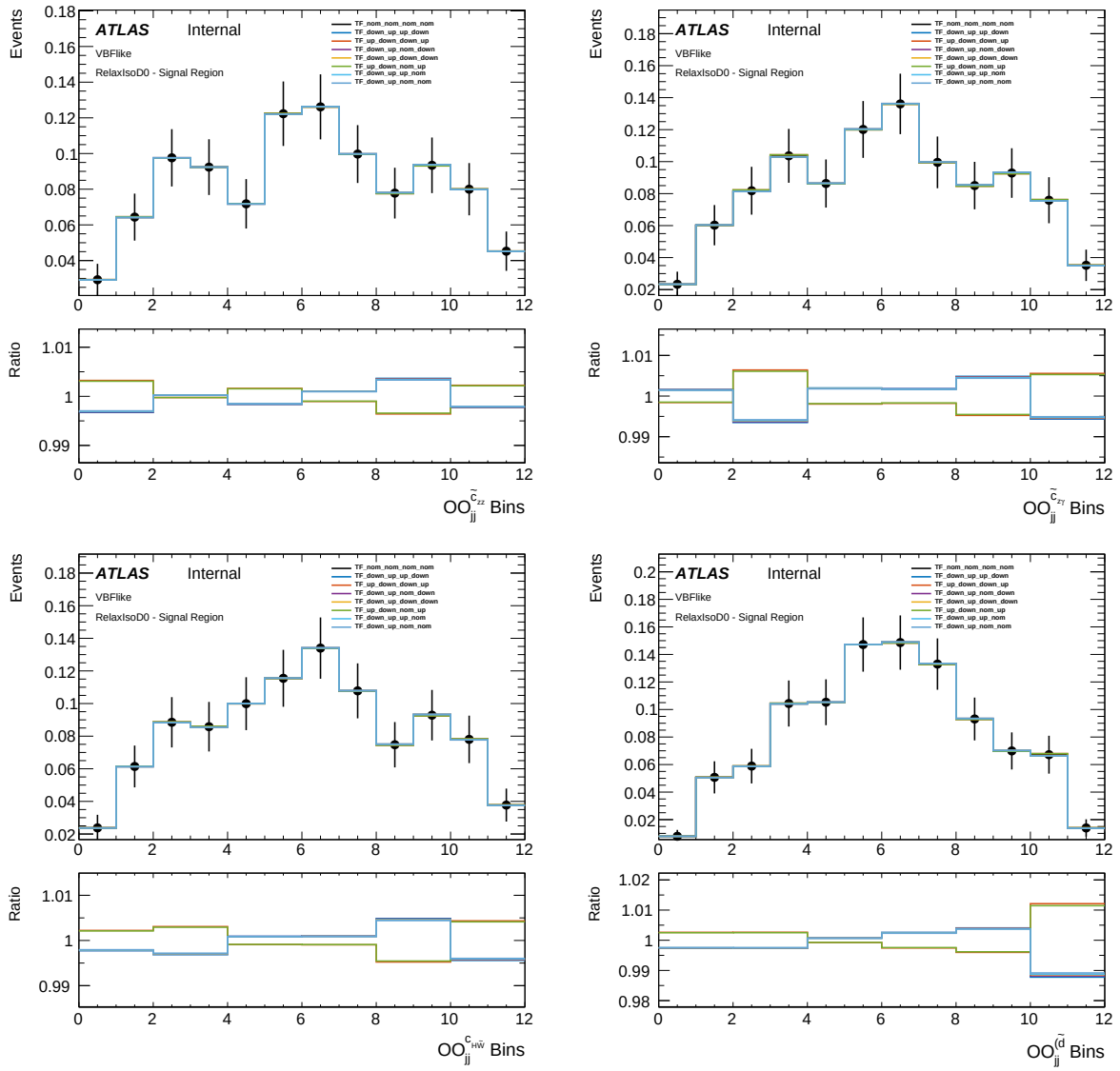


Figure D.1: Data-driven estimated OO_{jj} distributions for the reducible $Z + \mu\mu$ backgrounds. The upper panels show the distributions normalized to 1. The lower panels show the impact of systematical uncertainties on the shapes.

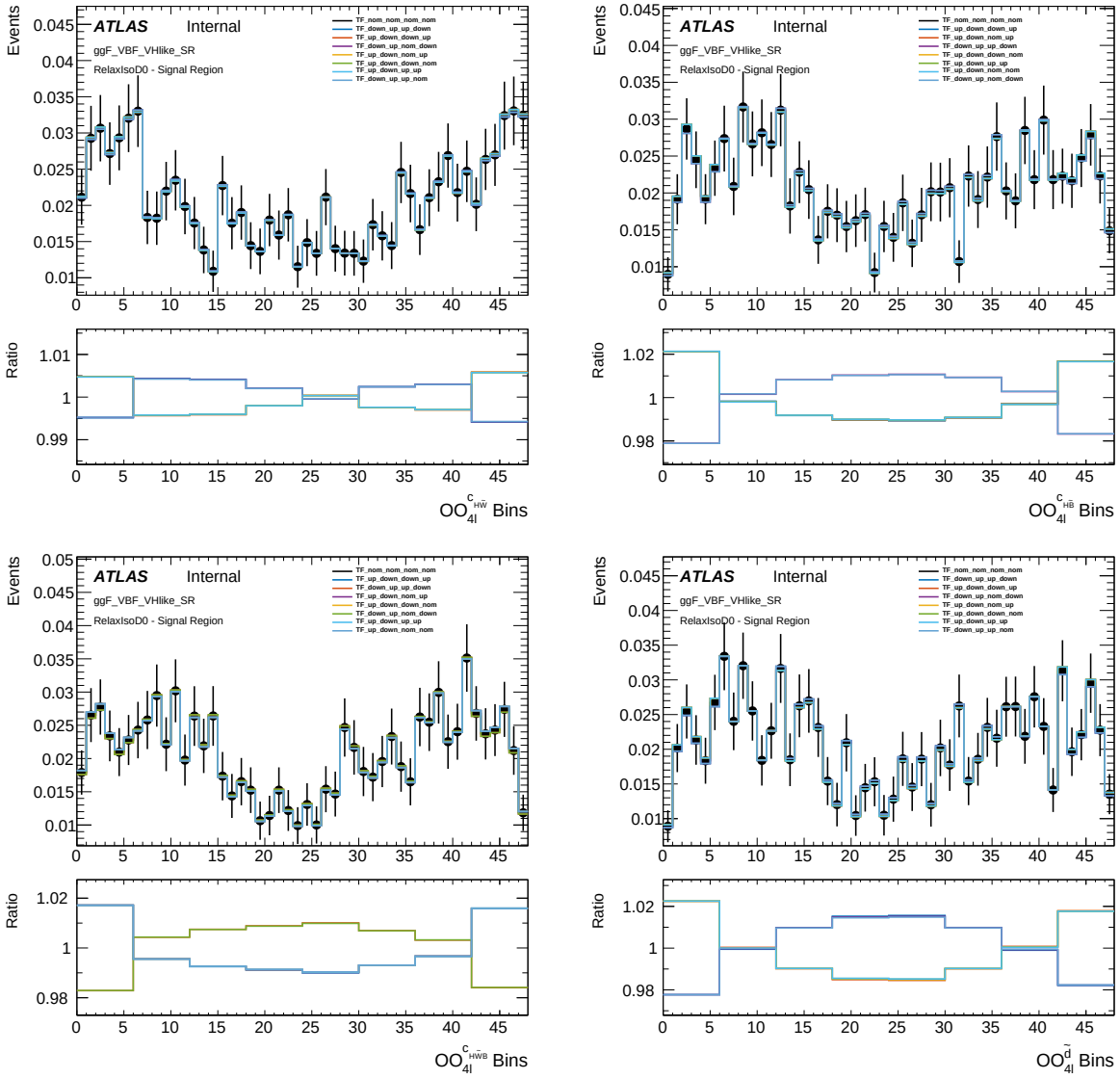


Figure D.2: Data-driven estimated $OO_{4\ell}$ distributions for the reducible $Z + \mu\mu$ backgrounds. The upper panels show the distributions normalized to 1. The lower panels show the impact of systematical uncertainties on the shapes. The Warsaw Basis observables are shown.

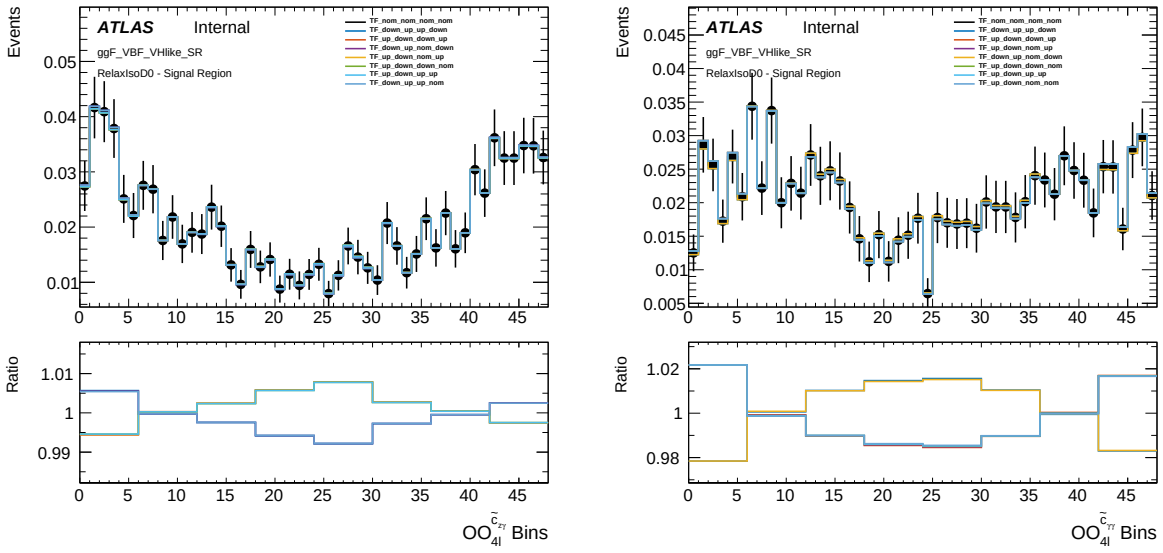


Figure D.3: Data-driven estimated $OO_{4\ell}$ distributions for the reducible $ll\mu\mu$ backgrounds. The upper panels show the distributions normalized to 1. The lower panels show the impact of systematical uncertainties on the shapes. The Higgs Basis observables are shown.

Table D.1: $Z + \mu\mu$ reducible backgrounds yield in the signal regions and control regions estimated through the data-driven method.

Category	Nominal	Statistical	Systematical
inclusive SR	4.94	0.13	+0.18 -0.18
VBF SR1	0.471	0.028	+0.061 -0.060
VBF SR2	0.168	0.018	+0.035 -0.034
VBF SR3	0.032	0.010	+0.015 -0.013
VBF SR4	0.006	0.006	+0.005 -0.005
VBF-depleted region	4.24	0.13	+0.16 -0.16
ZZ^* CR	10.43	0.19	+0.37 -0.37

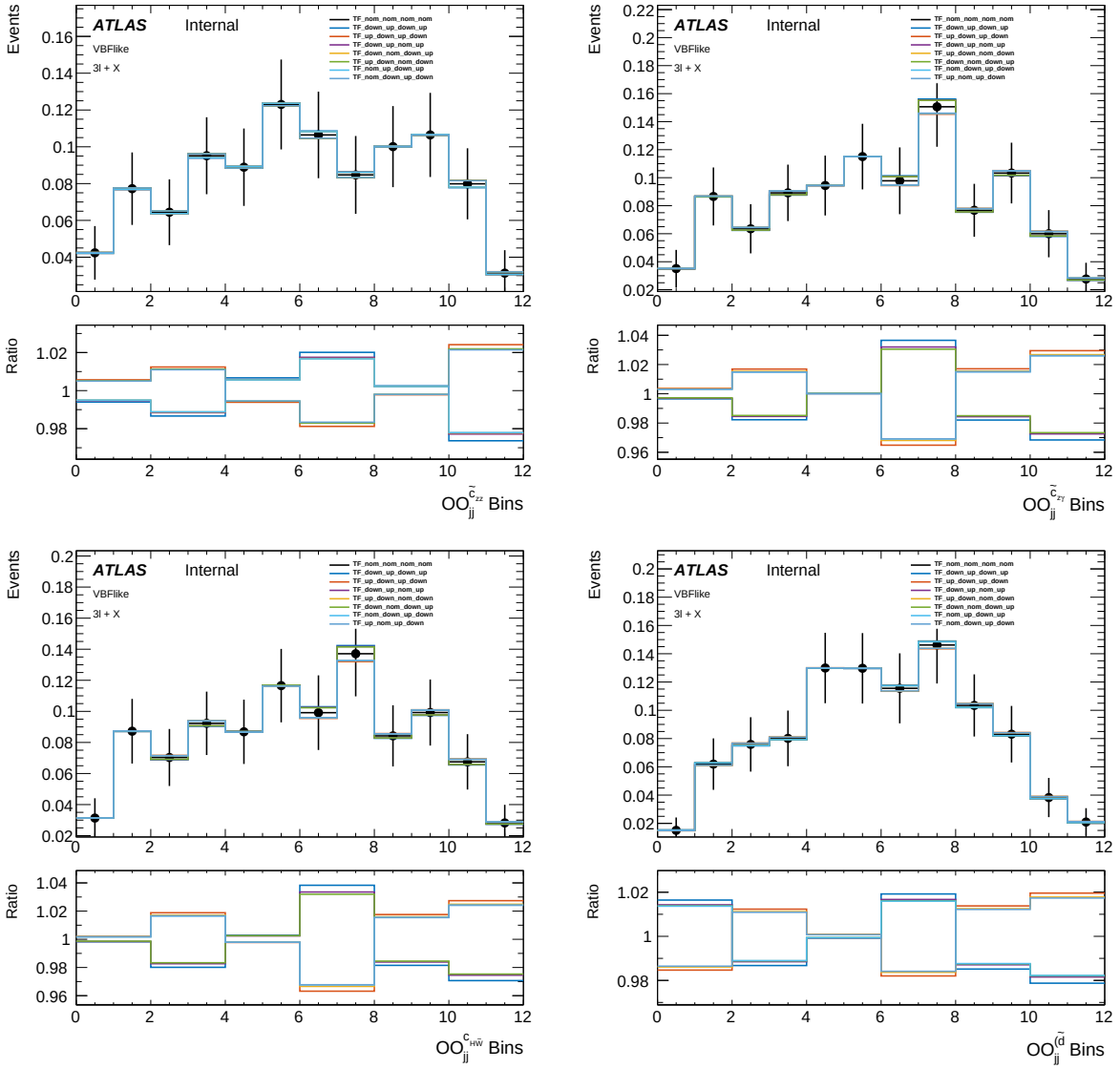


Figure D.4: Data-driven estimated OO_{jj} distributions for the reducible $Z + ee$ backgrounds. The upper panels show the distributions normalized to 1. The lower panels show the impact of systematical uncertainties on the shapes.

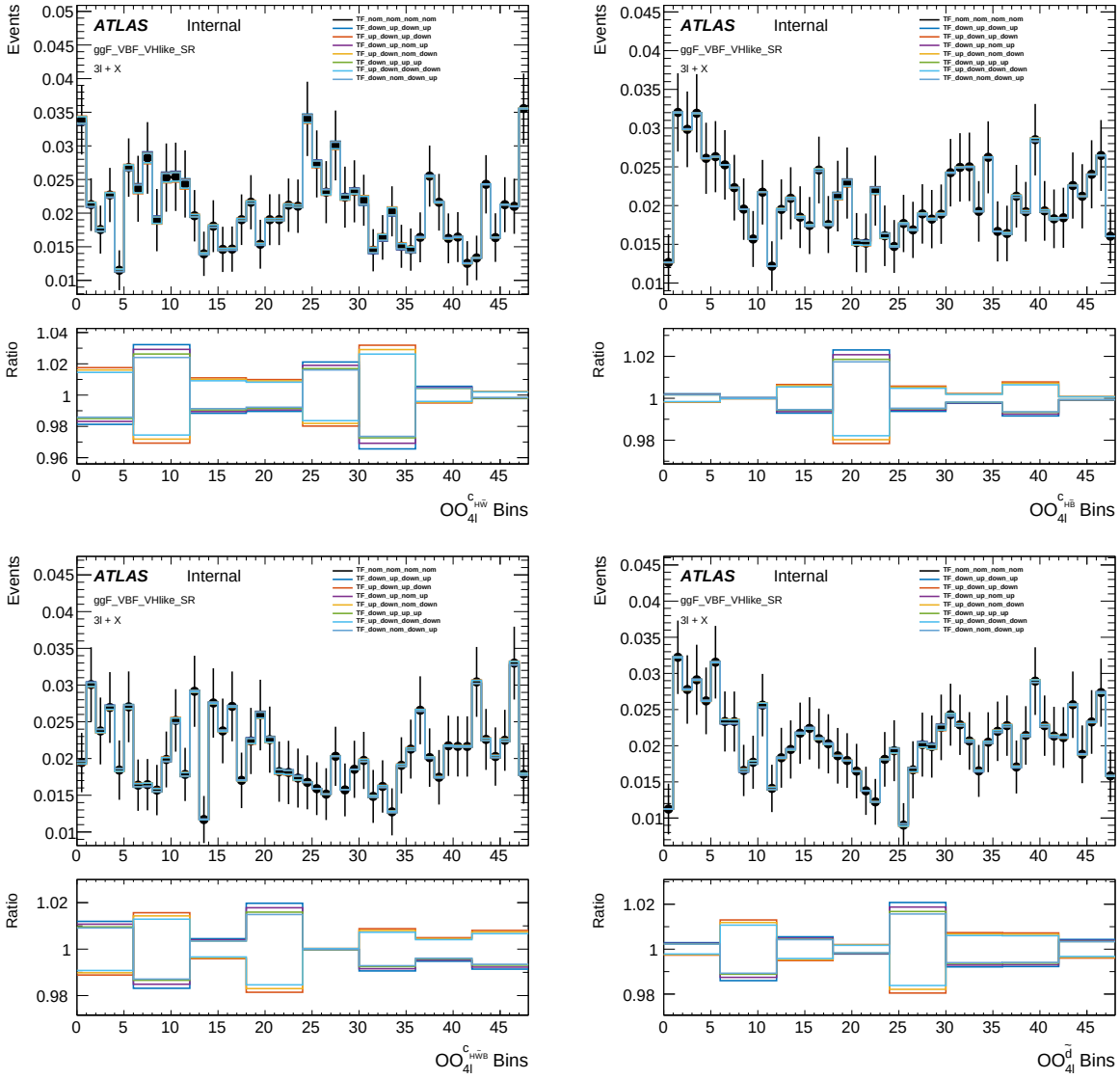


Figure D.5: Data-driven estimated $OO_{4\ell}$ distributions for the reducible $llee$ backgrounds. The upper panels show the distributions normalized to 1. The lower panels show the impact of systematical uncertainties on the shapes. The Warsaw Basis observables are shown.

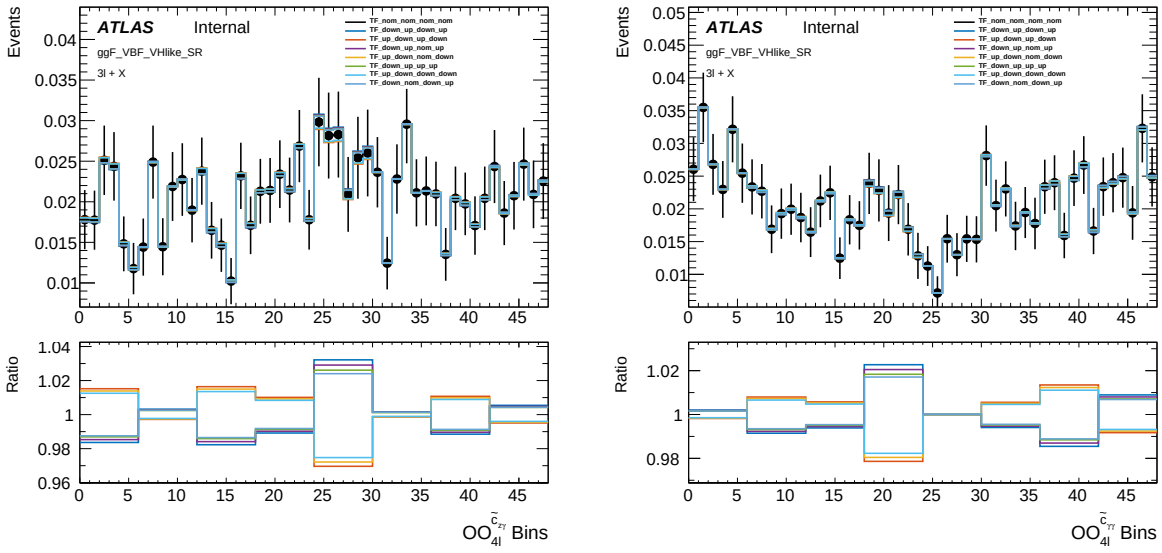


Figure D.6: Data-driven estimated $OO_{4\ell}$ distributions for the reducible $llee$ backgrounds. The upper panels show the distributions normalized to 1. The lower panels show the impact of systematical uncertainties on the shapes. The Higgs Basis observables are shown.

Table D.2: $Z + ee$ reducible backgrounds yield in the signal regions and control regions estimated through the data-driven method.

category	Nominal	Statistical	Systematical
inclusive SR	2.58	0.07	+0.13 -0.13
VBF SR1	0.157	0.014	+0.009 -0.008
VBF SR2	0.086	0.011	+0.008 -0.008
VBF SR3	0.0171	0.0046	+0.0017 -0.0015
VBF SR4	0.0021	0.0016	+0.0005 -0.0003
VBF-depleted region	2.41	0.08	+0.11 -0.11
ZZ^* CR	6.20	0.12	+0.31 -0.30

Appendix E

Interference with CP-even couplings

The CP violation searching in this study focuses on the measurement of the BSM CP-odd Higgs boson to Z bosons couplings. There are potential CP-even BSM couplings in the HZZ interaction. These couplings are not considered in the measurement. The check has been carried out for investigating the impact of the CP-even BSM couplings on the CP-odd coupling measurements. In this complimentary check, the hypothetical models are constructed without CP-even couplings. Different CP-even BSM couplings are added to the pseudo data fit on these models. The CP-even couplings tested include $c_{H\widetilde{W}} = \pm 1.5$, $c_{H\widetilde{B}} = \pm 0.5$ and $c_{H\widetilde{W}B} = \pm 1.0$. These are the 1σ limits measured in the STXS study [156].

CP-even observables impact in $H \to ZZ^* \to 4l$ measurement

For the check of the impact of CP-even couplings in the $H \to ZZ^* \to 4l$ study, the hypothetical models are prepared with the 3D morphing basis described in B. The likelihood fits for the pseudo datasets of the SM expectation and with different CP-even BSM couplings and no CP-odd couplings are shown in Figure E.1. The CP-even couplings have decreased the sensitivity and shifted the best-fit values. The shift from the CP-even couplings does not exclude the SM hypothesis.

The CP-even impact on the BSM CP-odd coupling fit is shown in Figure E.2 and Figure E.3. These CP-odd hypotheses are with the coupling values at the expected 2σ limit. The CP-even couplings are expected to make the best-fit value larger for the positive CP-odd couplings datasets and smaller for the negative CP-odd couplings datasets. The shift due to CP-even couplings is not very large.

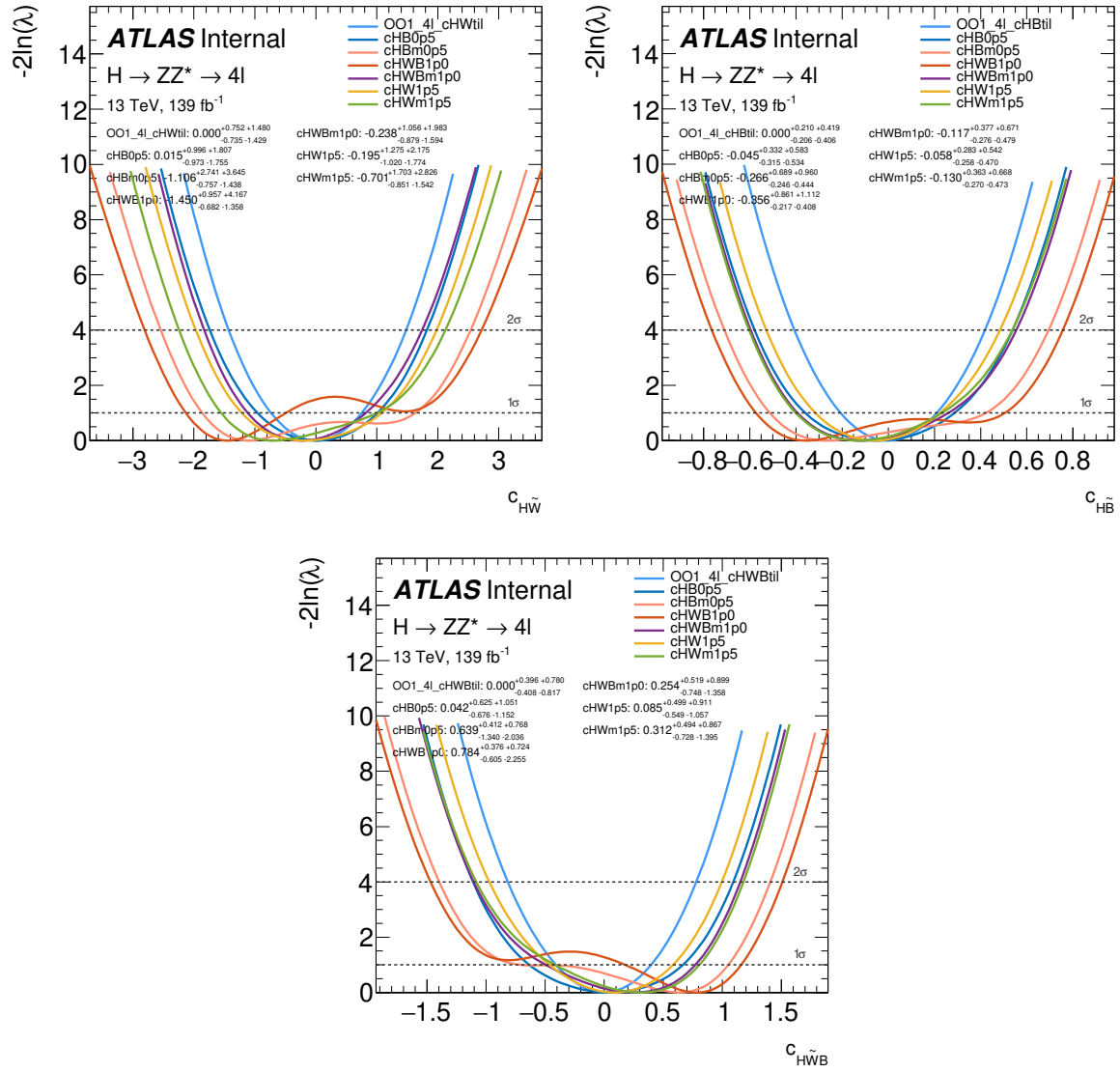


Figure E.1: $H \to ZZ^* \to 4l$ decay likelihood fit for the pseudo datasets of the SM expectation and with CP-even couplings and no CP-odd couplings to the ordinary hypothetical models from the morphing technique.

CP-even observables impact in VBF measurement

For the VBF production measurement, the impact of CP-even couplings was verified for the $\widetilde{c}_{zz}$ measurement in an earlier stage of the study. The hypothetical models are prepared with the 1D morphing basis with $\widetilde{c}_{zz}$ couplings $-4, -1.5, 0, 1.5, 4$. The comparison of $\widetilde{c}_{zz}$ likelihood for the datasets with and without CP-even couplings are shown in Figure E.4. The tests are done with the $\widetilde{c}_{zz}$ values of 0 and at expected $\pm 1\sigma$ limits. In the VBF production measurements, the presence of the BSM CP-even couplings does not cause significant bias to the bias-fit values.

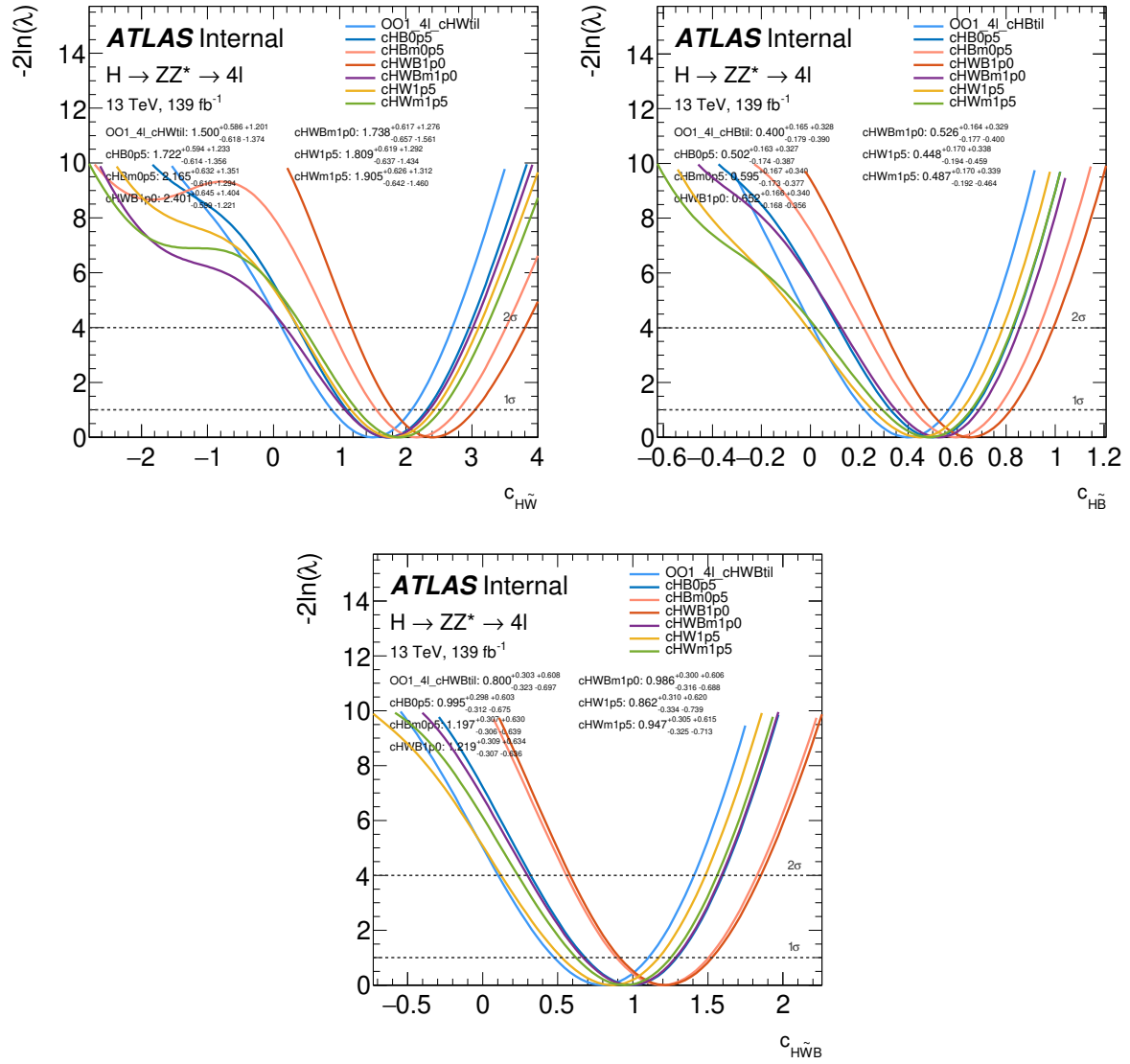


Figure E.2: $H \rightarrow ZZ^* \rightarrow 4l$ decay likelihood fit for the pseudo datasets with the CP-odd coupling values at expected $+2\sigma$ limit. The light blue curve is without CP-even couplings and the other curves are with additional BSM couplings.

The significance got smaller with the presence of BSM CP-even couplings.

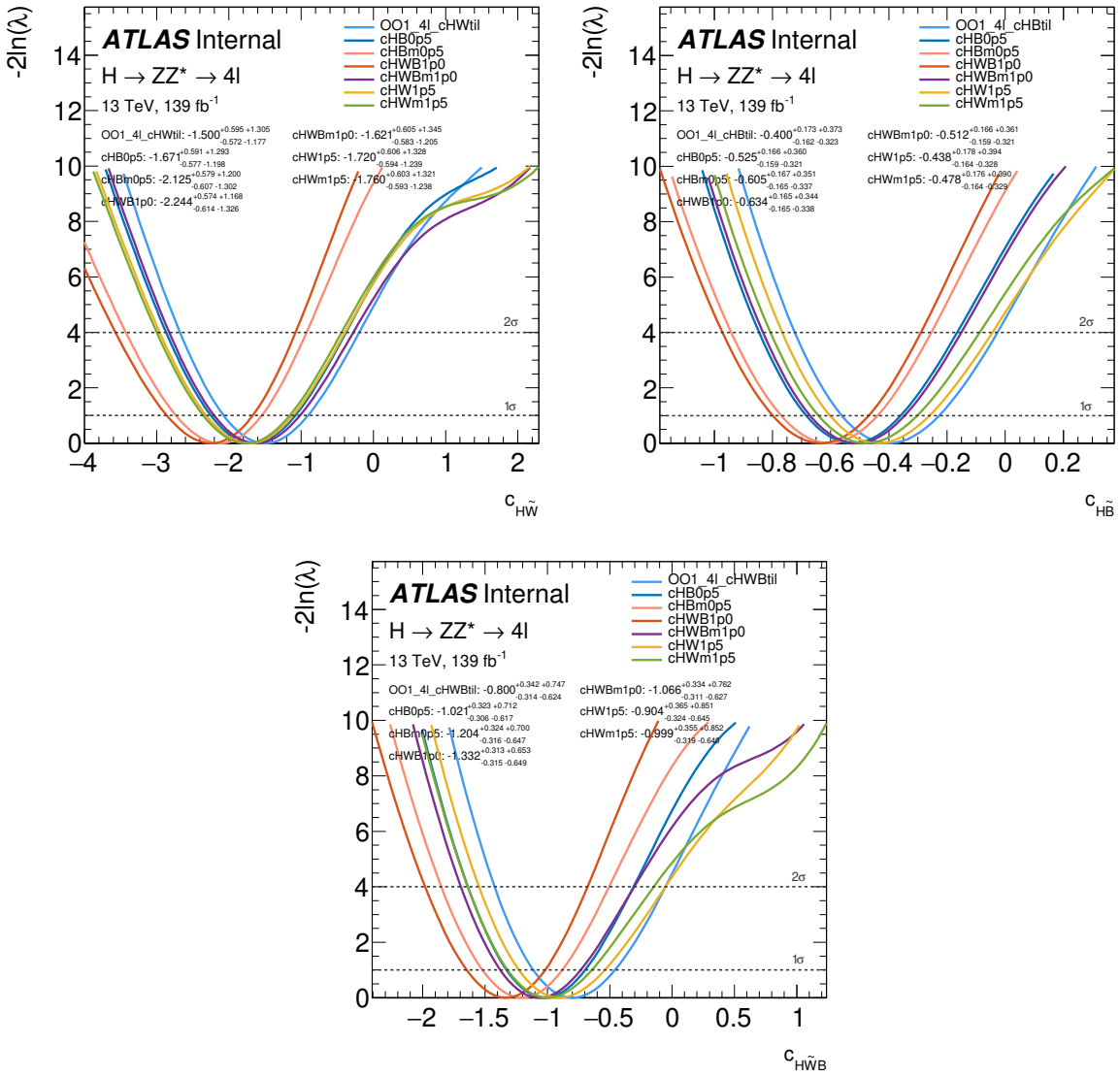


Figure E.3: $H \rightarrow ZZ^* \rightarrow 4l$ decay likelihood fit for the pseudo datasets with the CP-odd coupling values at expected -2σ limit. The light blue curve is without CP-even couplings and the other curves are with additional BSM couplings.

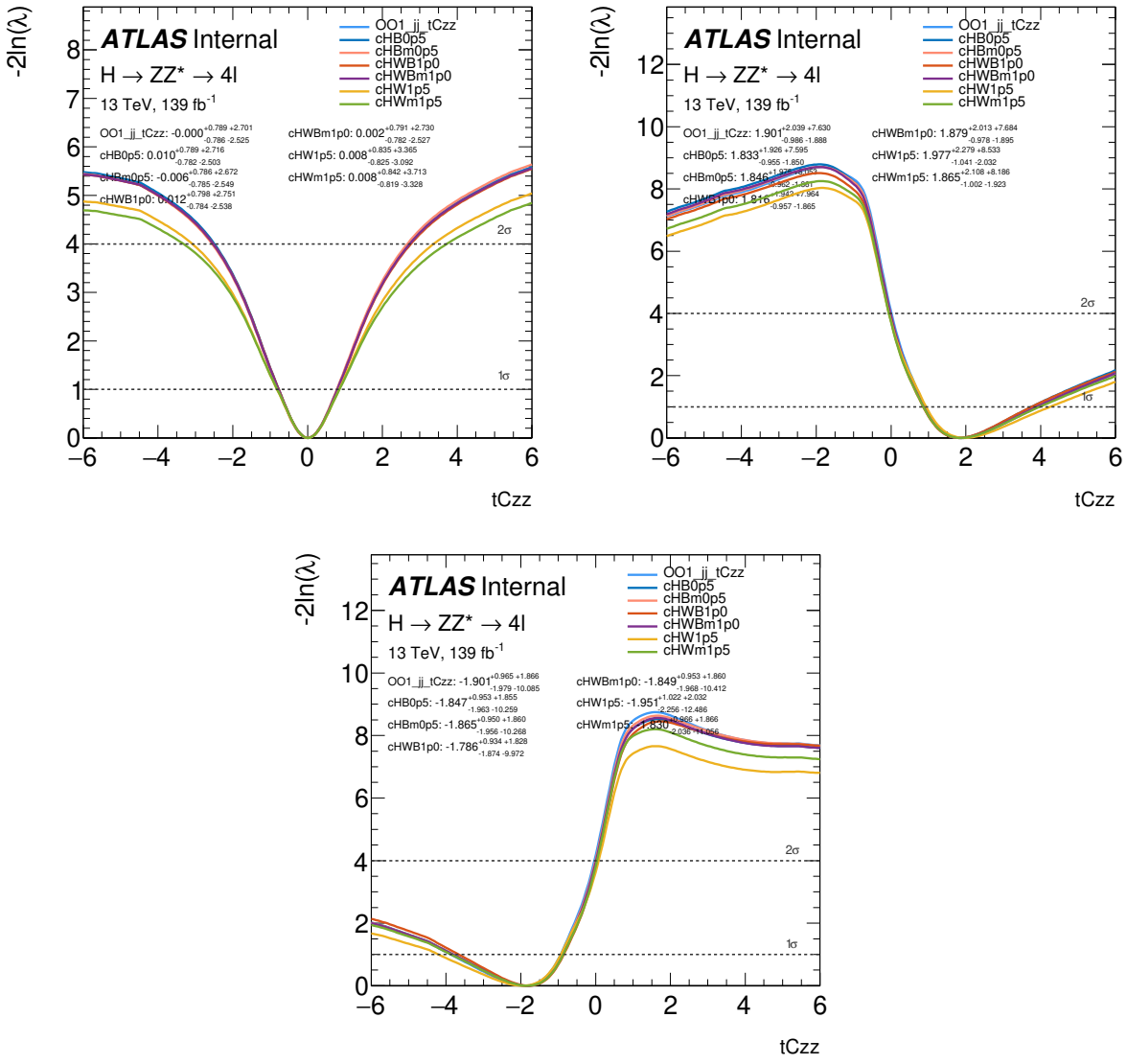


Figure E.4: VBF production likelihood fit for the pseudo datasets with the CP-odd coupling values for $\widetilde{c}_{zz} = 0$ and of values at $\pm 1\sigma$ limits. The light blue curve is without CP-even couplings and the other curves are with additional BSM couplings.

Appendix F

MC statistical uncertainties estimation

The statistical uncertainties in the MC simulation in this study are estimated through the bootstrapping method described in Section 6.7. In the likelihood fit for the MC statistical uncertainties test, the pseudo dataset is fit on 100 independent "MC toy" models for each coupling to measure and OO shape. In the VBF production measurements, the impacts on the $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ measurements are tested with $OO_{jj}^{\widetilde{c}_{zz}}$, $OO_{jj}^{\widetilde{c}_{za}}$ and $OO_{jj}^{c_{H\widetilde{W}}}$. In the $H \rightarrow ZZ^* \rightarrow 4l$ decay measurements, the impacts on Warsaw Basis couplings measurements are tested with $OO_{4\ell}^{\widetilde{c}_{aa}}$, $OO_{4\ell}^{c_{H\widetilde{W}}}$, $OO_{4\ell}^{c_{H\tilde{B}}}$ and $OO_{4\ell}^{c_{H\widetilde{W}B}}$ observables.

The $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ fit results with $OO_{jj}^{\widetilde{c}_{zz}}$ on MC toy models are shown in Figure F.1 as an example. The statistical uncertainties in the MC models do not introduce bias to the best-fit values in the VBF production measurements. The changes in the limits are small with the fit curves surrounding the scan on the original MC model. The $c_{H\widetilde{W}}$, $c_{H\tilde{B}}$ and $c_{H\widetilde{W}B}$ fit results with $OO_{4\ell}^{c_{H\widetilde{W}}}$ on the MC toy models are shown in Figure F.2 as an example. Similar to the VBF production test, the statistical uncertainties introduce no bias in the best-fit values and only small changes in the limits around the fit on the original MC model. The statistical uncertainties impact estimations with the other couplings show similar results.

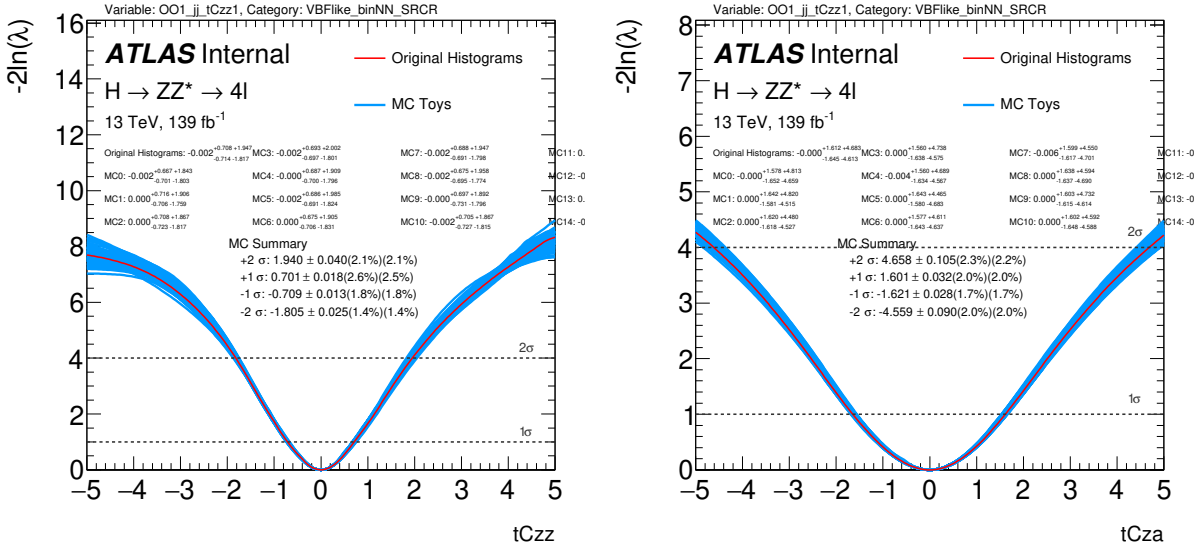


Figure F.1: Impact of the statistical uncertainties on the fit for VBF production measurement. The red curve is the nominal fit. The blue curves are the MC toys fit.

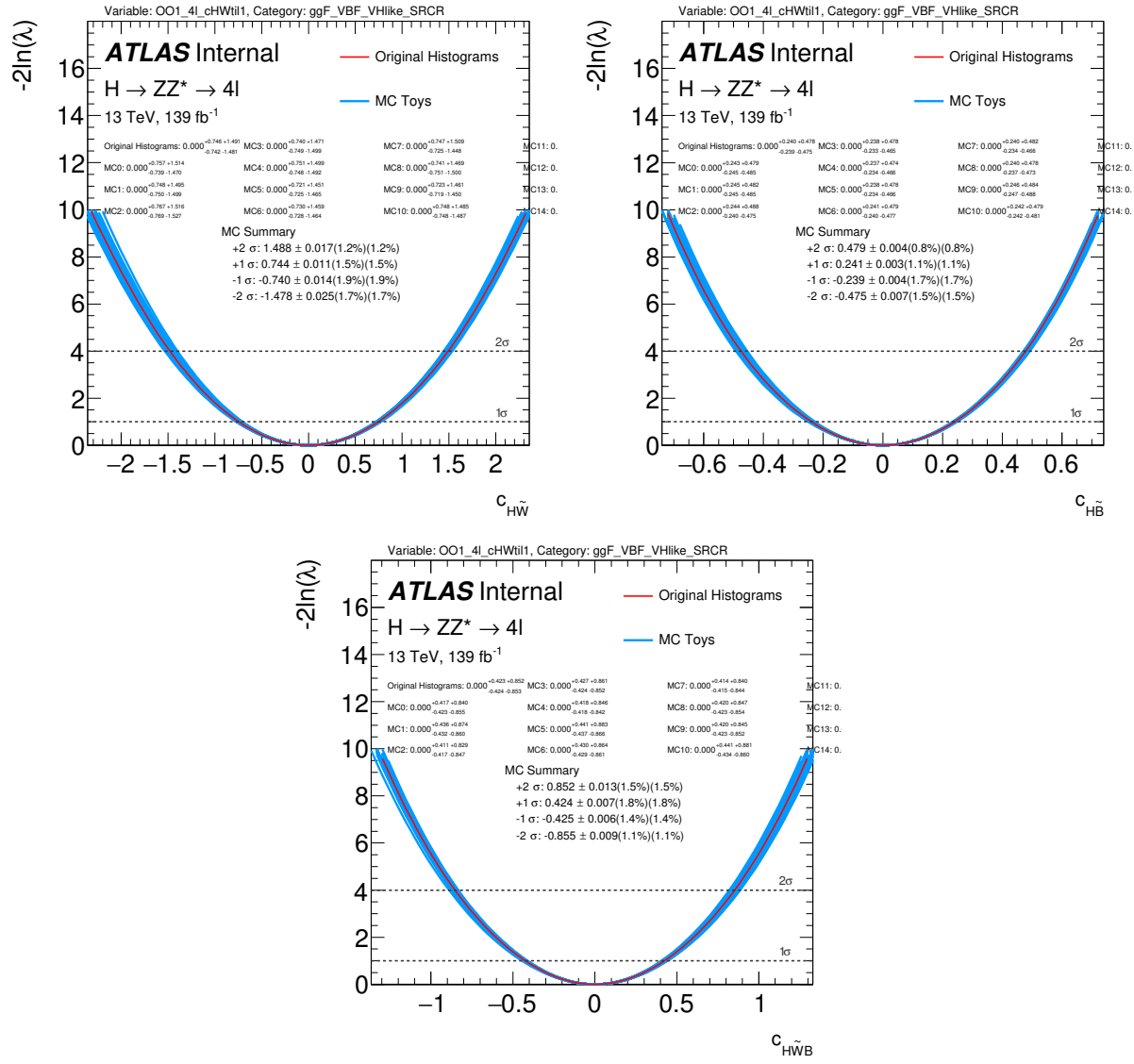


Figure F.2: Impact of the statistical uncertainties on the fit for four leptons decay measurement. The red curve is the nominal fit. The blue curves are the MC toys fit.

Appendix G

Higher order effects evaluation

In this study, the BSM samples with CP-odd couplings are generated by the SMEFT model with the MADGRAPH5_AMC@NLO generator [127]. The VBF Higgs boson production is modeled with only leading order QCD accuracy in this model. The missing higher-order QCD modeling is one of the dominant sources of systematical uncertainties. This is an additional complimentary check, where the missing higher-order effects are tested directly from the likelihood fit comparison with and without the NLO QCD modeling in the VBF measurements. In this test, the generator-level NLO VBF samples were prepared with the MADGRAPH5_AMC@NLO using the Higgs Characterisation model (HC) HC_NLO_X0_UFO [160]. The HC model is a framework describing the Higgs boson coupling properties. For the HZZ couplings, the HC provides equivalent modeling with the SMEFT couplings. The HC parameters are translated from EFT couplings [170]. This model allows the VBF modeling at NLO QCD accuracy.

The PDF4LHC was used in the generation through the ATLAS fragment MadGraph_NNPDF30NLO_Base_Fragment [124]. The MadSpin was used to model the decay process since the multi-body decay is not supported at the NLO generation [171]. The signal samples in this check are generated with only VBF processes only. The VH contribution came from the SM samples. The dipole showering option is used in this check, which is used in the generation of the SM POWHEG samples [172, 173]. This option eliminates the hadronization between the initial state and final state quarks. To control the same setup, the LO reference samples were generated with the HC_NLO_X0_UFO model and the same PDF set. The Higgs boson decay was modeled by MadSpin and with the dipole showering option as well.

In this check, the categorization took an older categorization scheme. The neural network score used was the same as the measurement. The events were split into 3 categories by NN score partition $0.0, 0.75, 0.9, 1.0$. The NN partitions of these samples are shown in Figure G.1. The LO generation had a similar partition ratio to the SM prediction. The NLO generation got a

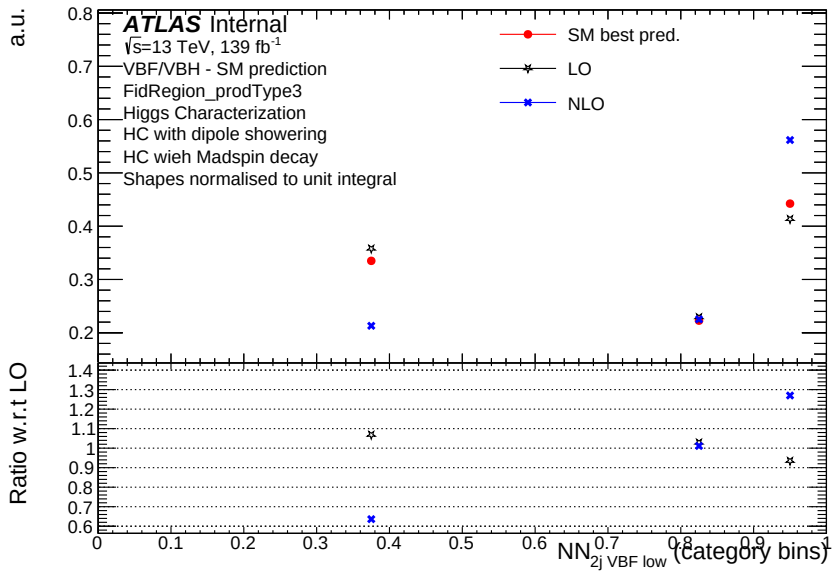


Figure G.1: The normalized neural network score NN_{VBF} ($2j$ Low) distributions for the LO and NLO HC generation and SM prediction comparison.

difference of about 40% in the first and last categories. The difference from the modeling leads to the need for separation normalization factor floating in NN categories.

The optimal observable distributions for NLO, LO generated samples and the SM expected POWHEG samples are shown in Figure G.2, G.3 and G.4 in VBF SR1, SR2 and SR3 respectively. The SM NLO and LO MADGRAPH5_AMC@NLO samples both have the OO distributions similar to the SM expectation from POWHEG.

In this study, the NLO impact on the variable distributions in BSM samples is checked as well. The LO and NLO NN score and OO distributions are compared on the samples with large $\widetilde{c}_{zz}$ and $\widetilde{c}_{z\gamma}$ values. The samples checked are with couplings $\widetilde{c}_{zz} = 4.98$ and $\widetilde{c}_{z\gamma} = 2.52$. The normalized NN split distributions are shown in Figure G.5. The green label shows the reweighed NN distribution for the NLO sample. The event weights are reweighted by the NLO-to-LO yield ratio in the SM samples shown in Figure G.1. The reweighting eliminates the NLO effect on the SM samples and presents the NLO effect on the BSM corrections. The distribution of the SMEFT sample used in the measurement is shown as a reference. There was a good agreement in the SMEFT sample, HC LO sample and reweighted NLO sample. This indicates the NLO accuracy in the VBF channel causes little variation in the BSM effect on the NN splitting.

The $OO_{jj}^{\widetilde{c}_{zz}}$ distributions for each VBF SR are shown in Figure G.6. In each NN category, the $OO_{jj}^{\widetilde{c}_{zz}}$ distributions had a good agreement between the NLO and LO HC samples and the referenced SMEFT sample. In populated bins of the inclusive VBF-enriched region, the reweighting process made the NLO distribution closer to the LO generation. The $OO_{jj}^{\widetilde{c}_{za}}$ distributions for

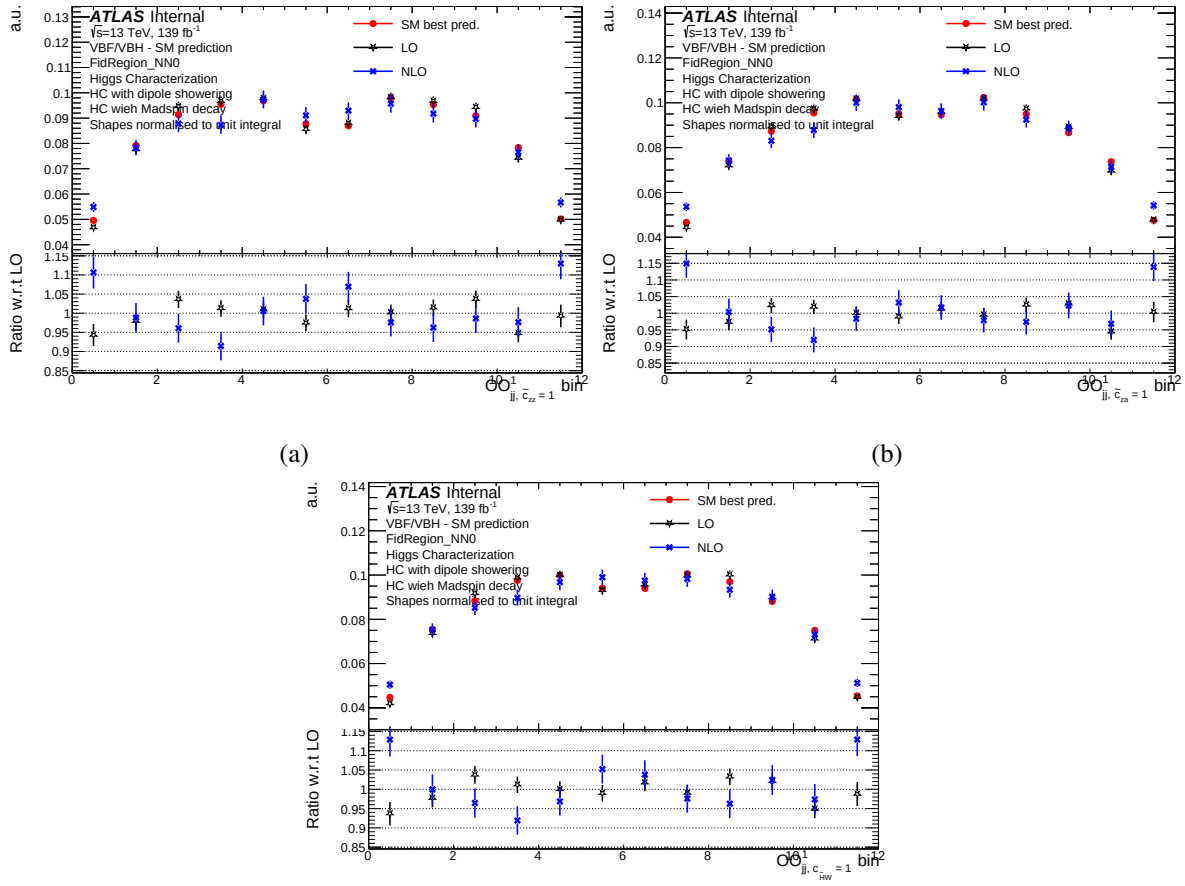
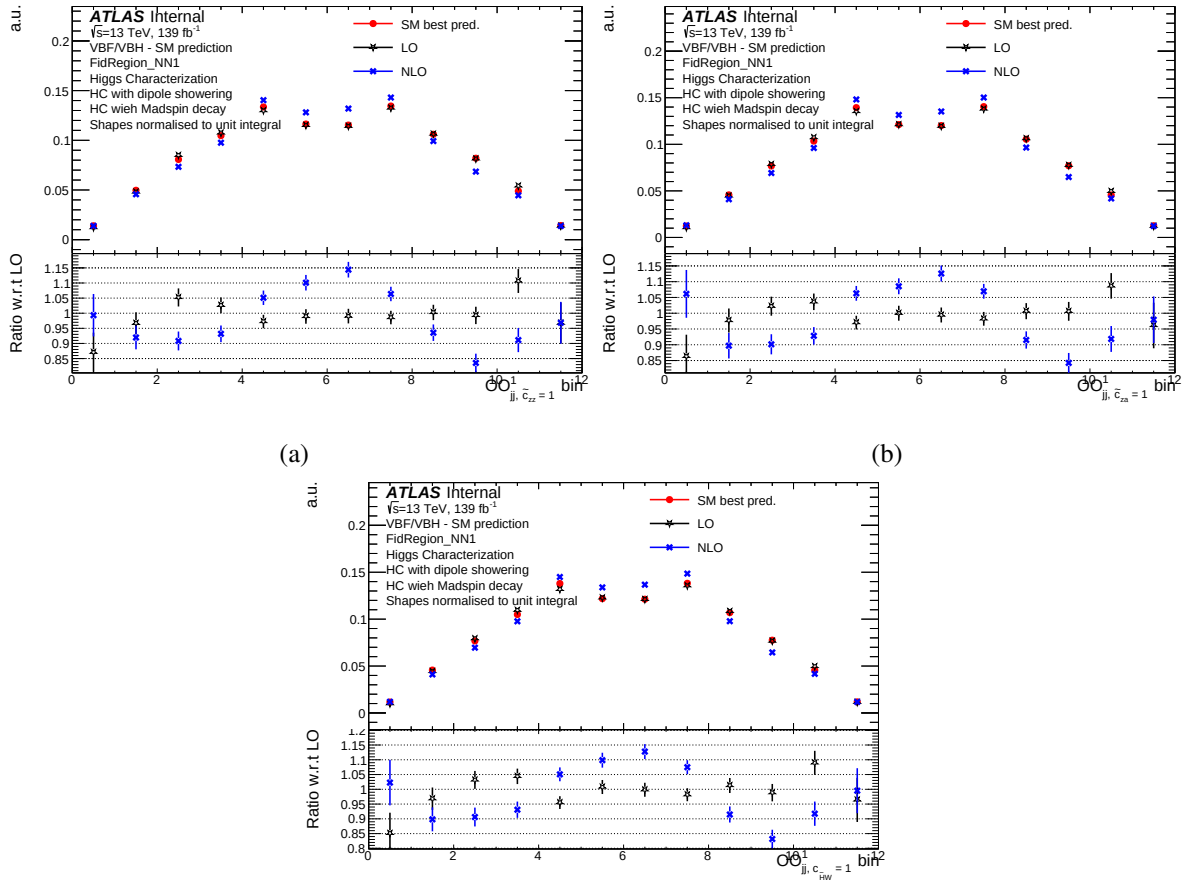


Figure G.2: The normalized $OO^{\tilde{c}_{zz}}_{jj}$, $OO^{\tilde{c}_{za}}_{jj}$, $OO^{c_{H\widetilde{W}}}_{jj}$ distributions for the LO and NLO HC samples and SM POWHEG prediction comparison in VBF SR1.

each VBF SR are shown in Figure G.7. The $OO^{\tilde{c}_{za}}_{jj}$ distributions for the LO and NLO samples and the SMEFT sample are in good agreement as well.

The LO and NLO $OO^{c_{H\widetilde{W}}}_{jj}$ distributions for the sample with large $c_{H\widetilde{W}}$ are compared as well. The samples checked are with the coupling $c_{H\widetilde{W}} = 4.00$. The NN partitions are shown in Figure G.8. The green label shows the NN distribution for the NLO sample reweighed by the SM difference Figure G.1 as well. There was a good agreement between the HC LO sample and the reweighted NLO sample. The NLO accuracy in the VBF channel causes little variation in the BSM effect on the NN splitting in the Warsaw basis as well. The $OO^{c_{H\widetilde{W}}}_{jj}$ distributions for each VBF SR are shown in Figure G.9. The $c_{H\widetilde{W}}$ distributions for the LO and NLO samples are in good agreement.

The likelihood fit comparisons for the NLO and LO samples are shown in Figure G.10. The differences between the LO expectation and NLO expectation are all negligible at the 1 σ limit

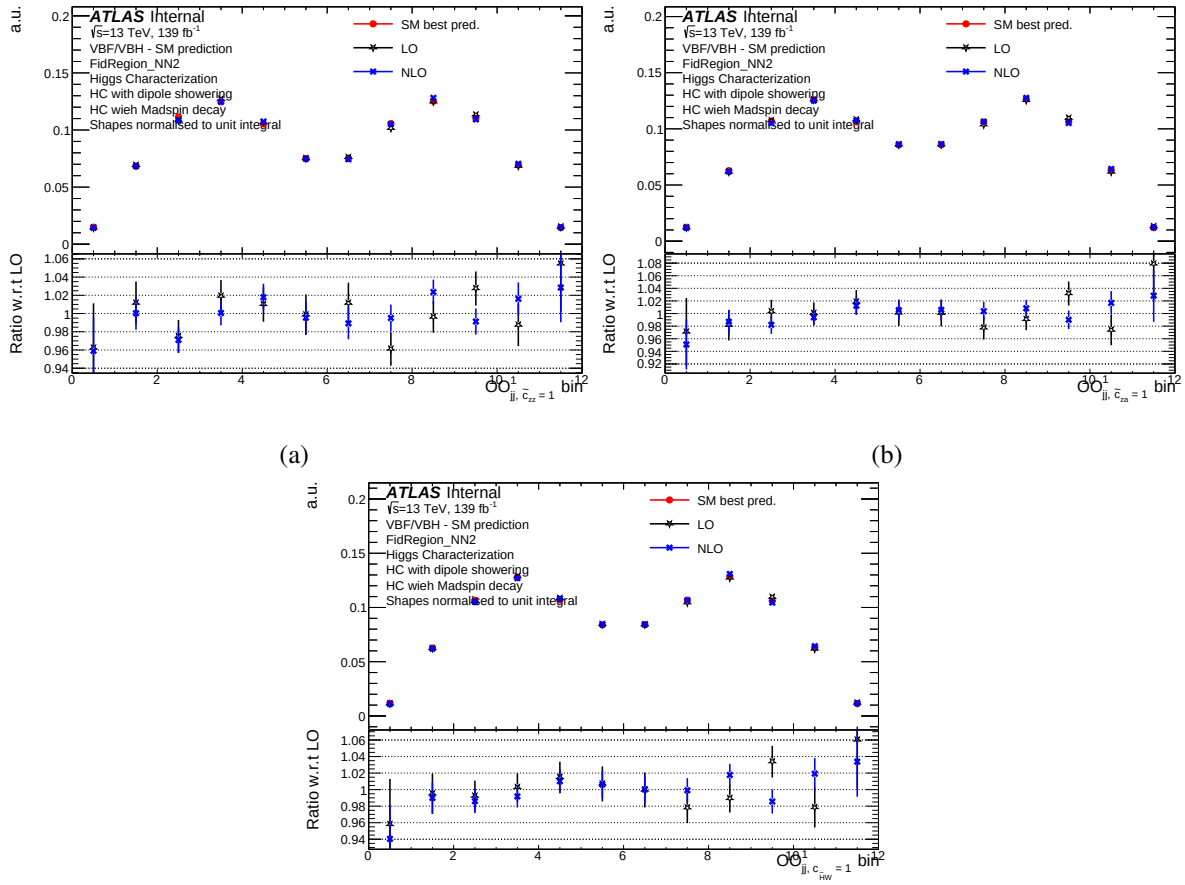


(a) (b)

(c)

Figure G.3: The normalized $OO^{\tilde{c}_{zz}}_{jj}$, $OO^{\tilde{c}_{za}}_{jj}$, $OO^{c_{H\widetilde{W}}}_{jj}$ distributions for the LO and NLO HC samples and SM POWHEG prediction comparison in VBF SR2.

in the plots. The small difference in the limits indicates the impact on the measurement from the NLO correction is negligible.



(a) (b)

(c)

Figure G.4: The normalized $OO^{\tilde{c}_{zz}}_{jj}$, $OO^{\tilde{c}_{za}}_{jj}$, $OO^{c_{H\widetilde{W}}}_{jj}$ distributions for the LO and NLO HC samples and SM POWHEG prediction comparison in VBF SR3.

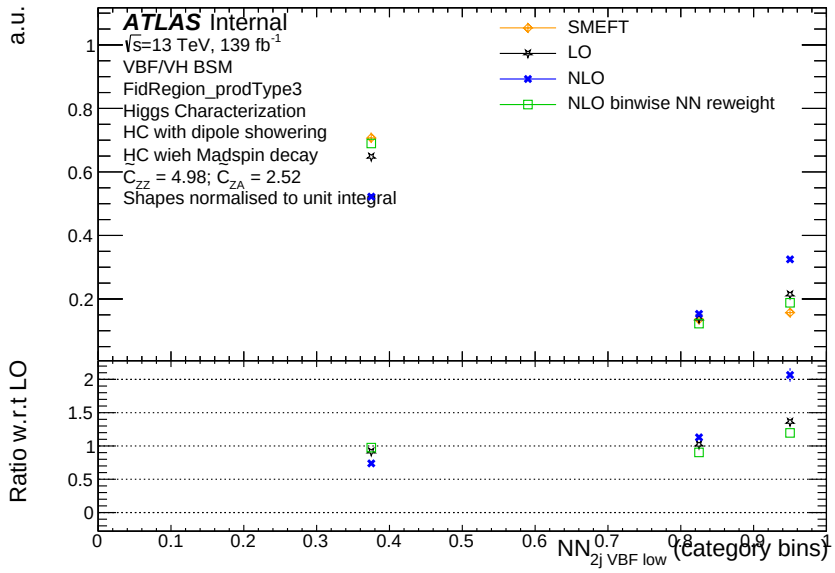


Figure G.5: The normalized NN_{VBF} ($2j$ Low) variables distributions for the LO and NLO HC samples with $\widetilde{c}_{zz} = 4.98$ and $\widetilde{c}_{z\gamma} = 2.52$. The distribution of the SMEFT sample used in the measurement is shown as a reference.

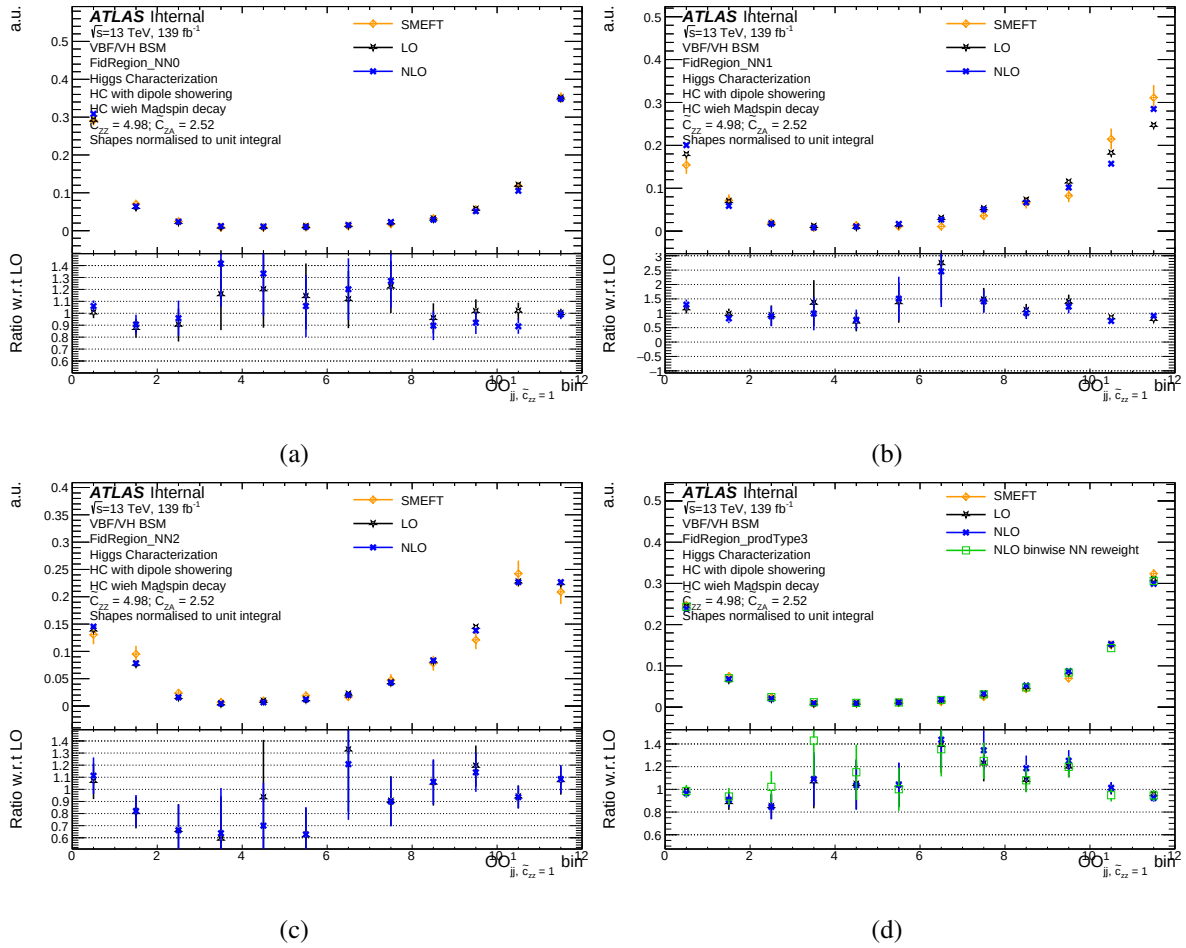


Figure G.6: The normalized $\widetilde{c}_{zz}$ distributions for the LO and NLO HC samples with $\widetilde{c}_{zz} = 4.98$ and $\widetilde{c}_{z\gamma} = 2.52$ in each VBF SR and the inclusive VBF-enriched region. The distribution of the SMEFT sample used in the measurement is shown as a reference.

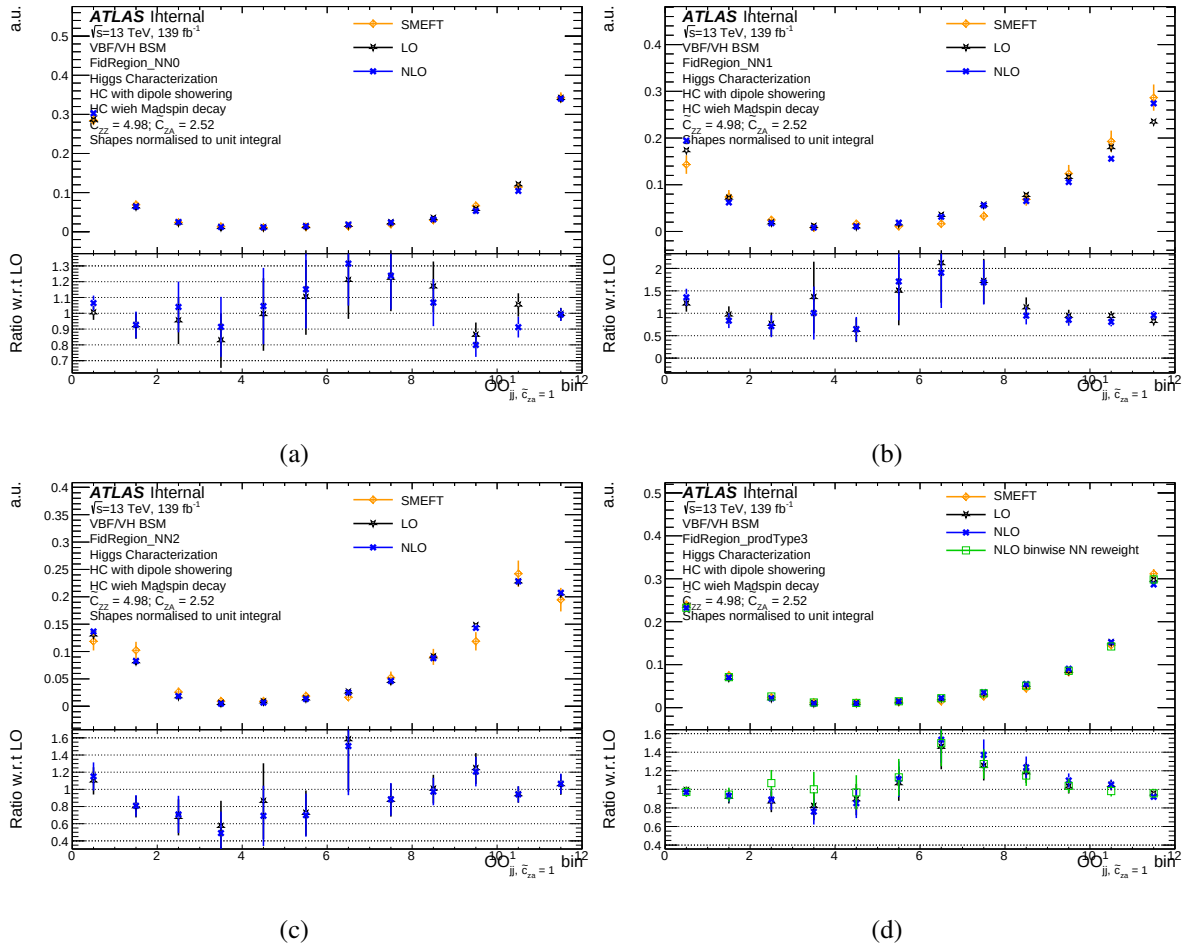
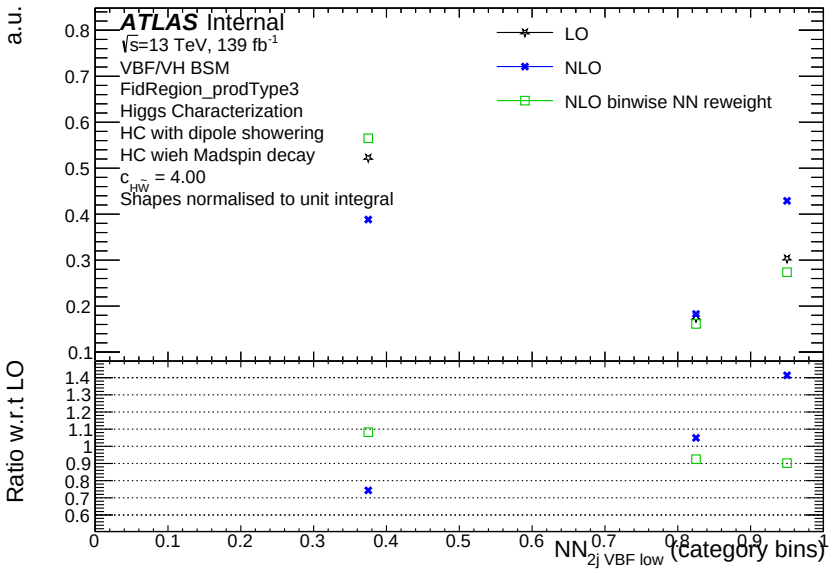


Figure G.7: The normalized $\widetilde{c}_{z\gamma}$ distributions for the LO and NLO HC samples with $\widetilde{c}_{zz} = 4.98$ and $\widetilde{c}_{z\gamma} = 2.52$ in each VBF SR and the inclusive VBF-enriched region. The distribution of the SMEFT sample used in the measurement is shown as a reference.



(a)

Figure G.8: The normalized NN_{VBF} ($2j$ Low) variables distributions for the LO and NLO HC samples with $c_{H\widetilde{W}} = 4.00$.

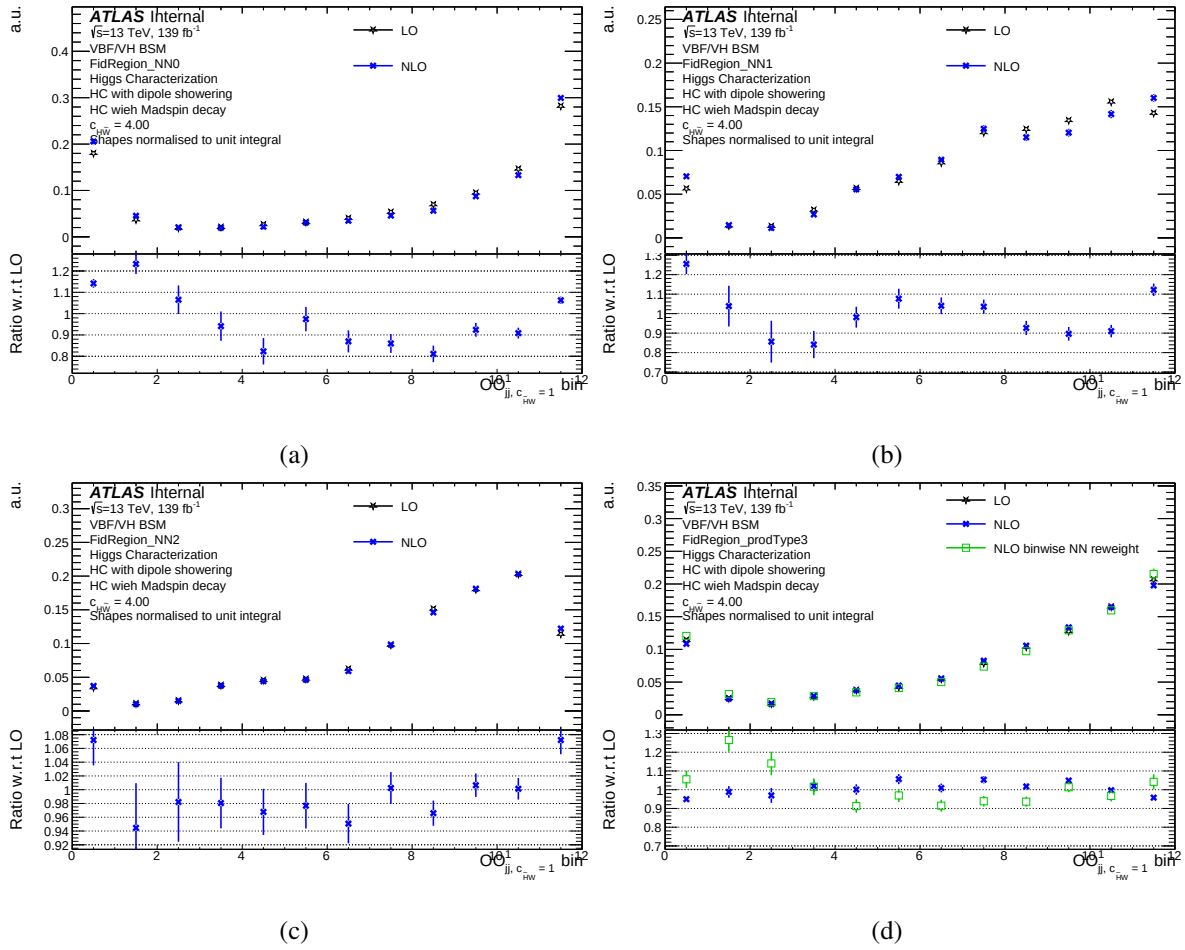


Figure G.9: The normalized $OO_{jj}^{c_{H\widetilde{W}}}$ distributions for the LO and NLO HC samples with $c_{H\widetilde{W}} = 4.00$ in each VBF SR and the inclusive VBF-enriched region.

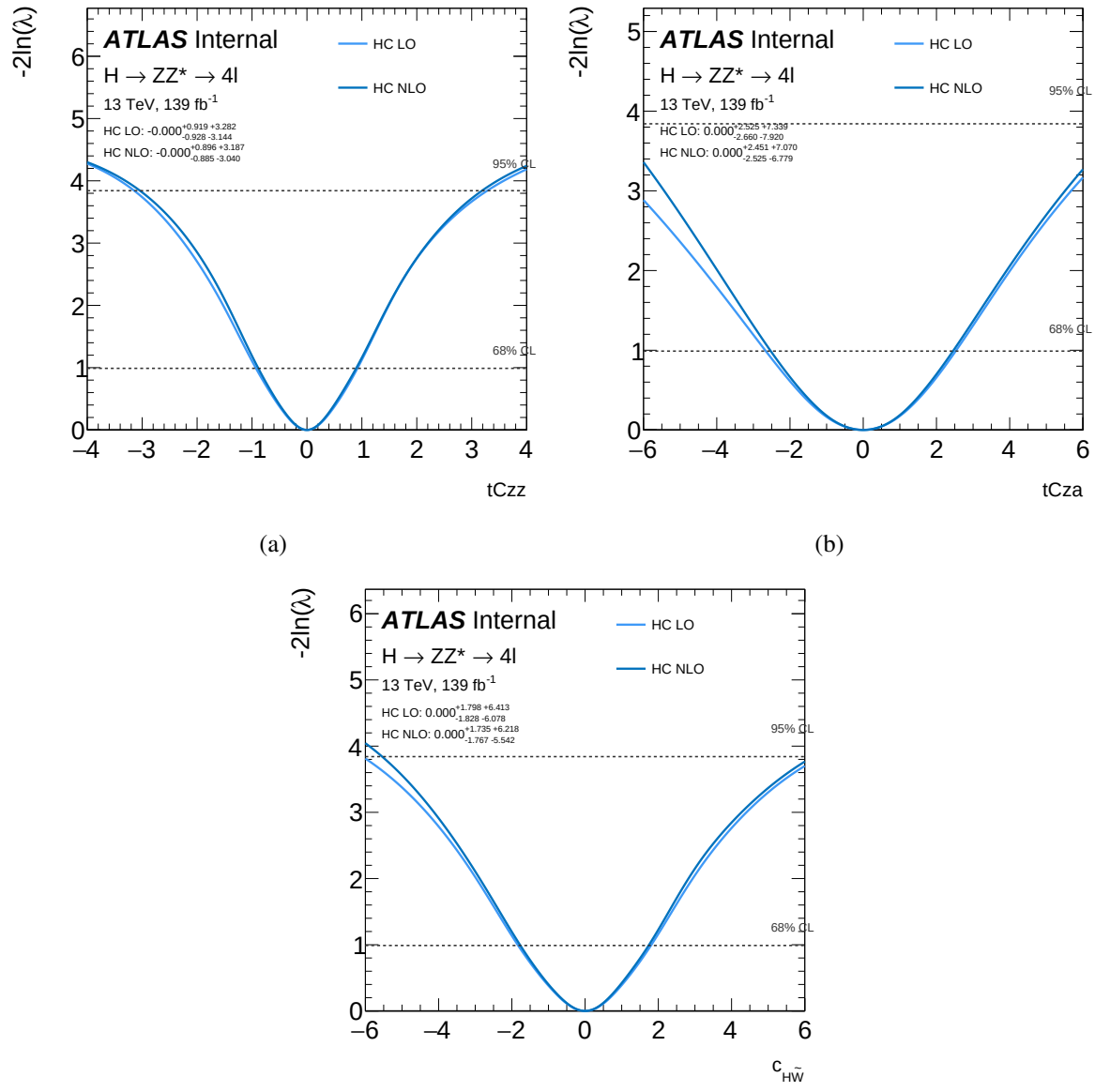


Figure G.10: The likelihood fit comparison of NLO and LO HC samples for $OO_{jj}^{\widetilde{c}_{zz}}$, $OO_{jj}^{\widetilde{c}_{za}}$ and $OO_{jj}^{c_{H\widetilde{W}}}$ along $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$ and $c_{H\widetilde{W}}$ axes respectively.

Appendix H

Observable distributions in data

The $OO_{jj}^{\widetilde{c}_{za}}$, $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4jj}^{\widetilde{d}}$ distributions of the ATLAS Run 2 data and the SM expectation in the NN categories are shown in H.1, H.2 and H.3. The distributions of these observables have similar patterns with the $OO_{jj}^{\widetilde{c}_{zz}}$ distributions. There are upward fluctuations on the right bins of the VBF SR1 for all these observables.

The $OO_{4\ell}^{c_{H\tilde{B}}}$, $OO_{4\ell}^{c_{H\widetilde{W}B}}$ and $OO_{4\ell}^{\widetilde{d}}$ distributions of the ATLAS Run 2 data and the SM expectation in the inclusive Higgs boson signal region are shown in H.4 [19]. The $OO_{4\ell}^{\widetilde{c}_{za}}$ and $OO_{4\ell}^{\widetilde{c}_{aa}}$ distributions are shown in H.5.

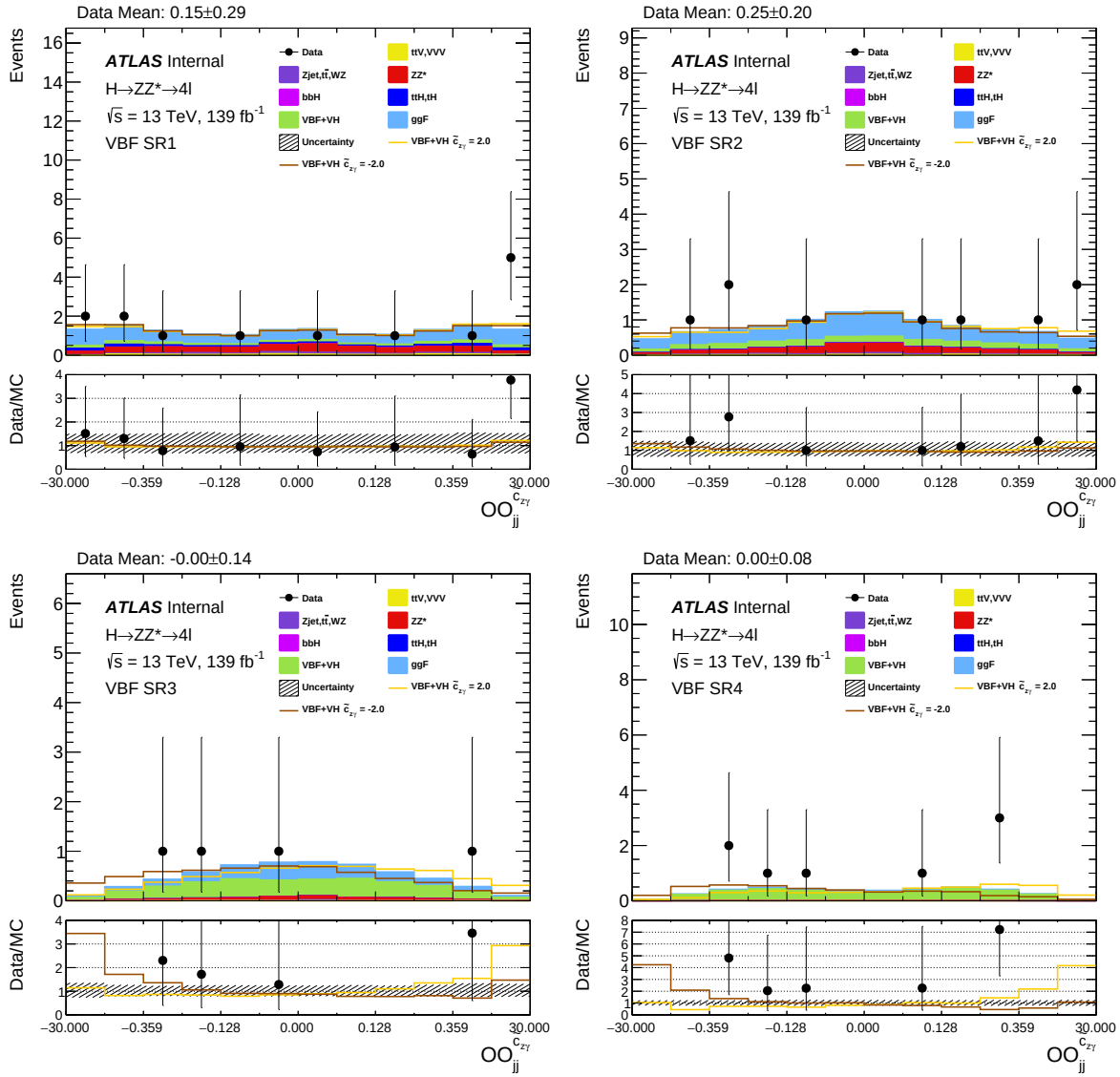


Figure H.1: $OO_{jj}^{\tilde{c}_{z\gamma}}$ distributions for the VBF-like NN signal regions of production side measurement. The dot represents unblinding data with statistical uncertainties. The green block represents the VBF/VH signal MC distributions. The other stacks represent different MC background components. The dashed area is the total uncertainties for the MC simulation.

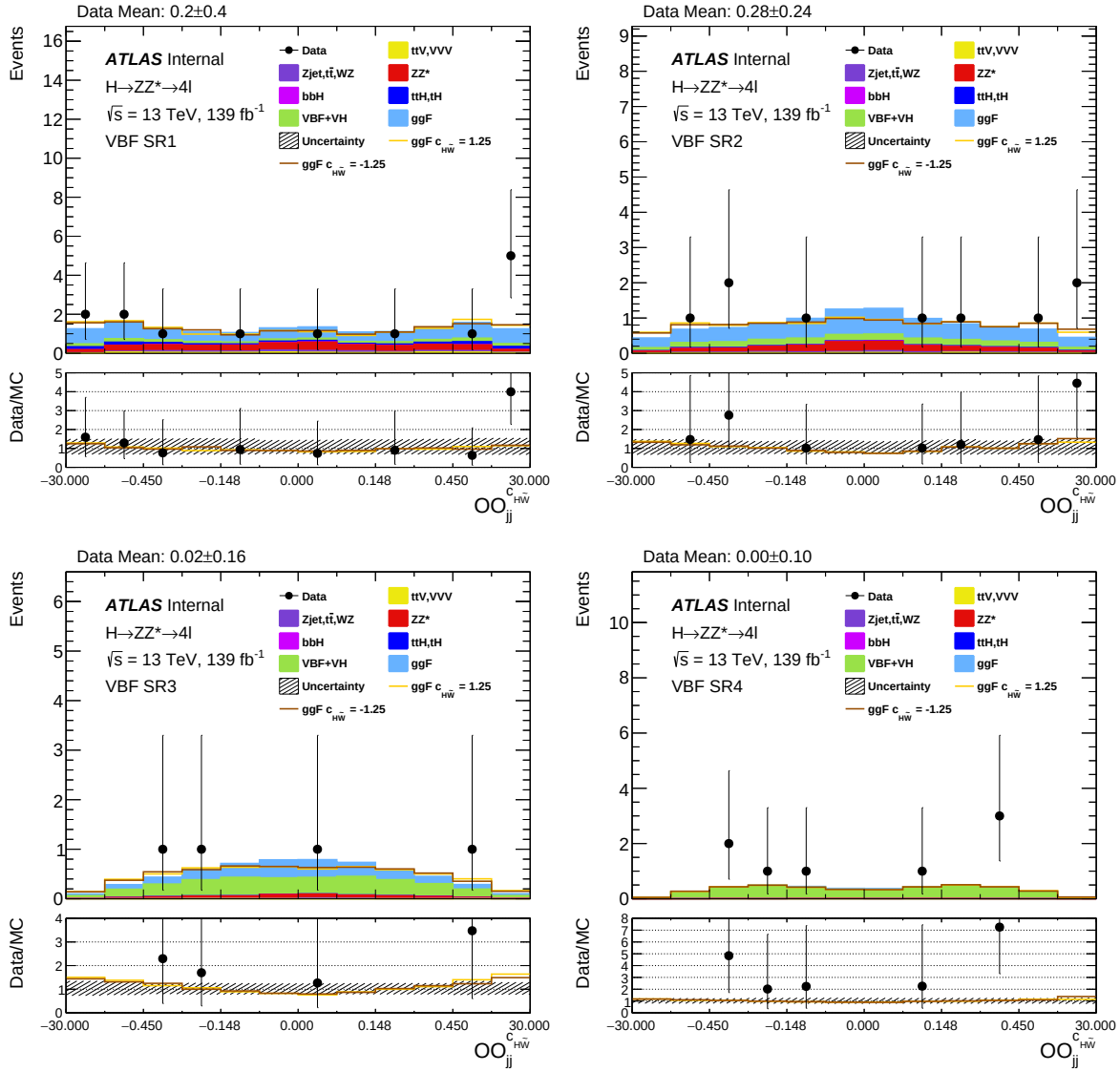


Figure H.2: $OO_{jj}^{c_{H\widetilde{W}}}$ distributions for the VBF-like NN signal regions of production side measurement. The dot represents unblinding data with statistical uncertainties. The green block represents the VBF/VH signal MC distributions. The other stacks represent different MC background components. The dashed area is the total uncertainties for the MC simulation.

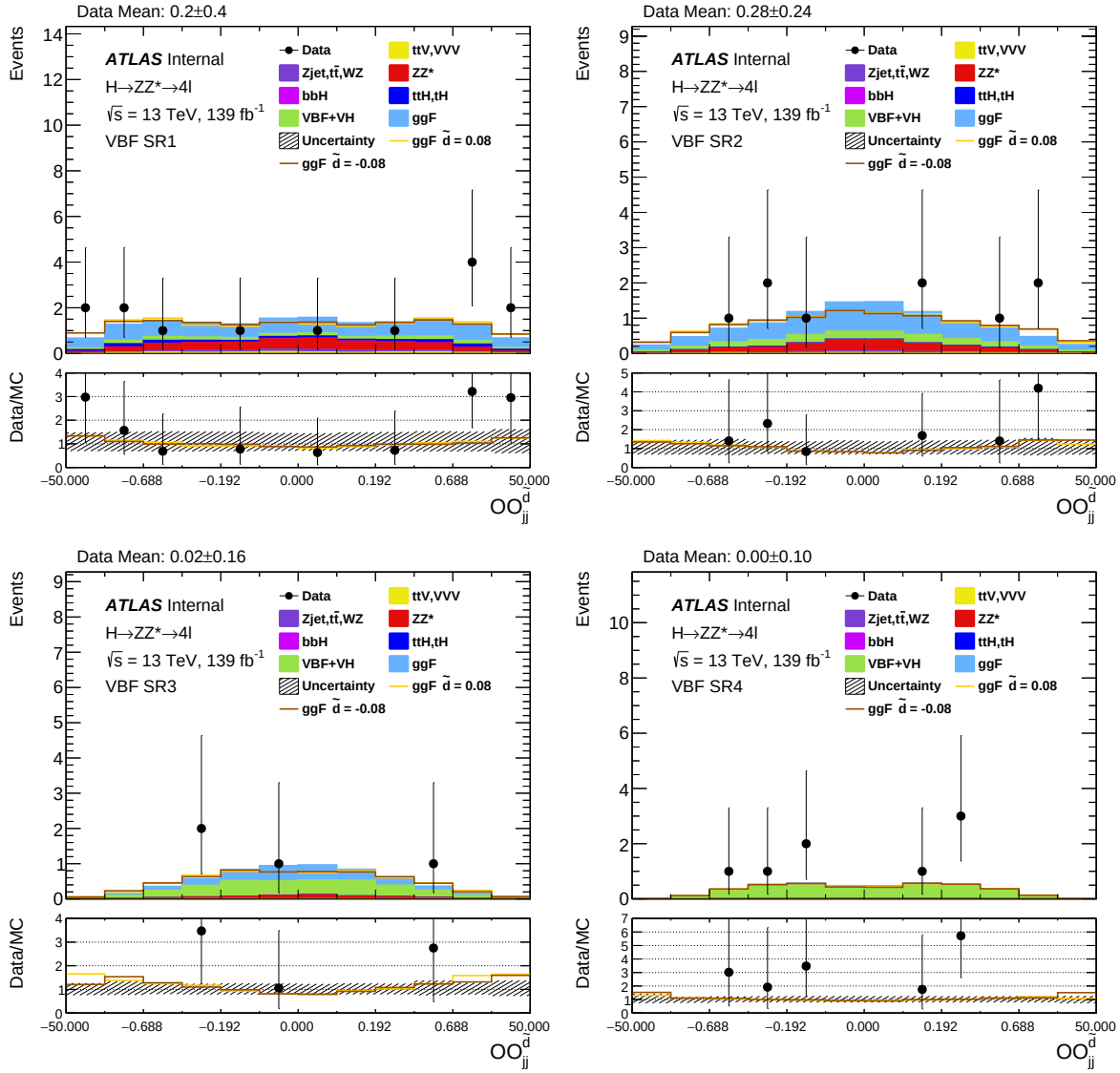


Figure H.3: $OO^{\tilde{d}}_{4jj}$ distributions for the VBF-like NN signal regions of production side measurement. The dot represents unblinding data with statistical uncertainties. The green block represents the VBF/VH signal MC distributions. The other stacks represent different MC background components. The dashed area is the total uncertainties for the MC simulation.

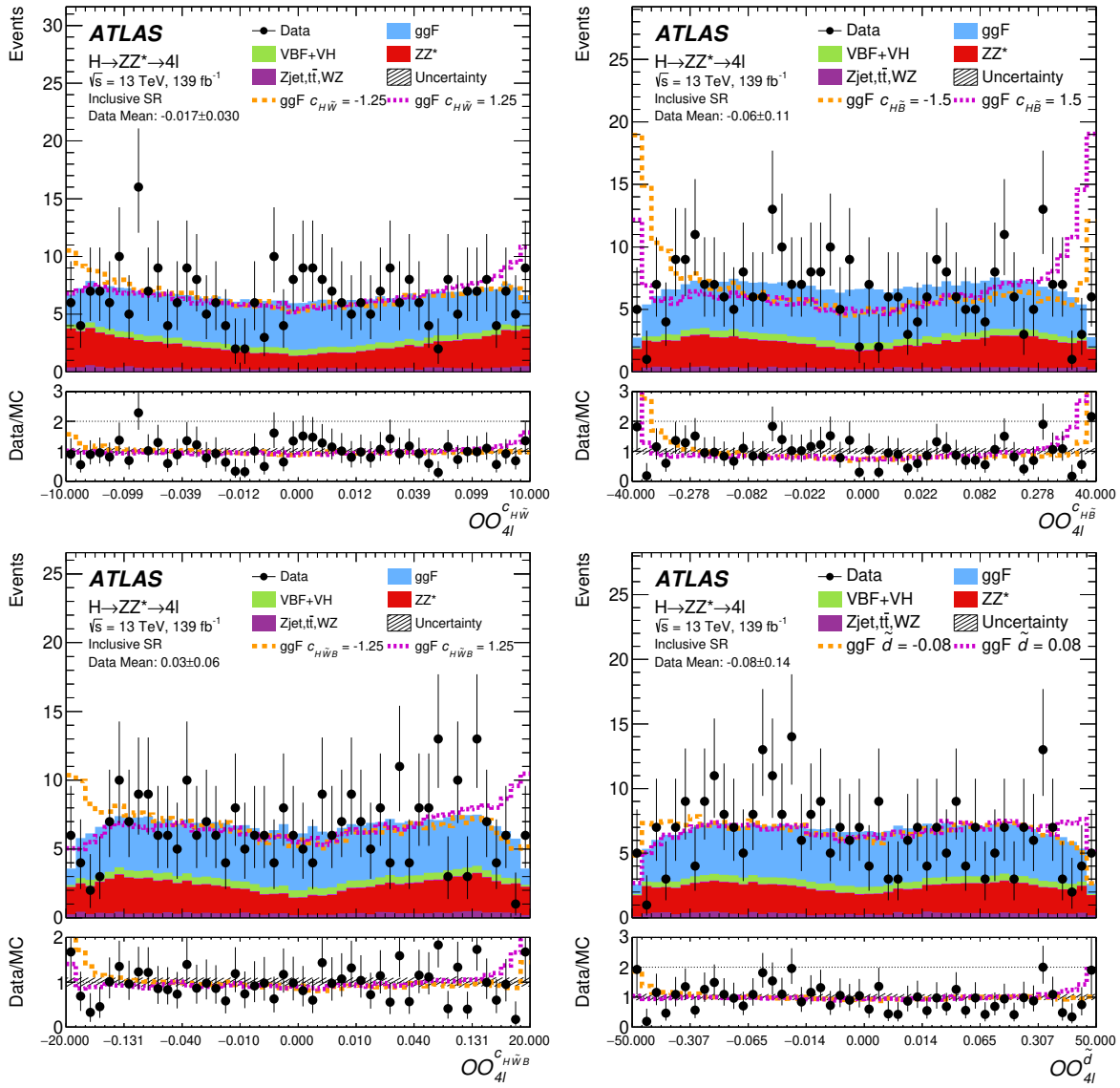


Figure H.4: $OO_{4\ell}^{c_{H\widetilde{W}}}$, $OO_{4\ell}^{c_{H\widetilde{B}}}$, $OO_{4\ell}^{c_{H\widetilde{W}B}}$ and $OO_{4\ell}^{\widetilde{d}}$ distributions for the Higgs boson signal regions of decay side measurement. The dot represents unblinding data with statistical uncertainties. The light blue blocks represent the dominated ggF signal MC components. The green blocks represent the VBF/VH signal MC components. The red blocks represent the non-resonant ZZ^* background MC components. The other stacks represent the Higgs boson signal and non-Higgs boson background MC components of small shares. The dashed area is the total uncertainties for the MC simulation.

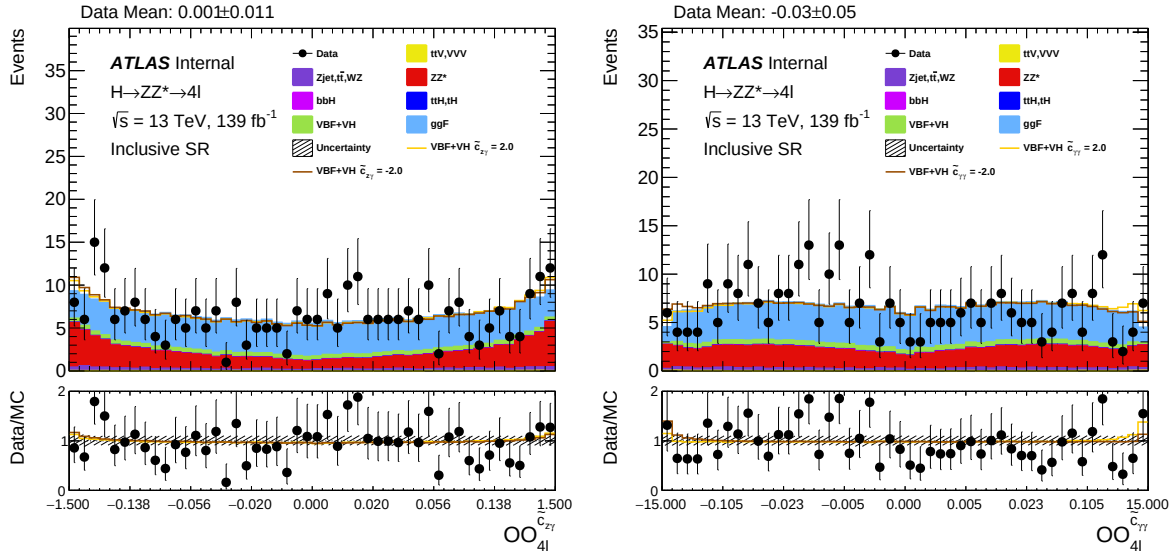


Figure H.5: $OO_{4\ell}^{\widetilde{c}_{za}}$ and $OO_{4\ell}^{\widetilde{c}_{aa}}$ distributions for the Higgs boson signal regions of decay side measurement. The dot represents unblinding data with statistical uncertainties. The light blue blocks represent the dominated ggF signal MC components. The green blocks represent the VBF/VH signal MC components. The red blocks represent the non-resonant ZZ^* background MC components. The other stacks represent the Higgs boson signal and non-Higgs boson background MC components of small shares. The dashed area is the total uncertainties for the MC simulation.

Appendix I

Likelihood measurement on data

The measurement of the couplings $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$ on the VBF production vertex are shown in Figure I.1. These likelihood fit curves on the ATLAS Run 2 data all have a positive shift in the best-fit values. This is consistent with the upward fluctuations in the right boundary bins in the data distribution shown in Appendix H. The results are consistent with the expectations of the Standard Model.

The measurement of the couplings $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ and $\widetilde{d}$ on the $H \to ZZ^* \to 4l$ decay vertex are shown in Figure I.2 [19]. The $\widetilde{c}_{z\gamma}$ and $\widetilde{c}_{\gamma\gamma}$ measurements are shown in Figure I.3 [19]. The measurement results are consistent with the Standard Model. The $c_{H\widetilde{W}}$ measurement on the ATLAS Run 2 data has a broader limit compared to the expectation. This can be explained by the upward fluctuations in the central part of the $OO_{4\ell}^{c_{H\widetilde{W}}}$ distribution. The measurements for the other couplings get better sensitivities compared to the expectations. This is consistent with the upwards fluctuation in the boundary bins.

The measurement of the couplings $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$ in the production and decay vertices combined are shown in Figure I.4. The measurement results all have a shift to right due to the contribution from the production OO distributions. The $c_{H\widetilde{W}}$ measurement shows a stronger impact from the $OO_{jj}^{c_{H\widetilde{W}}}$. The sensitivity is expected to be better than the single vertex measurements.

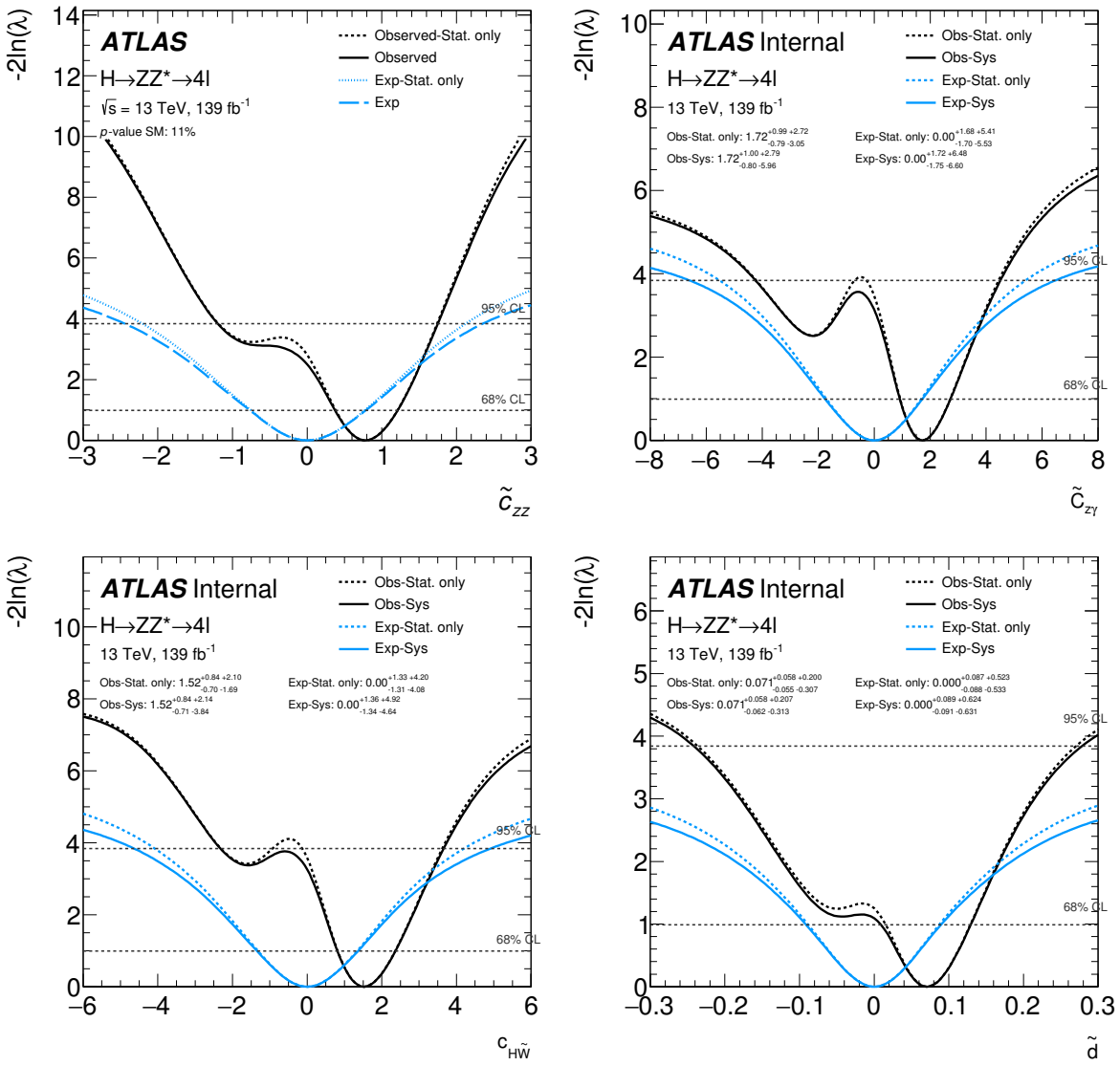


Figure I.1: Likelihood scans along $\widetilde{c}_{zz}$, $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$ axes on the SM expectation and ATLAS Run 2 data with $OO_{jj}^{\widetilde{c}_{zz}}$, $OO_{jj}^{\widetilde{c}_{za}}$, $OO_{jj}^{c_{H\widetilde{W}}}$ and $OO_{4jj}^{\widetilde{d}}$ respectively. The black curves are measured on ATLAS Run 2 data. The blue curves are measured on the SM expected pseudo-data. The solid curves include full systematical uncertainties in the scans. The dashed curves are the statistical uncertainties-only scans.

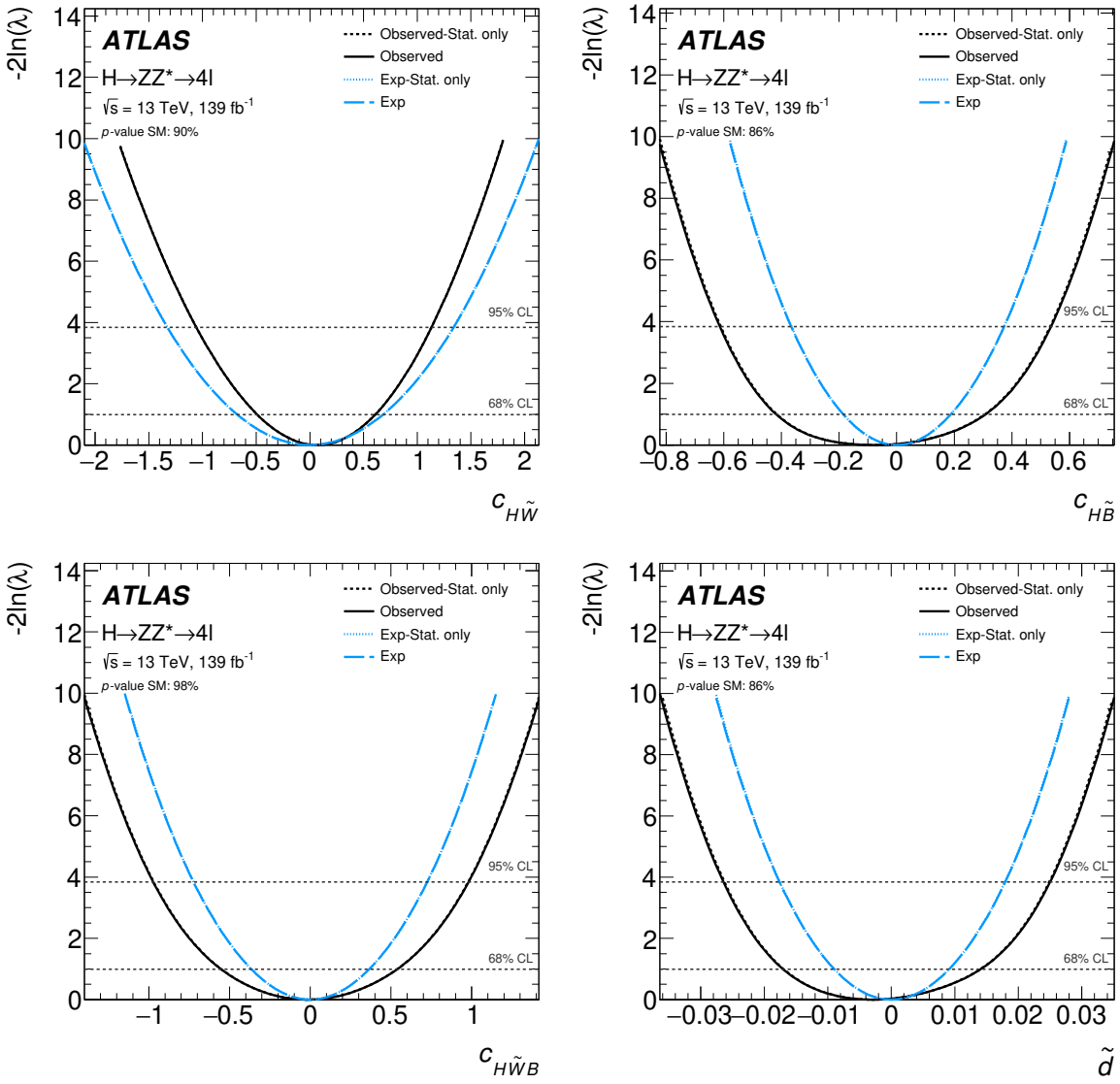


Figure I.2: Likelihood scans along $c_{H\widetilde{W}}$, $c_{H\widetilde{B}}$, $c_{H\widetilde{W}B}$ and $\widetilde{d}$ axes on the SM expectation and ATLAS Run 2 data with $OO_{4\ell}^{c_{H\widetilde{W}}}$, $OO_{4\ell}^{c_{H\widetilde{B}}}$, $OO_{4\ell}^{c_{H\widetilde{W}B}}$ and $OO_{4\ell}^{\widetilde{d}}$ respectively. The black curves are measured on ATLAS Run 2 data. The blue curves are measured on the SM expected pseudo-data. The solid curves include full systematical uncertainties in the scans. The dashed curves are the statistical uncertainties-only scans.

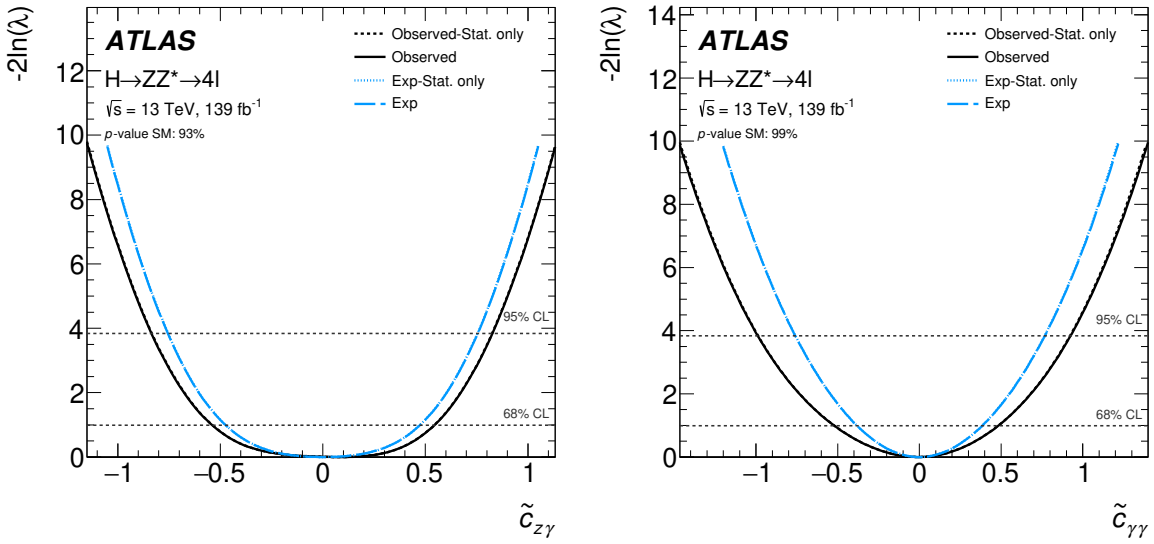


Figure I.3: Likelihood scans along $\widetilde{c}_{z\gamma}$ and $\widetilde{c}_{\gamma\gamma}$ axes on the SM expectation and ATLAS Run 2 data with $OO_{4\ell}^{\widetilde{c}_{za}}$ and $OO_{4\ell}^{\widetilde{c}_{aa}}$ respectively. The black curves are measured on ATLAS Run 2 data. The blue curves are measured on the SM expected pseudo-data. The solid curves include full systematical uncertainties in the scans. The dashed curves are the statistical uncertainties-only scans.

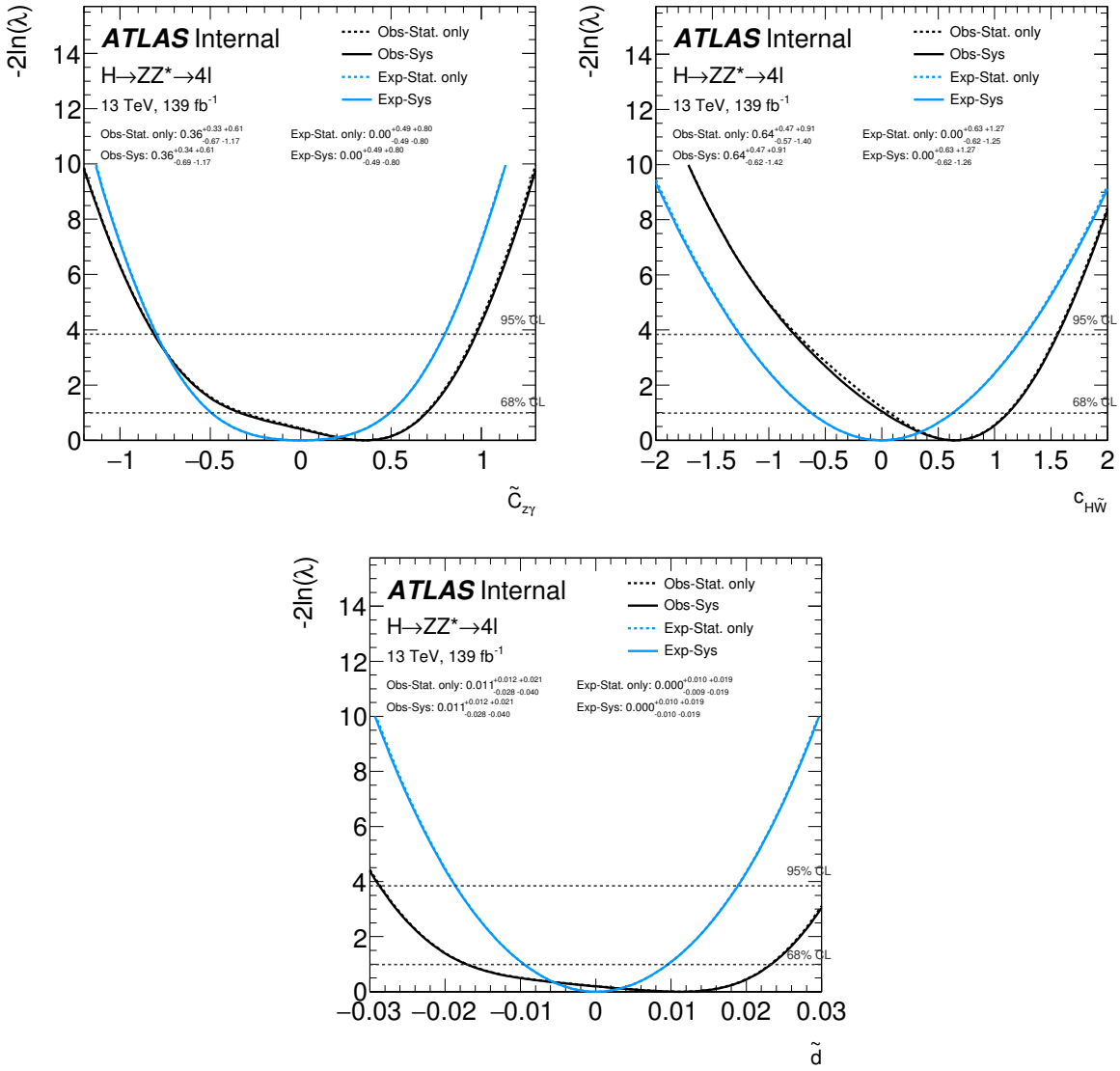


Figure I.4: Likelihood scan along $\widetilde{c}_{z\gamma}$, $c_{H\widetilde{W}}$ and $\widetilde{d}$ axes on the SM expectation and unblinding Run 2 data. The black curves are measured on unblinding data. The blue curves are measured on the SM expectation pseudo-data. The solid curves include full systematical uncertainties in the scan. The dashed curves are the statistical uncertainties-only scan.

Reference

[1] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett. B* 716 (2012), pp. 1–29. arXiv: 1207.7214 [hep-ex].

[2] CMS Collaboration. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". *Phys. Lett. B* 716 (2012), pp. 30–61. arXiv: 1207.7235 [hep-ex].

[3] Particle Data Group. "Review of Particle Physics". *PTEP* 2020.8 (2020), p. 083C01.

[4] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509.

[5] Steven Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266.

[6] ATLAS Collaboration. "Evidence for the spin-0 nature of the Higgs boson using ATLAS data". *Phys. Lett. B* 726 (2013), pp. 120–144. arXiv: 1307.1432 [hep-ex].

[7] ATLAS Collaboration. "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector". *Eur. Phys. J. C* 75.10 (2015). [Erratum: Eur.Phys.J.C 76, 152 (2016)], p. 476. arXiv: 1506.05669 [hep-ex].

[8] CMS Collaboration. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". *Phys. Rev. D* 92.1 (2015), p. 012004. arXiv: 1411.3441 [hep-ex].

[9] ATLAS Collaboration. "Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s}$ =13 TeV with the ATLAS detector". *Eur. Phys. J. C* 80.10 (2020). [Erratum: Eur.Phys.J.C 81, 29 (2021), Erratum: Eur.Phys.J.C 81, 398 (2021)], p. 957. arXiv: 2004.03447 [hep-ex].

[10] ATLAS Collaboration. "Measurement of the CP properties of Higgs boson interactions with τ-leptons with the ATLAS detector" (Dec. 2022). arXiv: 2212.05833 [hep-ex].

[11] ATLAS Collaboration. "CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector". *Phys. Rev. Lett.* 125.6 (2020), p. 061802. arXiv: 2004.04545 [hep-ex].

[12] ATLAS Collaboration. "Study of the CP property of the Higgs boson to electroweak boson coupling in the VBF $H \to \gamma\gamma$ channel with the ATLAS detector" (Aug. 2022). arXiv: 2208.02338 [hep-ex].

[13] CMS Collaboration. "Search for CP violation in ttH and tH production in multilepton channels in proton-proton collisions at $\sqrt{s}$ = 13 TeV" (Aug. 2022). arXiv: 2208.02686 [hep-ex].

[14] CMS Collaboration. "Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV". *JHEP* 06 (2022), p. 012. arXiv: 2110.04836 [hep-ex].

[15] WMAP Collaboration. "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters". *Astrophys. J. Suppl.* 148 (2003), pp. 175–194. arXiv: astro-ph/0302209.

[16] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35.

[17] Michael Dine and Alexander Kusenko. "The Origin of the matter - antimatter asymmetry". *Rev. Mod. Phys.* 76 (2003), p. 1. arXiv: hep-ph/0303065.

[18] Ilaria Brivio and Michael Trott. "The Standard Model as an Effective Field Theory". *Phys. Rept.* 793 (2019), pp. 1–98. arXiv: 1706.08945 [hep-ph].

[19] ATLAS Collaboration. "Test of CP-invariance of the Higgs boson in vector-boson fusion production and its decay into four leptons" (Apr. 2023). arXiv: 2304.09612 [hep-ex].

[20] J. J. Thomson. "Cathode rays". *Phil. Mag. Ser. 5* 44 (1897), pp. 293–316.

[21] Max Planck. "On the Law of Distribution of Energy in the Normal Spectrum". *Annalen Phys.* 4 (1901), p. 553.

[22] Albert Einstein. "Concerning an heuristic point of view toward the emission and transformation of light". *Annalen Phys.* 17 (1905), pp. 132–148.

[23] M. Born and P. Jordan. "Zur Quantenmechanik". *Z. Phys.* 34.1 (1925), pp. 858–888.

[24] M. Born, W. Heisenberg, and P. Jordan. "Zur Quantenmechanik. II." *Z. Phys.* 35.8-9 (1926), pp. 557–615.

[25] Paul A. M. Dirac. “Quantum theory of emission and absorption of radiation”. *Proc. Roy. Soc. Lond. A* 114 (1927), p. 243.

[26] S. L. Glashow. “Partial Symmetries of Weak Interactions”. *Nucl. Phys.* 22 (1961), pp. 579–588.

[27] Abdus Salam and John Clive Ward. “Electromagnetic and weak interactions”. *Phys. Lett.* 13 (1964), pp. 168–171.

[28] William J. Marciano and Heinz Pagels. “Quantum Chromodynamics: A Review”. *Phys. Rept.* 36 (1978), p. 137.

[29] Andrew Purcell. “Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules”. 35/2012 (2012), p. 10.

[30] W. Pauli. “Relativistic Field Theories of Elementary Particles”. *Rev. Mod. Phys.* 13 (1941), pp. 203–232. DOI: 10.1103/RevModPhys.13.203.

[31] Chen-Ning Yang and Robert L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. *Phys. Rev.* 96 (1954). Ed. by Jong-Ping Hsu and D. Fine, pp. 191–195.

[32] E. Fermi. “An attempt of a theory of beta radiation. 1.” *Z. Phys.* 88 (1934), pp. 161–177.

[33] T. D. Lee and Chen-Ning Yang. “Question of Parity Conservation in Weak Interactions”. *Phys. Rev.* 104 (1956), pp. 254–258.

[34] C. S. Wu et al. “Experimental Test of Parity Conservation in β Decay”. *Phys. Rev.* 105 (1957), pp. 1413–1414.

[35] Gargamelle Neutrino Collaboration. “Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment”. *Phys. Lett. B* 46 (1973), pp. 138–140.

[36] J. Goldstone. “Field Theories with Superconductor Solutions”. *Nuovo Cim.* 19 (1961), pp. 154–164. DOI: 10.1007/BF02812722.

[37] LHC Study Group. “The Large Hadron Collider: Conceptual design” (Oct. 1995). Ed. by Thomas Sven Pettersson and P. Lefèvre.

[38] Murray Gell-Mann. “A Schematic Model of Baryons and Mesons”. *Phys. Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3.

[39] Richard P. Feynman. “Very high-energy collisions of hadrons”. *Phys. Rev. Lett.* 23 (1969). Ed. by L. M. Brown, pp. 1415–1417.

[40] A. D. Martin et al. “Parton distributions for the LHC”. *Eur. Phys. J. C* 63 (2009), pp. 189–285. arXiv: 0901.0002 [hep-ph].

[41] Abdelhak Djouadi. "The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model". *Phys. Rept.* 457 (2008), pp. 1–216. arXiv: hep-ph/0503172.

[42] R. P. Feynman. "Mathematical formulation of the quantum theory of electromagnetic interaction". *Phys. Rev.* 80 (1950). Ed. by L. M. Brown, pp. 440–457.

[43] Richard P. Feynman. "An Operator calculus having applications in quantum electrodynamics". *Phys. Rev.* 84 (1951). Ed. by L. M. Brown, pp. 108–128.

[44] LHC Higgs Cross Section Working Group. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". 2/2017 (Oct. 2016). arXiv: 1610.07922 [hep-ph].

[45] ATLAS. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data". *JINST* 14.12 (2019), P12006. arXiv: 1908.00005 [hep-ex].

[46] ATLAS. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 81.7 (2021), p. 578. arXiv: 2012.00578 [hep-ex].

[47] John Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. "A Historical Profile of the Higgs Boson" (Jan. 2012). arXiv: 1201.6045 [hep-ph].

[48] LEP Working Group for Higgs boson searches. "Search for the standard model Higgs boson at LEP". *Phys. Lett. B* 565 (2003), pp. 61–75. arXiv: hep-ex/0306033.

[49] D0 Collaboration CDF Collaboration. "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data". July 2011. arXiv: 1107.5518 [hep-ex].

[50] ATLAS Collaboration. "Combined search for the Standard Model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". *Phys. Rev. D* 86 (2012), p. 032003. arXiv: 1207.0319 [hep-ex].

[51] CMS Collaboration. "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV". *Phys. Lett. B* 710 (2012), pp. 26–48. arXiv: 1202.1488 [hep-ex].

[52] Werner Herr, B. J. Holzer, and B. Muratori. "Design and Principles of Synchrotrons and Circular Colliders: Concept of Luminosity". *Elementary Particles - Accelerators and Colliders*. Ed. by S. Myers and H. Schopper. 2013, pp. 6.4-1 - 6.4–7.

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

[54] Particle Data Group. "Review of Particle Physics". *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[55] ATLAS Collaboration. "Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ decay channel using 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions recorded by the ATLAS detector at the LHC" (July 2022). arXiv: 2207.00320 [hep-ex].

[56] CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". *Phys. Lett. B* 805 (2020), p. 135425. arXiv: 2002.06398 [hep-ex].

[57] L. D. Landau. "On the angular momentum of a system of two photons". *Dokl. Akad. Nauk SSSR* 60.2 (1948), pp. 207–209.

[58] Chen-Ning Yang. "Selection Rules for the Dematerialization of a Particle Into Two Photons". *Phys. Rev.* 77 (1950), pp. 242–245.

[59] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". *Prog. Theor. Phys.* 49 (1973), pp. 652–657.

[60] Murray Gell-Mann and A. Pais. "Behavior of neutral particles under charge conjugation". *Phys. Rev.* 97 (1955), pp. 1387–1389.

[61] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". *Phys. Rev. Lett.* 13 (1964), pp. 138–140.

[62] KTeV Collaboration. "Observation of direct CP violation in $K_{S,L} \to \pi\pi$ decays". *Phys. Rev. Lett.* 83 (1999), pp. 22–27. arXiv: hep-ex/9905060.

[63] NA48 Collaboration. "A New measurement of direct CP violation in two pion decays of the neutral kaon". *Phys. Lett. B* 465 (1999), pp. 335–348. arXiv: hep-ex/9909022.

[64] Bruce Winstein and Lincoln Wolfenstein. "The Search for direct CP violation". *Rev. Mod. Phys.* 65 (1993), pp. 1113–1148.

[65] Belle Collaboration. "Observation of large CP violation in the neutral B meson system". *Phys. Rev. Lett.* 87 (2001), p. 091802. DOI: 10.1103/PhysRevLett.87.091802. arXiv: hep-ex/0107061.

[66] LHCb Collaboration. "Observation of CP Violation in Charm Decays". *Phys. Rev. Lett.* 122.21 (2019), p. 211803. DOI: 10.1103/PhysRevLett.122.211803. arXiv: 1903.08726 [hep-ex].

[67] C. Jarlskog. "Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation". *Phys. Rev. Lett.* 55 (1985), p. 1039. DOI: 10.1103/PhysRevLett.55.1039.

[68] Andrew G. Cohen, A. De Rujula, and S. L. Glashow. "A Matter - antimatter universe?" *Astrophys. J.* 495 (1998), pp. 539–549. arXiv: astro-ph/9707087.

[69] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. "Matter and Antimatter in the Universe". *New J. Phys.* 14 (2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012. arXiv: 1204.4186 [hep-ph].

[70] K. G. Wilson and John B. Kogut. "The Renormalization group and the epsilon expansion". *Phys. Rept.* 12 (1974), pp. 75–199.

[71] B. Grzadkowski et al. "Dimension-Six Terms in the Standard Model Lagrangian". *JHEP* 10 (2010), p. 085. arXiv: 1008.4884 [hep-ph].

[72] Steven Weinberg. "Baryon and Lepton Nonconserving Processes". *Phys. Rev. Lett.* 43 (1979), pp. 1566–1570.

[73] Andre de Gouvea, Juan Herrero-Garcia, and Andrew Kobach. "Neutrino Masses, Grand Unification, and Baryon Number Violation". *Phys. Rev. D* 90.1 (2014), p. 016011. arXiv: 1404.4057 [hep-ph].

[74] Frank F. Deppisch et al. "Falsifying High-Scale Baryogenesis with Neutrinoless Double Beta Decay and Lepton Flavor Violation". *Phys. Rev. D* 92.3 (2015), p. 036005. arXiv: 1503.04825 [hep-ph].

[75] Rodrigo Alonso et al. "Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology". *JHEP* 04 (2014), p. 159. DOI: 10.1007/JHEP04(2014)159. arXiv: 1312.2014 [hep-ph].

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

[77] G. Plass. "The LEP Project – Status and Plans". *IEEE Trans. Nucl. Sci.* 30 (1983). Ed. by L. S. Taylor, pp. 1978–1982.

[78] Yves Baconnier. "LHC: the Large Hadron Collider accelerator project" (1993). Ed. by Giorgio Brianti et al.

[79] Vittorio Frigo. "LHC structure.. Structure LHC." AC Collection. Legacy of AC. Pictures from 1992 to 2002. Jan. 1997. URL: http://cds.cern.ch/record/842611.

[80] "LHC Design Report Vol.1: The LHC Main Ring" (June 2004). Ed. by Oliver S. Bruning et al.

[81] ATLAS Collaboration. "ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN" (Dec. 1994).

[82] CMS Collaboration. "CMS, the Compact Muon Solenoid: Technical proposal" (Dec. 1994).

[83] LHCb Collaboration. "LHCb technical proposal: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays" (Feb. 1998).

[84] ALICE Collaboration. "ALICE: Technical proposal for a large ion collider experiment at the CERN LHC" (Dec. 1995).

[85] Esma Mobs. "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018" (Aug. 2018). General Photo. URL: https://cds.cern.ch/record/2636343.

[86] "LHC Design Report. 3. The LHC injector chain" (Dec. 2004). Ed. by M. Benedikt et al.

[87] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". *JINST* 3 (2008), S08003.

[88] A. Yamamoto et al. "The ATLAS central solenoid". *Nucl. Instrum. Meth. A* 584 (2008), pp. 53–74. DOI: 10.1016/j.nima.2007.09.047.

[89] ATLAS Collaboration. "Production and Integration of the ATLAS Insertable B-Layer". *JINST* 13.05 (2018), T05008. arXiv: 1803.00844 [physics.ins-det].

[90] Martin Aleksa and Marcella Diemoz. "Discussion on the electromagnetic calorimeters of ATLAS and CMS". *Nucl. Instrum. Meth. A* 732 (2013). Ed. by T. Bergauer et al., pp. 442–450. DOI: 10.1016/j.nima.2013.05.195.

[91] D. J. Mahon. "ATLAS LAr calorimeter performance in LHC Run 2". *JINST* 15.06 (2020), p. C06045. DOI: 10.1088/1748-0221/15/06/C06045.

[92] Antonio Salvucci. "Measurement of muon momentum resolution of the ATLAS detector". *EPJ Web Conf.* 28 (2012). Ed. by Gregorio Bernardi, Sandro De Cecco, and Yuji Enari, p. 12039. arXiv: 1201.4704 [physics.ins-det].

[93] "ATLAS muon spectrometer: Technical design report" (June 1997).

[94] S. Palestini. "The muon spectrometer of the ATLAS experiment". *Nucl. Phys. B Proc. Suppl.* 125 (2003). Ed. by F. L. Navarria, M. Paganoni, and P. G. Pelfer, pp. 337–345. DOI: 10.1016/S0920-5632(03)91013-9.

[95] Y. Arai et al. "ATLAS muon drift tube electronics". *JINST* 3 (2008), P09001. DOI: 10.1088/1748-0221/3/09/P09001.

[96] ATLAS. "Resolution of the ATLAS muon spectrometer monitored drift tubes in LHC Run 2". *JINST* 14.09 (2019), P09011. arXiv: 1906.12226 [hep-ex].

[97] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC" (June 2019).

[98] ATLAS TDAQ Collaboration. "The ATLAS Data Acquisition and High Level Trigger system". *JINST* 11.06 (2016), P06008.

[99] ATLAS Collaboration. "Performance of the ATLAS Trigger System in 2015". *Eur. Phys. J. C* 77.5 (2017), p. 317. arXiv: 1611.09661 [hep-ex].

[100] R. Achenbach et al. "The ATLAS level-1 calorimeter trigger". *JINST* 3 (2008), P03001. DOI: 10.1088/1748-0221/3/03/P03001.

[101] ATLAS Collaboration. "ATLAS data quality operations and performance for 2015–2018 data-taking". *JINST* 15.04 (2020), P04003. DOI: 10.1088/1748-0221/15/04/P04003. arXiv: 1911.04632 [physics.ins-det].

[102] ATLAS Collaboration. "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Eur. Phys. J. C* 81.8 (2021), p. 689. arXiv: 2007.02645 [hep-ex].

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

[104] ATLAS. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV". *Eur. Phys. J. C* 79.8 (2019), p. 639. arXiv: 1902.04655 [physics.ins-det].

[105] ATLAS. "Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2". *Eur. Phys. J. C* 77.10 (2017), p. 673. arXiv: 1704.07983 [hep-ex].

[106] Thijs G. Cornelissen et al. "The global chi**2 track fitter in ATLAS". *J. Phys. Conf. Ser.* 119 (2008). Ed. by Randall Sobie, Reda Tafirout, and Jana Thomson, p. 032013.

[107] T. Cornelissen et al. "The new ATLAS track reconstruction (NEWT)". *J. Phys. Conf. Ser.* 119 (2008). Ed. by Randall Sobie, Reda Tafirout, and Jana Thomson, p. 032014.

[108] ATLAS Collaboration. "Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung" (May 2012).

[109] J. Illingworth and J. Kittler. "A survey of the hough transform". *Computer Vision, Graphics, and Image Processing* 44.1 (1988), pp. 87–116. ISSN: 0734-189X.

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

[111] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". *Eur. Phys. J. C* 72 (2012), p. 1896. arXiv: 1111.6097 [hep-ph].

[112] Matteo Cacciari and Gavin P. Salam. "Dispelling the N^3 myth for the k_t jet-finder". *Phys. Lett. B* 641 (2006), pp. 57–61. arXiv: hep-ph/0512210.

[113] ATLAS Collaboration. "Jet reconstruction and performance using particle flow with the ATLAS Detector". *Eur. Phys. J. C* 77.7 (2017), p. 466. arXiv: 1703.10485 [hep-ex].

[114] V. Kostyukhin. "VKalVrt - package for vertex reconstruction in ATLAS." (Aug. 2003).

[115] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". *JHEP* 06 (2010), p. 043. arXiv: 1002.2581 [hep-ph].

[116] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". *JHEP* 11 (2007), p. 070. arXiv: 0709.2092 [hep-ph].

[117] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". *JHEP* 11 (2004), p. 040. arXiv: hep-ph/0409146.

[118] John M. Campbell et al. "NLO Higgs Boson Production Plus One and Two Jets Using the POWHEG BOX, MadGraph4 and MCFM". *JHEP* 07 (2012), p. 092. arXiv: 1202.5475 [hep-ph].

[119] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "MINLO: Multi-Scale Improved NLO". *JHEP* 10 (2012), p. 155. DOI: 10.1007/JHEP10(2012)155. arXiv: 1206.3572 [hep-ph].

[120] Keith Hamilton et al. "NNLOPS simulation of Higgs boson production". *JHEP* 10 (2013), p. 222. arXiv: 1309.0017 [hep-ph].

[121] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator". *JHEP* 05 (2015), p. 140. DOI: 10.1007/JHEP05(2015)140. arXiv: 1501.04637 [hep-ph].

[122] Stefano Catani and Massimiliano Grazzini. "An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC". *Phys. Rev. Lett.* 98 (2007), p. 222002. DOI: 10.1103/PhysRevLett.98.222002. arXiv: hep-ph/0703012.

[123] Keith Hamilton et al. "Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching". *JHEP* 05 (2013), p. 082. DOI: 10.1007/JHEP05(2013)082. arXiv: 1212.4504 [hep-ph].

[124] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". *J. Phys. G* 43 (2016), p. 023001. arXiv: 1510.03865 [hep-ph].

[125] Paolo Nason and Carlo Oleari. "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG". *JHEP* 02 (2010), p. 037. arXiv: 0911.5299 [hep-ph].

[126] Gionata Luisoni et al. "$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO". *JHEP* 10 (2013), p. 083. arXiv: 1306.2542 [hep-ph].

[127] 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), p. 079. arXiv: 1405.0301 [hep-ph].

[128] M. Wiesemann et al. "Higgs production in association with bottom quarks". *JHEP* 02 (2015), p. 132. arXiv: 1409.5301 [hep-ph].

[129] Richard D. Ball et al. "Parton distributions with LHC data". *Nucl. Phys. B* 867 (2013), pp. 244–289. arXiv: 1207.1303 [hep-ph].

[130] D. J. Lange. "The EvtGen particle decay simulation package". *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155.

[131] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual". *JHEP* 05 (2006), p. 026. arXiv: hep-ph/0603175.

[132] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". *Comput. Phys. Commun.* 178 (2008), pp. 852–867. arXiv: 0710.3820 [hep-ph].

[133] Enrico Bothmann et al. "Event Generation with Sherpa 2.2". *SciPost Phys.* 7.3 (2019), p. 034. arXiv: 1905.09127 [hep-ph].

[134] T. Gleisberg et al. "Event generation with SHERPA 1.1". *JHEP* 02 (2009), p. 007. arXiv: 0811.4622 [hep-ph].

[135] Richard D. Ball et al. “Parton distributions for the LHC Run II”. *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph].

[136] Tanju Gleisberg and Stefan Hoeche. “Comix, a new matrix element generator”. *JHEP* 12 (2008), p. 039. arXiv: 0808.3674 [hep-ph].

[137] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. “Scattering Amplitudes with Open Loops”. *Phys. Rev. Lett.* 108 (2012), p. 111601. arXiv: 1111.5206 [hep-ph].

[138] Steffen Schumann and Frank Krauss. “A Parton shower algorithm based on Catani-Seymour dipole factorisation”. *JHEP* 03 (2008), p. 038. arXiv: 0709.1027 [hep-ph].

[139] Stefan Hoeche et al. “QCD matrix elements + parton showers: The NLO case”. *JHEP* 04 (2013), p. 027. arXiv: 1207.5030 [hep-ph].

[140] B. Biedermann et al. “Electroweak corrections to $pp \to \mu^+\mu^- e^+ e^- + X$ at the LHC: a Higgs background study”. *Phys. Rev. Lett.* 116.16 (2016), p. 161803. arXiv: 1601.07787 [hep-ph].

[141] Benedikt Biedermann et al. “Next-to-leading-order electroweak corrections to the production of four charged leptons at the LHC”. *JHEP* 01 (2017), p. 033. arXiv: 1611.05338 [hep-ph].

[142] Andy Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. *Eur. Phys. J. C* 75 (2015), p. 132. arXiv: 1412.7420 [hep-ph].

[143] ATLAS Collaboration. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector”. *JHEP* 09 (2014), p. 145. arXiv: 1406.3660 [hep-ex].

[144] “ATLAS Pythia 8 tunes to 7 TeV data” (Nov. 2014).

[145] Ilaria Brivio, Yun Jiang, and Michael Trott. “The SMEFTsim package, theory and tools”. *JHEP* 12 (2017), p. 070. arXiv: 1709.06492 [hep-ph].

[146] Leif Lonnblad. “Correcting the color dipole cascade model with fixed order matrix elements”. *JHEP* 05 (2002), p. 046. arXiv: hep-ph/0112284.

[147] D. Atwood and A. Soni. “Analysis for magnetic moment and electric dipole moment form-factors of the top quark via e+ e- —> t anti-t”. *Phys. Rev. D* 45 (1992), pp. 2405–2413.

[148] M. Davier et al. “The Optimal method for the measurement of tau polarization”. *Phys. Lett. B* 306 (1993), pp. 411–417.

[149] Markus Diehl, Otto Nachtmann, and Felix Nagel. "Triple gauge couplings in polarized e- e+ —> W- W+ and their measurement using optimal observables". *Eur. Phys. J. C* 27 (2003), pp. 375–397. arXiv: hep-ph/0209229.

[150] Dan Guest, Kyle Cranmer, and Daniel Whiteson. "Deep Learning and its Application to LHC Physics". *Ann. Rev. Nucl. Part. Sci.* 68 (2018), pp. 161–181. arXiv: 1806.11484 [hep-ex].

[151] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. "Probing color singlet exchange in Z + two jet events at the CERN LHC". *Phys. Rev. D* 54 (1996), pp. 6680–6689. DOI: 10.1103/PhysRevD.54.6680. arXiv: hep-ph/9605444.

[152] Robert D. Cousins. "Lectures on Statistics in Theory: Prelude to Statistics in Practice" (July 2018). arXiv: 1807.05996 [physics.data-an].

[153] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[154] ATLAS Collaboration. "A morphing technique for signal modelling in a multidimensional space of coupling parameters" (2015).

[155] Muriel Pivk and Francois R. Le Diberder. "SPlot: A Statistical tool to unfold data distributions". *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. arXiv: physics/0402083.

[156] ATLAS Collaboration. "Evaluation of theoretical uncertainties for simplified template cross section measurements of V-associated production of the Higgs boson" (2018).

[157] Iain W. Stewart and Frank J. Tackmann. "Theory Uncertainties for Higgs and Other Searches Using Jet Bins". *Phys. Rev. D* 85 (2012), p. 034011. arXiv: 1107.2117 [hep-ph].

[158] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". *Eur. Phys. J. C* 76.4 (2016), p. 196. arXiv: 1512.01178 [hep-ph].

[159] *Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report*. Mar. 2018. arXiv: 1803.07977 [hep-ph].

[160] P. Artoisenet et al. "A framework for Higgs characterisation". *JHEP* 11 (2013), p. 043. arXiv: 1306.6464 [hep-ph].

[161] B Efron. "Bootstrap methods: another look at the jackknife". *The Annals of Statistics* 7 (1979), pp. 1–26.

[162] ATLAS Collaboration. "Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector". *Eur. Phys. J. C* 76.12 (2016), p. 658. DOI: 10.1140/epjc/s10052-016-4499-5. arXiv: 1602.04516 [hep-ex].

[163] CMS Collaboration. "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* 104.5 (2021), p. 052004. arXiv: 2104.12152 [hep-ex].

[164] "High-Luminosity Large Hadron Collider (HL-LHC) : Preliminary Design Report" (Dec. 2015). Ed. by G Apollinari et al.

[165] ATLAS Collaboration. "Prospects for measurements of tensor structure of the HZZ vertex in $H \to ZZ^* \to 4l$ decay with ATLAS detector" (2013).

[166] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties" (July 2013). Ed. by S Heinemeyer et al. arXiv: 1307.1347 [hep-ph].

[167] CEPC Study Group. "CEPC Conceptual Design Report: Volume 1 - Accelerator" (Sept. 2018). arXiv: 1809.00285 [physics.acc-ph].

[168] CEPC Study Group. "CEPC Conceptual Design Report: Volume 2 - Physics & Detector" (Nov. 2018). Ed. by João Barreiro Guimarães da Costa et al. arXiv: 1811.10545 [hep-ex].

[169] Qiyu Sha et al. "Probing Higgs CP properties at the CEPC" (Mar. 2022). arXiv: 2203.11707 [hep-ex].

[170] Adam Falkowski et al. "Rosetta: an operator basis translator for Standard Model effective field theory". *Eur. Phys. J. C* 75.12 (2015), p. 583. arXiv: 1508.05895 [hep-ph].

[171] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". *JHEP* 03 (2013), p. 015. arXiv: 1212.3460 [hep-ph].

[172] Stefan Höche et al. "A Study of QCD Radiation in VBF Higgs Production with Vincia and Pythia". *SciPost Phys.* 12.1 (2022), p. 010. arXiv: 2106.10987 [hep-ph].

[173] Barbara Jäger et al. "Parton-shower effects in Higgs production via Vector-Boson Fusion". *Eur. Phys. J. C* 80.8 (2020), p. 756. arXiv: 2003.12435 [hep-ph].