Universiteit van Amsterdam

VU University Amsterdam

MSc Physics and Astronomy

Track : Gravitation, Astro-, and Particle Physics

Master Thesis

Exploring the gg $\rightarrow$ ZH process for constraining Wilson coefficients within the SMEFT framework

---

Author:
Pranati Kharbanda

Supervisors:
Prof. Dr. Marcel Vreeswijk
Dr. Geoffrey Gilles

Second Examiner:
Prof. Dr. Wouter Verkerke

Nikhef

ATLAS EXPERIMENT

Abstract

The Standard Model Effective Field Theory (SMEFT) aims to investigate deviations from the Standard Model and place constraints on the free parameters of this theory, the Wilson coefficients. In the SMEFT framework, the process, $gg \to Z(\to l^+l^-)H(\to b\bar{b})$ receives a contribution from the dimension-6 operator $\mathcal{O}_{\phi t}$ whose associated Wilson coefficient, $c_{\phi t}$, is presently weakly constrained in the current state-of-the-art measurements. This thesis, for the first time, carries out feasibility studies in this process within this framework. We simulate and analyze events generated using the SMEFT model to identify key kinematic quantities which can be used to study these EFT effects. By effectively separating the most challenging backgrounds, which include $qq \to ZH$ and $pp \to Z + jets$, from the signal, we establish constraints on this Wilson coefficient at a 95% Confidence Level. Comparing our results with the current global fit values, we find a stricter constraint for the High Luminosity LHC era.

Acknowledgements

First and foremost I would like to thank my supervisor, Marcel, for giving me the chance to pursue this project. His clear explanations and unwavering support throughout the year have been invaluable. I would like to express my heartfelt gratitude to Geoffrey for his enthusiasm in engaging in physics discussions and consistently introducing me to new ideas, which have played a significant role in shaping my research journey. I am especially grateful to Wouter (H) for generously funding my trip to CERN which was truly an exceptional experience. Thanks to Wouter (V), Ivo and Patrick for their invaluable assistance in securing this scholarship. I would also like to thank Andrea (V), Andrea (GA) (in no particular order) and Oliver for being the best support team, specially during the last month. Andrea, thank you for always believing in me, and for your constant help every step of the way. I could not have asked for a more understanding office mate. Andrea, you have been a true inspiration for me. With helping me go beyond my analysis, by teaching me about the ITk detector, I am very grateful for all the things I have learned from you. Oliver, thank you for always encouraging me to ask questions, no matter how trivial or non-trivial they are. I am also grateful to Lydia for extending her help, specially during the PhD process. I would also like to thank Hannah, Zhuoran and Dylan for making my time at CERN so memorable and supporting me during my PhD interviews. A special thank you to all Nikhef ATLAS people for supporting me. Finally, I would like to extend thanks to my family for all the encouragement they have given me all these years.

Contents

Chapter 1

Introduction

The Standard Model of Particle Physics is our current most fundamental description of the universe and the elementary particles that the universe constitutes. It explains the interactions that these particles have in terms of mediation of force carriers. The theory exhibits unparalleled experimental accuracy, confirmed in numerous experiments around the world, one of them being the ATLAS experiment at the Large Hadron Collider. However, while explaining only three of the four fundamental forces, it is not a complete theory and faces some shortcomings.

There is a tremendous ongoing effort to search for physics beyond the SM. With a lack of evidence of direct detection of new particles, the focus is now shifted towards precision measurements to observe deviations from the SM predictions. One of the promising approaches to study these deviations is the model-independent framework, Standard Model Effective Field Theory (SMEFT). This theory introduces additional operators in the SM Lagrangian to account for new physics at high energy scales. These operators have couplings associated with them, which can be expressed in terms of the free parameters of this theory, the Wilson coefficients.

Studies have been performed for the interpretation of current measurements in the context of SMEFT, to place limits on these coefficients and quantify the effects of new physics. An interesting weakly constrained Wilson coefficient is $c_{\phi t}$, whose associated operator alters the Higgs, Z boson and top quark interactions. Being the most massive particles found in the SM, Higgs boson and top quark act as crucial probes to investigate physics beyond the SM. This thesis studies the $gg \to ZH$ process to interpret the measurements in terms of this coefficient. Enduring two challenging backgrounds, one of a much higher cross section, $pp \to Z + jets$, and a second one which has the same final state, $qq \to ZH$, this thesis addresses the challenge of separating these processes from $gg \to ZH$. After carrying out this separation, we impose constraints on this Wilson coefficient and verify if they compare well with the current global limits.

The remainder of the thesis is organized in the following way. Chapter 2 presents a brief introduction to the Standard Model and Standard Model Effective Field Theory. It also outlines the motivation behind selecting this process. Chapter 3 offers a short summary of the ATLAS experiment. Chapter 4 provides a comprehensive account of the studies conducted in this analysis, involving the identification of kinematic regions of interest to study EFT effects in these interactions. Chapter 5 describes the usage of deep learning algorithms for discriminating the signal from the backgrounds. Finally, we derive a sensitivity on the relevant Wilson coefficient in terms of a 95% Confidence Level exclusion limit in chapter 6. The thesis concludes with future outlooks in chapter 7.

Chapter 2

Theoretical Context

This chapter provides the theoretical basis for the rest of the thesis. It starts with an introduction to Standard Model and Standard Model Effective Field Theory in Sections 2.1 and 2.2. These sections are partly adapted from [1]. After this description, a motivation for the rest of the work is stated in Section 2.3.

2.1 Standard Model

The Standard Model (SM) is a theory model that formally defines and describes the interactions between all known particles. Particles are broadly categorized into two types based on their spin. Particles having integer spins are bosons, whereas particles with half an integer spin are fermions. The fermions come in three generations, where the generations only differ from each other in terms of the masses. Moreover, for each fermion particle, there is a corresponding antiparticle which carries the opposite (additive) charge. Fermions are further divided into leptons and quarks, depending on the forces they experience. The interactions of particles are mediated through the exchange of particles which are known as 'force carriers'.

The non-Abelian group that constitutes the SM is $SU(3) \times SU(2) \times U(1)$. The SU(3) group represents strong interactions between the quarks and the force carriers, gluons. The quantum number associated with particles which are charged under this group is the color charge. Due to the non-commutative structure of the group, the gluons interact with each other. The $SU(2) \times U(1)$ is the gauge group that dictates the electroweak interactions of the SM. The quantum charge for the SU(2) part is the third component of weak isospin. There are three gauge bosons, $W^{\pm}$ and the neutral Z bosons which are the mediators. Again, the non-commutative nature of the group leads to interactions between the bosons. U(1) group provides a description of the electromagnetic interaction between charged particles. For this group, photons are the mediators which are exchanged during such an interaction. Lastly, for the incorporation of masses in the SM, a scalar doublet, corresponding to the Higgs field, is added by hand to the SM Lagrangian, whose interactions with the fermions and bosons render them mass.

The Standard Model has been a tremendously successful framework, by guiding experimental efforts and providing accurate predictions for various particle physics phenomena since its development. Despite its achievements, there are some phenomena that the SM has not been able to explain fully. Some examples include the neutrino masses, the dark matter problem, and matter antimatter asymmetry.

2.2 Standard Model Effective Field Theory

To address the problems that the SM is not able to solve, many theories have been proposed. In the absence of discovery of resonances pointing towards new heavy particles, it is thought that the physics at low energies might be detached from the microscopic properties of a more fundamental UV theory at high energy scales. This implies that there exists heavy new physics at scales beyond the current LHC reach, which integrates out to give us interactions whose effects one can observe. To incorporate such interactions, Standard Model Effective Field theory is introduced. It is a theory which treats SM as a low energy approximation of such a fundamental UV theory, and is a model-independent way of parameterizing effects of new physics. It extends the SM Lagrangian in powers of an energy scale, Λ, corresponding to the energy scale of new physics, to include operators of higher dimensions. The operators consist of SM fields, and the Lagrangian obeys symmetries of the SM. The addition of these terms can either lead to new vertices or changes in the existing vertices of the SM. The Lagrangian is,

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i^{(5)}}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_i \frac{c_i^{(7)}}{\Lambda^3}\mathcal{O}_i^{(7)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4}\mathcal{O}_i^{(8)} + ... \tag{2.1}$$

The constants c_i, known as Wilson coefficients, determine the strength of the operators. '$\mathcal{O}$' are operators of increasing dimensional powers 'd', whose corresponding terms are suppressed by the energy scale Λ^{d-4}. A priori, the coefficients and the energy scale are unknown. If we fix the energy scale ($\Lambda = 1$ TeV for this thesis, as is commonly chosen for consistency), the only free parameters remaining in the theory are these Wilson coefficients which are to be found experimentally. Odd dimension operators violate lepton number conservation, introduce a CP asymmetry, and hence are typically ignored in SMEFT interpretations. Operators of dimension 8 (and higher) are further suppressed and currently unavailable in the current implementations of SMEFT framework. Thus, in this thesis, we only focus on dimension-6 operators.

Parameterization

The introduction of SMEFT operators in the SM Lagrangian can manifest itself in several ways. This includes the addition of effective vertices, or contact interactions that previously did not exist. In order to experimentally probe how large the effects of new physics are, we express the experimentally relevant quantity, the cross section, in terms of the Wilson coefficients. If we focus solely on dimension-6 operators, from the SMEFT Lagrangian introduced in Section 2.2, and confine ourselves to Feynman diagrams where only one SMEFT operator appears, we can rewrite the cross section as,

$$\sigma \propto |\mathcal{M}|^2 = |\mathcal{M}_{SM} + \sum_i \mathcal{M}_i^{(6)}|^2 \tag{2.2}$$

where $\mathcal{M}$ is the matrix element amplitude for the respective terms. Expanding this gives us three terms,

$$\sigma_{SMEFT} = \sigma_{SM} + \sigma_{int} + \sigma_{BSM} \tag{2.3}$$

where the first term corresponds to the terms originating from SM, while the second term is related to the interference between the SM and dimension-6 operators, and is suppressed by Λ^{-2}. The third term is a pure beyond Standard Model (BSM) term, resulting exclusively from additional operators and is suppressed by Λ^{-4}. This term encompasses contributions from multiple terms, e.g. pure dim-6 terms, interference of SM and dim-8 terms. However, this thesis only considers contributions from dim-6 operators. Rewriting equations (2.2) and (2.3), we arrive at

$$\frac{\sigma_{int}}{\sigma_{SM}} \sim \frac{\sum_i 2Re(\mathcal{M}^{\star}_{SM}\mathcal{M}^{(6)}_i)}{|\mathcal{M}_{SM}|^2} \sim \sum_i \alpha_i c_i \tag{2.4}$$

$$\frac{\sigma_{BSM}}{\sigma_{SM}} \sim \frac{\sum_{i,j} (\mathcal{M}^{(6)}_i)^{\star}\mathcal{M}^{(6)}_j}{|\mathcal{M}_{SM}|^2} \sim \sum_{i,j} \beta_{ij} c_i c_j \tag{2.5}$$

The constants α and β signify the strength corresponding coefficients, on the cross section. Finally, for $i = j$, i.e. if we consider only one SMEFT operator, which is the case in this thesis, we can write,

$$\implies \frac{\sigma_{c_i}}{\sigma_{SM}} = 1 + \alpha_i c_i + \beta_i c_i^2 \tag{2.6}$$

2.3 SMEFT interpretation in Higgs sector

2.3.1 Motivation

With SMEFT having a minimal set of assumptions, there is a tremendous ongoing effort to interpret all LHC measurements in terms of this framework. Fig. 2.1 shows the current 95% Confidence Limits of Wilson coefficients attained by a global fit performed on all data [2]. The magnitude of the limits vary significantly between the coefficients, and depends on the sensitivity of the measurements which are affected by the associated operators. The inclusion of some SMEFT operators has the impact of changing the distributions of kinematic observables, instead of a major change in the cross section, requiring us to investigate differential, in contrast to total, cross section measurements. For this study, we choose the Higgs sector to carry out SMEFT interpretations. This sector is explicitly chosen because this particle, owing to its special properties, acts as a crucial probe to explore new physics. It is one of the heaviest particles of the SM, rendering it sensitive to BSM physics at higher energy scales. It interacts with all the particles of the SM, and is the only particle with spin 0. In the SM Lagrangian, the Higgs fields is added by hand to provide mass to the other fields, and its recent discovery has opened several avenues to investigate its properties and research BSM scenarios.

Due to the lower production rates of this heavy particle, the operators which change its interactions are particularly difficult to put limits on. It thus becomes important to investigate processes which involve vertices with Higgs boson, in a dedicated manner and check if they can assist in placing stricter constraints on the respective coefficients. In this domain, the process we focus on is $gg \to ZH$, with Z decaying to two leptons and Higgs boson decaying to two b-quarks. Higgs boson production in association with a V boson (W or Z) is a prime process used to study Higgs bosons these days as the associated Z boson, decaying to leptons, can be reconstructed well in this channel. However, $gg \to ZH$ has never been measured independently because of its low cross section. It is studied under $pp \to ZH$, and no distinction is made between quark and gluon initiated processes. In this thesis, we examine the extent to which the separation of the two processes is possible.

From the EFT perspective, the $gg \to ZH$ process depends on $\mathcal{O}_{\phi t}$ (ϕ stands for Higgs field). This coefficient is weakly constrained ($\mathcal{O}(20)$), as highlighted with an arrow in Fig. 2.1. Fig. 2.2 displays some vertices which are modified by the associated operator. The interactions which are altered include those between top quark, Higgs, and Z (and W). This thesis, for the first time, performs feasibility studies on this process to derive the confidence limits on the coefficient $c_{\phi t}$ using only this process. Although this process receives contributions from other operators as well, we do not consider them since they are relatively more strongly constrained.

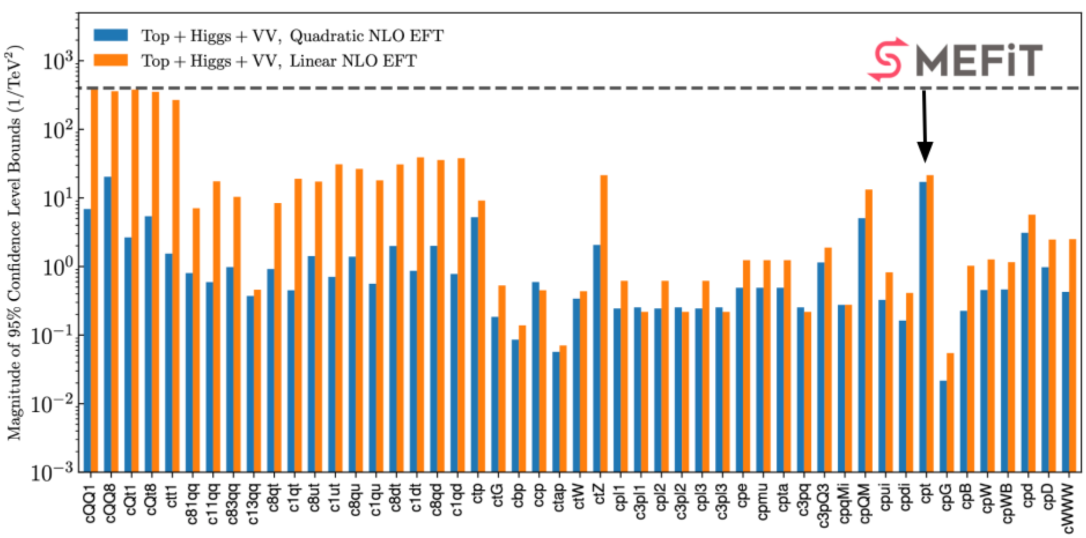


Fig. 2.1: Value of 95% Confidence Limit (CL) of Wilson coefficients obtained from a global fit to data. The operator for this analysis is highlighted (with an arrow). Adapted from [2]

2.3.2 Production Mode of the Higgs: $gg \to ZH$

Signal Process

We consider here the process, $gg \to ZH$. As can be seen in the two LO diagrams that contribute to this process in Fig. 2.3b and 2.3c, this process has the advantage of being sensitive to couplings of the Higgs boson between top-quark and Z boson. Since this process only occurs at NLO, it is studied in conjunction with the LO $qq \to ZH$. Because of the

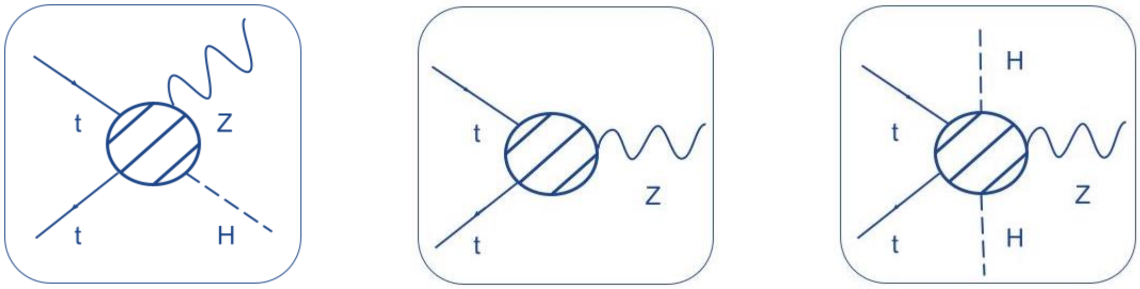


Fig. 2.2: Vertices modified due to SMEFT operators $\mathcal{O}_{tZ}$ (left and middle) and $\mathcal{O}_{\phi t}$ (left, middle and right)

dominance of the latter process, the variations in differential distributions of the physical observables of the two processes are not captured. It has been vastly demonstrated that the invariant mass spectrum of the ZH system and the p_T spectra of the Higgs and Z bosons are more boosted for the gluon initiated process than the quark initiated process [3, 4, 5]. This happens because of the top-dominated virtual loop in the gluon-gluon scattering (cf. Fig. 2.3c). The matrix element amplitude possesses a branch cut at $\sqrt{s} = 2m_t$ [3]. This enhances the cross section when $m_{ZH} \geq 2m_t$, because of the on-shell condition for the two top quarks being fulfilled in the scatterings $gg \to t\bar{t}$ and subsequently $t\bar{t} \to ZH$ [6]. This, in turn, leads to a harder p_T spectrum for this process, augmenting the cross section in the high p_T regime. The need to include this process in the analyses separately, instead of with $qq \to ZH$, becomes even more important if we wish to study EFT effects. As mentioned above, these effects appear as subtle deviations in these differential distributions. For this particular process, as we shall see, the EFT spectra for p_T^Z and m_{ZH} is even harder than the SM spectra. So, if this process is not included separately, the harder spectra in $pp \to ZH$ may be attributed to EFT effects, which may/may not be true. Other interesting aspects of this process are that the perturbation corrections coming from QCD are more prominent in gluon scattering than quark scattering, which amplify the cross section. Finally, the presence of the virtual loops makes it very sensitive to BSM physics wherein non-SM like particles could be present in these loops. Thus, the gluon initiated ZH is an important process, and we demonstrate here that there is information in the kinematic observables that can be used to separate it from $qq \to ZH$.

Considering the EFT effects, it is the top quark loop where the operator $\mathcal{O}_{\phi t}$ appears and SMEFT effects can be observed. There are other SMEFT operators that appear in this process which are not taken into account since they are relatively more strongly constrained. Due to the absence of this top loop in $qq \to ZH$ (cf. Fig. 2.3a), there is no contribution from $\mathcal{O}_{\phi t}$ in this process. The production of the Higgs and Z bosons with large transverse momenta enables us to explore this scenario and hence, is key for SMEFT studies considered in this thesis. The global EFT combination effort of ATLAS does not currently take into account the $gg \to ZH$ process [7], and the limits that are placed on this operator are only through indirect constraints. Thus, we wish to quantify the direct constraints in terms of the level of limits we can bring about through this single process.

Taking into account the decay products of the final state bosons, the channel we choose is,

$gg \to Z(\to l^-l^+)H(\to b\bar{b})$. We emphasize that the decay of Z to leptons providing a clean experimental signature (with effective triggering) combined with the decay of Higgs boson to bottom quarks having the highest branching ratio makes this a highly sensitivity (5σ established [8]) channel to study Higgs bosons. In fact, it is this channel that was used to establish the discovery of Higgs decay to two bottom quarks [9, 10].

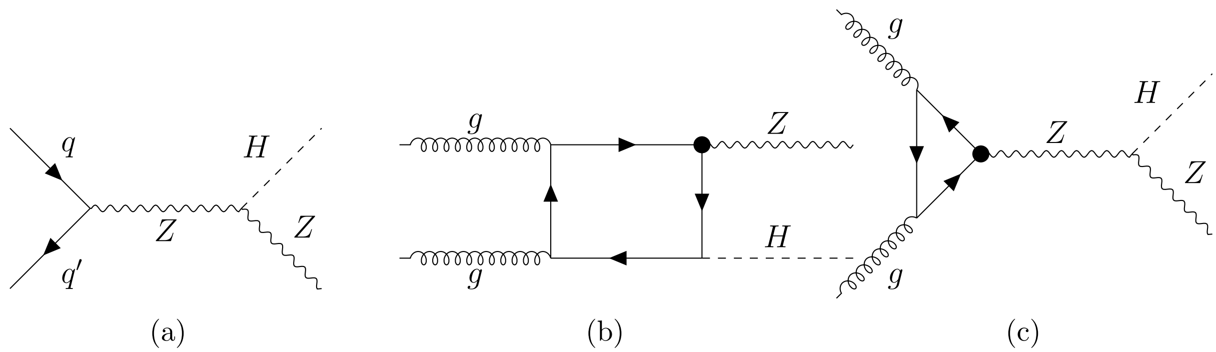


Fig. 2.3: Left: Feynman diagrams for the process, $qq \to ZH$. Middle: Triangle diagram for $gg \to ZH$. Right: Box diagram contributing to $gg \to ZH$

Background Processes

A background comprises of a process which shares the same final state as the signal we are considering. The major backgrounds that need to be considered are $qq \to ZH$ and $Z + jets$. Some of the other less dominant backgrounds are singletop, Diboson (WZ, ZZ, WW) and $\mathrm{t\bar{t}}$. Some Feynman diagrams for each process are given in Fig. 2.4 and Fig. 2.5. For obvious reasons, $qq \to ZH$ is a very challenging background, since the final state particles are exactly the same, and the cross section is $\mathcal{O}(10)$ times larger than $gg \to ZH$. The other most challenging background is $Z + jets$, for which, the cross section is much greater. The multiple jets arising from the initial state or parton showering can mimic the two b-jets coming from Higgs boson decay in $gg \to ZH$. For Diboson processes, the Z bosons can decay to two leptons and two b-quarks, imitating the signal. Similarly, the jets from single top quark and top-anti-top production can mimic the b-jets in $gg \to ZH$. A powerful kinematic observable to reject these backgrounds is the invariant mass of two b-tagged jets, which would not peak at $m_{bb} = 125$ GeV in these processes, whereas it would, in the signal. To perform background rejection, we use machine learning techniques, which are explored in Chapter 5.

Fig. 2.4: Some Feynman diagrams for V+jets and Diboson processes

Fig. 2.5: Some Feynman diagrams for single top and $t\bar{t}$ processes

Chapter 3

Experimental Setup

The Large Hadron Collider (LHC) [11] is the most powerful accelerator and collider in the world. It is a 27 km circular collider which accelerates protons and ions to very high speeds, which upon colliding produce highly energetic particles. Currently, it is operating at a peak center-of-mass energy of 13.6 TeV. This collider accommodates four collision points, wherein four detectors are housed. These four detectors, with their own specific goals, address a wide range of particle physics questions.

3.1 The ATLAS Detector

One of the two general multi-purpose detectors is the ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector [12]. With the measurements of 45 m in length and 25 m in diameter, it is the largest volume detector of the LHC. A cut-away illustration is displayed in Fig. 3.1. Surrounding the interaction point are the cylindrical barrel regions, which have several sub-detectors embedded in it, in a layer-like fashion, to detect the particles that are produced in the collisions, and provide a comprehensive measurement of their properties. The end cap structures are placed at the very end of these barrel regions, offering additional coverage. These sub-detectors consist of the inner detector, the calorimeters and the muon spectrometer which are discussed below. Before this discussion, we describe the coordinate system that the ATLAS experiment uses.

The coordinate system used is the right handed Cartesian system, with z axis along the beam. The x axis is towards the center of the LHC tunnel and y axis points upwards. Since the detector follows a rotational symmetry along the z direction, a cylindrical coordinate system is used to describe the properties of particles. The polar angle, ranging from $[0, \pi]$ is the angle from the z axis, and azimuthal angle, ranging from $[-\pi, \pi]$ describes angles along the z axis. Another angle used to describe the direction a particle follows is the rapidity (pseudo-rapidity) angle for massive (massless) particles. The rapidity is defined as, $y = \frac{1}{2}\frac{E+p_z}{E-p_z}$, where p_z is the momentum along the beam. The pseudorapidity is defined as, $\eta = -lntan(\theta/2)$. A key quantity which is measured is the transverse momenta which is the momenta in the x-y plane, p_T. Since the momentum along the beam axis is not known, it is the transverse momenta which is often studied. Moreover, p_T is Lorentz invariant with respect to the boost in the z-axis. Another quantity of interest is R, which is the distance in the (η, ϕ) plane, given by $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$.

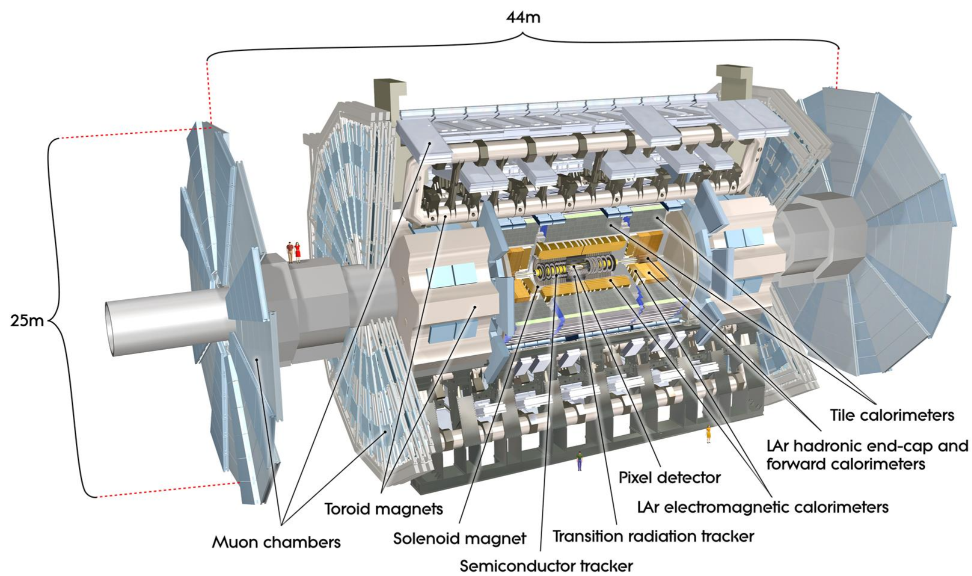


Fig. 3.1: Illustration of the ATLAS detector. Taken from [12]

Inner detector

The task of the inner detector is to perform tracking of charged particles and deduce their momenta. This is possible through the determination of the trajectory a particle takes, combined with information about the hits produced through ionisation when the particles interact with the low-dense materials lying in their path. The trajectory is reconstructed by seeing their paths in the 2T magnetic field from the superconducting magnets in which the inner detector is contained. It is further partitioned into three sub-detectors, the pixel detector, the semiconductor tracker and the transition radiation tracker, which have varying materials. Another function of this sub-detector is to perform vertexing (primary, secondary and tertiary) which helps in accurate determination of the location of the interaction point (IP). Lying nearest to the points where the protons collide, this sub-detector suffers from large amounts of radiation.

Calorimeters

The role of the calorimeters is to provide a measurement of the energy of particles. They are designed in such a way that they have alternating layers of passive and active material. The passive material initiates the showering, which is a cascade of particles which lose energy after covering each radiation length, while the active material detects these particles and carries out the measurement. There are two different calorimeters that are placed one after the other, electromagnetic (ECAL) and hadronic (HCAL), depending on the particles they measure the energy of. ECAL is involved with particles which shower electromagnetically and is made of liquid argon (passive) and lead (active), while HCAL plays the same role but with hadronically decaying particles and the material used here is scintillator tiles (passive) and steel (active). In the end caps, the materials are copper and liquid argon.

Muon Spectrometer

Most particles are contained after traversing the calorimeters, except muons and neutrinos. The muon spectrometer, positioned as the outermost layer in the system of sub-detectors, is used to track and detect muons. A distinct arrangement of magnets is employed for these purposes, in the barrel region and endcap region which cover different range of psuedorapidities. It is comprised of two chambers, the track and trigger chambers, which together are able to provide the measurement of the position and momentum of muons as well as identify the events that are interesting to be selected.

Trigger

With a bunch crossing of a few ns, the rate of occurrence of events is too high. This makes it computationally impossible to store all the events and thus a selection has to be made. This is carried out by a trigger system which comprises of three triggers and selects only interesting events. The first is a hardware trigger, the second a software trigger and a third which is done by the specific analyses. At each level of triggering, the reduction of the data to be recorded reduces by a significant amount.

Chapter 4

Phenomenological Studies of the $gg \to ZH$ process

This chapter provides a detailed account of the studies conducted on the process, $gg \to ZH$. The first section describes the process of event generation, followed by an analysis of these generated events using Rivet [13]. In the third section, the results of the analysis are presented.

4.1 Event Generation

With an abundance of data available from the detectors, simulations are needed to carry out exhaustive studies on them so they can serve as a guiding principle for interpretation and analysis of measured data. In particular, these simulations help us select a particular process which could be interesting to look at, details about the final states, and the backgrounds that would mimic the signal. The job of producing these simulations is done by Monte Carlo generators. Event generation consists of simulation of events of a process based on a theoretical model. The event files generated contain all the particles that are produced in a given process, with a complete description in terms of the locations of their production and decay, and energies and momenta. For a pp collision, the generation of events at the reconstruction level consists of three steps. The first of these is the production of the hard scatter process at the parton level, wherein the matrix element of the process is calculated. The particles generated at this level are then fed into a parton shower generator to produce parton showers and perform hadronisation. The third step consists of passing these showers and other final state particles through a detector simulation which imitates the conditions of a real-life detector, so the comparison to real data is direct.

Some Monte Carlo (MC) event generators commonly used in particle physics are MadGraph [14], Sherpa [15], Pythia [16], Herwig [17], and GEANT4 [18], and the choice varies depending on the target of the physics studies. For this study, the relevant event-generators are MadGraph(_aMC@NLO) and Pythia. MadGraph simulates pp collision truth-level events at LO and NLO accuracy, and the parton showering and hadronisation is performed by Pythia. Provided with a model (based on a selected theory) along with other inputs, MadGraph produces Feynman diagrams and computes cross sections of the generated process. A key feature is its ability to take as input BSM models enabling the exploration of BSM phenomenology. A plethora of BSM models are available interfaced with MadGraph like Higgs Doublet Model, Minimal Supersymmetric Standard Model and SMEFTatNLO.

SMEFTatNLO [19] is the model used here since the process $gg \to ZH$ requires one-loop matrix element computations. It realises dimension six SMEFT operators in the Warsaw basis, and the current implementation includes only two-fermion operators. The model is able to calculate the linear and quadratic contributions from SMEFT. It is assumed that the CKM matrix is unitary. Yukawa couplings of all fermions are ignored, except the top quark. The implementation uses five-flavour scheme. The flavour symmetry chosen is,

$$\mathrm{U}(2)_q \text{ x } \mathrm{U}(2)_u \text{ x } \mathrm{U}(2)_d \text{ x } [\mathrm{U}(2)_l \text{ x } \mathrm{U}(2)_e]^3$$

which treats the three generations of leptons the same, and the first two generations of quarks as the same. The third generation in the quark sector is treated differently to prioritize top quark operators [20]. Typically, specific flavour structures are targeted, owing to the presence of a large number of two- and four-fermion operators. The coefficient of interest with its operator definition is,

$$c_{\varphi t} \to i(\varphi^\dagger \overleftrightarrow{D_\mu}\varphi)(\bar{t}\gamma^\mu t)$$

with $\varphi^\dagger \overleftrightarrow{D_\mu}\varphi = \varphi^\dagger D^\mu \varphi - (D_\mu \varphi)^\dagger \varphi$ (φ corresponds to the Higgs field).

In this study, 6 $\times 10^5$ events of each process listed in Table 4.1 are simulated. The events are generated at a center-of-mass energy of 13 TeV and 13.6 TeV, using MadGraph5_aMC@NLO (version 2.6.5) [14] and the package SMEFTatNLO [19]. Three steps are involved to simulate events. First, restrict cards are created, which specify the value of the Wilson coefficient being considered (for SM, all values are set to zero), and information about the decays, masses of SM particles and the physical constants. Next, using these restrict cards, gridpacks are generated, which store the matrix element computation. This is done to avoid carrying out this calculation for each event, making the third stage much easier. Lastly, MadGraph5_aMC@NLO uses these matrix element calculations to simulate parton-level events and Pythia8 carries out the parton showering and hadronization. The events are produced in an EVNT format, which are then read by a Rivet routine to analyse the events. The commands used to generate the processes are summarized in Table 4.1, where q is defined

Table 4.1: Commands used for production of samples of different processes

gg → ZH	generate g g > h z QCD=2 QED=2 NP=2 [QCD]
qq → ZH	generate q q > h z QCD=2 QED=2 NP=2 [QCD]

as q = u, c, d, s, u∼, c∼, d∼, s∼, and p = g, q. The commands QCD, QED, and NP (new physics couplings) specify the coupling order of the respective interactions in the process and [QCD] accounts for next-to-leading order. For example, QCD=2 means the inclusion of all diagrams with QCD vertices <= 2. The final states this analysis looks at is $H \to b\bar{b}$ and $Z \to l^+l^-$, with $l = e, \mu$. With a branching ratio of 58.2% [21] (for m$_H$ = 125 GeV), the Higgs boson to $b\bar{b}$ is a natural choice, while Z boson to leptons provides a clean experimental signature. These decays are enforced in the Pythia showering commands, *''25:onMode = off''*, which enforces no decay of Higgs, followed by *"25:onIfMatch = 5 -5"* to impose the decay of Higgs only to b (anti-b) quarks : particle ID 5(-5). Similarly, decays of Z are forced to be electrons (positrons) or muons (antimuons) : Particle IDs: 11(-11) or 13(-13).

The value of Wilson coefficients considered are $c_{\phi t}$ = 1.0, 5.0, 10.0, 15.0, 20.0, and 50.0. As explained in Section 2.3, the current limits of these coefficients are $\mathcal{O}(20)$ [2], which is the value we use for carrying out the studies in this chapter. The remaining values are considered to compute the 95% Confidence Level exclusion limits on this coefficient in chapter 6. The cross sections of the processes (at $\sqrt{s}$ =13 TeV) are :

Table 4.2: Generated cross sections (pb) for various processes at $\sqrt{s}$ = 13 TeV

Process	Generated cross section (fb)
gg → ZH (SM)	52.58 ± 0.02
gg → ZH (SM + EFT ($c_{\phi t}$ = 1))	49.67 ± 0.04
gg → ZH (SM + EFT ($c_{\phi t}$ = 5))	39.11 ± 0.03
gg → ZH (SM + EFT ($c_{\phi t}$ = 10))	29.89 ± 0.03
gg → ZH (SM + EFT ($c_{\phi t}$ = 15))	25.23 ± 0.02
gg → ZH (SM + EFT ($c_{\phi t}$ = 20))	24.87 ± 0.01
gg → ZH (SM + EFT ($c_{\phi t}$ = 50))	158.1 ± 0.01
qq → ZH	723.6 ± 2.2

Comparing to theoretical predictions as presented by the LHC Higgs Cross Section Working Group in [22], the cross sections for $gg \to ZH$ and $qq \to ZH$ exhibit a deviation of ∼21% and ∼18% respectively. This is because of the difference in the orders of accuracy, for example, the former process is reported with a NLO (next-to-leading order) + NLL (next-to-leading logarithmic) accuracy, while MadGraph takes into account only NLO accuracy.

We observe from Table 4.2 that the $gg \to ZH$ process has a much smaller cross section than the $qq \to ZH$ process. The cross sections for SM + EFT processes decrease as the value of the Wilson coefficient increases, as compared to SM, due to a destructive interference of the SM term and the EFT term. The decrease occurs till $c_{\phi t}$ = 20, after which the quadratic term of EFT starts to dominate this linear term.

4.2 Analysis Strategy

4.2.1 Rivet Toolkit

RIVET (Robust Independent Validation of Experiment and Theory) [13] is a C++ framework used to validate and tune MC event generation and perform analysis on simulated events. Its biggest asset is its ability to produce distributions of kinematic quantities efficiently which can be measured directly in a process, which is why it has been extensively used to study SM and BSM processes.

HepMC events are fed into a Rivet routine, where they are analyzed to produce an output in the form of histograms. These histograms are stored in YODA (Yet more Objects for Data Analysis) format text files. A routine consists of three methods:

- The **init** method is called only during the initialization. Here, the final state projections are declared. Projections are objects which calculate the physical observables, like the final state particles that satisfy some definite criterion, for example particle IDs and kinematic properties. The output histograms to be produced, with detailed information about their binnings and labels, are also booked in this method.

- The **analyze** method forms the main body of the routine, and is carried out once for each event. Here, event selection is implemented and reconstruction of particles is done. Their kinematic properties are computed with predefined classes and functions and used to fill histograms. Each event has a weight associated to it, which accounts for the probability of a particular event to occur. During the filling of the histograms, these event weights are applied to each kinematic distribution under observation. Some relevant utility classes that are used in this analysis are Jets (depiction of a clustered jet), DressedLeptons (representation of leptons dressed with photons) and FourMomentum (class offering access to kinematic properties of particles).

- The **finalize** method is called after every run is finished, to normalize the histogram to a given value. In accordance with the purpose, either 'normalize' or 'scale' operations are used. To study shape effects of kinematic distributions, the former method is utilized which normalizes the area under the histogram curve to the set value. If the user wants to make comparisons with data, that is, normalize the histograms to the absolute cross section, the scale method is employed. The scaling factor used is (branching ratio $\times$) cross section $\times$ (sum of weights)$^{-1}$. In order to get the total number of events, this number is multiplied by the luminosity.

4.2.2 Reconstruction Algorithm

For the considered final state ($H \to b\bar{b}$ and $Z \to l^+l^-$), leptons and b-jets have to be reconstructed.

- Jets Reconstruction : Since jets are a collection of charged and neural particles originating from the fragmentation and hadronization of partons, their reconstruction can provide details of the kinematic properties of the original parton. The information from the calorimeters and tracker is utilized to retrieve jets, which is done by jet reconstruction algorithms. Jets have specific topology which makes them identifiable. There are two main types of algorithms typically employed, cone and sequential reconstruction algorithms. This analysis uses a sequential reconstruction algorithm named the anti-k_T algorithm [23]. It examines a list of particles, and clusters them together if they meet certain criterions. The criterion used is a distance measure. For each particle i, two distances are calculated, d_{ij} (with every other particle, j) and a beam distance d_{iB},

$$d_{ij} = min(p_{T,i}^{-2}, \mathrm{p}_{T,j}^{-2})\frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = p_{T,i}^{-2}$$

R here gives the jet radius, a measure of the extension of the cone around the jet. The algorithm starts by identifying the particle with the largest p_T. It computes all d_{ij}, selects the particle with the smallest d_{ij}, and combines these two particles to get another particle. If, however, d_{iB} is smaller, it calls this particle as a jet and moves on to the next particle in the list. The procedure is repeated until all particles in the list are exhausted and combined into some jet cluster. In ATLAS, $R = 0.4$ is used for resolved b-jets and $R = 1$ for boosted jets.

The routine used here begins with the definition of a FinalState object with $|\eta| < 2.5$, which represents all the final state particles in this pseudorapidity range. This cut exists because of the limit on the acceptance of the linear detector. The VetoedFinalState object is used to define the final state particles which will help to find jets. In this object, neutrinos and muons are removed. Neutrinos are generally not considered because it is difficult to reconstruct them, and muons are avoided so they are not misreconstructed as jets. The FastJets Projection is then used to define jets, specifying also the reconstruction algorithm and R to be used. We use $R = 0.4$ here. Kinematic cuts of $|\eta| < 2.5$ and $\mathrm{p}_T > 20$ GeV are enforced for maximal trigger efficiency and good energy resolution. Since we are only interested in b-jets arising from the decay of Higgs boson, we tag the identified jets by the 'btagged' function. This returns a bool value, reporting if the jets have at least one b-jet or not. Only those events are stored which contain two b-tagged jets. Selections are imposed on the invariant mass of the b-jets, so they correspond to a Higgs boson. For this reason, events with $m_{bb} > 140$ or $m_{bb} < 100$ are vetoed.

- Lepton Reconstruction : Leptons and photons, in general, are reconstructed using information from trackers and energy clusters in calorimeters. In the routine, the IdentifiedFinalState object is used to represent photons (to dress leptons), electrons and muons arising from the final state. They are recognised using their PIDs. The dressing of the leptons means that bare charged leptons are combined with photons surrounding them in a conical sense (defined by R). Their four vectors are added to give a dressed lepton. This is done to take into consideration the QED FSR. Then, PromptFinalState is used to recognise the final state particles which originate from the hard process. After this identification, dressed leptons are chosen with a cone of radius $R = 0.1$. Similar to jets, the standard cuts of $|\eta| < 2.5$ and $\mathrm{p}_T > 25$ GeV applied. Only those events are stored which contain one pair of electrons or one pair of muons. Similar to the b-jets, events with $m_{ll} > 101$ or $m_{ll} < 91$ are vetoed.

4.2.3 Kinematic observables of interest

The kinematic observables of the final state particles, e.g. the transverse momenta, invariant mass, rapidity, etc. are of primary interest to us. Our goal here is to find event quantities that are sensitive to EFT effects, and the ones that hold discriminating power to distinguish the process $gg \to ZH$ from the backgrounds. The LHC coordinate system uses p_T, η and ϕ

instead of p_x, p_y and p_z, as mentioned above. Knowing these three quantities, we can fully describe a particle's three momentum.

We examine the kinematic distributions for three processes, $gg \to ZH$ (SM), $gg \to ZH$ (SM + EFT ($c_{\phi t} = 20$)) and $qq \to ZH$. The only background considered here is the quark initiated ZH production because this background shares exactly the same final state of ZH and in terms of kinematic distributions, it is the most difficult background to separate. For these three processes, we investigate the following observables:

- $p_T \in$ (Z, H, ZH, b_1, b_2, l_1, l_2)
- m $\in$ ($l_1 l_2$, $b_1 b_2$, ZH)
- $\Delta\phi \in$ (HZ, $b_1 b_2$, $l_1 l_2$, $b_1 l_1$)
- $\Delta\eta \in$ (HZ, $b_1 b_2$, $l_1 l_2$, $b_1 l_1$)
- $\Delta R \in$ (HZ, $b_1 b_2$, $l_1 l_2$, $b_1 l_1$)

where b_1 (l_1) corresponds to the leading jet (lepton), i.e. jet (lepton) with higher transverse momentum, and b_2 (l_2) to the subleading jet (lepton).

In order to study correlations between different quantities, we also investigate the 2D distributions of all 28 combinations of the following 1D variables:

- m_{ZH}
- p_T^Z
- $\Delta\phi_{bb}$
- $\Delta\eta_{bb}$
- $\Delta\phi_{ZH}$
- $\Delta\eta_{ZH}$
- ΔR_{bb}
- ΔR_{ZH}

4.3 Results

4.3.1 Identification of Kinematics

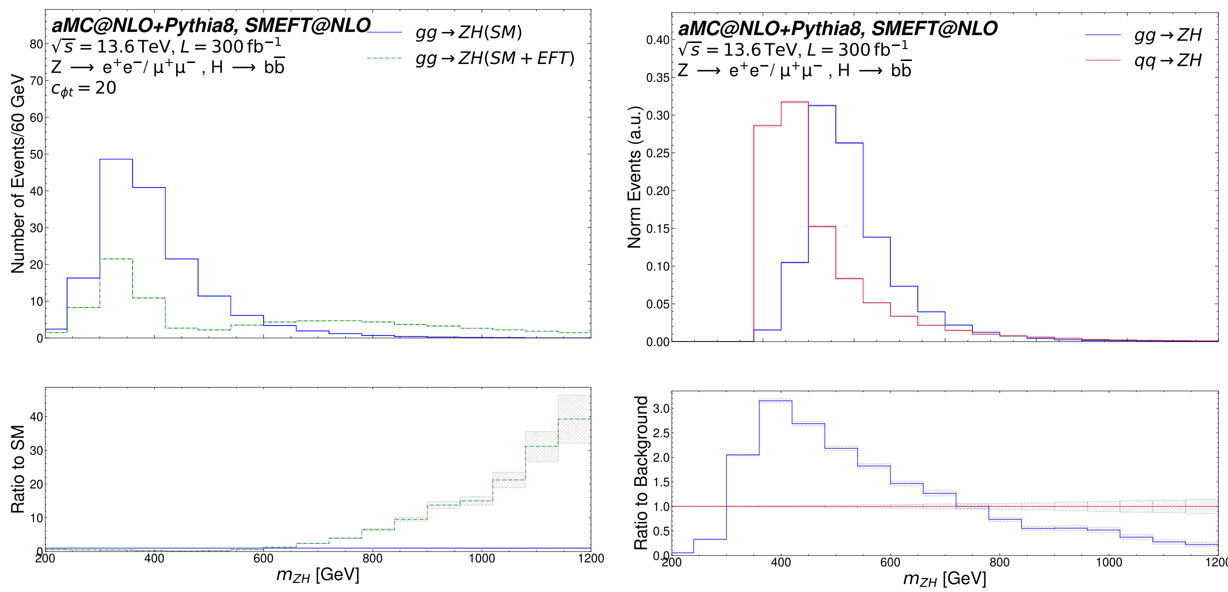
The results of this analysis presenting the kinematic distributions of the final state particles are discussed in this section. The centre-of-mass energy used is $\sqrt{s} = 13.6$ TeV, corresponding to Run-3 data, and the total luminosity is (expected to be) $L = 300$ fb^{-1}. We choose the value of $c_{\phi t} = 20$, which is consistent with the current limits, as explained in Section 2.3. At this value, there is a strong interplay between the linear (interference with SM term) and the quadratic terms of the associated EFT operator. In the plots given below, the left side distributions are targeted to isolate regions where EFT effects are notable in comparison to SM. For these plots, we use the normalization of the plots to be the total expected events, $N = L \times \sigma \times \epsilon \times BR$, where L signifies the luminosity, σ the cross section, ϵ the efficiency of the selected events and BR the branching ratio for the chosen process. The right side plots focus on $gg \to ZH$ (SM) and $qq \to ZH$, to highlight the differences which would be valuable for separation of these two processes. For these plots, we normalize the distributions so that the area under the curves is a fixed value ($= 1$). This is done so we can specifically focus on shape effects, since the distributions are still overwhelmed with the considered background as the cross section for quark initiated ZH production is much greater ($\mathcal{O}(10)$) than the gluon initiated ZH.

m_{ZH} and p_T^Z distributions

Upon the introduction of the $c_{\phi t}$ operator in the top quark loop, the top quark, Higgs boson and Z boson interactions change. EFT effects manifest themselves as a modification of the mediator particle's properties, and hence the energy scale and momentum transfer of the process. Any deviations in the distributions in the SM + EFT model from those of the SM should, thus, be attributed to a modification of the energy. Figure 4.1 shows the distributions for m_{ZH} for the three processes. This is a key variable because conservation of momentum implies $\sqrt{s} = m_{ZH}$, and any alterations in the former will be directly seen in the latter. Figure 4.1a displaying the invariant mass demonstrates that the spectrum for the SM + EFT model has an interesting region in the high mass range ($m_{ZH} > 600$ GeV) where there is a significant number of events, with an absence of similar events in the SM. This is exactly the manifestation of the changes explained previously. Figure 4.1b compares the same variable in the $gg \to ZH$ (SM) and $qq \to ZH$ processes. We observe here that the distribution for the $qq \to ZH$ process is shifted to give a softer spectrum. This happens because of the top-quark loop in the Feynman diagram of the $gg \to ZH$ process, when the two top quarks satisfy the condition of being on-shell, as was described in Section 2.3. The transverse momentum of the Z boson is correlated with the invariant mass of the ZH system. Being a directly measurable quantity, it is a principal variable to study. The transverse momentum of the Z boson, similar to m_{ZH}, shown in Figure 4.2a reveals that there is a considerable number of Z bosons which have a high momenta ($p_T^Z > 300$ GeV) in the SM + EFT model. On the right hand side, in Figure 4.2b, it is seen that the bulk of events in $qq \to ZH$ are populated in the lower momenta region.

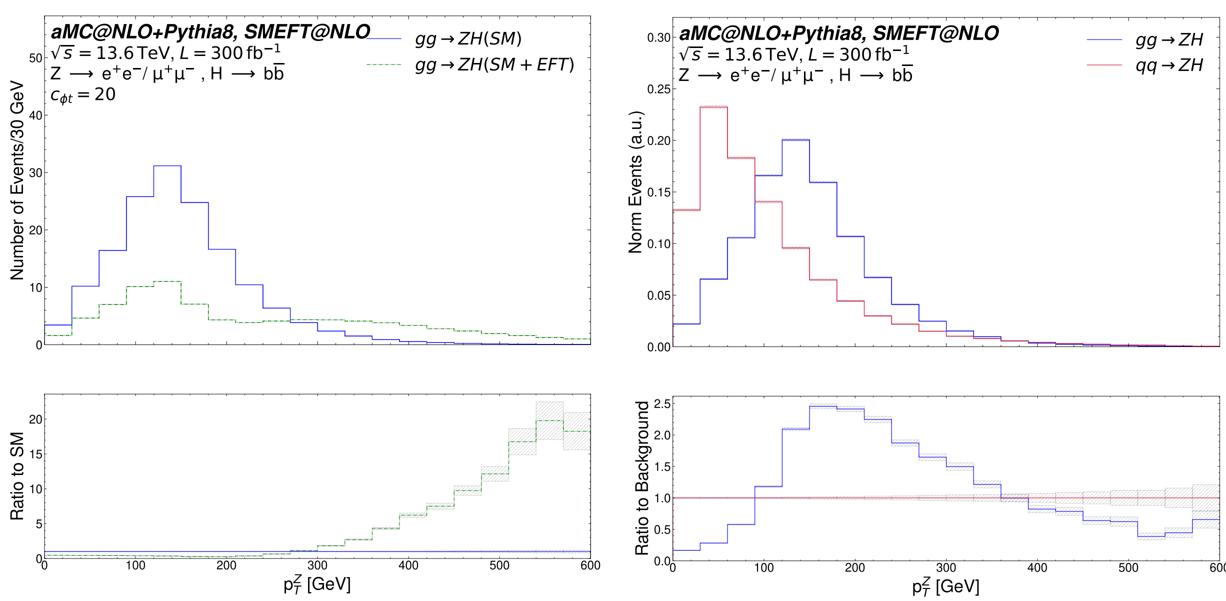
Angular distributions : $\Delta\phi_{bb}$

Next, an angular distribution of the $b\bar{b}$ system is presented. The plots for the two-lepton system emerging from the decay of the Z bosons are not showcased here. This is because the two final state bosons are symmetrical in terms of their transverse momenta and are produced back-to-back. No SMEFT operators are introduced in the decay of the Higgs and Z bosons, and so any digression observed in the distributions of the SM + EFT model from the SM in the $b\bar{b}$ system would be replicated in the l^+l^- system. We choose to present the plots for $b\bar{b}$ system since EFT effects are (slightly) more enhanced than in l^+l^-. Figure 4.3 shows the plot for the difference between the ϕ angles of the two bottom quarks. In Figure 4.3a, we see that in the SM + EFT model, the two quarks favor being emitted close to each other, in contrast to SM. This is due to the more boosted transverse momentum of their parent, the Higgs boson. This effect is repeated in the difference in the psuedorapidities, η, of the two quarks. Figure 4.3b with $\Delta\phi_{bb}$ for $gg \to ZH$ (SM) and $qq \to ZH$ depicts that for the latter process, the bottom quarks are emitted back-to-back, as opposed to the former process. Further, the distributions for this variable are striking because EFT effects and background contributions have their bulk of events in opposite regions of phase space. For the other variables, for example, the differences between the ϕ and η of the Higgs and Z bosons, the angular distributions exhibit only subtle deviations arising only from overall cross section normalization.



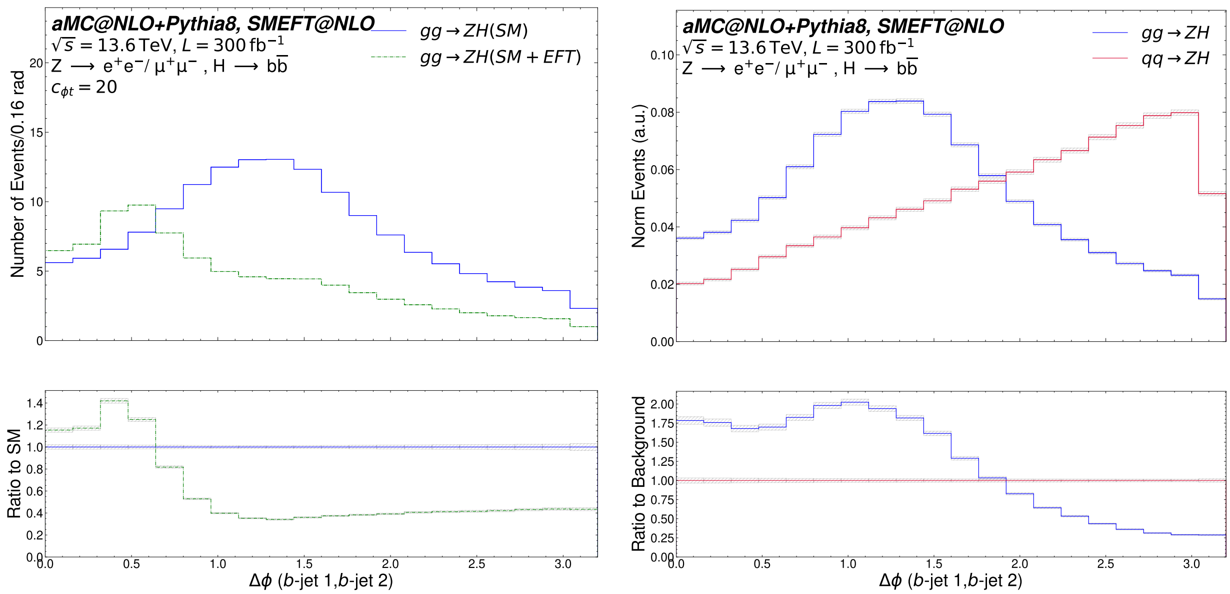
(a) (b)

Fig. 4.1: Invariant mass of ZH system, m_{ZH}, for $gg \to ZH$ (SM) and $gg \to ZH$ (SM+EFT) (left) and for $gg \to ZH$ (SM) and $qq \to ZH$ (right)



(a) (b)

Fig. 4.2: Transverse momentum of Z boson, p_T^Z, for $gg \to ZH$ (SM) and $gg \to ZH$ (SM+EFT) (left) and for $gg \to ZH$ (SM) and $qq \to ZH$ (right)



(a) (b)

Fig. 4.3: Difference in the azimuthal (ϕ) angle of the two bottom quarks, $\Delta\phi_{bb}$, for $gg \to ZH$ (SM) and $gg \to ZH$ (SM+EFT) (left) and for $gg \to ZH$ (SM) and $qq \to ZH$ (right)

2D distributions

Figures 4.4a, 4.5a and 4.6a report the expected correlation between m_{ZH} and p_T^Z for $gg \to ZH$ (SM), $gg \to ZH$ (SM + EFT) and $qq \to ZH$ respectively. These plots reaffirm the results that there are events in the higher mass and transverse momenta regime for $gg \to ZH$ (SM + EFT), as opposed to $gg \to ZH$ (SM). For the $qq \to ZH$ background, the spectra of mass and transverse momenta are shifted to the lower regimes. The plots are scaled to the total number of expected events. Similarly, figures 4.4b, 4.5b and 4.6b show the 2D plots for p_T^Z vs $\Delta\phi_{bb}$ for $gg \to ZH$ (SM), $gg \to ZH$ (SM + EFT) and $qq \to ZH$ respectively. In $gg \to ZH$ (SM + EFT), boosted events relating to Z bosons with higher momenta are the events where the decayed products of Z and Higgs bosons have small angles between them. Studying these plots highlights the part of the phase space to be scrutinised to explore EFT effects. For the $qq \to ZH$ process, when Z bosons have smaller transverse momenta, the decayed leptons are emitted with large angles between them.

4.3.2 Significance of a cut-based approach

These results show that a selection can be made in this 2D plane of p_T^Z vs $\Delta\phi_{bb}$ (or, equivalently, m_{ZH} vs $\Delta\phi_{bb}$) to reject this background to the greatest degree. A series of diagonal selections (cuts) is made in this plane to select the one that yields the highest significance for the two signals we consider, which are $gg \to ZH$ (SM) and $gg \to ZH$ (SM + EFT). The cut of $\Delta\phi_{bb} < (\frac{1.5}{300} \times p_T^Z + 1.5)$ gives a significance, $(\frac{S}{\sqrt{S+B}}) = 4.34$ for $S = gg \to ZH$ (SM) and $B = qq \to ZH$, and 2.57 for $S = gg \to ZH$ (SM+EFT) and $B = qq \to ZH$. The smaller value of the significance for the $gg \to ZH$ (SM + EFT) process is, of course, arising from the lower cross section. Here, we only include the $qq \to ZH$ background, and to make a prediction, we must incorporate all other backgrounds. This is done to only obtain a rough estimate and conclude if these observables can assist us in carrying out the task of separation of signal from backgrounds. The full significance studies are performed in the next chapter. In conclusion, we have discerned variables which have pointed to regimes that can be used to isolate EFT effects and reject the $qq \to ZH$ background.

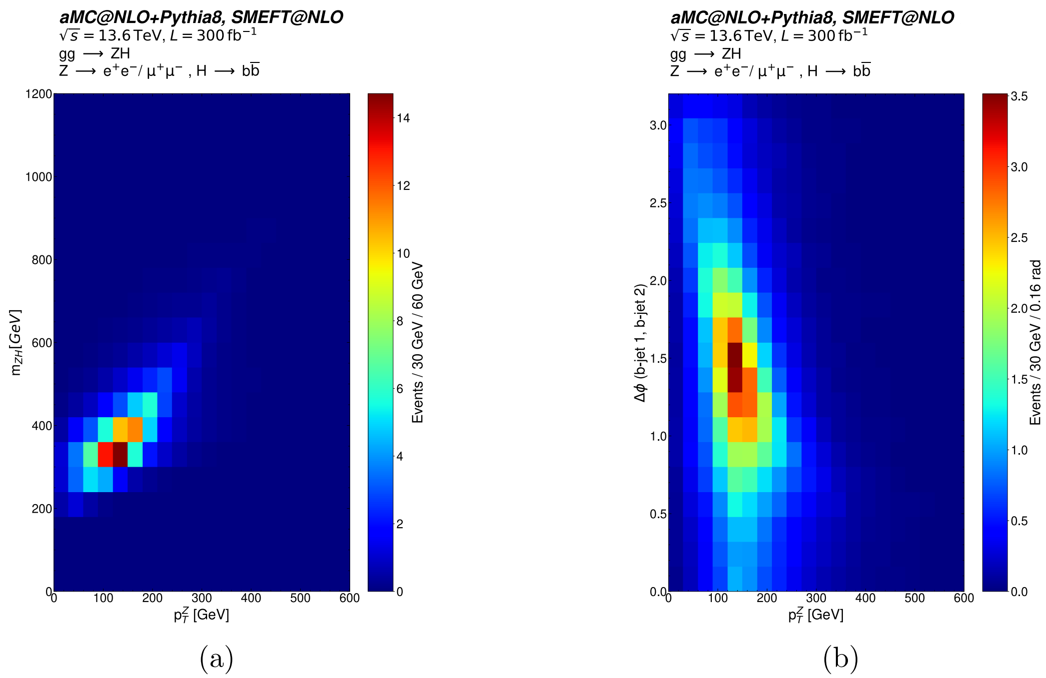
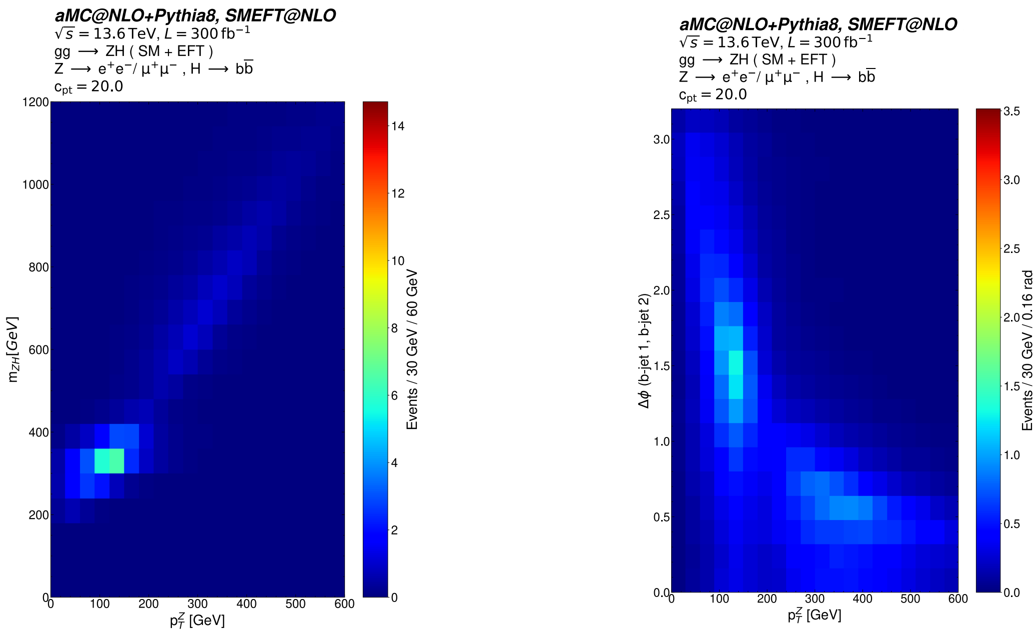
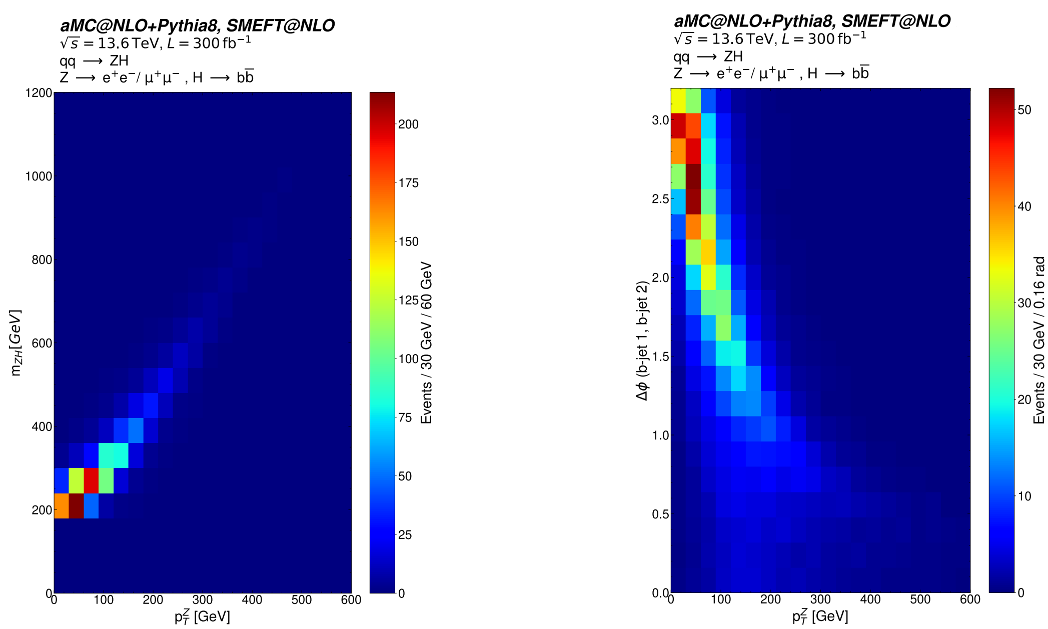


Fig. 4.4: 2D distribution between m_{ZH} and p_T^Z (left) and between p_T^Z and $\Delta\phi_{bb}$ (right) for $gg \to ZH$ (SM)



(a) (b)

Fig. 4.5: 2D distribution between m_{ZH} and p_T^Z (left) and between p_T^Z and $\Delta\phi_{bb}$ (right) for $gg \to ZH$ (SM+EFT)



(a) (b)

Fig. 4.6: 2D distribution between m_{ZH} and p_T^Z (left) and between p_T^Z and $\Delta\phi_{bb}$ (right) for $qq \to ZH$

Chapter 5

Background Rejection using Deep Learning Techniques

This chapter demonstrates the usage of deep learning techniques to separate the signal events from the background events. An overview about the motivation for background rejection is first presented in section 5.1, followed by some basics of the technique in section 5.2. Section 5.3 illustrates the dataset used for the signal and background samples. Sections 5.4 and 5.5 discuss the details of the algorithm used and final results.

5.1 Background rejection

In particle physics, for a discovery to be established, a global significance of 5σ in the signal is required over background fluctuations. This means that the probability of the observed signal being a statistical fluctuation in the background is 0.006% or less. Achieving this level of significance requires us to understand not only the physics of the signal process, but also the backgrounds which imitate the signal. And so, it is a necessity to examine the distributions of backgrounds precisely, to be able to identify regimes where the background can be separated from the signal as effectively as possible.

In order to do this, various techniques are used, which include usage of cut-based analysis or advanced machine learning algorithms. In the simplest cut-based analysis, event quantities, for example the transverse momenta, or the azimuthal angle, of the final states corresponding to signal and backgrounds are studied. Cuts are made in the phase space so we maximally reject the background. On the other hand, machine learning algorithms process the input features which are these kinematic distributions, identify complex correlations between them and use this information to discriminate between signal and background.

5.2 Neural Network Basics

One of the machine learning algorithms commonly used are artificial neural networks. A neural network (NN), inspired from the neuron network of the human brain, is a system which is used to process data using connections between several nodes. At its heart, it is a mapping which upon feeding some input features gives us an output. A typical network consists of an input layer, one or more hidden layers and an output layer.

We describe the basic architecture of a neural network, as shown in Fig. 5.1. To begin the process, some features (in our case, the kinematic distributions of particles), are fed as inputs to a network's input layer. Each feature acts as a node in this layer. Each node has a weight attached to it, which signifies its importance. A higher (lower) weight assigned to a feature indicates its greater (lesser) influence. An activation function is associated to each layer that has the role of making the output non-linear. Additionally, a bias term is applied to each node to introduce an offset in the output of the activation function. Thus, upon taking the weighted sum of the inputs and the bias term, the activation function decides whether a node is activated or not. The information from all these nodes is propagated to the next layer, called the hidden layer. This layer consists of a certain number of predefined number of nodes. The processing of information again occurs here, and weights are assigned to the nodes. This process is repeated for each subsequent hidden layer. Ultimately, the processed information is passed to an output layer. The output varies depending on the problem being addressed by the network. These weights dictate the decision-making process of the network and so, the goal of the network training is to optimise these weights to produce the best possible predictions.

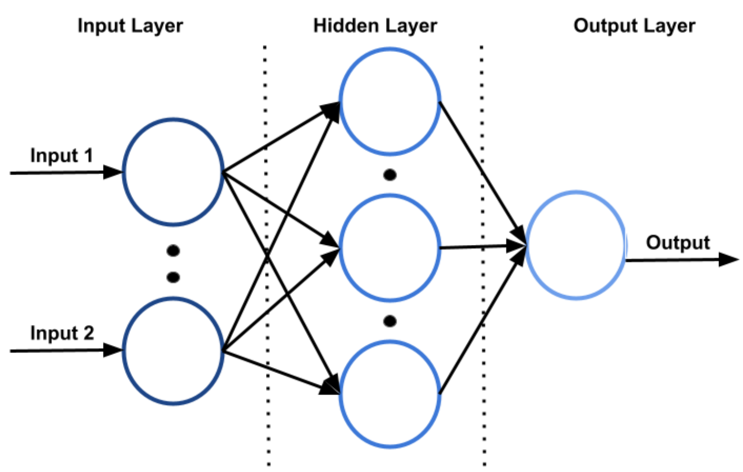


Fig. 5.1: Architecture of a simple neural network

The optimisation of the weights is done using backpropagation and gradient descent. After a first iteration of the computation of predictions, a 'loss function' is calculated, which expresses how good a network performs. It is typically a function of the difference between the prediction and the true value. After this computation, the backpropagation method uses the chain rule of calculus to calculate the gradient of the loss with respect to all the weights. Gradient descent is the algorithm which uses these gradients to update the weights by minimising the loss function.

A neural network can be used to solve a variety of problems, categorized into three types of learning: supervised learning, unsupervised learning and reinforcement learning. Supervised learning has two subcategories, namely classification and regression. The task of separating signal events from the background events falls into the category of supervised learning, and is seen as a binary classification problem. For this scenario, the NN is presented with some inputs and output labels. The network learns to predict the output of new unseen inputs.

5.3 Input Dataset

In this chapter, we use the ATLAS samples from the VH ($V = W$ or Z) analysis [8] as input data for the signal and backgrounds. A description of the samples is provided below.

5.3.1 ATLAS Samples

The ATLAS samples for the $pp \to VH \to (ll/l\nu/\nu\nu)b\bar{b}$ channel [8], with $l = e, \mu$ corresponds to data taken at the LHC at a centre-of-mass energy of $\sqrt{s} = 13$ TeV during the period 2015-2018. This data amounts to a total integrated luminosity of (139 ± 2) fb^{-1} [24]. The backgrounds include top quark ($\mathrm{t\bar{t}}$ and single top), $V+$ jets, and Diboson processes. The channels looked at are 0-, 1- and 2-leptonic final states. To reduce theoretical dependencies (models and uncertainties), and to enhance sensitivity, the Simplified Template Cross Section (STXS) framework [22] is used to carry out differential cross section measurements. In this framework, the phase space is partitioned into various bins of kinematic regions defined for some event quantities. This assists in isolating regions with increased sensitivity to BSM effects. The variables chosen for this channel are the transverse momenta of the V boson, p_T^V and the number of jets, n_j, produced. The bins correspond to $\mathrm{p}_T^V, \in$ (0, 75, 150, 250, 400, ∞) GeV and $\mathrm{n}_j \in (0, 1, \geq 2)$. In the current implementation, $qq \to ZH$ and $gg \to ZH$ are combined into one measurement because of lack of experimental sensitivity in separating these two processes. For the same reason, the first bin with $\mathrm{p}_T^V < 75$ GeV is not considered.

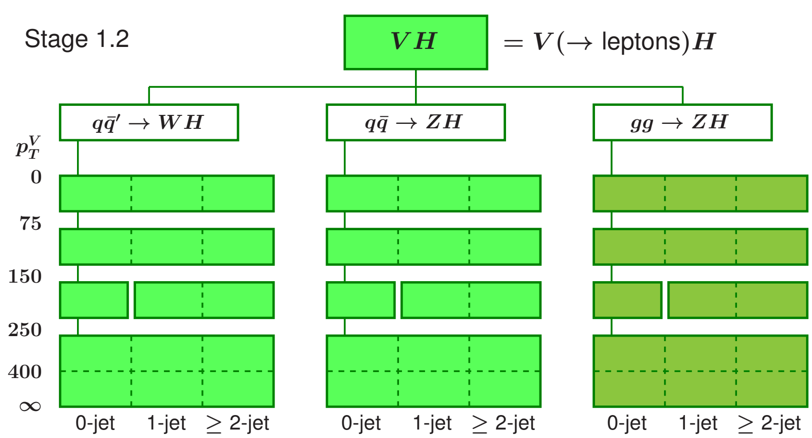


Fig. 5.2: Schematic representation of Stage 1 of STXS binning scheme for VH samples [25]

Details about the generators of the signal and background processes are listed in Fig. 5.3. For this analysis, we emphasize that the relevant channel is that of 2-leptons. The selected events must meet the following criteria :

- Exactly 2 leptons of the same flavour and opposite charges passing basic identification and isolation criteria

- For at least one lepton, $p_T > 27$ GeV
- 81 GeV $< m_{ll} <$ 101 GeV
- Exactly 2 b-tagged jets
- For leading b-tagged jet, $p_T > 45$ GeV
- For jets, $p_T > 20$ GeV if $|\eta| < 2.5$ and $p_T > 30$ GeV if $2.5 < |\eta| < 4.5$

Process	ME generator	ME PDF	PS and Hadronisation	UE model tune	Cross-section order
Signal, mass set to 125 GeV and $b\bar{b}$ branching fraction to 58%					
$qq \to WH \to \ell\nu b\bar{b}$	Powheg-Box v2 [54] + GoSam [57] + MiNLO [58, 59]	NNPDF3.0NLO$^{(\star)}$ [55]	Pythia 8.212 [45]	AZNLO [56]	NNLO(QCD)$^{(\dagger)}$+ NLO(EW) [60-66]
$qq \to ZH \to \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg-Box v2 + GoSam + MiNLO	NNPDF3.0NLO$^{(\star)}$	Pythia 8.212	AZNLO	NNLO(QCD)$^{(\dagger)}$+ NLO(EW)
$gg \to ZH \to \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg-Box v2	NNPDF3.0NLO$^{(\star)}$	Pythia 8.212	AZNLO	NLO+ NLL [67-71]
Top quark, mass set to 172.5 GeV					
$t\bar{t}$	Powheg-Box v2 [72]	NNPDF3.0NLO	Pythia 8.230	A14 [73]	NNLO+NNLL [74]
s-channel single top	Powheg-Box v2 [75]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [76]
t-channel single top	Powheg-Box v2 [75]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [77]
Wt	Powheg-Box v2 [78]	NNPDF3.0NLO	Pythia 8.230	A14	Approximate NNLO [79]
Vector boson + jets					
$W \to \ell\nu$	Sherpa 2.2.1 [48, 80, 81]	NNPDF3.0NNLO	Sherpa 2.2.1 [82, 83]	Default	NNLO [84]
$Z/\gamma^* \to \ell\ell$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NNLO
$Z \to \nu\nu$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NNLO
Diboson					
$qq \to WW$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NLO$^{(\ddagger)}$
$qq \to WZ$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NLO$^{(\ddagger)}$
$qq \to ZZ$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NLO$^{(\ddagger)}$
$gg \to VV$	Sherpa 2.2.2	NNPDF3.0NNLO	Sherpa 2.2.2	Default	NLO$^{(\ddagger)}$

Fig. 5.3: Monte Carlo generators used for production of signal and background processes. Adapted from [8]

.

5.3.2 EFT Reweighting

Motivation

The unavailability of ATLAS official EFT samples for the (scan of) Wilson coefficients makes it necessary to use an alternative approach. The MadGraph samples obtained as described in Chapter 4 serve a dual purpose. First, they act as an inspiration for kinematic variables which are most relevant, i.e. discriminating for EFT effects and can be used to execute the task at hand. Secondly, they are used to derive correction factors to be applied to the official ATLAS $gg \to ZH$ samples in order to mimic EFT effects.

Technique

During event generation, each event appears with a certain Monte Carlo (MC) weight, factoring in the probability of its occurrence in real-life. When constructing histograms for specific kinematic variables, these MC weights are used to populate the histogram bins instead of simply counting the number of events. This makes it possible to construct EFT samples using the SM samples, by modifying the per-bin MC weights of the latter. In this study, we make use of the 2D distribution of p_T^Z and $\Delta\phi_{bb}$. As established, p_T^Z is directly correlated with m_{ZH}, and any changes due to EFT contributions would be directly seen in this variable. We also identified $\Delta\phi_{bb}$ as an important angular quantity that has the potential to capture additional EFT effects. To create the EFT samples from the SM samples, we divide this 2D phase space into a 20 × 20 grid and take the ratio of the MC weights of the (truth-level) EFT sample and the (truth-level) SM sample bin-by-bin. This ratio of the MC weights gives us the correction weights, which are applied to the ATLAS (reconstructed-level) $gg \to ZH$ samples to get the EFT samples (refer equation (5.1)) for this process. We make the assumption that the detector efficiencies that translate the truth level information to reconstructed level information remain the same for the EFT and SM samples.

$$\left(\frac{\sigma_{EFT}}{\sigma_{SM}}\right)_{Truth} \times (\sigma_{SM})_{Reco} = (\sigma_{EFT})_{Reco} \tag{5.1}$$

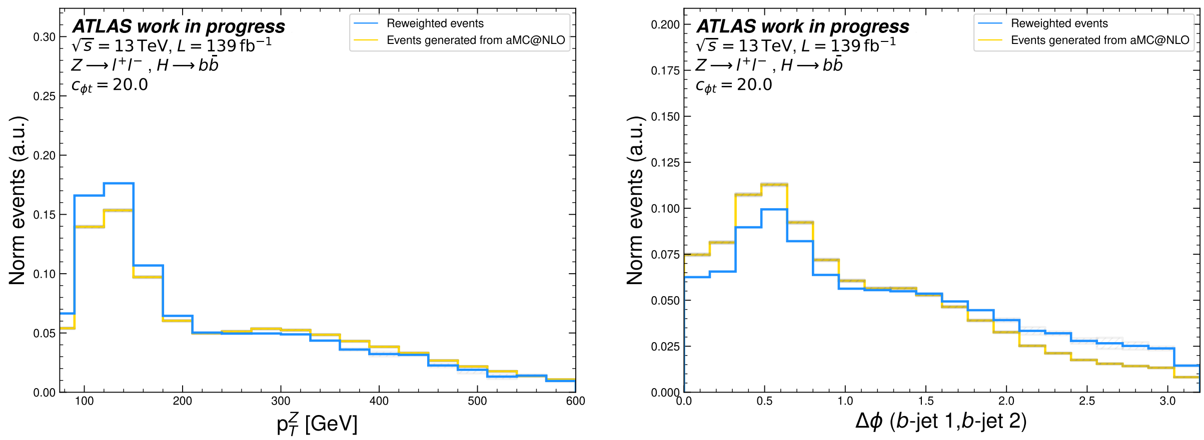
Comparison of Results

We present the results of the application of the reweighting technique to get EFT $gg \to ZH$ samples from ATLAS SM $gg \to ZH$ samples. This is done by drawing a comparison between the kinematic distributions of samples generated from MadGraph5_aMC@NLO+Pythia8 and samples obtained through the reweighting technique.

In the last chapter, the centre-of-mass energy was set to $\sqrt{s}$ =13.6 TeV and luminosity to L = 300 fb^{-1} to get the number of events. To align with the current analysis, we regenerate these events from MadGraph5_aMC@NLO+Pythia8 so that they correspond to $\sqrt{s}$ = 13 TeV and the luminosity is set to L = 139 fb^{-1}. We confirm that there are no shape differences introduced when switching between these energies. Fig. 5.4 depicts the p_T^Z and $\Delta\phi_{bb}$ distributions for the two different kinds of samples. The overall effect is replicated in the ATLAS samples. A point worth mentioning is that in the event generation process using Madgraph, the generated events are produced at the truth level, without incorporating detector simulation. On the other hand, ATLAS samples are generated at the reconstructed level, where the simulation of the detector is applied. The discrepancy that arises might be due to this fact, however, it does not impact our results. Moreover, when comparing, we should be mindful of the fact that the definitions of the b-jet differ in the two levels of construction. Since we used p_T^Z and $\Delta\phi_{bb}$ to perform the reweighting, it is expected that these distributions exhibit the closest match. To ensure that the reweighting technique accurately reproduces the effects in the other variables, we evaluate all the other event quantities listed in section 4.2.3. In Fig. 5.5, we present two such distributions, $\Delta\phi_{ZH}$ and $\Delta\eta_{ZH}$. The plots show that the distributions of the two samples align well with each other.

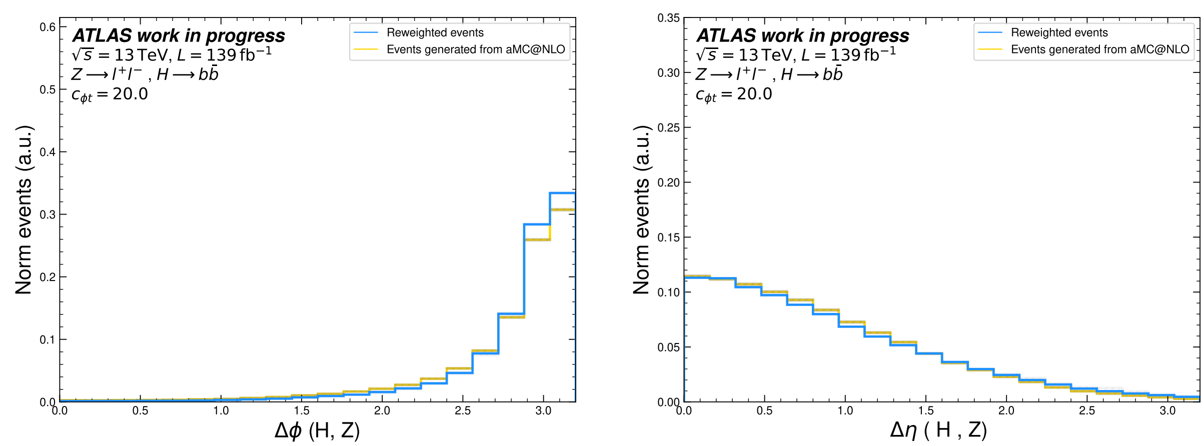
5.4 Algorithm: Deep Feed-Forward Neural Network

Feed forward neural networks are trained to separate the signals $gg \to ZH$ (SM) and $gg \to ZH$ (SM + EFT) from the backgrounds, $qq \to ZH$ and $pp \to Z + jets$. Our initial focus is on the $qq \to ZH$ background because this is the more challenging background, in terms of similarities in kinematics. We use two different networks, for $gg \to ZH$ (SM) vs $qq \to ZH$ and $gg \to ZH$ (SM + EFT) vs $qq \to ZH$, to assess their robustness in both cases.



(a) (b)

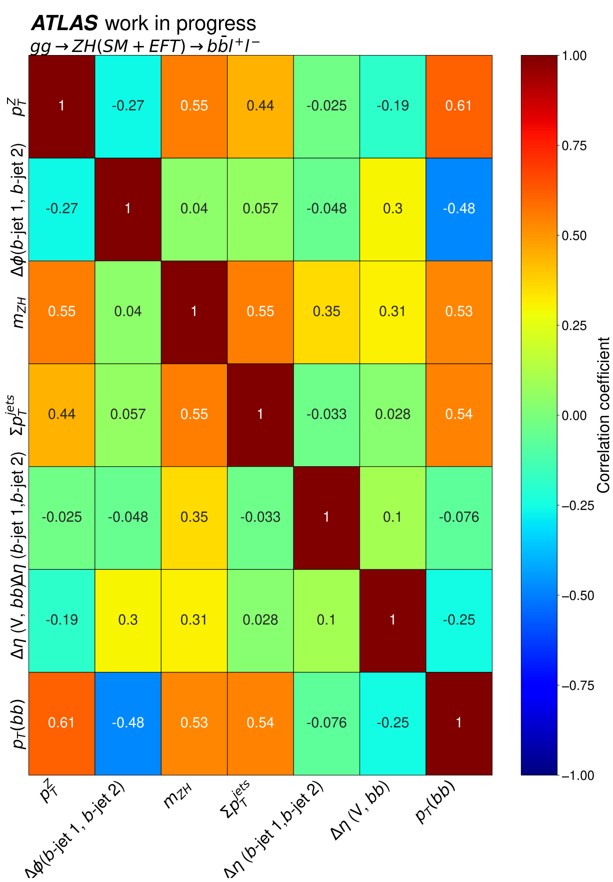
Fig. 5.4: Comparison of distributions generated by MadGraph5_aMC@NLO+Pythia8 and by reweighting of ATLAS samples



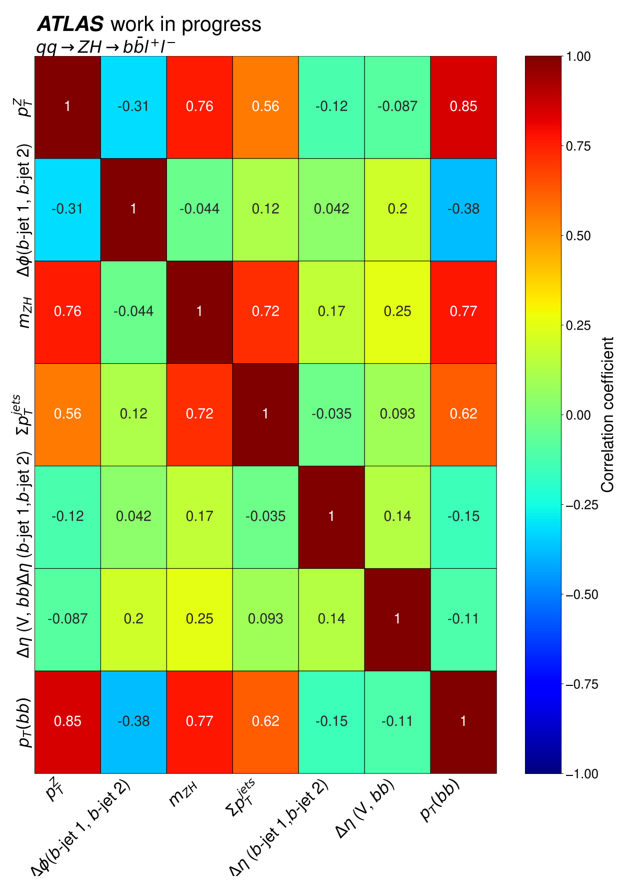
(a) (b)

Fig. 5.5: Comparison of distributions generated by MadGraph5_aMC@NLO +Pythia8 and by Reweighting of ATLAS samples

Subsequently, we test the SM and EFT samples on both these networks and select the more effective one. Given that the $qq \to ZH$ background shares the exact same final state as the signal, it poses a considerable challenge for separation. Therefore, a dedicated training specifically targeting this background becomes necessary. The VH analysis currently merges these two processes into a single one, but there is a desire and intention to eventually separate them. As a result, one of the objectives of this chapter is to quantify the extent to which this background can be rejected from the signals. To comprehend how similar these two processes are, we plot a correlation matrix between event quantities for both the processes, shown in Fig. 5.6. It is evident that the correlation matrices exhibit striking similarities with only minor differences. The matrices, however, do not detect any non-linear correlations. A neural network is expected to capture these subtle variations and learn how to differentiate between the two. Additionally, $pp \to Z + jets$ is a dominant background with its sizeable cross section. Thus, we train a network against this background as well and investigate how it performs. We show its correlation matrix in Fig. 5.7. It is seen that there are some differences between this process and the $gg \to ZH$ process. In order to begin the task of separation, we first rank the input features to identify the most relevant kinematic variables for these ATLAS samples, drawing inspiration from the previous chapter.



(a)



(b)

Fig. 5.6: Correlation matrix for $gg \to ZH$ (SM) (left) and $qq \to ZH$ (right) processes

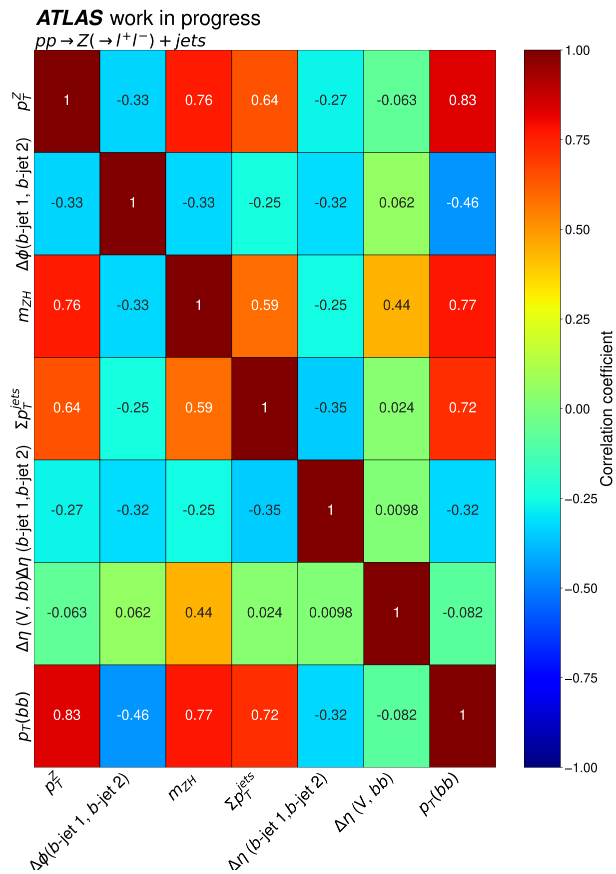
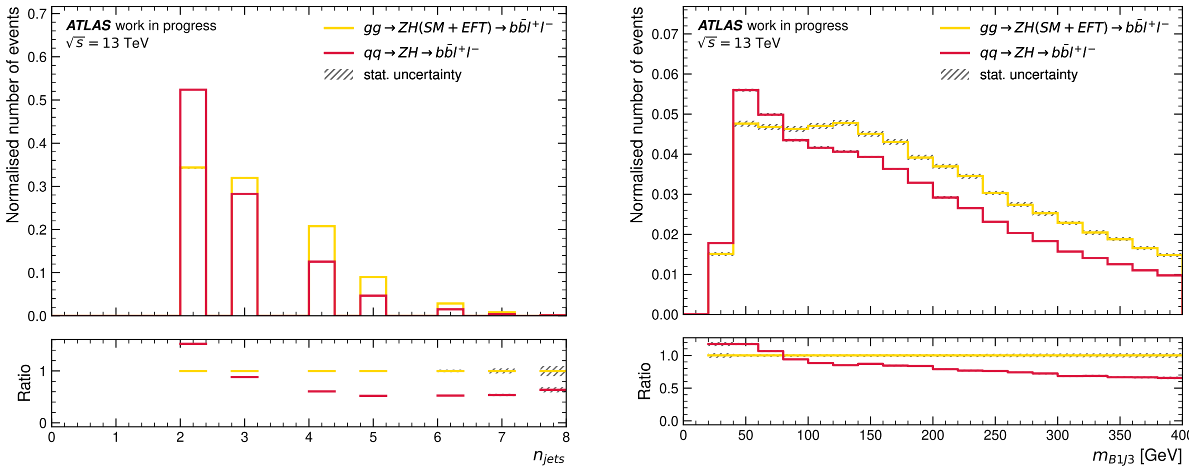


Fig. 5.7: Correlation matrix for $Z + jets$ process

5.4.1 Ranking of input features

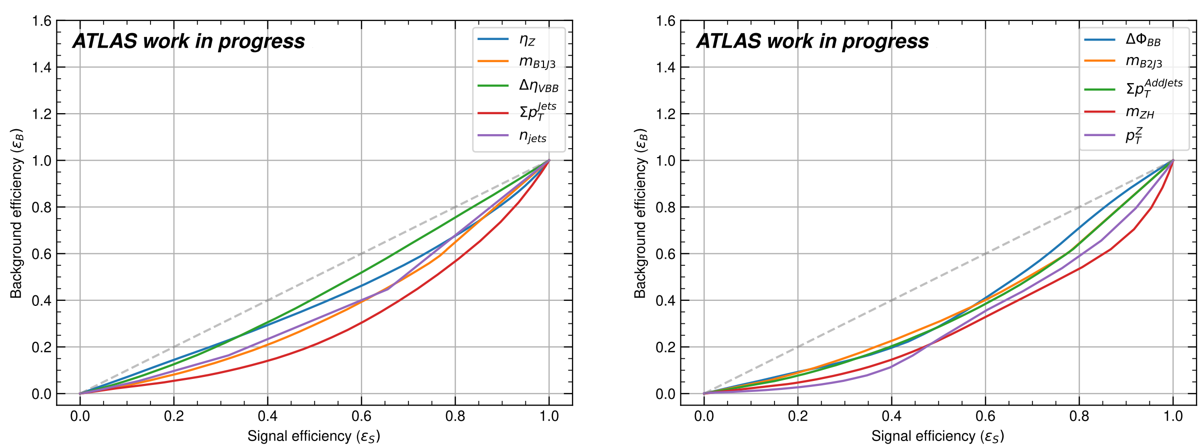
For the ranking of the variables, the signal and background we use are $gg \to ZH$ (SM + EFT) and $qq \to ZH$ respectively. From the list of variables we have, we carefully choose a total of 10 input features which hold the largest discriminating power. This is done by examining the kinematic distributions of the quantities, and investigating if there are significant variations between the signal and background. By ranking and selecting these 10 features, we aim to simplify the network by gradually eliminating features one-by-one and evaluating its performance at each step. In contrast to the Madgraph_aMC@NLO and Pythia8 generated samples we had in the previous chapter, where we only included 2 (b-)jets in the final state, the ATLAS VH samples include additional jets, arising from the splitting of gluons or quarks in the initial state. Since the initial state varies, the variables relating to these jets are particularly important. Two such quantities are number of jets, n_{jets}, and invariant mass of leading b-tagged jet and a third jet, $m_{b_1 J_3}$, shown in Fig. 5.8a and 5.8b. The first figure depicts that the probability for additional jets to be emitted in a gluon initiated process is more than for a quark initiated process. This is due to the distinct Casimir factor (color factor) that appear in the expression of the radiation emitted from quarks and gluons [26]. The second figure shows that this quantity, related to additional jets, holds discriminating powers.

For every variable, successive cuts are placed on the respective distributions to get the signal and background efficiency. The efficiency is defined as $\epsilon = \frac{N^{selected}}{N^{total}}$. The cut-based analysis



(a) (b)

Fig. 5.8: Distribution for the number of jets (left) and invariant mass of leading b-tagged jet and a third jet (right) for $gg \to ZH$ (SM + EFT) and $qq \to ZH$ processes



(a) (b)

Fig. 5.9: Ranking of input features for $gg \to ZH$ (SM + EFT) sample

is done in a 1-dimensional plane, and disregards any potential correlations between the variables, which could potentially be recognised by a neural network. Figure 5.9 shows the signal efficiency vs background efficiency curves obtained for the 10 variables. The y = x line (shown in grey) corresponding to a straight lines implies a complete overlap of the two distributions leading to no separation. The larger the area between this line and the curve for any variable is, the more pronounced is the difference between the distributions, leading to a better separation. To quantify the ranking, the area between the y = x curve and the 1D curve is calculated, and presented in Table 5.1. Here, inv. means invariant and add. stands for additional. The table and the figure demonstrate that the most discriminant variables are m_{ZH} and variables related to the transverse momenta of the final jets. This is exactly what was established in the previous chapter. As established, apart from these quantities, variables which are associated to the initial state, for example, $m_{b_1j_3}$, $m_{b_2j_3}$ and n_{jets} are essential inputs for achieving maximum separation. Lastly, we have the angular quantities which are less discriminating than the other ones. This is in contrast to what we saw in the last chapter, and the discrepancy arises because of the cut imposed on $p_T^Z > 75$ GeV in these samples. This cut eliminates the events wherein the 2 b-jets appear back-to-back. Although this cut already rejects a huge background of $qq \to ZH$, but the discriminating power is now lost for these angular distributions.

Table 5.1: Ranking of kinematic variables for selection of inputs

Ranking	Kinematic Variable	Definition	Area (%)
1	m_{ZH}	Inv. mass of (Z, H)	20.41
2	$\sum p_T^{jets}$	Sum of p_T of b-tagged jets	19.60
3	p_T^Z	p_T of Z	19.16
4	$\sum p_T^{addjets}$	Sum of p_T of add. jets	14.25
5	m_{b1j3}	Inv. mass of (leading b-jet, leading add. jet)	13.79
6	m_{b2j3}	Inv. mass of (subleading b-jet, leading add. jet)	13.14
7	n_{jets}	Total number of jets	12.31
8	$\Delta\phi_{bb}$	$\Delta\phi$ between tagged b-jets	12.09
9	η_Z	η of Z	9.01
10	$\Delta\eta_{Zbb}$	$\Delta\phi$ between Z and tagged b-jets	6.15

5.4.2 Specifics of the model

Hyperparameters

Hyperparameters are constants that dictate the performance of a neural network, and need to be tuned to enhance it as much as possible. The main hyperparameters to be optimised include the learning rate, layer count, number of nodes within each layer, batch size of the input samples, the dropout rate, and epoch number. The learning rate governs the speed at which the network updates the weights, specifying the magnitude of the step taken along the slope of the loss function to approach its minimum. The batch size controls the count of input data which the network handles before it updates the weights. The dropout rate, on a random basis, sets a fraction of the nodes to zero to minimise the dependence of the network on a particular node. The role of this parameter is to prevent overfitting. The total number of times the network processes the full training data is given by the epoch size.

Model architecture and specifications

We adopt a feed-forward neural network using TensorFlow [27] and Keras API [28]. The network architecture is described as follows. The input layer consists of a number of nodes which is equivalent to the number of features. If set to be True, a preprocessing of the input tensor is initialised after the input layer, which standardises the input distributions such that their mean is 0 and their variance is 1. This is done so that all the features are scaled in the same range and there is no dominance of a particular feature. The activation function is set to be Rectified Linear Unit ($f(x) = max(0, x)$). 'N' number of hidden dense layers (i.e. fully connected so that all nodes of a previous layer are connected to all nodes of the next layer) are added to the network, inter-mediated by a dropout layer. After the last hidden and its subsequent dropout layer is the output layer for which the activation function is set to softmax ($f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$). This function assigns to each event, the probability of belonging to each class. For binary classification, the output is a two dimensional array containing the probabilities of the input data belonging to either the signal or the background class. The argmax function (f = arg max(g(x))) is used to get the predictions of a particular event. The optimiser and loss function selected are the Adam optimiser [29] and categorical crossentropy, suitable for multi-class classification tasks.

Optimization of Hyperparameters

For the input data, the total statistics we have are 8×10^5 events of $gg \to ZH$, 5×10^6 events of $qq \to ZH$, and 5×10^7 events of $Z + jets$. The events are normalized so that they correspond to a luminosity of 139 fb^{-1}. For the training, we use a sample size of 8×10^5 events for the signal and backgrounds. The shards method of TensorFlow [27] is used to split this sample into two equal halves, training and test data. The entirety of the statistics is used as the validation set. For assignment of labels to the data, the signal and backgrounds events are one-hot encoded. Shuffling of the dataset is done before feeding the events into the network. The input array is formed by concatenating the signal and background events. To ensure that each batch contains a balanced number of events from both classes, shuffling with an appropriate step size (equal to the total number of events) is necessary.

Prior to evaluating the network's performance, the initial step involves confirming the smoothness of the loss curve in relation to the number of epochs and checking for signs of overfitting. Overfitting is detected when the loss curves of the training and test sets begin to diverge beyond a certain epoch. The tuning of the hyperparameters can have drastic effects on the loss curve and it is imperative to find an optimal set of constants for which the training and test losses converge and reach a plateau.

A grid scan of the hyperparameters is performed :

- learning rate $\in (10^{-2}, 5 \times 10^{-3}, 10^{-3}, \mathbf{5 \times 10^{-4}}, \mathit{10^{-4}}, 5 \times 10^{-5}, 10^{-5}, 5 \times 10^{-6}, 10^{-6})$

- batch size $\in (32, 64, 128, \mathit{1 \times 10^3}, \mathbf{2 \times 10^3}, 5 \times 10^3)$
- number of layers $\in (1, 2, 3, 4, \mathit{5}, 6, \mathbf{7}, 8, 9, 10)$
- number of nodes per layer $\in (10, \mathit{20}, 30, \mathbf{50})$

Here, the italic and bold values correspond to the best achieved values of these parameters for $gg \to ZH$ (SM) and $gg \to ZH$ (SM+EFT) (for separation from $qq \to ZH$). The learning rates are varied in steps of 5, from 10^{-2} to 10^{-6}. If the rate is too small, a larger number of epochs are required to reach the minimum, or it may not converge at all. If the rate is too big, it might bounce around the minimum and never reach convergence. The training samples are divided into several batches, which the network handles at a time, and the weights are updated after each batch. This has several advantages like computational efficiency and ensuring the stability of the network. Typically the batch sizes chosen are of the order of $\mathcal{O}(10)$ to $\mathcal{O}(100)$. This ensures that there are less number of samples in each batch so that the updating of weights takes place more frequently, leading to better performance of the network. But in this case, owing to the fact that $gg \to ZH$ and $qq \to ZH$ have similar distributions, the batch size is kept large enough so that the network is able to differentiate between the signal distribution and statistical fluctuations in the background. The number of epochs are fixed to be 300 or 10^3, since the optimal learning rate is small and it takes a larger number of epochs to reach the minimum of the loss function. The layer count determines the depth of the network, the deeper it is, more complex the network becomes. It is seen that a highly complex network is required for an adequate performance.

Figures 5.10 and 5.11 show the loss and accuracy curves for the $gg \to ZH$ (SM) vs $qq \to ZH$ and $gg \to ZH$ (SM + EFT) vs $qq \to ZH$ respectively. Both figures demonstrate an ideal loss function. For the SM case, we choose $n_{epochs} = 300$ since the plateau is reached with this number of iterations, while for the EFT scenario, we extend it to 1000 epochs. The loss and accuracy plots complement each other. Accuracy is defined as the ratio of correct predictions (obtained from the argmax function) to the total number of predictions. Notably, from the accuracy curve, we observe that the network exhibits better discrimination between $gg \to ZH$ (SM) and $qq \to ZH$ (final accuracy = 66%) compared to $gg \to ZH$ (SM + EFT) and $qq \to ZH$ (final accuracy = 68%). This is anticipated since we established in Chapter 4 that the populated regions in the phase space differ more significantly when comparing $gg \to ZH$ (SM + EFT) to $qq \to ZH$ than $gg \to ZH$ (SM) to $qq \to ZH$. As no over-training is observed between training and validation (test), the dropout factor is set to zero and no regularizers are applied. We observe a kink in both the curves which could imply that the model transits from a first minima to a second minima. This happens when the learning rate is small, and the model is stuck in the first minima for a long time before it reaches the true minimum.

5.5 Performance of Neural Network

The neural network outputs the probability of an event being a signal or background. A discriminant can be constructed from these quantities,

$$D = log(\frac{P_{signal}}{P_{background}})$$

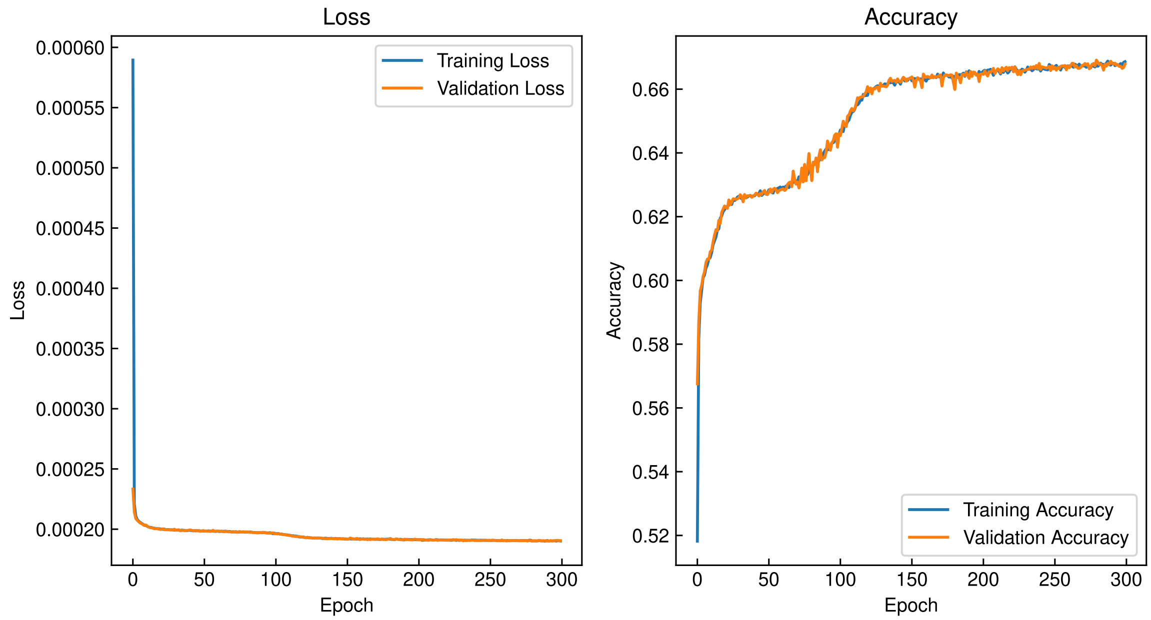


Fig. 5.10: Loss Function for the optimal set of hyperparameters for $gg \to ZH$ (SM) vs $qq \to ZH$

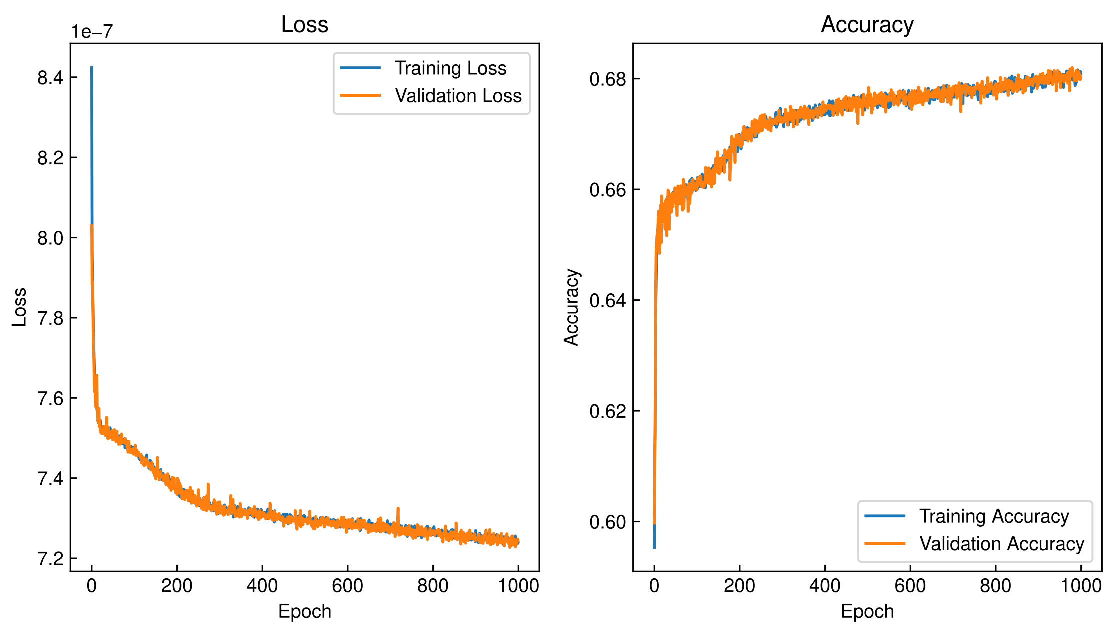


Fig. 5.11: Loss Function for the optimal set of hyperparameters for $gg \to ZH$ (SM + EFT) vs $qq \to ZH$

The aim of a network is to have two distinct peaks for the two discriminants, D_{signal} and $D_{background}$, and subsequently place a cut on this distribution so that only signal events remain. Alternatively, we can simply look at the quantity, P_{signal} for both classes of events and follow a similar approach. In order to evaluate the performance of a neural network, we utilize the area under the ROC (receiver operating characteristic) curve, known as AUC, as the figure of merit. The ROC curve employed in this context is a plot between background rejection (1 - background efficiency) and signal efficiency. A higher AUC corresponds to a superior network performance, as it reflects a larger area under the curve.

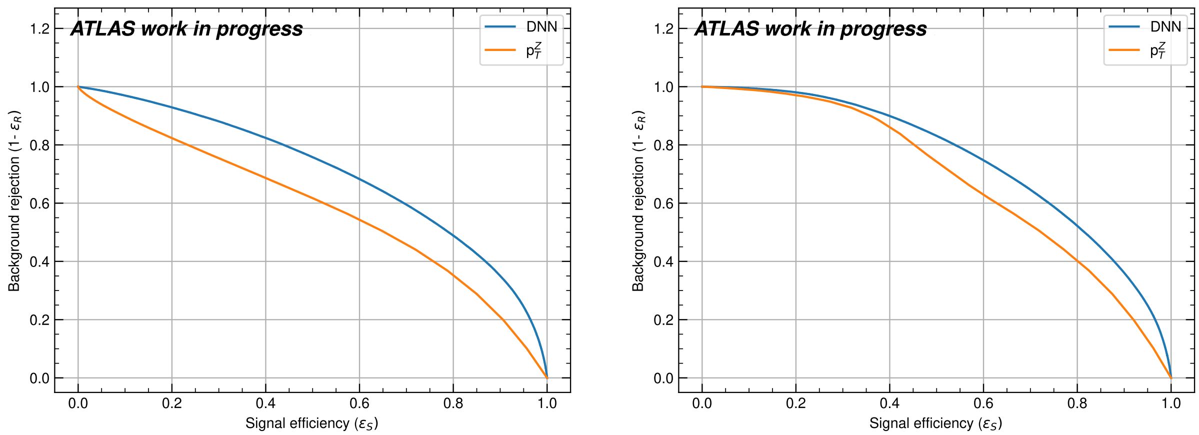
5.5.1 $qq \to ZH$ rejection

We begin by incorporating 7 features in the network, according to the ranking made above. Incremental addition of features to the network from 7 to 10 is done and it is confirmed that the network's learning capability improves. Fig. 5.12a shows the final ROC curve for $gg \to ZH$ (SM) vs $qq \to ZH$ for 10 features. We use a cut on p_T^Z as a baseline and observe that the network outperforms a cut-based analysis done on p_T^Z, therefore, justifying the usage of a neural network. A calculation of the AUC gives 70.4% and 74.9% for SM and SM + EFT processes respectively. To draw a comparison between these two networks, we validate the SM and SM + EFT samples on both of them and inspect which one is a better compromise. Fig. 5.13a shows the comparison of the networks for the SM sample. As anticipated, the network which is trained dedicatedly for the SM sample has a slightly higher AUC (5 %) in contrast with the one trained for SM + EFT. Fig. 5.13b shows the same plot for SM + EFT sample, but the difference between the performance for different networks is larger here (12 %). This disparity arises due to the SM-specific network's limited ability to capture the distinctive tail of the distributions of p_T^Z and m_{ZH} present in the $gg \to ZH$ (SM + EFT) process. Consequently, the network trained on the EFT sample emerges as the preferred choice.

For this network, we display its output in Fig. 5.14, which is the probability of an event ($gg \to ZH$ (SM + EFT) or $qq \to ZH$) being a signal. For the background events, we see a distinct peak at values near 0.3, which means that the network is able to classify these events. For the signal events, we do not observe a definite peak, but rather a plateau from values of 0.5 to 0.9. The network is able to classify a certain number of events with high probability, but this number is quite small, indicated by the uncertainty. These events belong to the tail of the distributions we mentioned before. The plot demonstrates that placing a cut on this distribution at 0.6 (or higher) will aid us in rejecting a majority of the background.

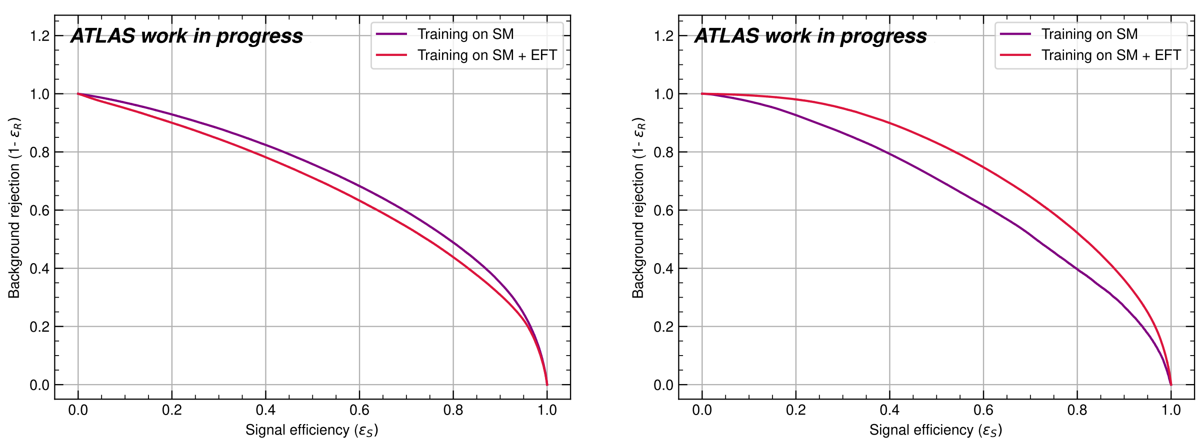
5.5.2 $Z + jets$ rejection

We proceed by employing a neural network to reject the dominant background, namely $Z + jets$. For this network, we use two additional input features which are the invariant mass of the b-tagged jets, m_{bb}, and polarization of the Z boson. Since m_{bb} has a peak at the Higgs boson mass for $gg \to ZH$ while it is a flat function for $Z + jets$, it is an important discriminating variable. The choice of including polarization of the Z boson in the network is inspired from the VH analysis [8]. Before feeding into the network, the weights of the events are normalized to 1. This is an important step to prevent data imbalance due to the



(a) (b)

Fig. 5.12: Left : ROC Curve depicting background rejection vs signal efficiency trained and validated for $gg \to ZH$ (SM) vs $qq \to ZH$. Right : ROC Curve depicting background rejection vs signal efficiency trained and validated for $gg \to ZH$ (SM + EFT) vs $qq \to ZH$



(a) (b)

Fig. 5.13: ROC Curve for $gg \to ZH$ SM samples (left) and $gg \to ZH$ (SM + EFT) (right) validated on a network trained to separate $gg \to ZH$ (SM) vs $qq \to ZH$ and $gg \to ZH$ (SM + EFT) vs $qq \to ZH$

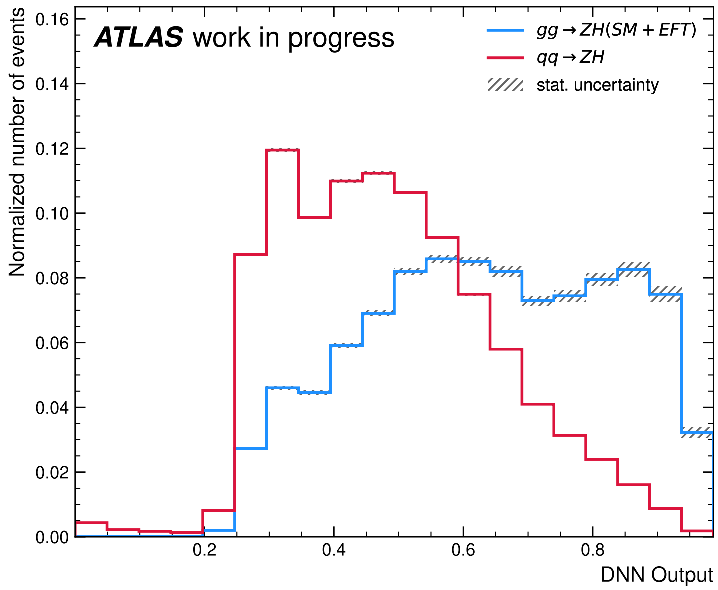


Fig. 5.14: Output of the neural network: Distribution of the probability of being a signal for signal and background events

large difference in the cross sections. We again perform a grid scan of hyperparameters as the previous network, and present the loss and accuracy curves in Fig. 5.15 for the optimal set (learning rate = 10^{-3}, batch size = 128, number of layers = 10, number of nodes per layer = 50). For this set, the network exhibits the expected behaviour, as is demonstrated in the loss curve. The accuracy now is 75%, which aligns with what is anticipated, considering the significant variation of the event quantities in this process compared to $gg \to ZH$. Fig. 5.16 showcases the ROC Curve and output of this network. The AUC is now 82%, surpassing the previous case. Furthermore, if we examine the output, we notice a distinct peaks for the background events and signal events. This indicates the network's enhanced ability to distinguish the two processes. Once again, we see that the network is quite confident ($Output > 0.8$) in recognising the signal for a certain regime of phase space, where the transverse momenta of the Z boson is high. We verify that the inclusion of the two additional variables enhances the performance of the network.

5.5.3 2D distributions of DNN outputs

Having employed two networks each of which targets to separate the two most challenging backgrounds, we now see how the signal events and background events populate the 2D phase space formed by the outputs of the two DNNs. To maximally reject the $Z+jets$ background, we place a cut on $100 < m_{bb} < 140$ GeV. Fig. 5.17a, 5.17b and 5.18 show these distributions for the processes $gg \to ZH$ (SM + EFT), $qq \to ZH$ and $pp \to Z+jets$ respectively. The first figure shows that the signal events are populated in the right hand corner, which is anticipated because this region corresponds to the regions where DNN output > 0.8 for both the neural networks. The 5.17b and 5.18 figures show that the $qq \to ZH$ and $pp \to Z+jets$ backgrounds have the majority of events populated near the middle and lower middle part of the 2D phase space. Moreover, there are less number of events in the top right corner where

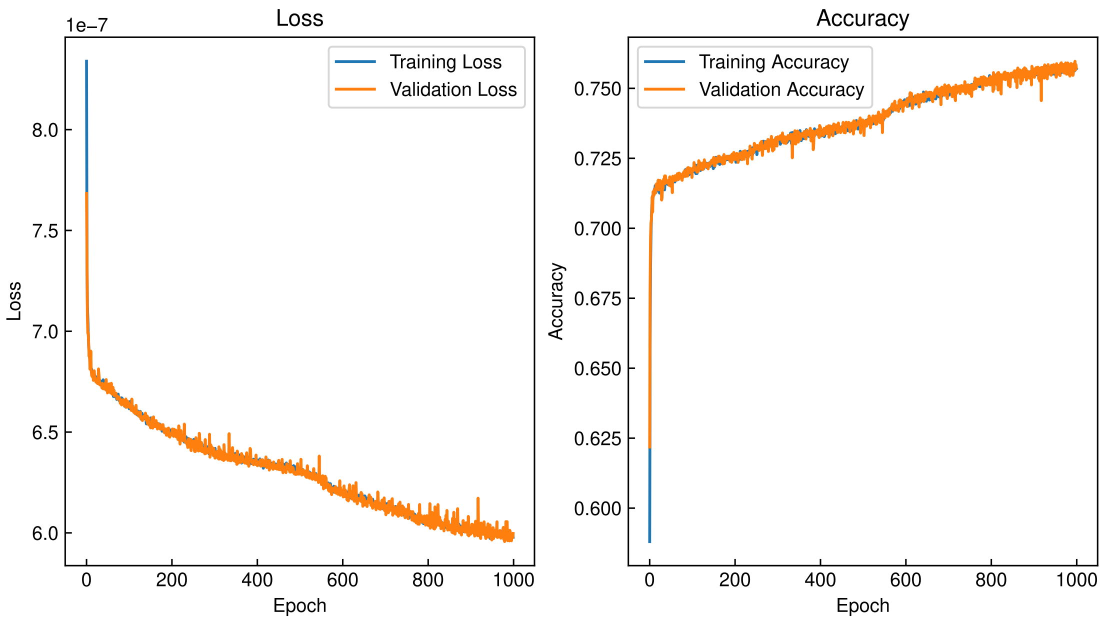
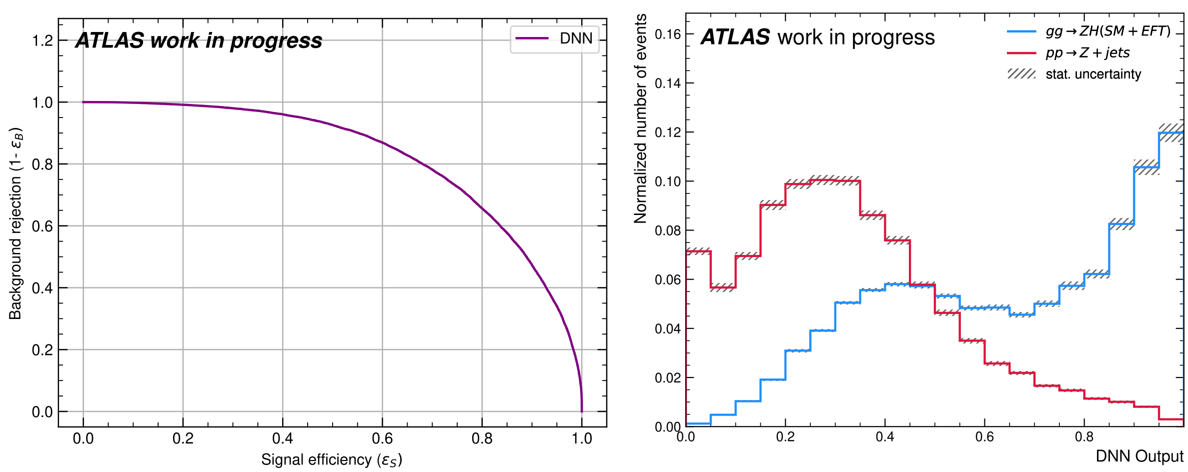


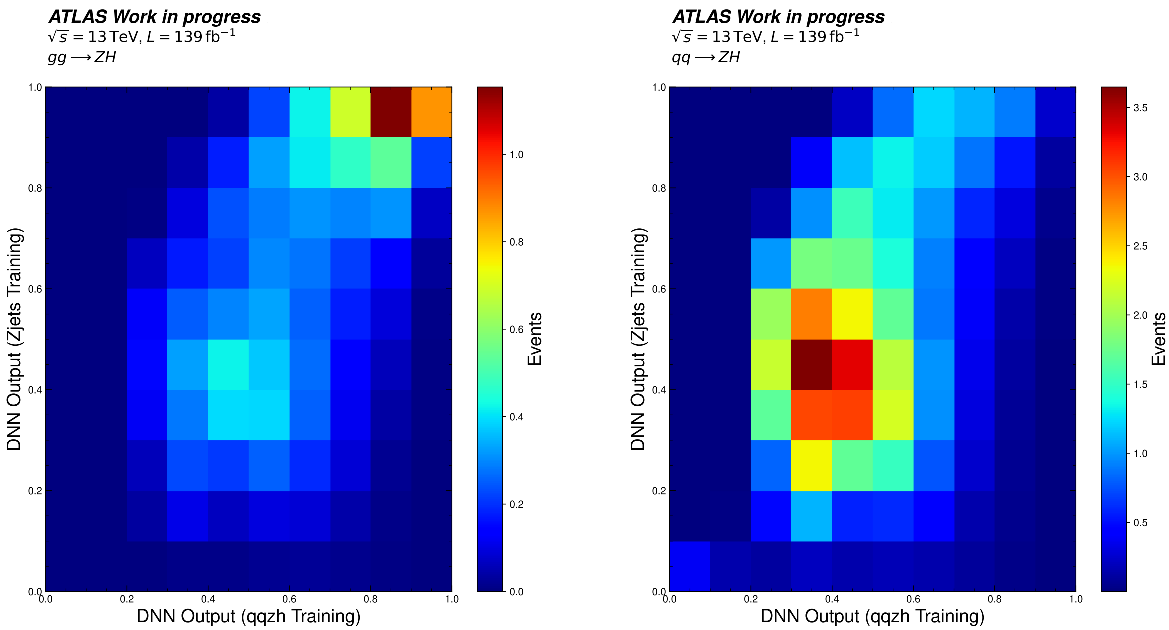
Fig. 5.15: Loss Function for the optimal set of hyperparameters for $gg \to ZH$ (SM + EFT) vs $Z + jets$



(a) (b)

Fig. 5.16: ROC Curve (left) for a network trained for and validated on $gg \to ZH$ (SM + EFT) vs $Z + jets$ and output of the neural network (right)

the signal events reside. Hence, we have identified a region in this phase space which could be useful in separating the signal events from the background events. In the next chapter where we derive the 95% Confidence Level exclusion limits, we will concentrate on these 2D heatmaps.



(a) (b)

Fig. 5.17: 2D distribution of the output of the two neural networks validated on $gg \to ZH$ (SM + EFT) (left) and $qq \to ZH$ (right)

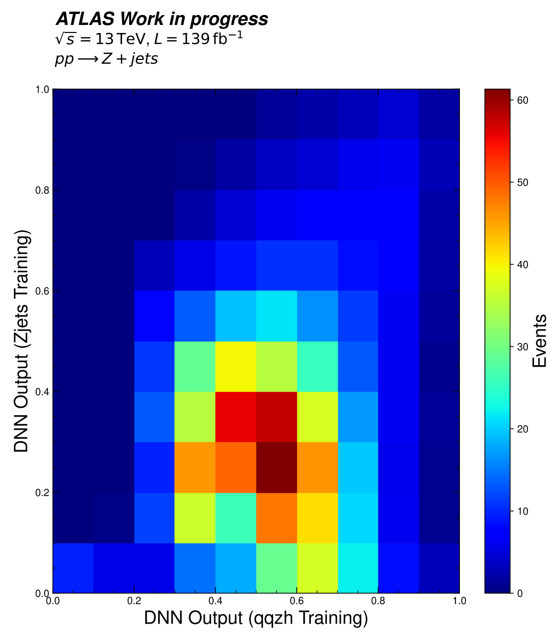


Fig. 5.18: 2D distribution of the output of the two neural networks validated on Z+jets

Chapter 6

Constraints on Wilson Coefficient, $\mathbf{c}_{\phi t}$

Equipped with the two neural networks outlined in chapter 5, we can utilize their outputs to impose constraints on the Wilson coefficient, $c_{\phi t}$, in terms of 95% Confidence Level (CL) exclusion limits, which is the main focus of this chapter.

6.1 Constraining $c_{\phi t}$: 95% CL limit

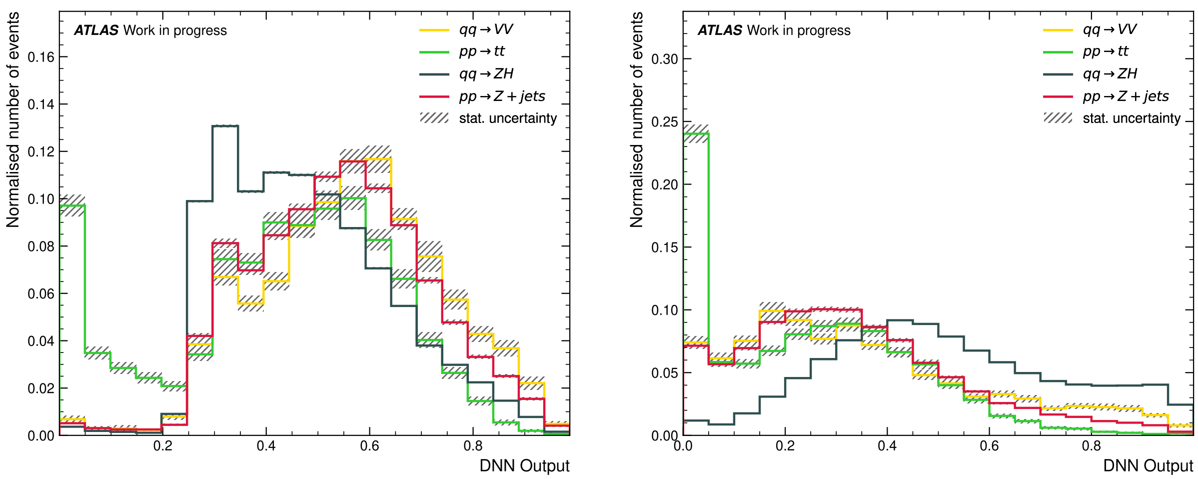
By analyzing the outputs of the two DNNs, we can obtain an estimate on the number of signal and background events. This information is instrumental in establishing 95% CL limits on $c_{\phi t}$. So far, we have focused only on the two most challenging backgrounds. However, to obtain an accurate estimation of the 95% CL limits, it is necessary to consider all the backgrounds discussed in Sec. 2.3.2. To achieve this, we must get the neural network predictions for the additional backgrounds. Additionally, in order to impose these constraints, we also need these predictions for the other EFT samples with varying $c_{\phi t}$ values.

6.1.1 Predictions of other backgrounds

The other backgrounds in consideration consist of $t\bar{t}$, diboson, single-top, WH and $W+jets$. The preselection process applied to the ATLAS samples effectively removes a significant number of these background events. Furthermore, concentrating on the high p_T^Z regime ($p_T^Z > 150$ GeV) and implementing an additional selection on the invariant mass of the two b-tagged jets, these backgrounds are substantially suppressed. On these selected samples, we get their predictions from both the optimised neural networks by validating them on the networks. Fig. 6.1a and 6.1b show the distributions of the DNN outputs for the major backgrounds, $\mathrm{t\bar{t}}$, diboson, $Z+jets$ and $qq \to ZH$, as the remaining backgrounds are considered negligible, on networks optimised for discrimination against $qq \to ZH$ and for $Z+jets$ respectively. Although not depicted in the plots, the small contributions of other backgrounds are taken into account for the final estimation.

For the $qq \to ZH$ optimized training, the diboson and Z+jets backgrounds exhibit, in general, a similar trend, with a noticeable peak around values of approximately 0.6. The $\mathrm{t\bar{t}}$ background observes a first peak at very small values, and a second peak at values of 0.5. These peaks mean that the network is not able to identify these samples as backgrounds. This is expected since the network learns to differentiate the subtle differences between $gg \to ZH$

and $qq \to ZH$, rendering it unable to classify events with distinct distributions. Secondly, we use the network optimised for $Z + jets$ to validate these samples, and the results are shown in 6.1b. The backgrounds, $t\bar{t}$ and diboson, now peak at lower values now and thus the task of discrimination between signal events and background events would particularly benefit from this training. The background, $qq \to ZH$, also has a peak near 0.4 but still has some events in the tail of the distribution, which is, again, expected. To gain insight into the total number of events that remain, we examine the prediction of the absolute number of events after being validated on the $Z + jets$ optimised network in Fig. 6.2, as the previous plots solely illustrate the distribution shapes. The figure reveals that the dominant backgrounds in the last bins, where the signal would peak, are $Z + jets$ and $qq \to ZH$, which reassures the need to have two dedicated networks for these backgrounds.



(a) (b)

Fig. 6.1: DNN output distribution of background events for neural networks optimised for $qq \to ZH$ background (left) and $Z + jets$ background (right)

6.1.2 Predictions of other EFT Samples

The results of the validation of EFT samples in terms of the output of the DNNs is shown in Fig. 6.3a and 6.3b for $qq \to ZH$ and $Z + jets$ optimised networks respectively. Both graphs indicate that the networks possess lower discrimination power from the background when $c_{\phi t} < 15$, since they learn to differentiate the signal from the background based on features unique to the EFT distributions. Although these signatures begin to appear for $c_{\phi t}$ values greater than 0, the effects become large enough to change the distributions from the SM distribution for values larger than $c_{\phi t} > 15$. For small values of $c_{\phi t}$, the kinematic distributions closely resemble those of the SM, and hence, there is no major difference in the output. The performance improves for higher values due to the network being trained on $c_{\phi t} = 20$.

Having obtained an estimation of the background events and EFT samples in this 2D plane, we can now move forward and impose constraints on $c_{\phi t}$.

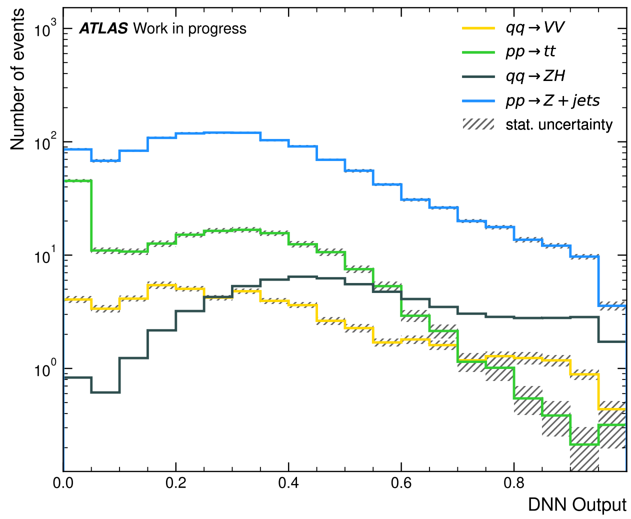


Fig. 6.2: DNN output distribution of the total number of background events for neural networks optimised for $Z + jets$ background

6.1.3 χ^2 Fit

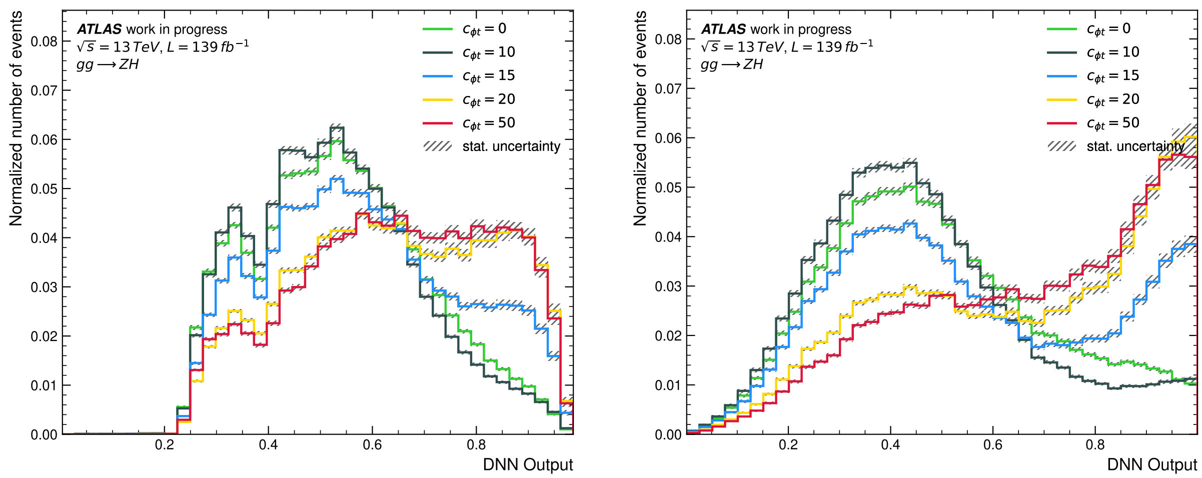
A χ^2 test is a standard test performed in statistics to see how well some measured data competes against a theoretical model, i.e. expected data. It is defined as,

$$\chi^2 = \sum_i \frac{(f_i(c)^{Theor} - f_i(c)^{Data})^2}{(\sigma_{f_i}^{Data})^2} \tag{6.1}$$

where the sum is over all the bins and f_c is some function whose values we are comparing. In our context, we rewrite Eq. (6.1) as,

$$\chi^2(c_{\phi t}) = \sum_i \frac{(N_i^{c_{\phi t}} - N_i^{PD})^2}{(\sqrt{N^{PD}}_i)^2} \tag{6.2}$$

where the sum is over all bins and $N_i^{c_{\phi t}}$ is the total number of events (signal + background) corresponding to an EFT sample with a certain $c_{\phi t}$ value in each bin, 'i'. 'PD' here means pseudodata. The χ^2 function can be seen as the log-likelihood function for a Gaussian distribution, where N^{PD} denotes the mean and the variance is given by N^{PD}. Assuming that the 'pseudodata' here corresponds to the EFT sample with $c_{\phi t} = 0$, we can evaluate this function for all other $c_{\phi t}$ values. Following this, the minimum of the plot of χ^2 vs $c_{\phi t}$ gives us the best fit value and the values corresponding to $\chi^2 = 3.841$ will yield 95% CL exclusion limits for $c_{\phi t}$. The (binned) number of events are extracted from the outputs of the two DNNs. This is done by calculating the number of background + signal events in each bin of the 2D heatmap formed by the outputs of the DNN. We take the luminosity to be 300 fb^{-1}, corresponding to (expected) Run3 data. Following the procedure described yields the χ^2 function which is shown in Fig. 6.4. We observe a minimum at $c_{\phi t} = 0$, which is expected since we chose the value of $c_{\phi t} = 0$ for pseudodata, and obtain the 95% CL exclusion limits, $c_{\phi t}^{95} = 26.3$. We observe that for values of $c_{\phi t} = 10$ to 20, the function remains flat. The sudden rise after $c_{\phi t} = 20$ is because of the increase in the cross section, owing to the dominance of the



(a) (b)

Fig. 6.3: DNN output distribution of EFT samples for neural networks optimised for $qq \to ZH$ background (left) and $Z + jets$ background (right)

quadratic term. If we scale this luminosity to 4000 fb^{-1}, we get the exclusion limit to be $c_{\phi t}^{95} = 4.06$ which is a much stricter constraint. A point worth mentioning is that this function preserves the shape effects of the distribution of the signal events and background events. If there exists unusual fluctuations leading to unphysical shape effects in the background, this would be propagated in the result.

Assuming that the 'pseudodata' here corresponds to the EFT sample with $c_{\phi t} = 20$, we can get a sensitivity for the measurement if the true value is $c_{\phi t} = 20$. We do this by again calculating the χ^2 function for all $c_{\phi t}$ values under this new assumption. The results are shown in Fig. 6.5. We observe a minimum at $c_{\phi t} = 20$, which is anticipated because of the chosen pseudodata sample. We obtain the 95% CL limits, $c_{\phi t}^{95} = 16.05$ and $c_{\phi t}^{95} = 22.70$ if $c_{\phi t}^{true} = 20$. We observe that for lower values of $c_{\phi t}$, the function remains flat till $c_{\phi t} = 10$, since the distributions for these samples are very similar to each other but quite different from $c_{\phi t} = 20$. Again, sudden rise after $c_{\phi t} = 20$ is due to the fact that the cross section rises, because of the dominance of the quadratic term.

6.1.4 Comparison to global fits

If we compare the value of $c_{\phi t}^{95} = 26.3$ to that from the global fit, which is 18, we find that for Run3, the current limit is a bit more strict that what we have obtained. However, for HL-LHC era, we have $c_{\phi t}^{95} = 4.06$, which is much stricter. While the value from the global fits obtained arises from only indirect constraints, we exploit information from only a single process here which brings a direct constraint on this coefficient.

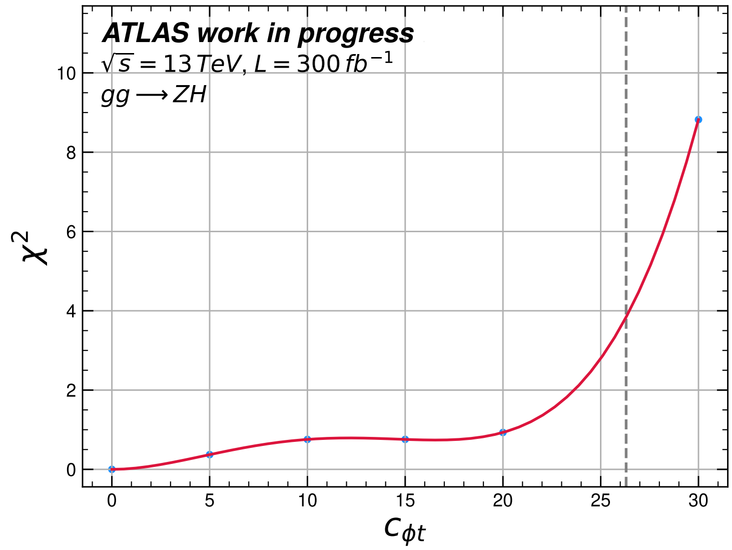


Fig. 6.4: χ^2 distribution as a function of $c_{\phi t}$ if pseudodata is the EFT sample with $c_{\phi t} = 0$, to yield 95% CL exclusion limits

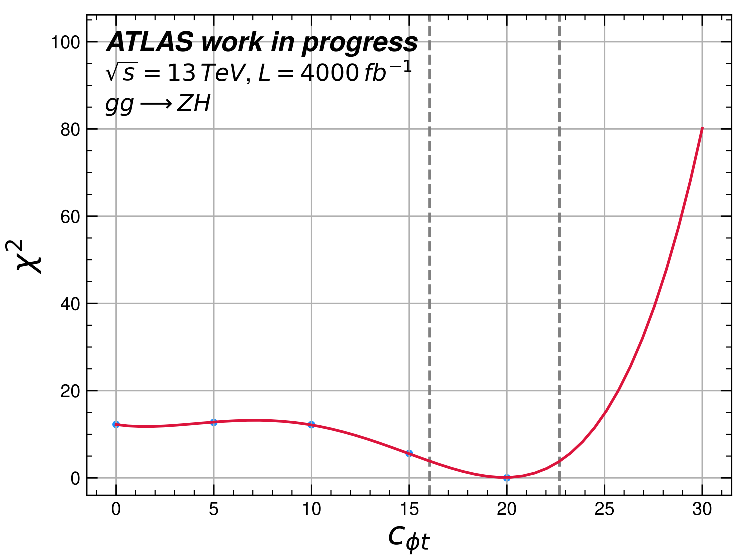


Fig. 6.5: χ^2 distribution as a function of $c_{\phi t}$ if pseudodata is the EFT sample with $c_{\phi t} = 20$, to yield sensitivity of the measurement

Chapter 7

Conclusions and Outlook

The Standard Model (SM) provides a description of the elementary particles and the interactions they experience. Although it is a marvel of modern particle physics, there exists some phenomena which are left unexplained by this theory model. In order to describe physics beyond the SM, one of the theories proposed is the Standard Model Effective Field Theory (SMEFT). Treating the SM as an EFT, this theory expresses the strength of new physics beyond the SM in terms of some coefficients, named the Wilson coefficients. This thesis carried out feasibility studies of the $gg \to ZH$ process to derive a constraint on the Wilson coefficient, $c_{\phi t}$. We first presented phenomenological studies performed on this process with a goal of identifying key kinematic variables which can be probed to study EFT effects, and assist in isolating the signal from one of the challenging backgrounds, $qq \to ZH$. Currently, due to lack of sensitivity, the $gg \to ZH$ and $qq \to ZH$ processes are studied together. We demonstrate that there is information in the kinematic distributions of these processes which can be utilized to separate them eventually. Additionally, a dominant background is $pp \to Z + jets$, because of its very large cross section. Thus, we employed neural networks to reject these two backgrounds. Using the outputs of these two neural networks, we derived direct constraints in terms of exclusion limits on $c_{\phi t}$. We arrived at the conclusion that we are able to place a constraint on the 95% exclusion limit for a luminosity of 300 fb^{-1}. On comparison with the current global fit limits, the global limits are more strict than the one we have obtained.

Having derived these constraints, we stress that we have a long way to go before we reach this sensitivity experimentally. For future studies, the following considerations should be made. Considering the two main backgrounds for which neural networks were employed, we did not take into account that the background $pp \to Z + jets$ is modified by the addition of the operator, $\mathcal{O}_{\phi t}$. Moreover, training neural networks in dedicated bins of p_T^Z might be beneficial in recognising the bins where the sensitivity for these limits is the highest. Additionally, we employed two different neural networks to separate two backgrounds but convoluting them in a sophisticated and effective manner can help increase their power and lead to better signal versus background rejection. Finally, we took a simple statistical approach to get the value of the exclusion limits but a comprehensive statistical analysis must be performed, while also taking into account the uncertainties.

References

[1] Mark Thomson. *Modern particle physics*. Cambridge University Press, 2013.

[2] Jacob J Ethier, Giacomo Magni, Fabio Maltoni, Luca Mantani, Emanuele R Nocera, Juan Rojo, Emma Slade, Eleni Vryonidou, and Cen Zhang. Combined smeft interpretation of higgs, diboson, and top quark data from the lhc. *Journal of High Energy Physics*, 2021(11):1–97, 2021.

[3] Christoph Englert, Matthew McCullough, and Michael Spannowsky. Gluon-initiated associated production boosts higgs physics. *Physical Review D*, 89(1):013013, 2014.

[4] Benoît Hespel, Fabio Maltoni, and Eleni Vryonidou. Higgs and z boson associated production via gluon fusion in the sm and the 2hdm. *Journal of High Energy Physics*, 2015(6):1–29, 2015.

[5] Ken Mimasu, Verónica Sanz, and Ciaran Williams. Higher order qcd predictions for associated higgs production with anomalous couplings to gauge bosons. *Journal of High Energy Physics*, 2016(8):1–27, 2016.

[6] Long Chen, Gudrun Heinrich, Stephen P Jones, Matthias Kerner, Jonas Klappert, and Johannes Schlenk. Zh production in gluon fusion: two-loop amplitudes with full top quark mass dependence. *Journal of High Energy Physics*, 2021(3):1–22, 2021.

[7] Marc Escalier. Combined higgs boson measurements and their interpretations with the atlas experiment. *arXiv preprint arXiv:2306.01850*, 2023.

[8] ATLAS Collaboration et al. Measurements of and production in the -bb decay channel in collisions at 13tev with the atlas detector. 2021.

[9] Morad Aaboud, Georges Aad, Brad Abbott, B Abeloos, DK Abhayasinghe, SH Abidi, OS AbouZeid, NL Abraham, H Abramowicz, H Abreu, et al. Observation of h→ bb decays and vh production with the atlas detector. *Physics Letters B*, 786:59–86, 2018.

[10] Albert M Sirunyan, Armen Tumasyan, Wolfgang Adam, Federico Ambrogi, Ece Asilar, Thomas Bergauer, Johannes Brandstetter, Marko Dragicevic, Janos Erö, A Escalante Del Valle, et al. Observation of higgs boson decay to bottom quarks. *Physical review letters*, 121(12):121801, 2018.

[11] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of instrumentation*, 3(08):S08001, 2008.

[12] ATLAS Collaboration, G Aad, E Abat, J Abdallah, AA Abdelalim, A Abdesselam, O Abdinov, BA Abi, M Abolins, H Abramowicz, et al. The atlas experiment at the cern large hadron collider, 2008.

[13] Christian Bierlich, Andy Buckley, Jonathan Mark Butterworth, Christian Holm Christensen, Louie Corpe, David Grellscheid, Jan Fiete Grosse-Oetringhaus, Christian Gutschow, Przemyslaw Karczmarczyk, Jochen Klein, et al. Robust independent validation of experiment and theory: Rivet version 3. *SciPost physics*, 8(2):026, 2020.

[14] Johan Alwall, R Frederix, S Frixione, V Hirschi, Fabio Maltoni, Olivier Mattelaer, H-S Shao, T Stelzer, P Torrielli, and M Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):1–157, 2014.

[15] Tanju Gleisberg, Stefan Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02):007, 2009.

[16] Torbjörn Sjöstrand, Stefan Ask, Jesper R Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O Rasmussen, and Peter Z Skands. An introduction to pythia 8.2. *Computer physics communications*, 191:159–177, 2015.

[17] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[18] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, Pedro Arce, Makoto Asai, D Axen, Swagato Banerjee, GJNI Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[19] Céline Degrande, Gauthier Durieux, Fabio Maltoni, Ken Mimasu, Eleni Vryonidou, and Cen Zhang. Automated one-loop computations in the smeft. *arXiv preprint arXiv:2008.11743*.

[20] JA Saavedra, C Degrande, G Durieux, F Maltoni, E Vryonidou, C Zhang, D Barducci, I Brivio, V Cirigliano, W Dekens, et al. Interpreting top-quark lhc measurements in the standard-model effective field theory. *arXiv preprint arXiv:1802.07237*, 2018.

[21] Abdel Djouadi, Jan Kalinowski, and Michael Spira. Hdecay: A program for higgs boson decays in the standard model and its supersymmetric extension. *Computer Physics Communications*, 108(1):56–74, 1998.

[22] Daniel de Florian, D Fontes, J Quevillon, M Schumacher, FJ Llanes-Estrada, AV Gritsan, E Vryonidou, A Signer, P de Castro Manzano, D Pagani, et al. *arXiv: Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. Number arXiv: 1610.07922. Cern, 2016.

[23] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.

[24] Luminosity determination in *pp* collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. 6 2019.

[25] Nicolas Berger, Claudia Bertella, Thomas P Calvet, Milene Calvetti, Valerio Dao, Marco Delmastro, Michael Duehrssen-Debling, Paolo Francavilla, Yacine Haddad, Oleh Kivernyk, et al. Simplified template cross sections-stage 1.1. *arXiv preprint arXiv:1906.02754*, 2019.

[26] Alexander Kovner and Urs Achim Wiedemann. Gluon radiation and parton energy loss. In *Quark–Gluon Plasma 3*, pages 192–248. World Scientific, 2004.

[27] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[28] François Chollet et al. Keras. https://keras.io, 2015.

[29] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.