Adversarial Neural Networks for Associated Top Pair and Higgs Boson Production in the Di-Photon Decay Channel

Emily Petrova Takeva

A thesis presented for the degree of
Doctor of Philosophy

The University of Edinburgh
United Kingdom
May 2022

CERN-THESIS-2022-341
//2022

To my grandmother Pavlina
Thank you for being my inspiration
with your intelligence, strength and generosity!
Липсваш ми, бабо!

Abstract

The main topic of this thesis is classification with Adversarial Neural Networks, which are for the first time used in an analysis targeting final states in which the Higgs boson decays to pairs of photons ($H \to \gamma\gamma$). The analysis uses 139 fb^{-1} of proton-proton collision data recorded at $\sqrt{s} = 13$ TeV by the ATLAS experiment at the Large Hadron Collider, and targets the associated top pair and Higgs boson production (ttH). Backgrounds with non-resonant photon pairs such as multi-jet or top-antitop pair production in association with photons are mainly rejected by using the photon kinematics. The signal is extracted from a fit of the di-photon invariant mass ($M_{\gamma\gamma}$) distribution, which consists of a narrow signal peak on the top of a substantial background. Using the kinematic variables of the photons for the classification causes the background $M_{\gamma\gamma}$ distribution to peak at the Higgs boson mass value, due to these variables being correlated with $M_{\gamma\gamma}$. This sculpted background distribution is hard to parametrise with a simple functional form needed for the background fit to the $M_{\gamma\gamma}$ sidebands. The novel adversarial neural network approach developed in this thesis enables designing a classification discriminant independent of $M_{\gamma\gamma}$, which removes the sculpting, while keeping the classification efficiency optimally high.

Additionally, work towards the creation of a programme to deal with the interpolation of energy values used for the incident single particles in the fast calorimeter simulation of the ATLAS experiment, which is used to date, is also presented in this thesis.

Lay Summary

The Standard Model is the theory able to describe all fundamental interactions apart from gravity. In the Standard Model conjecture, fundamental particles obtain mass through the coupling with the Higgs boson. The aim of the thesis is precision measurement of the coupling of the top quark, the Standard Model particle with the largest mass. This measurement could reveal deviations from the theory predictions, hinting that the standard model is not the ultimate theory.

The data for the measurement is provided by the Large Hadron Collider (LHC), the world's largest particle accelerator which collides protons with extremely high energies in a 27 km underground tunnel located close to Geneva. In these collisions, particles, that cannot be found in the usual matter around us, such as the top quark and the Higgs boson, can be created. Four main detectors are used to observe the collisions. The thesis uses data taken by the ATLAS detector, one of the two experiments which descries this coupling.The ATLAS experiment records about a thousand high energy collisions each second, and each of these collisions contains thousands of particles. Sophisticated computer algorithms are therefore required to analyse the collision events and draw conclusions from them.

This thesis investigates a novel machine learning algorithm (artificial intelligence), called adversarial neural network (ANN). The ANN is used to measure the top quark coupling in events containing two top quarks and a Higgs boson. The Higgs boson has a short lifetime and is studied by the particles into which it decays. The thesis uses the Higgs decays to two photons, which is one of the most distinct Higgs decays that ATLAS can observe.

An adversarial neural network consists of two neural networks with different objectives, nevertheless working towards the same final goal: a precise measurement of the top quark coupling. The first network is developed to remove events, which look like they are containing two top quarks and the Higgs boson, even though they do not actually contain these particles. This process is not 100% efficient and events which evade it are "sculpted" to look like the actual events containing two top quarks and the Higgs boson. This sculpting substantially degrades the precision of the measurement. This problem is addressed by the second network, which is tasked with removing the sculpting. With this system of two networks, the adversarial method finds the balance between the removal of unwanted events and sculpting. The ANN is unique compared to other algorithms developed by ATLAS, because it is designed to find the optimal balance. This is shown to match the highest precision top quark coupling measurements.

Additionally, this thesis deals with a separate problem related to the simulation of the ATLAS detector. Simulation is a crucial factor in studying the ATLAS detector's response for all physics measurements, which are computationally possible. The general ATLAS simulation is detailed and computationally expensive, so fast simulations, such as AtlFast calorimeter simulation, which is discussed in detail in this thesis, are essential. Simulation events used to be generated with specific energies in simulation while In reality, the particles can have any energy. So an energy interpolation method to address this issue is presented in this thesis and included in the AtlFast3 fast simulation.

Declaration

I declare that this thesis has been composed by myself and that the work has not be submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis, where reference has been made to the work of others.

The work presented in Chapter 5.2 was published in Computing and Software for Big Science (Springer) in 2021 [1] . This study was conceived by all of the authors. I, personally, carried out the work in 5.2.3, 5.2.7 and D.

Acknowledgments

First and foremost, I want to thank all the people, without whom, the work which led to my PhD offer for a position associated with the University of Edinburgh and CERN, both my dream institutions to work at, would not have been possible. In chronological order: my very first physics teacher at Sofia High School of Mathematics (СМГ): Ms Shishkova, who saw potential future for me in science, which I had not even considered at the time. Благодаря Ви, госпожо Шишкова! My second and main high-school teacher Mr. Bahchevanov, who nominated me for a fully funded trip to CERN, which was when I realised what direction I wanted my future career to take. You are, to date, one of the best people I know, you were my inspiration and you provided a stepping stone for my future success in the field. Благодаря Ви сърдечно, господин Бахчеванов!

I also want to thank my very first mentor in the UK: Prof. Nigel Watson! I am fully aware, that you hear this all the time by all the students you have helped, but once more can not hurt. You are an absolutely wonderful, pure hearted human being and an incredible physicist and professional! I can not find the words to describe my gratitude towards you, they all seem plain in comparison to how I feel! You treated a poor foreign girl with a lot of non-academical problems, like a complete equal and you always gave advice with care and love! For what simple words are worth, thank you, from the bottom of my heart! Thank you also to another brilliant supervisor, who I met in Oxford for my summer internship at RAL: Dr.Pawel Majewski!

I want to thank the incredible professionals I met in Edinburgh. Franz, Victoria, Matt and Alex, thank you for giving me a chance! You will never know the full personal story, but I want you to know, you will always remain the people, who provided the gateway to an incredible permanent change in my life. Thank you! My supervisor: Liza. Thank you for answering all my questions and being there for me academically, in both successful and difficult times of my PhD! Liza, I sincerely thank you for all the explanations, help and advice and for taking the role of not just a mentor, but also at times a personal psychologist. Your calm, pragmatic approach always balanced me out and made me feel better!

Last, but by no means least, I want to express my personal gratitude to those who were there for me, in times, when others were nowhere to be found. My mother. Благодря за всичко, мамо! And Jean-Christophe! Merci et je t'aime!

Glossary

SM	Standard Model of particle physics
BSM	Beyond the Standard Model
LHC	Large Hadron Collider
ATLAS	A Toroidal LHC Apparatus
pp	proton–proton collision
MC	Monte Carlo
QCD	Quantum Chromodynamics
EM	Electromagnetic
FCS	Fast Calorimeter Simulation
AF	AtlFast calorimeter simulation
G4	Geant4 simulation toolkit
NTNI	Non-Tight or Non-Isolated; refers to data with NTNI photons
SS	Spurious Signal; background misidentified as signal
ML	Machine Learning
NN	Neural Network
ANN	Adversarial Neural Network
ROC	Receiver Operating Characteristic curve
JSD	Jensen-Shannon Divergence

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The theory, which combines the relevant fundamental forces and the elementary particles involved is the Standard Model of particle physics. Although, it is a set of laws and equations, which describe many physical phenomena, there is a lot we still do not know and need to understand. The process of understanding what we do not yet know include two main branches in particle physics: discovering new particles and improving already made measurements about known particle and interactions, which leads to either confirming theoretical expectations of the Standard Model or implying the existence of new, yet unknown physics. The analysis described in this thesis is under the second branch. It deals with improving the understanding of a fundamental interaction called the Yukawa interaction by examining one of the Higgs boson production mode (from a pair of top quarks) and one of its decay signatures (to two photons) with the final goal of contributing to the coupling strength measurement between the top quark and the Higgs boson (top-Higgs Yukawa coupling). When using the kinematic variables of the final state photons in the analysis chain, an additional separation power between the signal ($ttH(H \to \gamma\gamma)$) and the background (all other processes) is provided. They are particularly important for the rejection of backgrounds, which contain photons in their final state. An example of such a background process is the production of a pair of photons from a pair of top quarks: $tt\gamma\gamma$.

The photon kinematic variables have not been used in previous iterations of the analysis, because selections using these variables result in a large sculpting of the background distribution: (50-60)%. This sculpting results in a background shape that mimics closely the signal shape. The signal peak is the Higgs boson's mass peak, which is expected to be seen as a two sided distribution around the measured Higgs mass of 125.35 GeV. The sculpting made it impossible to see the signal peak after background rejection.

In this thesis, a machine learning method using Adversarial Neural Networks (ANN) is proposed to deal with the above issue, while keeping the signal acceptance and background rejection optimal. In the adversarial approach used, there are two networks, which work together towards the same goal, but with different objectives. Those objectives are the optimal signal and background classification, which the first neural network is tasked with, and the complete removal of background sculpting, which is the task of the second network. The ANNs have been successfully used in ATLAS analyses targeting resonance decays to large-radius jets [2] and Higgs decays to b-quarks ($H \to bb$) [3]. In this thesis, ANN is for the first time used in an analysis targeting $H \to \gamma\gamma$ decays.

The performance of the ANN matches that of the scaled network used by ATLAS at the time of the studies in this thesis, where the transverse momenta and energy of the Higgs candidates are scaled by the di-photon invariant mass. This is used as a bench-mark comparison for the analysis presented here, in particular to compare the final sensitivity obtained, and to quantify the level of remaining background sculpting. In the final states with at least one lepton (leptonic decay channel), the ANN performance is found to be satisfactory and comparable to that of the scaled network. In the lower sensitivity final states with no leptons (hadronic decay channel), categories with a high ANN discriminant cut could only be fitted with functions using many free parameters, but they contained too few expected background events to constrain these parameters in a fit. A modification of the adversary architecture (less nodes, smaller number of Gaussian Model Mixture components) and more training data alleviates the sculpting in this case. The full update of the results with such architecture is beyond the scope of this thesis. In hadronic categories with lower discriminant cut values, the Adversarial network performed comparably well to the Scaled network. The ANN setup developed in this thesis is thus applicable to the ttH, $H \to \gamma\gamma$ classification in most categories and could be a viable scientific solution to any resonant physics channel, which suffers from sculpting issues after background rejection while maintaining high signal rates.

Chapter 2

Theoretical Background

2.1 The Standard Model

Particle physics deals with the smallest constituents of our Universe called fundamental particles and with the interactions between them. The theory, which combines these particles and the forces between them is called the Standard Model [4]. It is a relativistic non-abelian gauge theory and has a total of twelve gauge bosons: the photon, three weak bosons and eight gluons. It comprises all non-gravitational physical laws as we know them. Any deviation from it would be considered new physics, not just unknown to us, but also bringing new fundamental meaning to our Universe. For this exciting reason, physicists at CERN and other particle physics research oriented institutions, are providing a united daily effort to measure all possible properties of those particles, as precisely as possible and compare them with the theoretical expectations provided by the Standard Model. There are three families of elementary particles in Standard Model - leptons, quarks and gauge bosons There are 12 elementary fermions, six quarks and six leptons and 4 gauge bosons [4].

2.1.1 Fundamental Interactions

Generally, it seems as if the world around us as we know it consists of mostly just a few fundamental particles. All matter consists of atoms, which are further comprised of neutrons and protons in their core and electrons around that core. But the more we understand all the physical processes involved, the more all particles appear to have interesting and important roles. The way electrons are bound to the nucleus is a low-energy scale **electromagnetic** property and described by the part of the Standard Model theory called Quantum Electrody-

namics (QED) [5]. The force, which is responsible for how protons and neutrons are bound in the nucleus is called the strong nuclear force and is the force corresponding to the **strong interaction**. The part of the SM theory, which deals with that is called Quantum Chromodynamics (QCD) [6]. Another fundamental interaction is the **weak interaction**, which deals with β decays and here appears another important particle: the neutrino ν. The next fundamental interaction is **gravitation** not included in the SM. It is typically much weaker than others; the gravitational force between a proton and an electron is about 10^{40} times weaker than the electromagnetic force. This brings scientists a lot of new ideas of potential undiscovered dimensions and new physics to be discovered, but for now we use it to describe the attraction of objects to one another, which is particularly important when talking about objects with astronomical scale, for example in space. The Yukawa process happens from electroweak symmetry breaking and the Higgs mechanism, which is the most important theory for this thesis. It describes the interaction strength between the Higgs boson field and fermion particles, which is responsible for the fermion particle masses.

At higher energy scales, even more detailed (and small) structures are observed, including smaller particles. In the experimental particle physics, the most fundamental constituents are currently believed to be the quarks and the leptons. The proton consists of two valence up quarks and one down quark $p(uud)$ and the neutron of two down quarks and one up quark $n(ddu)$. Together the electron-neutrino, electron, up and down quarks are known as the first generation of fundamental particles. They are each considered to be point-like. There are three generations in total (Figure 2.1), where each of the other two consists also of four particles, which differ in mass and their decays but are otherwise identical to the ones in the first generation.

2.2 Strong Interactions

Strong interactions are mediated by the exchange of massless particles called gluons. The theory of how those interactions happen is called Quantum Chromodynamics (QCD) [6]. Gluons interact with quarks and other gluons through the colour force. There are three colours in QCD: red, green and blue, and three corresponding anti-colours. Due to colour confinement all freely observed particles are colour neutral.

One explanation for colour confinement is the fact that two colour-connected quarks attract each other when pulled apart (Figure 2.2 a). If an electron-positron

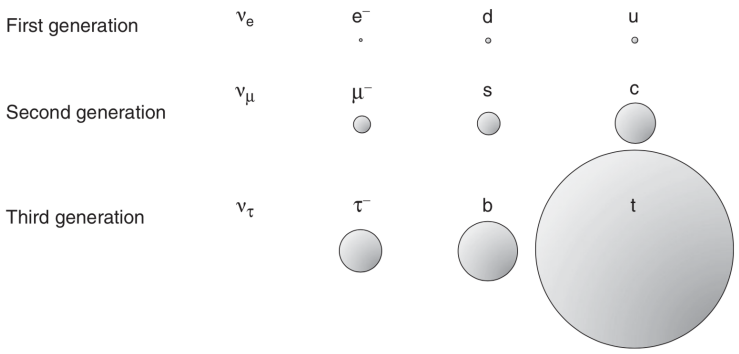


Figure 2.1: The three generations of fundamental fermions. The particles not shown are the anti-particles with opposite charge. First generation with the electron neutrino ν_e, electron e^-, down quark d and up quark u. Second generation with the muon neutrino ν_μ, the muon μ^-, the strange quark s and the charm quark c. Third generation with the tau neutrino ν_τ, the tau τ^-, the bottom quark b and the top quark t [7].

pair in QED is pulled apart by increasing the distance between the electron and the positron, the field lines between them would spread out (Figure 2.2 b). This is different in QCD, where if two quarks are pulled away from each other (Figure 2.2 c), the field lines are confined to a tube between the quarks and the force between the quarks is very large regardless of the separation. This means that it will require an infinite amount of energy to separate two quarks [4].

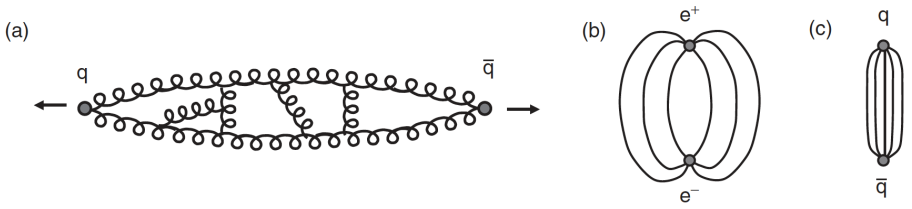


Figure 2.2: Fields in QED vs. QCD when particles moved apart. (a) attraction, (b) field lines spreading out, (c) field lines confined to a tube [4].

When a quark anti-quark pair is produced (say from the process $e^+e^- \rightarrow q\bar{q}$), the two initially free quarks, which are back to back cannot be observed individually, but rather each is observed as a collimated spray of colourless particles. The process is known as hadronisation and all its stages are shown in Figure 2.3. Hadronisation is the mechanism by which quarks and gluons produced in hard processes form the hadrons that are observed in the final state. First, the quark-antiquark pair is produced (i), followed by them separating where the QCD field between them is contained in a tube-like narrow shape (ii), then as they separate further, the potential energy becomes sufficient for the formation of new quark-

antiquark pairs (iii), more pairs are produced (iv) and finally the energy of the quarks is low enough for them to form a spray of colourless hadrons.

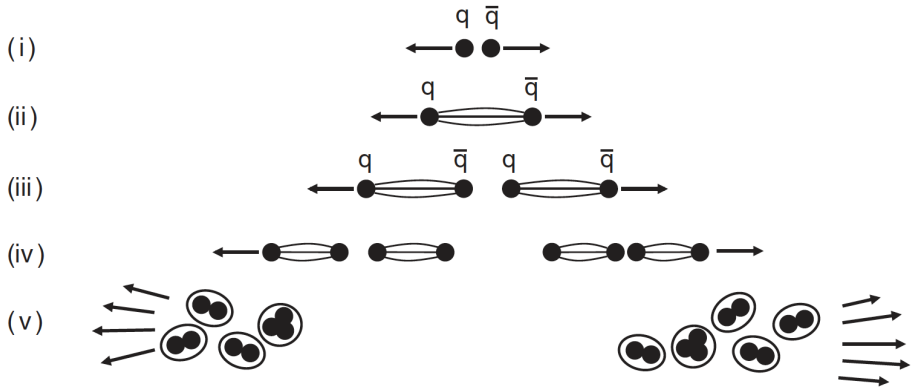


Figure 2.3: The five stages of producing hadrons [4].

Splitting functions [8] in QCD describe exactly how energy is shared between partons. They provide the mechanism for handling un-cancelled collinear divergences, which arise from the radiation of massless partons from one of the incoming partons taking part in a scattering process.

2.2.1 Kinematics

In hadron-hadron collisions, several different kinematic variables are used to describe the interactions. The most common QCD processes at the LHC are $pp \to jj + X$, where j denotes hadrons grouped in a jet. Frequently used kinematic variables are the angle of the two jets with respect to the beam axis (z-axis) and the component of the momenta in the plane transverse to the beam axis (xy plane), referred to as the transverse momentum. The pseudorapidity is a more convenient way for physicists to describe the angle of a particle relative to the beam axis, because the differences in pseudorapidities are Lorentz invariant under boosts along the longitudinal axis. This is an important variable in particle physics as colliding particles carry differential longitudinal momentum, which leads to different longitudinal boosts under different reference frames. The azimuthal angle is measured from the xy plane, around the beam. These measurements contribute to the final measurement of different cross-sections for interactions, which give information about both well known and new physical processes [4].

The direction of the beam axis is defined as z and the momentum transverse

to that axis is given as:

$$p_T = \sqrt{{p_x}^2 + {p_y}^2} \tag{2.1}$$

In processes like $pp \to jj + X$, the two resulting jets are then boosted in the direction of the beam in the detector. The rapidity is given in terms of the energy E of the resulting particle (a jet in the example here) and the momentum parallel to the beam axis, p_z, by:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.2}$$

In the majority of cases at the LHC, the energies with which the detector operates are so large, that the masses of some of the particles can be neglected. In such cases, the variable used is the pseudorapidity:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right). \tag{2.3}$$

2.2.2 Cross section

The cross section, σ, is a measurement of the probability that an event occurs. Due to the finite proton size, elastic scattering at high momentum scales are unlikely and inelastic reactions where the proton breaks up dominate. The cross-section for a proton-proton event must consider the parton distribution functions (PDFs), i.e., the probability density for finding a parton with a certain longitudinal proton momentum fraction, x, at momentum scale Q [9]. PDFs encode information about the proton's deep structure. The cross-section is therefore expressed as:

$$\sigma = \int dx_1 dx_2 f_i(x_1; Q^2) f_j(x_2; Q^2) \hat{\sigma_{ij}}(x_1, x_2, Q^2). \tag{2.4}$$

Individual Q can be extracted from a set of structure function measurements. Gluons are not measured directly, but carry about $1/2$ the proton's momentum. An example of PDFs at two specific values of Q^2 is showed in Figure 2.4, which shows the NNLO PDFs at scales of $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2, including the associated one-sigma (68%) confidence-level uncertainty bands. The gluon PDFs are largest and this is why ggF is the dominant production process at the LHC. The figures show the probability density functions weighted by x of the various partons. At small x, gluon is the most likely for both scales Q.

To calculate the rate or cross-section of a certain set of final state particles (say

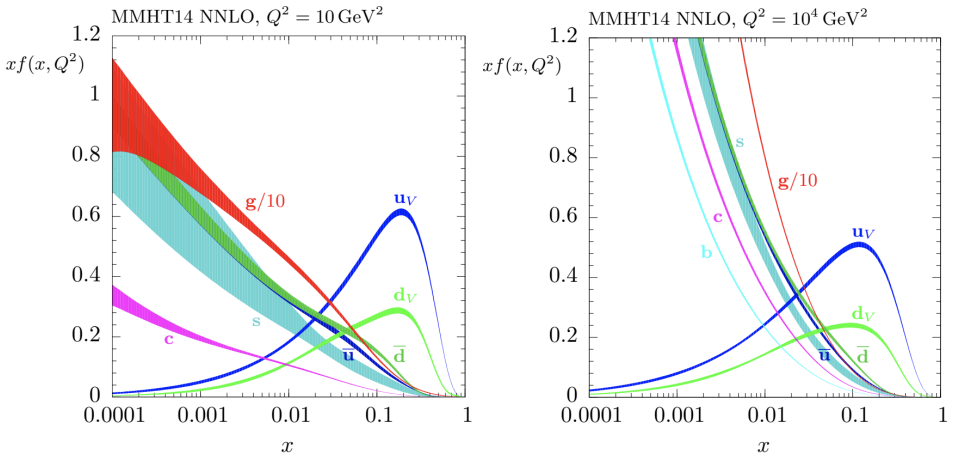


Figure 2.4: Example NNLO PDFs at scales of $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right) [9].

$jj+X$), in principle all possible combinations which lead to this final state, as well as all orders of QCD diagrams, would need to be calculated and the summation of all those contributions leads to the final cross-section of the process. If we just calculate the lowest-order term in QCD coupling, that is called a leading order (LO) diagram, If only the lowest-order term in QCD coupling is calculated, that is called a leading order (LO) diagram , if the second is calculated, it is next to leading order (NLO) [10]. In QCD the coupling strength is larger than in QED and therefore higher orders must be considered to obtain accurate predictions.

2.2.3 Phenomenology

The strong coupling, α_s, runs with the energy such that it is larger at lower momentum transfers than at higher momentum transfers. This changeable behaviour along with the complexity of the cross-section calculation are some of the main challenges the LHC physics faces when examining the properties of proton-proton events. In the modern age of QCD phenomenological calculations, ATLAS is one of the collaborations, which have proven to provide excellent comparison with the theoretical evidence we have for the existence of the $SU(3_{colour})$ non-Abelian (non-commutative group operation) gauge symmetry.

2.3 Electroweak interactions

The theory of the electroweak interaction [4] unifies two of the fundamental interactions, the weak and electromagnetic forces which at low energies seem very

different.

The W boson [1] is a spin-1 particle with a mass of 80.433 ± 0.009 GeV [11]. It has either a positive or a negative electric charge and a lifetime of $\approx 3 \times 10^{-25}$ s. It can decay to a lepton and an anti-neutrino with a branching ratio of about 33%, or to a quark and anti-quark pair. The three possible positively charged leptonic decays correspond to the three flavours of leptons and are as follows: $W^+ \to e^+\nu_e$, $W^+ \to \mu^+\nu_\mu$ and $W^+ \to \tau^+\nu_\tau$, where e^+ is a positron, μ^+ and τ^+ and ν_l is the corresponding neutrino. The lowest order Feynman diagram for the example process $W^- \to e^-\bar{\nu}_e$ can be seen in Figure 2.5.

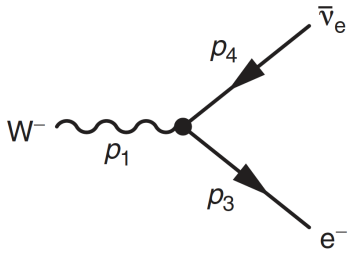


Figure 2.5: Feynman diagram of lowest order for one of the possible decay channels of the W boson. Conservation of momentum means that $p_1 = p_3 + p_4$, where p_1, p_3 and p_4 are the momenta used to describe the motion of the particles [4].

Sheldon Glashow, Abdus Salam, and Steven Weinberg all contributed to the unification of the weak and the electromagnetic interaction and created what is known as the Weinberg–Salam theory [12] [13] [14]. The weak interaction belongs to an SU(2) local gauge symmetry and after the unification with the electromagnetic interaction, that becomes the SU(2) × U(1) gauge group, which as before describes the exact transformations on the fields (W_1, W_2 and W_3 for the three SU(2) gauge bosons of the weak and B for the electromagnetic interaction), under which the dynamics of the system would not change. In the Standard Model, all particles associated with this unified interaction (W, Z bosons and the photon γ) are produced through processes called "spontaneous symmetry breaking" and "the Higgs mechanism" and are discussed further in Section 2.4.

The W_3 and B in the unified interaction are coalesced to the neutral Z boson and photon γ and considering the mixing angle θ_W:

$$\begin{bmatrix} \gamma \\ Z^0 \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} B \\ W_3 \end{bmatrix} \tag{2.5}$$

[1] W could be either W^+ or W^-.

The other two bosons involved (W_1 and W_2) combine to produce the charged massive bosons $W^{\pm}$:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.6}$$

The Lagrangian (described in more detail in Section 2.4) of this interaction before symmetry breaking contains a term, $\mathcal{L}_g$, which depends on the interaction between the three W bosons and B, the kinetic term for the Standard Model fermions, $\mathcal{L}_f$, the Higgs field, $\mathcal{L}_h$ and the Yukawa interaction with fermions, $\mathcal{L}_y$:

$$\mathcal{L}_{EW} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_h + \mathcal{L}_y \tag{2.7}$$

After electroweak symmetry breaking, the Lagrangian for the electroweak interaction includes the Higgs boson and takes a different form, due to the electroweak symmetry breaking. It is given by:

$$\mathcal{L}_{EW} = \mathcal{L}_k + \mathcal{L}_n + \mathcal{L}_c + \mathcal{L}_h + \mathcal{L}_{hv} + \mathcal{L}_{wwv} + \mathcal{L}_{wwvv} + \mathcal{L}_y \tag{2.8}$$

$\mathcal{L}_k$ is the kinetic term and includes all the mass terms and all the quadratic terms, $\mathcal{L}_n$ and $\mathcal{L}_c$ are the terms corresponding to the neutral and charged currents, $\mathcal{L}_{hv}$ corresponds to the Higgs interactions with the gauge vector bosons, $\mathcal{L}_{wwv}$ is for the gauge three-point self interactions and $\mathcal{L}_{wwvv}$ for the gauge four-point self interactions.

2.4 The Higgs Mechanism

The Higgs Mechanism [4] and the associated Higgs boson allow for the W and Z bosons to acquire mass after the electro-weak symmetry breaking as well as being responsible for the fermion masses.

To understand the basics of the Higgs mechanism, one should have a good understanding of the concept of the Lagrangian of the Standard Model. We start with the Lagrangian for classical systems, i.e., those which are present in our everyday experiences. The general equation used to describe any physical system, where T is the kinetic energy and V the potential is:

$$L = T - V \tag{2.9}$$

Replacing that with the Lagrangian density instead gives us the equivalent for a continuous system (instead of discrete coordinates), where we now need to

consider fields ϕ instead of points:

$$L \rightarrow \mathcal{L}(\phi_i, \partial\phi_i). \tag{2.10}$$

If we take as an example a free non-interacting scalar field with the Lagrangian would be given with the same equation as a spin-0 particle in Quantum Field Theory (QFT):

$$\mathcal{L}_S = \frac{1}{2}[(\partial_\mu\phi)(\partial^\mu\phi)) - m^2\phi^2]. \tag{2.11}$$

When we have an interacting scalar field, on the other hand, combinational terms (such as $\phi\bar{\phi}A^\mu$, where A^μ denotes the electromagnetic field) on top of the ones above, need to be considered in the new Lagrangian.

For a scalar field with a potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$, where the first term represents the mass of the particle and the second the self-interactions of the scalar field (say for a four-point interaction vertex), the Lagrangian now becomes:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4. \tag{2.12}$$

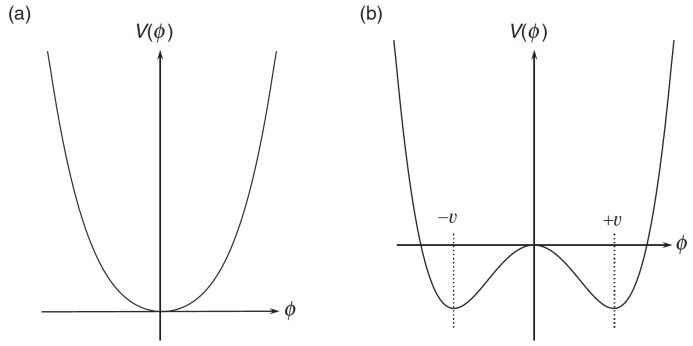


Figure 2.6: 1D potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ of a real scalar field ϕ for a) $\mu^2 > 0$ and b) $\mu^2 < 0$. If $\lambda \geq 0$, the minimum of $V(\phi)$ is at 0 in the case a). In the case b) we have two non-zero minima: $\pm\nu$ [4].

A simplified 2d case for the potential is illustrated in Figure 2.6. In the second scenario, 2.6 b) $\mu^2 < 0$, the term proportional to Φ^2 can no longer be interpreted as the mass term and there are two minima at non-zero vacuum expectation values $-\nu$ and $+\nu$. This choice between two local minima breaks the symmetry of the Lagrangian, which is what we call spontaneous symmetry breaking. To deal with this symmetry breaking, the mass term needs to be represented in a form, which still describes a massive scalar field. To do that, we express the field as excitations about the minimum $\phi = \nu + \eta$, or in other words, as small

perturbations. This leads to:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \frac{1}{2} m_\eta^2 \eta^2 - V(\eta). \tag{2.13}$$

The case of a complex scalar field which has a local U(1) symmetry (potential illustrated on Figure 2.7) is similar and the Lagrangian ends up as:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \frac{1}{2} m_\eta^2 \eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - V(\eta, \xi). \tag{2.14}$$

In the above, the scalar part of the field is denoted with η as before and is massive but now there are extra terms corresponding to the complex part of the field ξ, which is massless. This is called the "Goldstone boson".

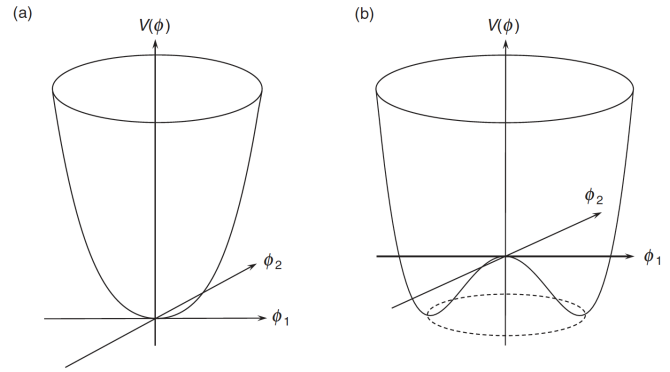


Figure 2.7: Potential $V(\phi)$ for a complex scalar field $\phi = \phi_1 + i\phi_2$. a) $\mu^2 > 0$, when both the real and complex fields are 0, so is the potential. b)$\mu^2 < 0$, the potential has an infinite set of minima $|\phi|^2 = \nu^2$ [4].

The full complex field Lagrangian of the Higgs mechanism after considering all the gauge transformation properties and the gauge field B and dealing with the broken symmetry can be written as:

$$\begin{aligned}\mathcal{L} = &\underbrace{\frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda\nu^2 h^2}_{\text{massive h scalar}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2\nu^2 B_\mu B^\mu}_{\text{massive gauge boson}} \\ &\underbrace{+ g^2\nu B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2}_{\text{h, B interactions}} - \underbrace{\lambda\nu h^3 - \lambda\nu h^3 - \frac{1}{4}\lambda h^4}_{\text{h self-interactions}}.\end{aligned} \tag{2.15}$$

This final Lagrangian describes fully the Higgs mechanism and the way particles acquire masses through their interactions with it. It is for a new quantum field; Higgs field and a massive gauge boson B associated with the U(1) local gauge symmetry. It has four main parts. The first part is the kinetic term for a massive scalar field, which is denoted with h, with constants for the self-interaction term λ^2 and the vacuum state ν, which sets the scale for the masses

of both the gauge boson and the Higgs boson. The second part is about the massive gauge boson with its kinetic term $F_{\mu\nu}F^{\mu\nu}$ and potential term $B_\mu B^\mu$ of the gauge field with coupling g. The third part describes the interactions between the massive scalar field h and the gauge field B and the final fourth part is about the self-interactions of the massive scalar field h. After the symmetry breaking and gauge transformation, the Goldstone field no longer appears. It has been replaced by the longitudinal polarisation state of the massive gauge field B (η is replayed by ν and h). Equation 2.15 is an example for how gauge bosons get their mass. Fermions instead would require a separate term.

2.5 The Higgs Boson

In 2012, the ATLAS experiments at CERN announced the observation of a new particle with measured by ATLAS [15] 5.9σ sensitivity and $m_H = 126.0 \pm 0.4(\text{sys}) \pm 0.4(\text{stat})$ GeV mass (confirmed by the CMS experiment [16]), which behaved in every way like a neutral scalar boson. It corresponded to the theoretical SM prediction for the Higgs boson - the most important missing (at that time) piece in the electroweak symmetry breaking scientific mystery. Being such a big part of the very fundamentals of our universe, responsible for the masses of fermions and gauge bosons in a local gauge invariance theory, electroweak symmetry and quark mixing, the discovery of the Higgs boson can easily be described as one of the most important physics discoveries of our century [17]. There were three problems with the electroweak theory which were resolved with the introduction of the Higgs boson. First, the massive gauge bosons and massive fermions were not allowed in the theory. The second problem comes from the fermions. The left-handed and right-handed fields helicity states ψ_R and ψ_L change differently under the gauge invariance and the symmetry is broken. The third problem is the violation of unitarity by for example WW-scattering at high energies [18]. All these three problems are resolved with the introduction of a new field which keeps the Lagrangian invariant. This new term is what we call the Higgs mechanism which is based on the neutral scalar Higgs boson.

2.5.1 Higgs couplings and Decay

The SM Higgs boson couplings are proportional to the masses of the coupled particles. The strongest couplings are to the decays to W and Z bosons, t and b quarks and τ leptons. The decay to two photon has a very small branching ratio $(2.28 \pm 0.01) \times 10^{-3}$ for $m_H = 125$ GeV, but is particularly attractive because of

the possibility to unravel new physics beyond the Standard Model and because of the large signal yield due to the high photon reconstruction and identification efficiency at ATLAS. The signal manifests itself as a narrow peak, due to the excellent ATLAS calorimeter resolution, on top of a smooth falling background [19].

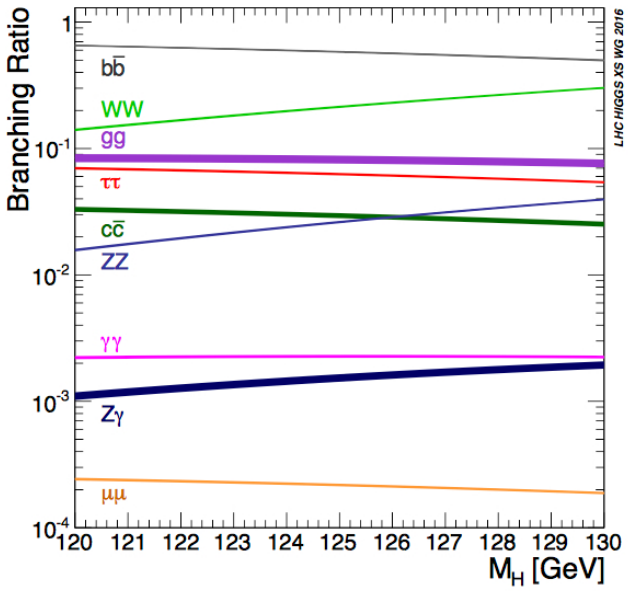
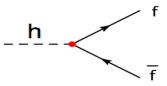


Figure 2.8: Higgs boson branching fractions for the mass region 120 - 130 GeV [20].

The branching ratios for the different Higgs decay channels are given using Figure 2.8 for the Higgs mass region (120-130) GeV and the branching ratios for the SM Higgs boson mass [17] [21] in Table 2.1. The decay rates to fermions, gauge bosons, gluons and photons can be calculated:

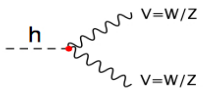


$$\Gamma(h \to f\bar{f}) = \frac{N_c}{8\pi v^2} m_f^2 m_h \sqrt{1-x} \qquad (2.16)$$

where $x = \frac{4m_f^2}{m_h^2}$, m_h is the mass of the Higgs boson, m_f of the fermion, N_c is the number of colours of the quarks and $v = 246GeV$, $g = 2\frac{M_W}{v}$

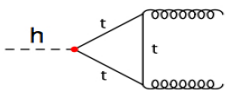
Process	Branching ratio [%]
$H \to bb$	$57.7^{+3.2\%}_{-3.3\%}$
$H \to WW$	$21.5\ ^{+4.3\%}_{-4.2\%}$
$H \to gg$	$8.57\ ^{+10.2\%}_{-10.0\%}$
$H \to \tau\tau$	$6.32^{+5.7\%}_{-5.7\%}$
$H \to \gamma\gamma$	$0.228\ ^{+5.0\%}_{-4.9\%}$
$H \to Z\gamma$	$0.154\ ^{+9.0\%}_{-8.8\%}$
$H \to \mu\mu$	$0.022\ ^{+6.0\%}_{-5.9\%}$

Table 2.1: SM branching ratios for all Higgs boson decay channels for a Higgs boson mass of 125 GeV. Uncertainties: QCD corrections were calculated by scale dependence of the width resulting from a variation of the scale by a factor 2 or from the size of known omitted corrections. EW corrections were calculated based on the known structure and size of the NLO corrections [22].



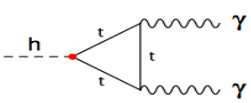
$$\Gamma(h \to VV) = \frac{g^2}{64\pi M_W^2} m_h^2 S_{VV} (1 - x + \frac{3}{4}x^2)\sqrt{1-x} \tag{2.17}$$

where $x = \frac{4m_V^2}{m_h^2}$ and $S_{WW} = 1, S_{ZZ} = \frac{1}{2}$ and g is the gauge coupling



$$\Gamma(h \to gg) = \frac{\alpha_s^2}{72\pi v^2} m_h^2 [1 + (\frac{95}{4} - \frac{7N_f}{6})\frac{\alpha_s}{\pi} + ...]^2 \tag{2.18}$$

where α_s is the coupling strength and N_f the number of fermions



$$\Gamma(h \to \gamma\gamma) = \frac{\alpha^2}{256\pi^3 v^2} m_h^3 [\frac{4}{3}\Sigma_f N^{(f)} e_f^2 - 7]^2 \tag{2.19}$$

where e_f is the fermion's EM charge.

2.5.2 Higgs Production

There are four main ways for the SM Higgs boson to be produced in proton-proton collisions. The dominant one is gluon fusion (Figure 2.9 a), following by vector boson fusion (Figure 2.9 b), then associated Higgs and electroweak boson production (WH and ZH), also called Higgsstrahlung (Figure 2.9 c) and the least likely, the associated Higgs and top quark pair production ttH (Figure 2.9 d).

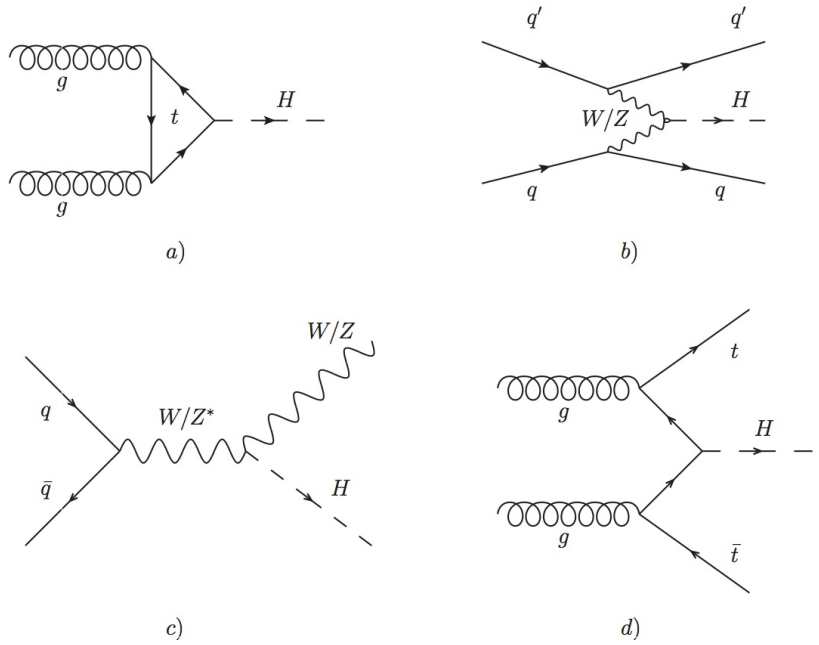


Figure 2.9: Feynmans diagram for the main production modes of the Higgs boson in pp collisions, where (a) is the dominant gluon fusion ggH,(b) vector boson fusion VBF, (c) associated Higgs and electroweak boson production WH, ZH, and (d) associated Higgs and top quark pair production ttH.

The most probable way for the Higgs to be produced is for two gluons to collide and forming a triangular W boson, top or bottom quark loops (Higgs coupling is proportional to mass and t and b quarks are heavy). It happens $\approx 80\%$ of the time and is ≈ 10 times more likely than vector boson fusion (VBF) [23] [24].

The second most likely production mode of the Higgs boson at LHC is VBF - two quarks collide to produce two virtual W or Z bosons which produce the Higgs boson alongside two quarks. For this process, the high mass behaviour of the Higgs has been one of the interesting recent studies as it is key for setting the upper limit of m_H and as it gives knowledge about the scattering of longitudinal vector bosons which is possible due to electroweak symmetry breaking [25].

Figure 2.9 c shows Higgs-Strahlung - two quarks collide to produce a virtual W or Z boson, which if energetic enough, emits a Higgs boson. This mechanism

is quite interesting in the high mass range of the virtual boson, because of the possibility to tag the boson and reconstruct the Higgs decay to two bottom quarks by using jet-substructure techniques [24].

Finally, there is the channel of interest in this thesis - top quark pair production, ttH. Although with the smallest cross section, this channel is of significant importance for the direct measurement of the top-Higgs Yukawa coupling. More about the current status of ttH can be found in the following section.

The total cross-sections for the five most likely production processes can be found in Table 2.2.

Process	σ at $\sqrt{s} = 13$ TeV [pb]	Uncertainty [%]
ggF	48.580	$^{+4.56}_{-6.72}$
VBF	3.782	$^{+0.43}_{-0.33}$
WH	0.943	$^{+0.5}_{-0.7}$
ZH	0.178	$^{+3.8}_{-3.1}$
ttH	0.516	$^{+6.0}_{-9.5}$

Table 2.2: Cross sections σ for all known Higgs production channels, calculated at Higgs mass of $m_H = 125$ GeV [26].

2.5.3 Higgs Couplings Constraints by ATLAS

In order to check for deviations in the bosonic and fermionic couplings of the Higgs boson from the SM predicted Higgs couplings, coupling modifiers are introduced and the κ-framework [27] used for their evaluation. The coupling modifier κ for decay mode j is defined as follows:

$$\kappa_j^2 = \frac{\sigma_j}{\sigma_j^{SM}} \tag{2.20}$$

σ is the measured cross section and σ_j^{SM} is the SM cross section. A coupling modifier κ_j for a production or decay process via the coupling to a particle j is defined as:

$$\kappa_j^2 = \frac{\sigma_j}{\sigma_j^{SM}} \quad \text{or} \quad \kappa_j^2 = \frac{\Gamma_j}{\Gamma_j^{SM}},$$

where σ_j is the measured cross-section, Γ_j the partial decay width into a pair of particles j, and $\sigma_j^{SM}, \Gamma_j^{SM}$ their SM values.

A common scaling can be assumed for all fermions κ_F, as well as for the electroweak bosons κ_V. "The best fit points and the 68% and 95% confidence

level (CL) intervals" [28]. The 68% and 95% CL likelihood is shown in Figure 2.10. Results are compatible with the SM prediction.

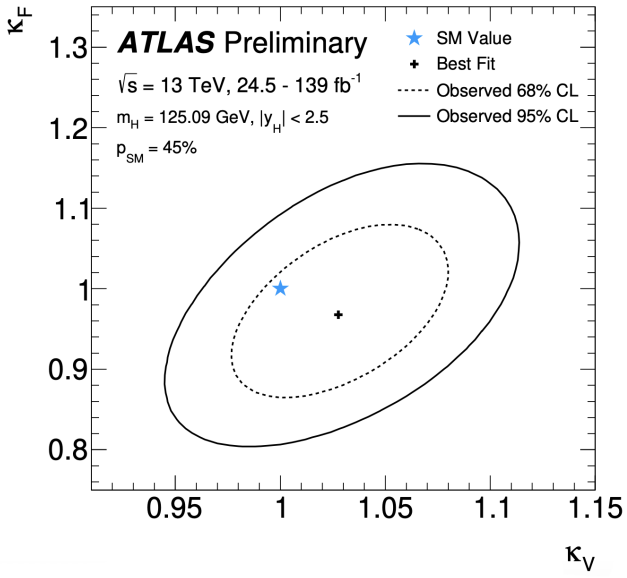


Figure 2.10: Fermion (κ_F) and boson (κ_V) coupling modifiers, obtained in the combination of Higgs boson production and decay measurements by the ATLAS collaboration [29]. The best-fit value, 68% and 95% CL contours and the SM value are shown.

To probe contributions of new particles though loops, the effective coupling strengths to photons and gluons κ_γ and κ_g are measured. Both κ_γ and κ_g are measured to be compatible with the SM expectation Figure 2.11.

2.6 Yukawa interaction

The Yukawa interaction was initially developed to model the strong force between hadrons or to describe the nuclear force between nucleons and pions. Later on it was expanded to describe the coupling between the Higgs field and the fermion fields [30].

All interactions between the Higgs field and the fermions are named Yukawa interactions. A generic interaction Lagrangian between a scalar doublet and the fermion fields is given by:

$$\mathcal{L}_Y = -\frac{1}{\sqrt{2}}(\nu+H).[h^i_e(\bar{e^i_L}e^i_R+\bar{e^i_R}e^i_L)+h^i_u(\bar{u^i_L}u^i_R+\bar{u^i_R}u^i_L)+h^i_d(\bar{d^i_L}d^i_R+\bar{d^i_R}d^i_L)]+h.c. \tag{2.21}$$

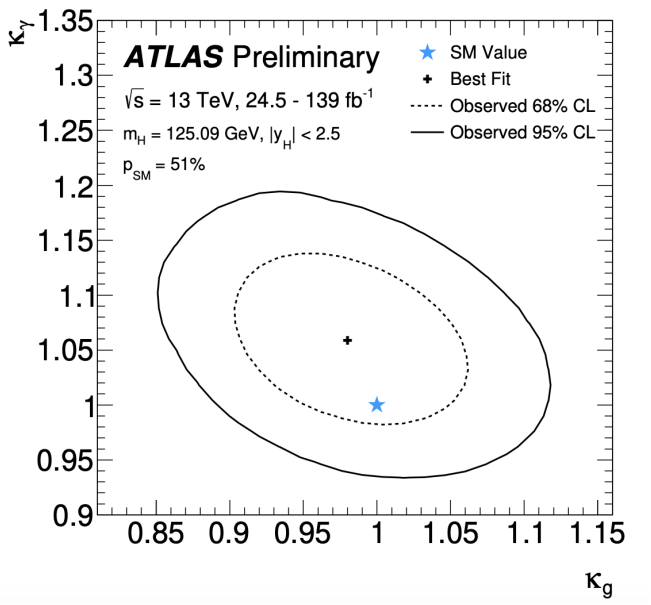


Figure 2.11: Photon (κ_γ) and gluon (κ_g) effective coupling modifiers, obtained in the combination of Higgs boson production and decay measurements by the ATLAS collaboration [29]. The best-fit value, 68% and 95% CL contours and the SM value are shown.

The three fermion families are denoted with u, d and e, H is a scalar field with non-zero vacuum-expectation value and denotes the Higgs boson, subscripts L and R correspond to left and right handed fermions. The two terms of Equation 2.21 are controlled by the Yukawa coupling h_f^i and show that the interaction of the Higgs boson with the fermions is proportional to their masses [31] [30] [32].

An essential ingredient of the Standard Model are the Cabibbo mixing of d and s quarks as well as other fermion flavour mixing. In the quark sector, the rotation to the mass eigenstate basis introduces the mixing among the families. Masses among the three flavours and on each fermion types always follow the following hierarchical order:

$$m_d(M) \ll m_s(M) \ll m_b(M) \tag{2.22}$$

Fermion mixing is given by:

$$M_f = \nu Y_f = \nu(Y_{f,b} + Y_{f,s} + Y_{f,d}), \tag{2.23}$$

where ν the non-vanishing vacuum expectation value of the neutral component of the Higgs field and Y_f is the fermion family coupling to the Higgs field. The mixing between the weak eigenstates of the d, s and b quarks is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [33]:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}$$

A non-diagonal element of the CKM matrix results in the coupling of the W boson to two quarks, which belong to two different fermion families. The matrix can be fully described by the 3 mixing angles, which control the mixing between the families and a parameter called the complex phase, responsible for the CP-violation.

The important Yukawa coupling, for this thesis, is that between top quark (largest mass in the SM) and the Higgs boson. If the final measurement is different from that expected given the top quark mass, that would mean new physics, which hasn't been fully explored yet and with every improvement of its measurement, we come closer to the full understanding of this.

2.7 Top quark and ttH production

Top quarks have an extremely short lifetime, and decay through the weak interaction into a W boson and a bottom quark with a branching ratio of almost 100%. In the ttH channel, there are two W bosons from the t quarks. There are three possible cases for the decay of the two W bosons; di-leptonic in which both decay to a lepton and a neutrino, semi-leptonic in which one decays to jets and fully hadronic in which both W bosons decay to hadrons. The branching ratios for the three cases are given in Table 2.3. In this thesis, the semi-leptonic and the di-leptonic channels are combined in one leptonic dataset.

1	2	3
45.7%	43.8%	10.5%

Table 2.3: The $t\bar{t}$ branching ratios for three cases: 1. fully hadronic $t\bar{t} \rightarrow W^+b, W^-\bar{b} \rightarrow q\bar{q}b, q\bar{q}\bar{b}$, 2. semi-leptonic $t\bar{t} \rightarrow W^+b, W^-\bar{b} \rightarrow q\bar{q}b, l\nu\bar{b}$ and 3. di-leptonic $t\bar{t} \rightarrow W^+b, W^-\bar{b} \rightarrow l\nu b, l^-\bar{\nu}\bar{b}$ respectively [34].

The di-leptonic top decay (Figure 2.12) is particularly challenging. First, it has a small branching fraction (Table 2.3), second there are two undetectable neutrinos, but with only one parameter for the missing transverse energy $\mathrm{E_T^{miss}}$.

The $ttH(H \rightarrow \gamma\gamma)$ channel, the subject of this thesis is of particularly high importance due to several factors, which will be discussed in more detail in 2.8.2.

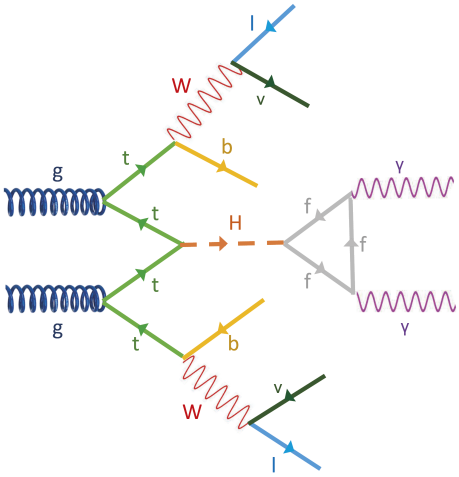


Figure 2.12: Leading order Feynman diagram for the di-leptonic $ttH(\gamma\gamma)$ channel, in associated Higgs and top production. Higgs decays to two photon through a fermion loop.

The Feynman diagrams for this process are shown in Figure 2.13 along with the three different decay possibilities for the W boson. As previously mentioned, the data are collected into only two categories: fully hadronic and leptonic (both single and di-lepton decays).

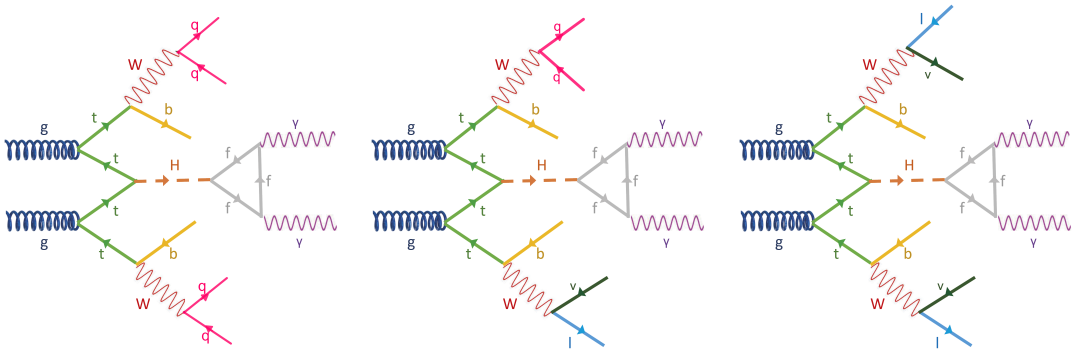


Figure 2.13: Example Feynman diagram of the production and decay channel interest of this thesis: $ttH(H \rightarrow \gamma\gamma)$. The Higgs boson does not decay to two photons directly but rather through a fermion loop. From left to right: fully hadronic, semi-leptonic and di-leptonic channels.

The ttH production cross-section at $\sqrt{s} = 13$ TeV is predicted to be: $\sigma_{ttH}^{SM} = 507^{+35}_{-30}$ fb, where the uncertainty corresponds to the combined scale and PDF uncertainty.

The cross-section has been calculated at NLO QCD and NLO EW accuracies [20].

2.8 Measurements of ttH production

2.8.1 Observation of ttH production

The ttH production was observed in 2018 with 80 fb^{-1} of LHC Run 2 data [35]. Figure 2.14 shows the ttH production cross sections measured in the individual Higgs boson decay modes at the time of the ttH observation at the LHC. The Higgs multi-lepton analysis, which targets $H \to WW^*$, $H \to ZZ^*$ and $H \to \tau\tau$ decays had the lowest total uncertainty of about 40% [35], with equal contributions from systematic and statistical uncertainties.

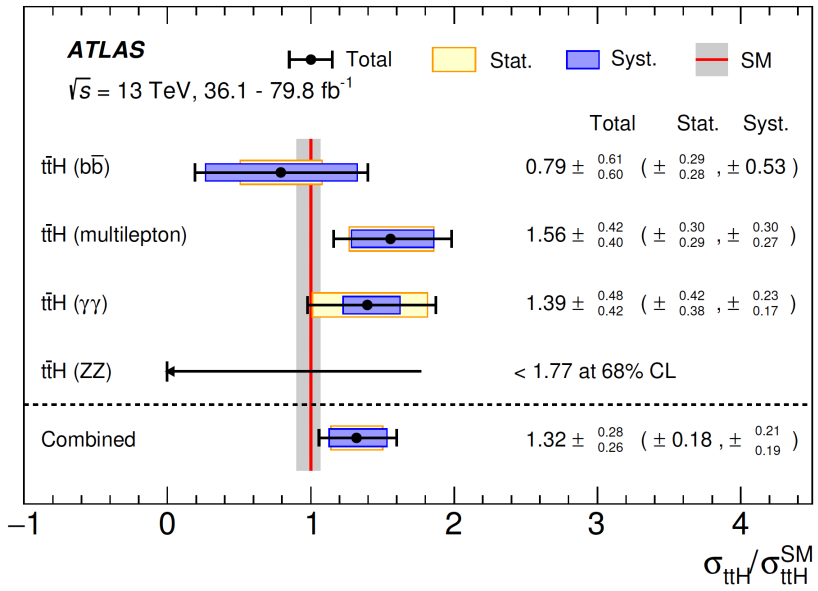


Figure 2.14: The combined ttH production cross-section over the SM prediction, as well as cross-sections measured in the individual decay modes of the Higgs boson as measured for the observation of the ttH production in 2018. Ratios of the measured values to the SM prediction are shown. The black lines show the total $\pm 1\sigma$ uncertainties, and the bands indicate the statistical and systematic uncertainties. The red vertical line indicates the SM cross-section and the grey band represents the uncertainties due to missing higher-order corrections [35].

2.8.2 Latest ttH measurements and prospects

As expected from extrapolating the results of the ttH observation described in the previous section, with the full Run 2 luminosity of 139 fb^{-1}, the $H \to \gamma\gamma$ decay channel provides the highest sensitivity [36]. Both ATLAS and CMS reached the total uncertainty of about 20% and observed the ttH production in the $H \to \gamma\gamma$ decays at 5 standard deviations (5σ) from the background-only

hypothesis [37] [38].

The latest measurements of the ttH signal strength μ, defined as the measured ttH cross-section divided by the SM cross-section value, are stated in Table 2.4 with their total uncertainties. The combined result is the 2018 observation of the ttH production. It uses Run 2 data, but does not yet include the latest measurements in $H \to bb$, $H \to \gamma\gamma$ and $H \to ZZ$ decay channels, which are listed separately.

Channel	Best-fit μ		Sensitivity		Ref.
	Observed	Expected	Observed	Expected	
$H \to b\bar{b}$	$0.43^{+0.36}_{-0.33}$	1.0 ± 0.6	1.3σ	3.0σ	[39]
$H \to \gamma\gamma$	$0.92^{+0.27}_{-0.24}$	$1.0^{+0.8}_{-0.6}$	4.7σ	5.0σ	[40]
$H \to ZZ^* \to 4l$	$1.7^{+1.7}_{-1.2}$	$1.0^{+3.2}_{-1.0}$	1.0σ	0.8σ	[41]
Combined	1.32 ± 0.27	1.0 ± 0.3	5.8σ	4.9σ	[35]

Table 2.4: Signal strength (μ) and sensitivity of the ttH cross-section measurements by ATLAS in the different Higgs boson decay channels. The combined result uses Run 2 data, but does not yet include the latest measurements in $H \to bb$, $H \to \gamma\gamma$ and $H \to ZZ \to 4l$ decay channels, which are listed separately.

With the LHC Run 2 data, the $H \to \gamma\gamma$ decay channel provides the highest accuracy measurement of ttH production.

There are several physical factors behind the high sensitivity of the $H \to \gamma\gamma$ measurements compared to the other Higgs boson decay channels. The $H \to \gamma\gamma$ decay channel had an uncertainty of about 45% [35], dominated by the statistical uncertainty. The $H \to bb$ decay channel had an uncertainty of about 60%, and was dominated by systematic uncertainty [42]. The $H \to bb$ decay channel has the highest branching ratio of $\approx 58\%$. As the ttH production cross-section is only about 507 $\pm$40 fb, this was advantageous in Run 1 and early Run 2 ttH searches. The b-jets are identified with an efficiency of about 70%, and jets are reconstructed an energy resolution of about 50 $\%/\sqrt{E}$, which yields a Higgs mass peak with the width of about 10 GeV. In this broad peak, contamination from irreducible background from $ttbb$ production is large. The modelling uncertainty of this $ttbb$ background limits the sensitivity of ttH measurements in the $H \to bb$ decay channel [43] [44].

In the $H \to$ multi-lepton decay channel, the branching ratio is about 6% and the signal identification is high due to the presence of leptons. Most signal events contain missing transverse energy from τ decays. As the E_T^{miss} resolution is low, the resulting Higgs peak is relatively broad, with relatively large contamination

from the irreducible non-resonant ttW background. The modelling uncertainty of this background limits the sensitivity of ttH measurements in the multi-lepton decay channel [45] [46].

Experimentally, the $H \to ZZ^*$ decay channel is reconstructed with even narrower mass peak than the $H \to \gamma\gamma$ decay channel. However, it has a branching ratio of only 0.012% and is therefore expected to remain limited by the statistical uncertainty even at the future high-luminosity LHC, as shown on Figure 2.15.

Finally, in the $H \to \gamma\gamma$ decay channel, the branching ratio is about 0.23%, but the final state photons provide high identification efficiency. As all Higgs decay products are reconstructed with high resolution ($\sigma_E/E \approx 10\%/\sqrt{E}$), the resulting Higgs peak is narrow, and the non-resonant $tt\gamma\gamma$ background contamination small. The background can be readily estimated from the fit to the data side-band. The systematic uncertainties are low, and the Run 2 data-set provides sufficient statistics for an accurate measurement.

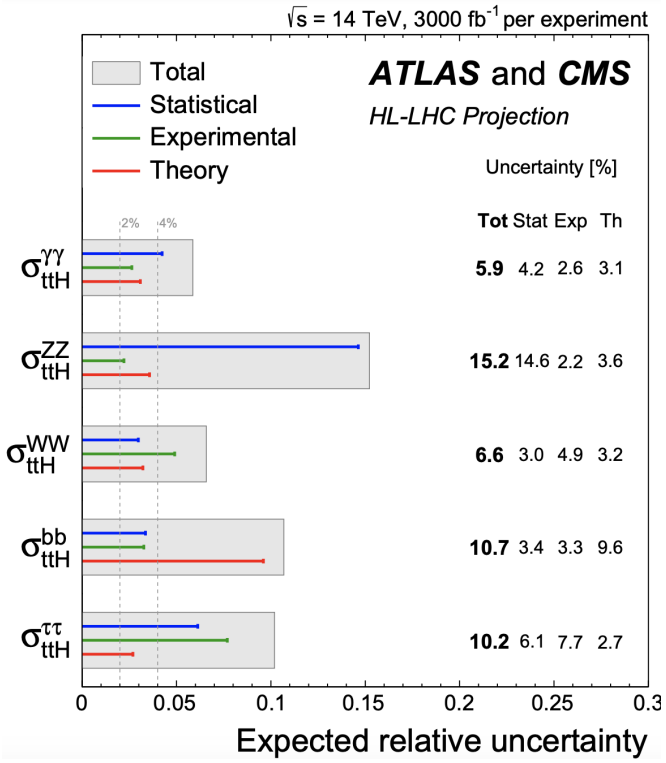


Figure 2.15: Expected uncertainties on the ttH production cross sections for ATLAS and CMS at the future high-luminosity LHC. For each measurement, the total uncertainty is indicated by a grey box while the statistical, experimental and theory uncertainties are indicated by a blue, green and red line respectively. In addition, the numerical values in % are also reported [47].

2.9 Beyond the Standard Model

Despite the Standard Model being able to describe all known physical phenomena in one theory, there are still unanswered questions. Examples include [7]: why do the fermion masses follow the observed hierarchy? Why is there far more matter than antimatter in the observable universe? What is dark matter? How to include gravity?

Some compelling evidence for potential beyond SM physics in the dark matter sector are the tangential velocity calculations of large objects in our Universe and also some calculations in the cosmic model related to the Cosmic Microwave background (CMB) [48], which both confirm the existence of dark matter but so far have not given specifics for what the majority of it consists of or how to detect it. The leading candidates for dark matter are weakly interacting massive particles (WIMPs) [49], axions [50] and sterile neutrinos but as to date, there is no clear evidence supporting either. WIMPs appeared for a long time as a perfect dark-matter candidate, as new particles at the weak-interaction mass scale (10 GeV to 1 TeV) would be produced naturally with the right relic abundance in the early Universe [51]. There was also hope for the resolution of the hierarchy problem. Unfortunately, no detection of other than the Higgs boson particles at the electroweak scale has been made so far [52]. The cosmological models also predict the existence of dark energy [53].

Another example for compelling evidence for potential beyond SM physics is the fact that CP violation, which we have measured (eg. in [54] and [55]) and theorised so far in the SM does not seem to be sufficient enough to explain the observed matter-antimatter asymmetry of the Universe. Parity symmetry is the invariance of physics under a discrete transformation, which changes the sign of the space coordinates [56] [57] [58]. Charge symmetry is the existence of a particle with an opposite charge for every particle but with exactly the same properties and violation is the lack of the existence of an exact pair.

Additionally, the SM predicts all charged leptons to have identical EW interaction strengths (lepton universality). Evidence has been observed by the LHCb collaboration at CERN for the breaking of lepton universality [59].

Other examples can be found in the lack of success in the attempts for full unification of all forces (gravity is much weaker than the others), extra dimensions, which are possible mathematically, the unexplained difference of mass of the neutrinos in comparison to all other fermions etc.

In conclusion, there's a lot more to be discovered and with the increased data in Run 3, high-luminosity LHC, and future colliders, we can expect some

revolutionary discoveries in the future.

Chapter 3

The ATLAS Experiment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [60] is the world's largest and highest energy particle accelerator. It was first operational in the year of 2008 with first collection of data in the autumn of 2009. It consists of a 27 km main ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles. During its first run (colliding bunches of protons and collecting the data), the protons were collided with a centre of mass energy of $\sqrt{s}$ = 7 TeV (2011) and $\sqrt{s}$ = 8 TeV (2012) and in the second run with a centre of mass energy of $\sqrt{s}$ = 13 TeV (2015-2018), which makes it the most powerful collider in the world built to date. The full ATLAS 13 TeV Run 2 data corresponds to an integrated luminosity of 139 fb^{-1}. The trigger menu improved due to the increase of the instantaneous luminosity and the number of pile-up interactions. Peak instantaneous luminosities ranged from 0.5×10^{34} cm^{-2}s^{-1} to 2.1×10^{34} cm^{-2}s^{-1} and 28–60 pile-up collisions [61]. These pile-up collisions are multiple *pp* interactions in the same bunch crossings.

The CERN accelerator as a whole (Figure 3.1) consists of a system, which prepares and accelerates the protons to their injection energy in the LHC, in which they are further accelerated, focused and collided. The preparation includes stripping the electrons from the hydrogen atoms using an electric field, accelerating in a linear trajectory (LINAC2) to an energy of 50 MeV and then further accelerating in the three synchrotron systems (BOOSTER) to 1.4 GeV. Guiding the particles, which have high energy and flux, requires extremely powerful superconducting magnets. An extraordinary cooling system for those bending magnets is necessary. The magnets have to be cooled to -271.3°C (a temperature lower than the average temperature of the Universe) and for that purpose,

a distribution system with liquid helium is connected to the magnets.

The LHC consists of eight straight parts and eight circular arcs connected to one another. The protons are accelerated in the RF cavities in one of the straight components. The bunches in the LHC are held in the circular orbit by the bending magnets. There are 1232 main dipoles, each 15 metres long and weighing 35 tonnes. They bend the particle trajectories, while quadrupole magnets for focusing. Quadrupoles have four magnetic poles arranged symmetrically around the beam pipe to squeeze the beam either vertically or horizontally. The two beams of protons circulate around the LHC in opposite directions, colliding in the locations of the different experiments: ATLAS [62], CMS [63], LHCb [64] and ALICE [65]. The beam dump system consists of 15 fast extraction magnets (MKD),15 magnetic septa (MSD) and 10 dilution kickers (MKB) together with the various control system elements [66].

3.2 The ATLAS Detector

ATLAS [62] is the largest of the LHC detectors and used for various purposes including searches and measurements in the Higgs boson sector, searches for new BSM particles eg. possible candidates for dark matter, precision Standard Model measurements (W mass, top quark physics etc.), heavy ion physics and the investigation of the matter/antimatter asymmetry through CP-violation. The coordinate system used by ATLAS is right-handed, with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. The detector consists of four main parts: the Inner Detector (ID), surrounded by a 2 T magnetic field, the Electromagnetic liquid-argon Calorimeter (ECAL), Hadronic Calorimeter (HCAL) and Muon Spectrometer (MS).

The ID consists of three sub-detectors: Pixel, Semiconductor Tracker and Transition Radiation Tracker (TRT). It provides efficient and precise tracking measurements of the kinematics of charged particles through examining their trajectories. The magnetic field bends the charged particles and creates curvatures, the measurements of which provide momentum and charge of the particles [67]. The TRT also provides electron identification via transition radiation measurements.

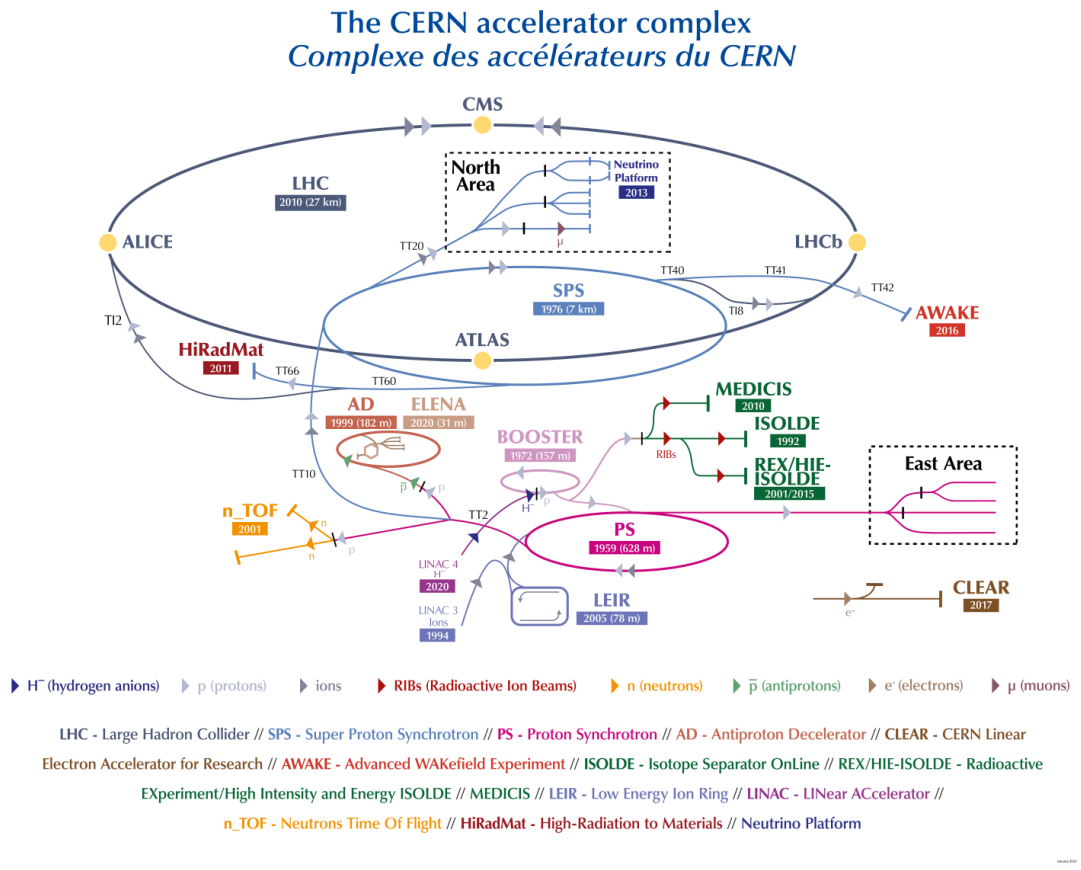


Figure 3.1: The accelerator complex at CERN is a succession of machines that accelerate particles to increasingly higher energies. Each machine or accelerator boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence. The last stage is the Large Hadron Collider (LHC), where four major experiments are installed: ALICE, ATLAS, CMS, LHCb. They use detectors to analyse the myriad of particles produced by collisions in the accelerator [60].

The ECAL is used for electron and photon identification and measurements, missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) and jet measurements. The HCAL detects mainly hadrons that interact via the strong and electromagnetic force and is predominantly made out of iron as an absorber and scintillating tiles as an active material.

The MS is the outer-most sub-detector due to the penetrating power of muons. It complements the calorimeters and the tracker to identify and reconstruct muons. The reconstruction is done through looking for the hit patterns in the different layers of the MS, creating segments and combining them together to build the track candidates [68].

3.2.1 Inner detector

The inner detector, ID [69], shown in Figure 3.2 is a part of the detector, with which particle tracks are found. The detector requires high-precision measurements to achieve the momentum and vertex resolution needed to study fundamental physics processes. That requirement is met by pixel and Semiconductor (SCT) trackers combined with straw tubes in the transition radiation tracker (TRT).

The ID acts as dead material in front of the calorimeter, which reduces the calorimeter's resolution. To ensure the required precise tracking, several measuring points are needed along the particle's trajectory. This leads to the need of multiple tracking layers. Therefore, the innermost detector must have a very fine granularity. This minimises the occupancy, while maximising the impact parameter (distance of closest approach of the track to the collision point) resolution. This maximising of the resolution leads to the overall improvement of performance and the decrease in number of fake hit assignments. This inner most part is called the pixel detector technology. It consists of three barrel layers concentric with the beam line and centred on the interaction point.

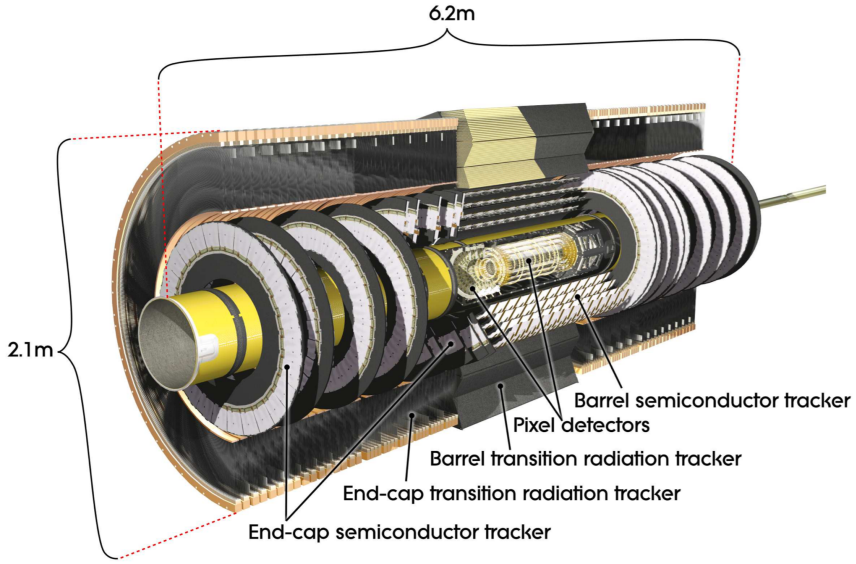


Figure 3.2: Cut-away view of the ATLAS Inner Detector [62]

The pixel detector is 1.3 m long and provides a three-hit system for particles with $|\eta| < 2.5$ [70]. The three layers are situated at radius of 50.5 mm, 88.5 mm and 122.5 mm from the beam pipe respectively. It consists of 1700 identical modules with 80 million pixels. The radiation hardness requirement is 50 MRad [71].

The inner most pixel layer is called the Insertable B-Layer (IBL) [72] and has been operational since the beginning of Run 2 data taking. Providing additional hit information at the closest position to the beam collision point, the IBL significantly improves the performance of b-jet tagging, resolving decay vertices for b, c and τ.

The semiconductor tracker [73] provides additional precision position measurements for the reconstruction of charged particle tracks after the pixel detector. It covers $|\eta| < 2.5$ and consists of four barrel layers and nine end-cap disks. Each of these consists of two module layers, placed at 40 mrad stereo angle between the direction of their respective microstrip sensors. In this way, the SCT provides 2D position measurements.

The Transition Radiation Tracker (TRT), or the most outer part of ID [74], consists of $\approx$ 350 000 drift/straw tubes, 4 mm each. It provides 36 space points in $\eta < 2$ and for $p_T > 0.5$ GeV/c. It detects transition radiation x-ray photons in an Xe-based gas mixture, which provides electron identification capability.

The final high-precision of measurements and pattern recognition in both $R-\phi$ and z polar coordinates is achieved by combining the SCT tracker information with the TRT hits at larger radii. The straw hits at the outer radius are an important contribution to the momentum measurement. The tracker also helps the calorimeters with the electron identification by detecting transition-radiation photons in the xenon-based gas mixture of the straw tubes.

3.2.2 Calorimeters

Electromagnetic and hadronic showers

Calorimeters in particle physics are devices, which measure the deposited energies of particles traversing the calorimeter's material. Most particles enter the calorimeter and initiate particle showers, depositing either their full energies or a sample of their energies. Simulations of calorimeters include a certain particle injection into the calorimeter medium and the study of the processes, which follow. A view of the calorimeters is presented in Figure 3.3. ATLAS calorimetry has what is called sampling structure. This means that the active signal generation and passive particle absorption are performed in two separate media. The alternative would be a single material for both, which is called a homogenous structure. It consists of the hadronic end-cap $1.5 <| \eta |< 3.2$, the electromagnetic barrel $| \eta |< 2.5$ (lead and liquid argon), the electromagnetic end-cap (also lead and liquid argon), the forward calorimeter $3.1 <| \eta |< 4.9$ (copper-tungsten

and liquid argon) and the hadronic barrel calorimeter (iron and scintillating tile) covering the pseudorapidity range $|\eta| < 1.7$.

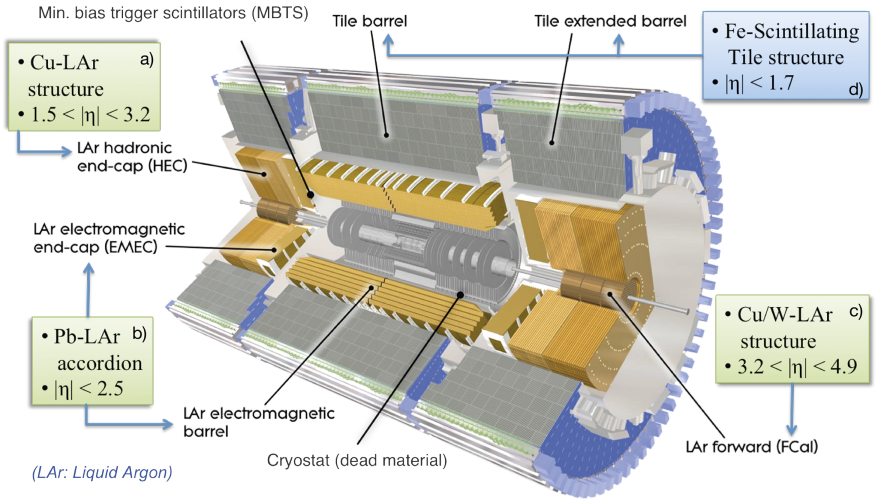


Figure 3.3: ATLAS calorimeter components; a) hadronic calorimeter (HCAL) end-cap, b) electromagnetic calorimeter (ECAL) barrel and end-cap, c) forward calorimeter (FCAL) and d) hadronic calorimeter (HCAL) barrel [62].

In the work presented in Chapter 5.2, three types of particles were used for the generated simulation samples - electrons, photons and pions. At low energies, photons lose energy through Compton scattering and the photoelectric effect and electrons through ionisation and electron capture. For electron interactions above 10 MeV, the dominant mechanism through which energy is lost is bremsstrahlung (when a charged particle loses energy, by emitting photons, as a result of being deflected by another charged particle), for photons at energies of 1.02 MeV and above it is pair production (e^-e^+). Shower shape parameters are energy dependent, but the underlying processes (pair production and bremsstrahlung) become energy independent above 1 GeV. Showers initiated by electrons develop initially in a different way than those initiated by photons. The shower development is described by a radiation length X0. The X0 is the specific length traversed in a material. For electrons 1 X0 is defined as the material passed until their energy falls to 1/e of their initial energy. In Figure 3.4 the energy loss for high energy electrons and photons traversing 5 X0 is shown. It shows the fraction of energy deposited, it can be seen that electrons lose approximately 21% on average and photons 14.8%, but the width of the e- distribution is much narrower [75]. After the secondary photons have been produced, the process repeats until the energy of the final electron reaches a critical value, ϵ. The mean energy deposition $\langle E(x) \rangle$

is represented using the radiation length X_0:

$$\left\langle E(x) \right\rangle = E_0 e^{\frac{x}{X_0}} \tag{3.1}$$

for electrons, where $E(x)$ is the energy, x the distance, E_0 the initial energy and X_0 the radiation length and

$$\left\langle I(x) \right\rangle = I_0 e^{\frac{7}{9}\frac{x}{X_0}} \tag{3.2}$$

for photons, where $\left\langle I(x) \right\rangle$ is the mean intensity, I_0 is the initial intensity. Photons reach 1/e of their initial intensity after a distance of $\frac{7}{9}\frac{x}{X_0}$ [76].

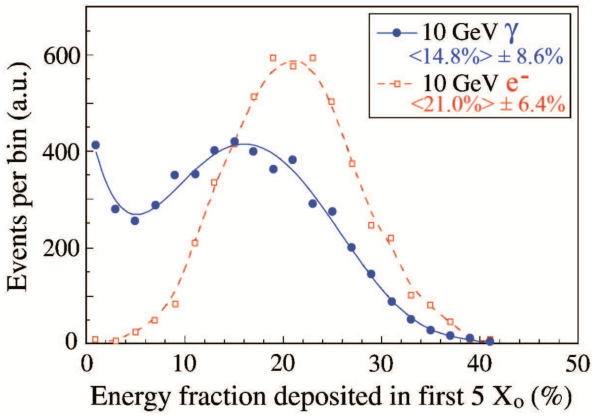


Figure 3.4: Distribution of energy fraction deposition in 5 X_0 for e^- (red) and for γ (blue). The mean energy fraction deposited is higher for the electron, as expected [75].

To understand the behaviour of the pions, a grasp of hadron calorimetry is also needed. Hadronic showers are more complicated and have longer radiation lengths than electromagnetic showers due to the more complex hadronic and nuclear processes such as excitation, nuclear capture, nucleon evaporation, spallation etc. Hadronic showers also have an electromagnetic component due to the neutral pions π^0 which decay to two photons. Protons and neutrons from a pion induced shower are released from the nucleus. The binding energy has to be provided, therefore the fraction of the shower energy needed for that is invisible and its contribution to calorimeter signal has be accounted for. The ratio between visible EM and visible hadronic energy is a crucial part in the understanding and improving the resolution of a hadronic calorimeter. The visible energy for electrons is given by:

$$E^{\mathrm{e}}_{\mathrm{vis}} = E\eta_e \tag{3.3}$$

where η_e and η_h are the efficiencies for observing purely electro-magnetic and purely hadronic signals and E is the incident energy. The visible energy for pions is given by:

$$E^{\pi}_{\text{vis}} = \eta_e(F_{\pi_0} + \frac{\eta_h}{\eta_e}F_h)E, \tag{3.4}$$

where F_{π_0} and F_h are the pion and the hadronic fractions.

Electromagnetic Calorimeter

In the Electromagnetic Calorimeter (ECAL), photons, electrons and positrons interact with the inner material to produce showers. The architecture shown in Figure 3.3 consists of a cylindrical barrel centred on the beam and two end-caps. It has an accordion geometry for gaps and absorbers. The active medium is liquid argon gas (LAr) cooled by a cryostat to a temperature of 88 K with the purpose of keeping it in its liquid form. The passive material is lead [77].

Hadronic Calorimeter

In the Hadronic Calorimeter (HCAL) shown in Figure 3.3, sprays of particles interact with the material inside and produce hadronic showers. It consists of three parts: the Tile [78], LAr hadronic end-cap and the LAr forward calorimeters [79]. The Tile calorimeter has an absorber material made of steel and the LAr hadronic end-cap has an absorber material made of copper. The LAr forward calorimeter (FCAL) consists of three modules in each end-cap, where the innermost module also has copper and the two outer modules have tungsten as absorber materials.

Resolution

There are many factors that contribute to the deterioration of the response eg. noise from electronics, material changes, instrument effects etc. The energy resolution is expressed as:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{3.5}$$

where $\oplus$ is the symbol for the quadratic sum. The first term in Equation 3.5 is due to statistical fluctuations in the shower development. In homogeneous calorimeters, the intrinsic fluctuations are actually smaller than the statistical prediction due to the fact that the energy deposited in the active medium does not fluctuate event by event. This is measured by the Fano factor [80]. In

sampling calorimeters on the other hand, the deposited energy does fluctuate event by event, due to the fluctuations of the number of particles which traverse the active layers. Therefore, sampling calorimeters have lower energy resolution compared to homogeneous calorimeters [62] [76].

The second (noise) term is due to electronic noise in the detector. When a signal is collected in the form of light, for example with photomultipliers, the noise term is smaller than when collected by charge. The noise term decreases linearly with increased energy of the incident particle. In sampling calorimeters, the noise term can be decreased by increasing the sampling fraction and therefore improving signal to noise ratio.

The third (constant) term includes calibration inhomogeneities, imperfections in geometry of the detector, leakages in longitudinal energy component and energy loses in dead material. As the energy increases, the other terms decrease and the constant term dominates the energy resolution.

The energy resolution of the current ATLAS ECAL can be seen in Figure 3.5. The performance goal for the ECAL is $\sigma_E/E = 10\% \oplus 0.7\%$ over a range of $|\eta| < 3.2$ and for the HCAL: $\sigma_{p_T}/E = 50\%/\sqrt{E} \oplus 3\%$ in the barrel and end-cap and $\sigma_{p_T}/E = 100\%/\sqrt{E} \oplus 10\%$ in the forward region [81]. These resolutions are after noise subtraction and fitting only the stochastic and constant terms. The relative resolution improves with energy, which means that particles with higher E_T are measured more accurately.

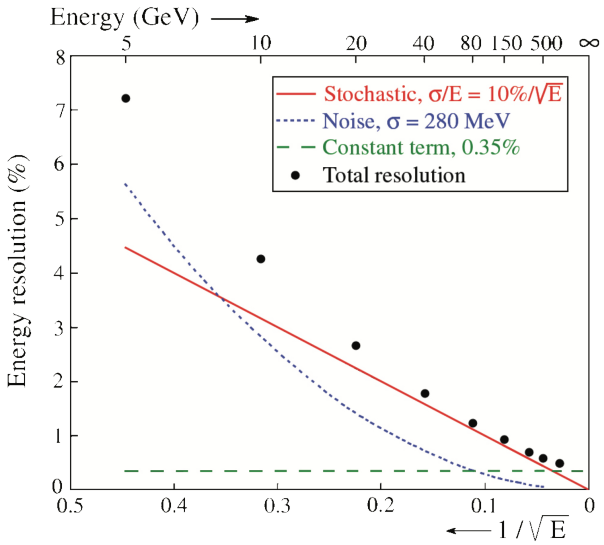


Figure 3.5: ATLAS EM calorimeter energy resolution with all contributions - stochastic (red), noise (blue) and constant (green) terms. The total resolution is obtained from electron test-beam data, and the contributions from a fit to this data [75].

3.2.3 Muon System

The muon spectrometer system [62] (Figure 3.6) is based on magnetic deflection of muon tracks in three large superconducting air-core toroid magnets. The performance of the bending power of the magnets is calculated by $\int Bdl$, where B is the magnetic field component normal to the muon direction. Two of the magnets are located in the end-caps of the muon system and one in the barrel region. The barrel toroid provides 1.5 to 5.5 Tm of bending power in the pseudorapidity range of $0 < |\eta| < 1.4$ and the two end-cap toroids 1 to 7.5 Tm in the range $1.6 < |\eta| < 2.7$. The region between is called transition region and has a lower magnetic field strength. The measurements of track coordinates are done with Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs). The trigger system consists of Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs), which are responsible for bunch-crossing identification, p_T thresholds and the measurements of the muon coordinates in the direction orthogonal to the one determined by the precision-tracking chambers.

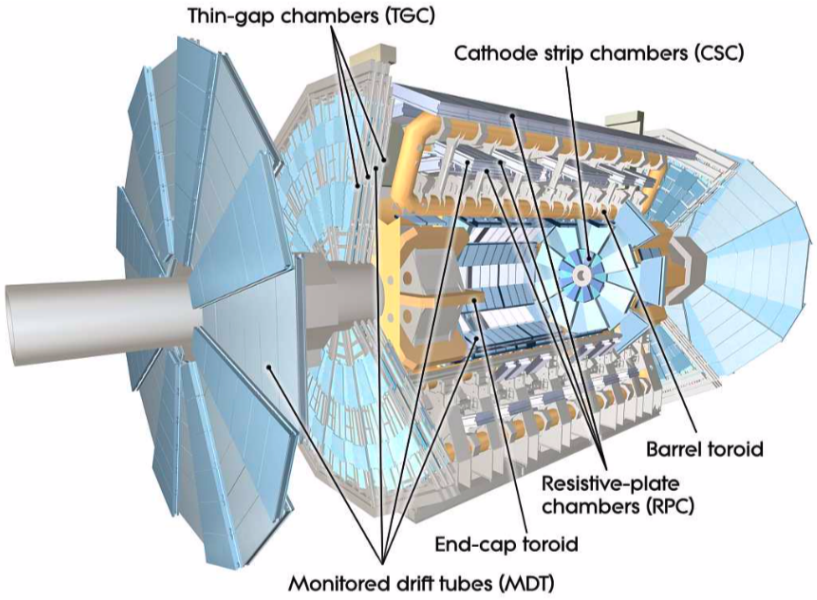


Figure 3.6: Sketch of ATLAS muon system. The detector size is about 22 m in diameter and 44 m in length [62].

The expected resolution of the muon spectrometer is illustrated in Figure 3.7 with respect to the transverse momentum of the muons. The resolution for muons with p_T around 100 GeV is 4%, which increases to 10% at 1 TeV. For $p_T < 100$ GeV, the dominant process is multiple scattering and for $p_T > 100$ GeV, calibration and alignment of the spectrometer are the main contributors in

the momentum resolution. The muons cross three layers of MDT chambers for sagitta measurements. The measurements related to the tracks are performed with a resolution of ≈ 40 μm. For muons with $p_T = 1$ TeV, the resultant sagitta is ≈ 500 μm and for the highest expected resolution of 10%, the sagitta is ≈ 50 μm [82]. The higher the momentum, the more precision is required in the measurement of the sagitta.

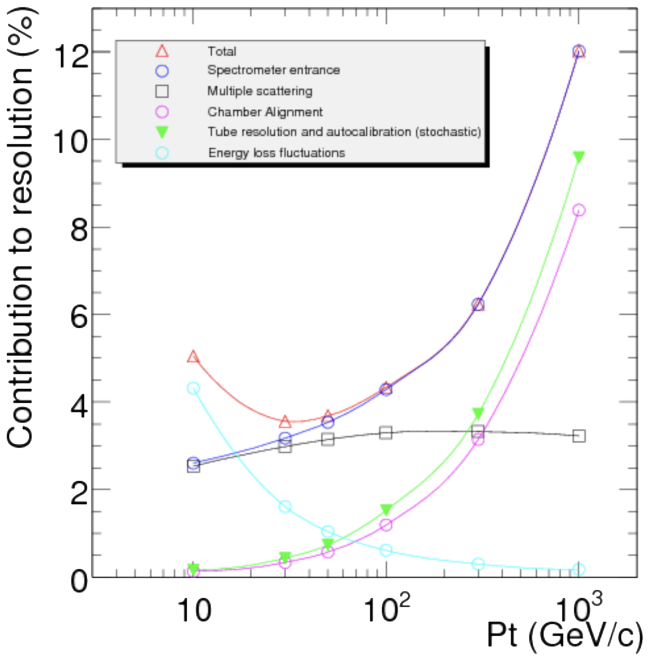


Figure 3.7: Resolution of the muon spectrometer for the different contributing factors [82].

3.2.4 Forward Detectors

The forward detectors of ATLAS are located in the forward region. Two of them: LUCID (LUminosity measurement using Cherenkov Integrating Detector) in pseudorapidity range $5.6 < |\eta| < 5.9$ [83] and ALFA(Absolute Luminosity For ATLAS) in pseudorapidity range $10.6 < |\eta| < 13.5$ [84], determine the luminosity (number of events) delivered to the experiment and the third: ZDC (Zero Degree Calorimeter) in pseudorapidity range $|\eta| > 8.3$ [85] determines the centrality of heavy-ion collisions [62]. LUCID measures instantaneous luminosity at $z = \pm 17$ m and ALFA at $z = \pm 240$ m. The main idea of the forward detectors is to measure elastic scattering at very small angles, so that the calculations of the absolute luminosity at ATLAS can be made with the appropriate precision [86].

3.2.5 Trigger and readout

The trigger system is responsible for event selection to significantly reduce the rate of collecting data, which can be up to 1 TB/s. It consists of three levels: L1, high-level L2 trigger and Event Filter, where L1 includes muon and calorimeter triggers and L2 and Event Filter are combined in HLT and deal with Regions of Interest and whole event physics respectively [62]. The trigger can be adjusted to a set of conditions, which vary with respect to the types of events of interest, and therefore the type to be thrown out. The general trigger constraints used for Run 2 are the maximum L1 rate of 100 kHz (75 kHz in Run 1) defined by the ATLAS readout capacity and an HLT average rate of 1 000 Hz (400 Hz in Run 1), defined by the off-line computing capacity [87]. The primary event processing occurs at CERN in a what is called Tier-0 facility. The RAW data starts on site at CERN and is the copied to different sites around the world, which are called Tier-1 facilities. At this stage of the process, scientists process and analyse the data and then, it is copied further to Tier-2 facilities [88].

The read-out system is responsible for transferring data from the detector, the configuration and control of the hardware and software components and the conversion of the detector's responses into human storable electronic information. The read-out system receives and temporary stores the data in the local buffers. The selected events passing both levels of the triggers are then transferred to the event-building system and the event filter for final selection. Those resultant events are passed to the CERN computing centre. The data acquisition system is called TDAQ. It takes 2.5 μs for the signal from the detector to reach L1, therefore all the data, which reaches in that 2.5 μs needs to be processed, while $\approx$ 100 other bunch crossings take place in the mean time. This is the reason for ATLAS to have buffers as part of their front-end electronics. An important occurrence is time between measurements, which is called "dead time". The data flow in the ATLAS sub-detector acquisition systems needs to be controlled in order to prevent information losses. To minimize dead time, during which no data can be recorded, TDAQ has a parallel processing technique. At the end of Run 2, the simple dead time setting was four bunch crossings, which corresponds to an inefficiency of about 1% for a L1 rate of 90 kHz [61].

Chapter 4

Analysis strategy and Datasets

The thesis targets $ttH(H\gamma\gamma)$ final states. The signature of the Higgs boson in the di-photon decay channel is a narrow peak in the smoothly falling di-photon invariant mass $M_{\gamma\gamma}$ distribution. The width of the peak is consistent with the resolution of the detector and is typically between 1 GeV and 2 GeV, depending on the kinematics in the event. The mass and event rate of the Higgs boson can be inferred from the fits of the $M_{\gamma\gamma}$ distribution.

Backgrounds with non-resonant photon pairs such as multi-jet production in association with photons or tt production in association with photons, can be rejected mainly by using the photon kinematics. The signal is extracted from an $M_{\gamma\gamma}$ fit with a narrow peak on the top of a substantial background. Using the kinematic variables of the photons will sculpt the background, or otherwise said, a background peak will appear exactly where the Higgs boson mass peak in $M_{\gamma\gamma}$ is. This thesis, therefore, introduces a scientific way to de-correlate the cuts set on photon variables from $M_{\gamma\gamma}$ in order to remove the sculpting of the background and achieve smaller uncertainties. The state-of-the-art $ttH(H\gamma\gamma)$ analyses [43] have used simpler, approximate procedures to avoid the sculpting, such as scaling the photon kinematics with $M_{\gamma\gamma}$.

4.1 Monte Carlo signal and background simulated data

Monte Carlo (MC) simulations are used to develop the $ttH(H\gamma\gamma)$ analysis and estimate the expected sensitivity. The first step is the event generation, in which matrix element events are produced, showered and hadronised. The event generation relies on MC four vector description generators, which are writ-

ten separately by third parties and interfaced to the ATLAS software framework [89]. The different processes used in this thesis were generated using different MC software packages. The signal ttH is produced using *PowhegBox v2* generator [90] [91] [92], at next to leading order (NLO) in the strong coupling constant α_s. The *NNPDF3.0nlo* set of parton distribution functions is used [93]. *Pythia8.230* [94] with *A14* tune [95], is used for parton shower and hadronisation. The decays of bottom and charm hadrons are simulated using a different generator called *EvtGen v1.6.0* [96].

Following the event generation, the signal events are passed through the ATLAS simulation infrastructure, using the Geant4 toolkit [69]. They are normalised to the inclusive cross-section of $\sigma = 0.51$ pb calculated at 13 TeV and Higgs mass of 125.09 GeV [97] and the branching ratio for the Higgs decaying to two photons of 0.227% [20].

The main background process is assumed to be from the $tt\gamma\gamma$ production, for which the matrix element events were generated with *MadGraph5 v2.3.3* [98] at LO in QCD, using *NNPDF2.3lo* parton distribution function. The shower and hadronization are the same as for the ttH signal events. The detector response is simulated using a fast parametric simulation of the ATLAS calorimeter [69]. In all simulated samples, pile-up events are modelled with *Pythia 8.186* using *A3 tune* [99].

4.2 Real data collected with the ATLAS detector

The data is from proton-proton collisions at the centre-of-mass energy of $\sqrt{s} =$ 13 TeV. It was collected with the ATLAS detector at CERN between 2015-2018. This data is called Run 2 data, as it is collected during the second data collecting period of time after major detector upgrades. Events in which the calorimeters or the inner detector were not fully operational are excluded, using the data quality requirements in Ref. [100]. After the data quality requirements, the data has a preliminary integrated luminosity of 139.0 ± 2.4 fb^{-1} [101] and an average number of interactions per bunch crossing of 33.7.

Events used for this study are required to pass a di-photon trigger with thresholds of > 35 GeV and > 25 GeV for the leading and sub-leading photon respectively. The trigger uses shower shape information from the calorimeter to identify the photons. In 2015-2016, photons were required to pass the loose photon identification criterion at the trigger level. In 2017-2018 the medium identification requirement was used, to cope with the higher instantaneous luminosity. After

passing the trigger di-photon trigger, the events were required to contain at least one primary vertex, and the offline photons had to match the photons identified by the trigger.

Chapter 5

Simulation

Simulation is an important part of particle physics analyses at CERN. It is needed for three general purposes: to allow the detection efficiency to be measured, to study the performance of future detector designs before construction and to ease the work of physicists by providing them with the possibility for applying numerous methods to manipulate the data from the detector, without actually using the physical detector itself.

This chapter is devised into two sections and ten subsections. Section 5.1 describes the general ATLAS simulation and the fast simulations used for less CPU-intensive running of physical processes in the detector. Section 5.2 describes the fast calorimeter simulation AtlFast3, which is the next generation of high-accuracy fast simulation in ATLAS and which combines parametrisation based approaches and machine learning techniques. The subsections of Section 5.2 are ordered as follows: simulation data samples used in Section 5.2.1, energy parametrisation in Section 5.2.2 , energy interpolation, which the author's main personal contribution to AtlFast3 in Section 5.2.3, all final corrections, made through validation by comparing with the general full Geant4 simulation in Section 5.2.4 and reconstruction of physical objects in Section 5.2.5. The physics list and performance studies, which also include personal contributions are described in Sections 5.2.6 and 5.2.7. Finally, conclusions on the AtlFast3 fast calorimeter simulation are given in Section 5.2.8.

5.1 ATLAS simulation

To study the ATLAS detector's response for as many physics scenarios as computationally possible, a detailed simulation was developed to carry events from the event generation level through to the final step of outputting the data in

the same format as the data received from the physical detector (Raw Data Object type converted to byte stream) [69]. The ATLAS simulation uses the Athena framework [102] for the software architecture and the Geant4 toolkit [103] [104] for the simulation of passing particles through the detector geometry [105]. The ATLAS simulation operates on three general levels: generation of the particles, interactions of the particles and their decay particles with the detector, pileup and digitisation of the output energy depositions in the sensitive parts of the detector geometry. The reason for making sure that the formats produced for simulated and real data are the same is because the simulated data, just like the real data, should be run through the same ATLAS trigger and reconstruction packages to avoid biases. The computational power required for the full chain is large, so several fast simulations were developed separately. The fast simulations have simplified detector geometry and various response parametrisations for analyses, which don't require the full complexity of the detector or to facilitate Monte Carlo to data tuning. Moreover, with latest improvements of fast simulation, there is an ongoing discussion for using them as the default for analyses [69].

The full chain of the ATLAS simulation is shown in Figure 5.1. The first step is at generator level, where particles are produced in a format called HepMC [106]. This is an object oriented event record written in C++ and used specifically for high energy physics simulations [106]. The process of particle generation also includes particle filtering, where only particles of interest can be chosen. Event generation jobs can be run for several thousands of events at a time. The detector is not described at that level, because only prompt decays are dealt with at this stage, stable particles are stored and propagated through the detector during the simulation stage at Geant4. The physical processes of the propagated particles through the detector are also described with Geant4. The energies deposited in the sensitive parts of the detector are recorded as "hits" and the stored information about each event (called truth) includes the momentum, decay time and tracks of incoming and ongoing particles.

The output with the hit information goes through the digitization stage next. The noise of the detector is considered at this stage and digits are produced. The first level trigger (hardware trigger) and all information is outputted in a Raw Data Object (RDO) format. Hypotheses are evaluated, but no events are discarded.

Finally, the ATLAS high level trigger (HLT) and reconstruction use the RDO output files. Apart from truth information, which is simulation specific, the reconstruction is identical for simulation and data.

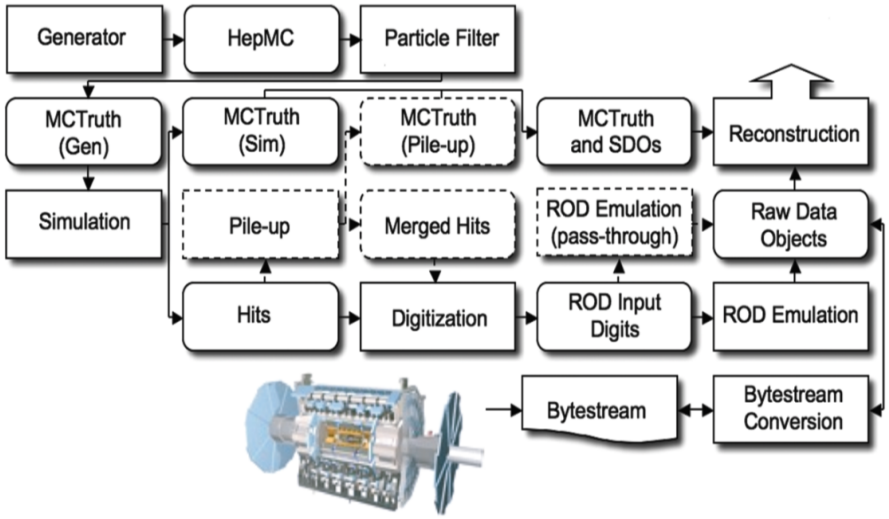


Figure 5.1: Flow of the ATLAS Simulation software [69].

5.1.1 Physics Lists

Physics lists are collections of numerical models of the interactions of particles in Geant4. A single model is usually not enough for the description of a complex process, as it only involves a specific type of interaction and for a specific range, which is why the use of several models together is required. The lists used by the ATLAS collaboration are provided by Geant4 with the exception of the transition radiation model, which was added by the ATLAS collaboration. Physics lists used by ATLAS are `FTFP_BERT` (since 2014, i.e. for Run 2), `QCSP_BERT` (before 2014) and `QGSP_BERT_HP` was used for specialised neutron studies, eg. for cavern background.

The currently used physics list `FTFP_BERT` [107] has a hadronic and an electromagnetic component. The hadronic package includes the Fritiof string Precompound (FTFP) and the Bertini intranuclear cascade model (BERT). The electromagnetic package includes step limiting multiple Coulomb scattering. Transportation processes dominate the inner detector simulation and electromagnetic physics processes dominate the calorimeter simulation, due to the large number of soft electrons, positrons and photons within the showers.

5.1.2 Fast Simulations

The simulation of physical processes in ATLAS depends on Monte Carlo based creation of events. The Monte Carlo (MC) method deals with the idea of solving mathematical, statistical and more importantly for the studies in this thesis, physical problems, which have many degrees of freedom. It does that by random

sampling to evaluate the integrals numerically, which can not be solved analytically. The simulation of the development of particle showers in the events is very CPU (Central Processing Unit) intensive, so much so, that it takes 90% of the overall simulation time. Therefore, fast simulations are essential to simulate enough MC events to match the real data with limited computing resources. The slowest software component of the full ATLAS simulation is the calorimeter part. This led to the development of the alternative fast simulation tool called Fast Calorimeter Simulation (FCS).

5.2 Fast Calorimeter Simulation

The calorimeter of the ATLAS experiment (described in Chapter 3.2.2) measures the energies deposited from the resultant particles of collisions and decays. The fast simulation used for this purpose is the AtlFast3 or Fast Calorimeter Simulation (FCS). Performance studies of a much improved parametrization-based Fast Calorimeter Simulation can be found in Appendix D. The particles modelled in FCS are photons (γ), electrons ($e^{\pm}$) and pions ($\pi^{\pm}$) for the physical processes in the electromagnetic (EM) and hadronic (HCAL) parts of the calorimeter. Input particles that are used for the parametrisation are produced with discrete values of the logarithm of their momenta in range (0.064 - 4) TeV and uniformly distributed in η bins with size 0.05 up to $\eta < 5$. The corresponding η values for the different parts of the detector are shown in Figure 5.2. The detector response for the input particles is simulated using the original Geant4 simulation of the ATLAS detector under Run 2 conditions, but with electronic noise, cross talk between neighbouring cells and dead cells turned off. The current AtlFast3 includes parametrisation-based and a machine learning based parts, where the former deals with the parametrisation of the single particle calorimeter simulation and the later is currently used for improving the performance of single pions in the momentum range of 16 GeV to 256 GeV. An energy interpolation is used to make sure that both are consistent with one another and that the transition is smooth.

5.2.1 Simulation of Reference Samples for AtlFast3

All AtlFast3 fast simulation samples were compared to the full Geant4 (G4) ATLAS simulation. [103] The Geant4 samples produced used Geant4 version 10.1.3. The physics list used was *FTFP-BERT-ATL* [108], [109], [110], which uses the Bertini intra-nuclear cascade model below 9 GeV and transitions to

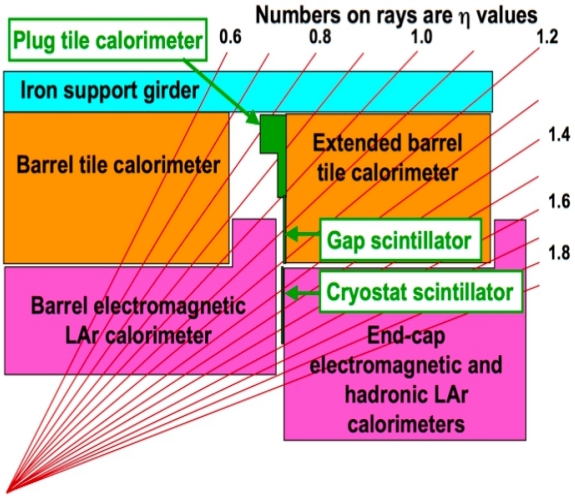


Figure 5.2: Pseudorapidity values for the ATLAS calorimeter.

the Fritiof model with a pre-compound model for 12 GeV and higher [1]. The "frozen showers" technique, which is used for the full ATLAS Geant4 simulation in general, to simulate electromagnetic showers [69] is omitted in the Geant4 reference samples.

All reference Geant4 samples were produced for single particles with coordinates $|z| < 3550$ mm outside of the TRT and a cylinder around them with a radius r = 1148 mm. The particles are produced with a uniform distribution in ϕ and without the spread of the LHC beam in z, to correspond to the simplifications adopted in AtlFast3. All showers are parametrised from the the particles entrances to the calorimeters. The hadronic showers are parametrised with positive and negative pions $\pi^{\pm}$ and the electromagnetic showers with electrons $e^{\pm}$ and photons γ. Corrections were made to ensure that all particles perform optimally. The differences in shower development due to the lack of the modelling of the beam spread were considered negligible [1]. Other differences between the produced Geant4 reference samples and the used samples for ATLAS analyses in general (full ATLAS simulation) include the omission of any cross talk or read-out noise from electronics and smaller simulation step.

For a complete comparison to AtlFast3, the simulation samples included the same energy and pseudorapidity ranges and steps between them and an energy interpolation 5.2.3 was used to generalised the results for any available real-life possibility. The energy range for the incoming single particles is 16 MeV $< E <$ 4.2 TeV, with steps of powers of 2 and the pseudorapidity of $|\eta| \leq 5$ with steps of 0.05.

After all comparisons are performed, several corrections are applied to the

AtlFast3 responses, which are described in 5.2.4.

5.2.2 Energy Parametrisation

The energies deposited in the different layers of the ATLAS calorimeter are very strongly correlated to one another, which makes it difficult to simulate them correctly. A method using principal component analysis (PCA) was developed to de-correlate the longitudinal deposited energies between the different layers of the calorimeter. The principal components of the data are a sequence of directional vectors, where every vector is orthogonal to the previous one. These components are eigenvectors of the data's covariance matrix. The directions of the vectors constitute an orthonormal basis, in which the different dimensions of the data space are linearly uncorrelated. This is also referred to as a "change of basis". A Geant4 single particle input sample is used with a particular η bin and particular fraction of the total energy in each layer. Using an inverse error function, the energy distributions for all layers are then converted into Gaussian distributions. A PCA matrix is created using the information given by the Gaussian distributions. The distributions are then changed so that they use bins with the same number of events per bin, after which a PCA is performed again, to further de-correlate the distributions of the different layers from one another [111].

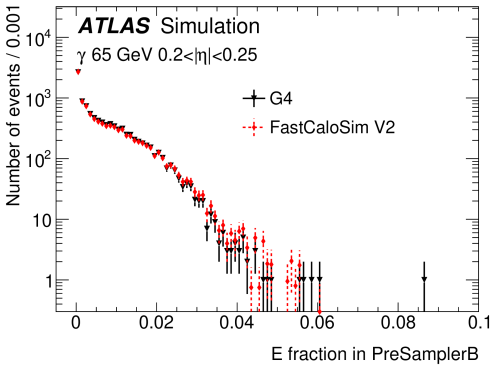


Figure 5.3: Comparison of the energy parametrisation in FCS with G4. PreSampler [1].

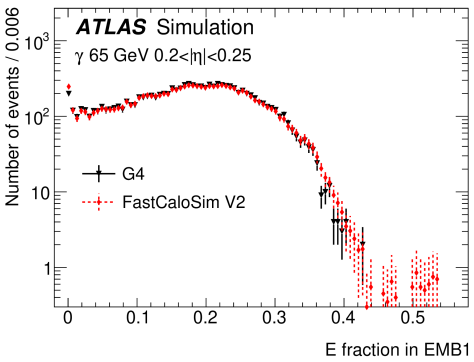


Figure 5.4: Comparison of the energy parametrisation in FCS with G4. EMB1 (first layer) [1].

In the simulation, a similar PCA procedure is applied but in reversed order. A random bin is assigned for each simulated particle until there's a certain random number of events per bin, the distribution of which is Gaussian. The random numbers are rotated using a PCA matrix, resulting in transformed Gaussian

distributions. The error function is then used to make these distributions (which are at this stage uncorrelated) correlated with each other. After small corrections are applied to the resulting distributions to approximate them as much as possible to the expected quantities, the final energy distributions are obtained for each layer. If the particle does not deposit any energy in the calorimeter, it is assigned to the so called "0 bin".

The full longitudinal energy parametrisation results for two example parts of the calorimeter from validations with respect to the full Geant4 (G4) simulation are shown in Figures (5.3 and 5.4). The examples given in the Figures are for 65 GeV photons in the pseudorapidity range $0.2 < |\eta| < 0.25$. There is a general agreement between the G4 and FCS in the separate components.

5.2.3 Energy Interpolation

As mentioned earlier, in FCS, three types of particles are parametrised and used - photons, electrons and charged pions. Photons and electrons are used to parametrize electromagnetic showers and pions are used to parametrize hadronic showers. In the simulation at the time of this analysis (2019-2020), generated samples were available with specific energies of the incident particle with energies between 64 MeV and 4 TeV and 100 equidistant pseudorapidity $|\eta|$ slices in the full range $0 < |\eta| < 5.0$ [111]. Having only discrete values is insufficient as in reality, the particles entering the calorimeter will have a continuous range of kinetic energies. A software package to address this issue by interpolating between the different energies was written in C++ and included in the full ATLAS simulation infrastructure Athena.

To investigate what type of functionality (linear or more complex) was needed for the interpolation, the mean of the simulated total distribution of the energies deposited in all calorimeter cells over the energy of the sample was plotted against the energy of the sample. The energy loss was observed in order to check for potential irregularities. An example for the individual energy distributions of the incident particle is shown for a single pion for three of the used energies in Figure 5.5.

The tails of the distribution are asymmetric, therefore each and every entry has to be considered when calculating the mean. A perfect detector's response would be an ideal Gaussian shaped distribution, but a realistic one gives a more complicated shape. It consists of a Gaussian core, that models the detector resolution, with tails, which are non-Gaussian and parametrize the effect of photon radiation by the final state particles in the decay. The asymmetrical tails are

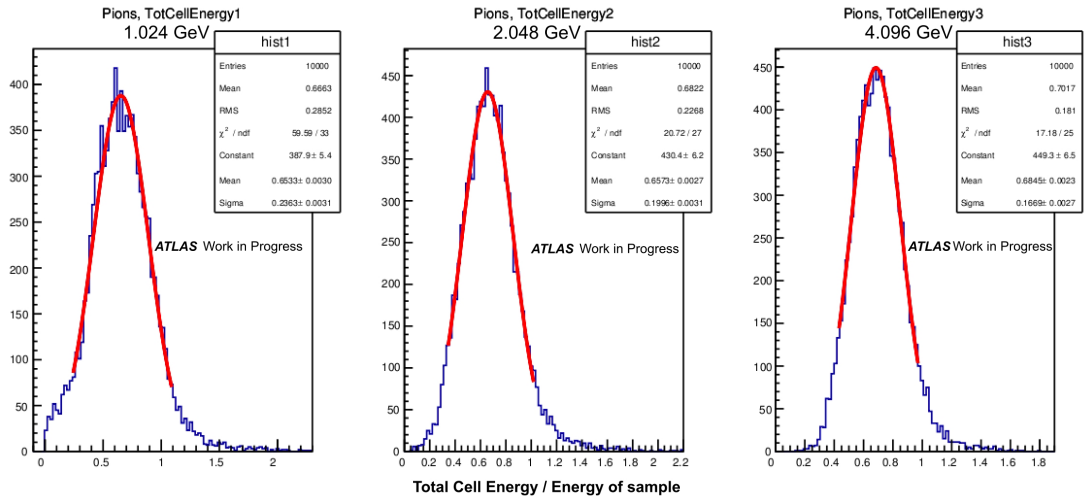


Figure 5.5: Individual pion distributions of the ratio between the total cell energy and the energy of the sample for incident energies of 1.024 GeV (left), 2.048 GeV (middle) and 4.096 GeV(right) of the pion with Geant4. Fits to the peak were made with a Gaussian function.

due to fluctuations (shower leakages) in the energy resolution of the calorimeter. They tend to occur in the low-energy side of the distributions because energy is escaping from the active detector volume.

The expected behaviour at lower energies (Figure 5.6) is for the fraction of the total cell energy over the energy of the sample to increase. The expected behaviour at higher energies (different for the different particles) is for the above mentioned ratio to approach and plateau at 1. These expectations were met for the photon and the electron samples. The corresponding plots for three example pseudorapidity ranges are shown for single pion, photon and electron samples. There is a clear loss of energy. This is due to the material in front of the calorimeter. In the full simulation, the electron comes from the interaction point or from photon conversions, which means the electron passes through material before the calorimeter. For photons, an energy loss of $3-5\%$ and for pions of $15-30\%$ is considered normal.

The expected plateau at one is seen for the photon and electron but not for pions in the investigated energy range. This is because most of the energy of electrons and photons is visible, which is not the case for pions. Some fraction of the energy is used for nuclear excitations and some to release nucleons from nuclei. This binding energy contributes to the invisible energy of pions. It was also noted that as expected, the shape varies in the transition region between barrel and end-cap (≈ 1.47, black line in Figure 5.6). The different patterns for the pions sample gave a first indication of a potentially more complicated energy

interpolation shape needed for the pion energies.

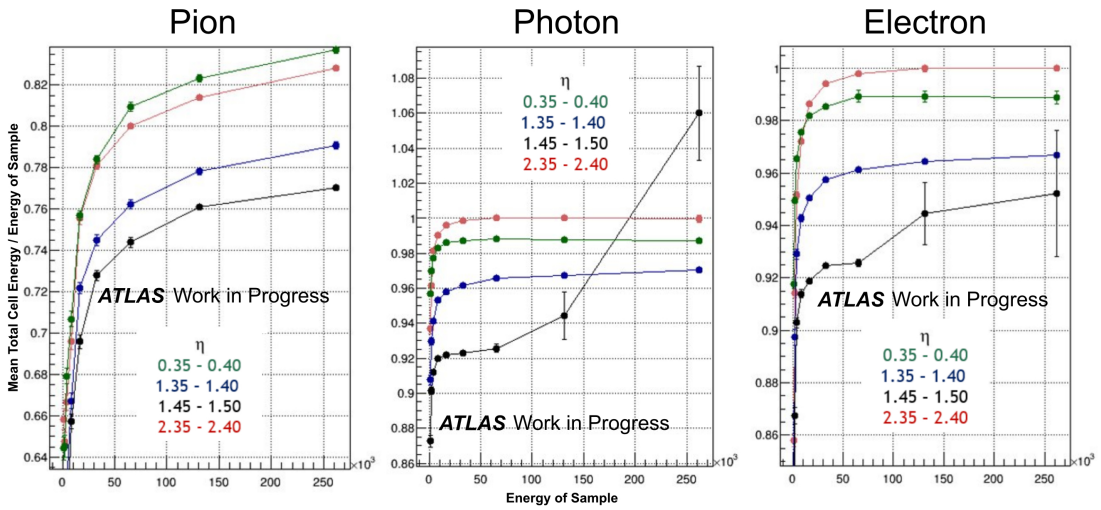


Figure 5.6: Investigation of the energy distributed in all cells for pion (left), photon (middle) and electron (right) single particle generated samples with Geant4 for various η ranges.

The plots created for all generated samples were used for the finalisation of the currently used energy interpolation package for FCS. A spline fit functionality was adopted, where the parametrisation is based on the logarithm of the total cell energy of the particle was selected. The closer the logarithm of the cell energy is to the logarithm of the energy of sample, the more likely it is that this parametrisation is picked. In Figure 5.7, the right plot shows constant behaviour of the energy response in a certain energy range and a jump to the next parametrization.

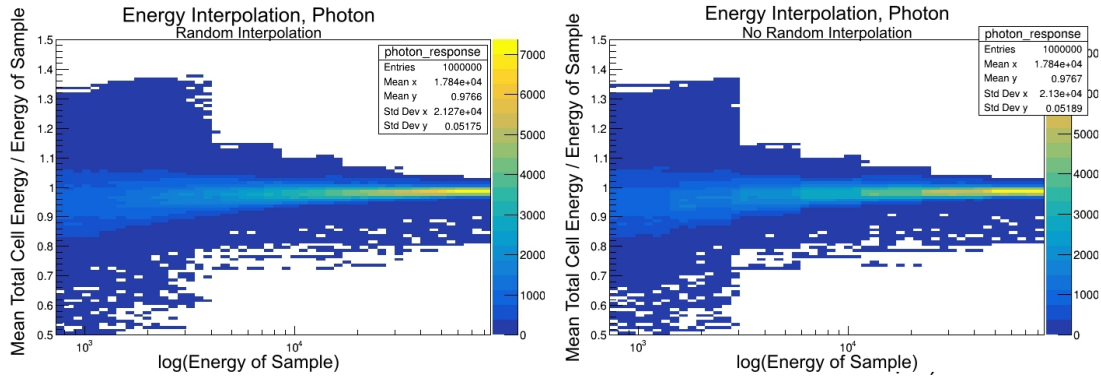
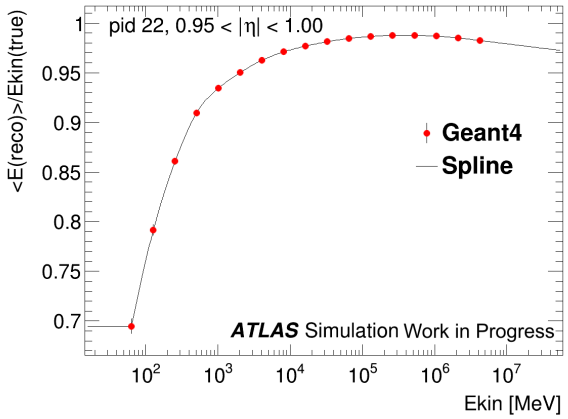


Figure 5.7: Comparison of the photon response with (left) and without (right) random choice for the interpolation between neighbouring parametrizations. Probability chosen based on log(energy of sample). Credit for plots: Michael Duehrssen (CERN).

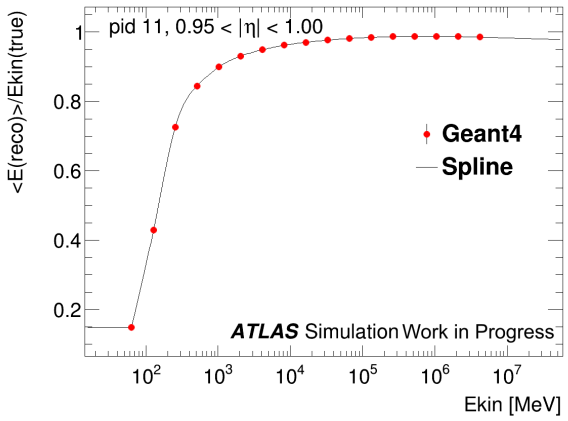
Three examples for the implemented spline fits patterns for electrons, photons

and pions are shown in Figure 5.8 for a) photon and b) electron and Figure 5.9 with an example pseudorapidity range of $0.95 < \eta < 1.00$. Spline fits for photons and electrons were relatively straightforward to create as the overall shape on a linear scale appeared to be linear for these two particles. The expected plateau effect of the spline fit at higher energies can be seen for electrons, but a drop of the kinetic energy was observed for photons. This is a result of particularly high energy photons, which escape the calorimeter and end up in other parts of the detector.

Pions required a more complicated spline fit shape which is shown in Figure 5.9. The pion spline fits have three notable features: a higher than expected energy of the first used sample, a resonance-like shape between 10^2 and 10^3 MeV and a drop at high energies. The higher than expected energy of the first sample was investigated further by the collaboration and only appears in some cases randomly, so it was concluded to not be a real physical effect. The resonance-like part of the spline fit is an artefact of the sampling structure of the calorimeter. Pions have a specific narrow range of energies, where most of their energy is deposited in the pre-sampler, which is calibrated using only high energies. This leads to high multiplication factors when calibrating the detector's response, which are correct for high pion energies, but not perfectly correct for particles which deposit most of their energies in the pre-sampler part of the detector. Finally, the drop at higher energies ($> 10^6$ MeV) in the last bin occurs due to longitudinal leakage into the muon system or pions, which did not finish their path in the hadronic calorimeter but instead continued and decayed in further parts of the detector.



(a) Photon



(b) Electron

Figure 5.8: Spline fit covering the kinetic energy interpolation using Geant4 samples of (a) photons and (b) electrons with energy 64 MeV $<$ E $<$ 4 TeV in $0.95 < \eta < 1.00$.

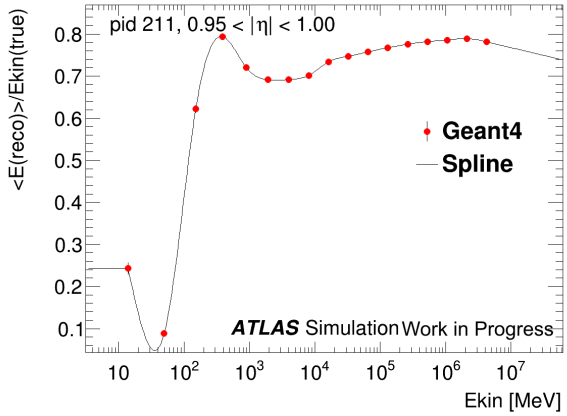
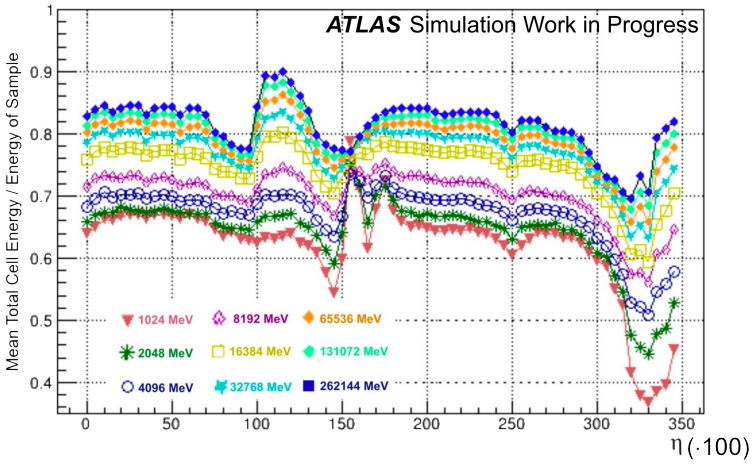
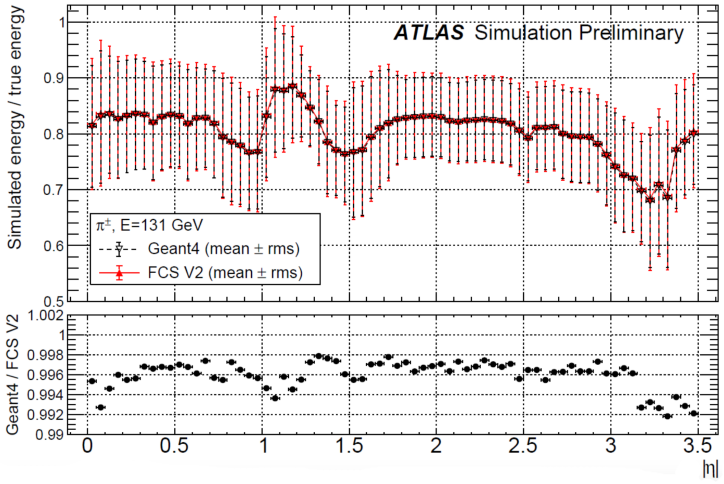


Figure 5.9: Spline fit for a pion covering the kinetic energy interpolation using samples with energy 64 MeV $<$ E $<$ 4 TeV in $0.95 < \eta < 1.00$.

In several different η ranges, a slightly fluctuating spline fit was seen, which is why the change of efficiency with respect to a change in the pseudorapidity range of the different incident samples was also investigated. A comparison was made with the pattern expected from Geant4 (Figure 5.10). Overall, the expected close similarity of FCS and Geant4 was seen. A noticeable bottleneck effect was seen at η of around 1.55 -1.60 in the FCS plot, which was thought to be an effect of the too fine granularity of the data points.



(a) FCS

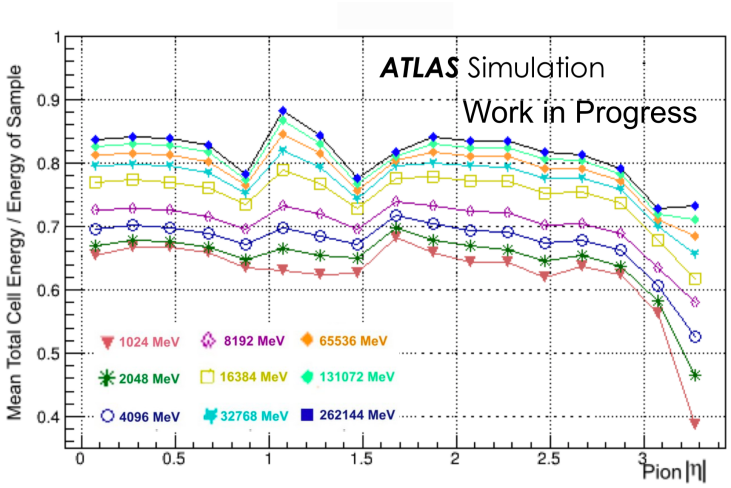


(b) Geant4. Plot thanks to Jana Schaarschmidt

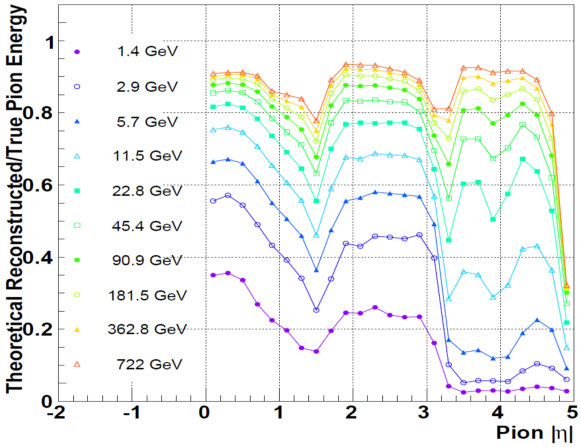
Figure 5.10: Comparison between FCS efficiency as a function of η for a range of particle energies shown in different colours (a) and full Geant4 for a single energy of 131 GeV (b).

The change of the efficiency with respect to the pseudorapidity was also compared to what was found for a previous release of the simulation (Figure 5.11). The bottleneck effect seemed reduced with the reducing of granularity of the

data points. Approximately 30% difference of efficiency for low energies was observed(Figure 5.10 a). Another difference was the dip around 0.8 in η, which was not to be seen on the plot for the previous release of the Athena simulation infrastructure. Those differences were attributed to differences in material description, densities and sampling fractions inside the calorimeter, as well as the noise threshold. The later was electronic noise dominated before 2011 (in the previous release) and cell-signal baseline fluctuations created pile-up dominated after 2011 (including 21.0.73 release used in this analysis).



(a) FCS recent release 21.0.73

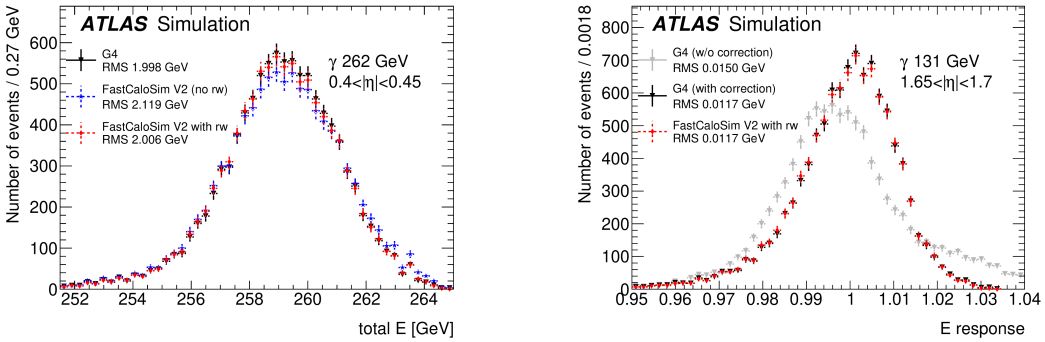


(b) FCS previous release: 13.0.20. Plot thanks to Sven Menke

Figure 5.11: Comparison between current (a) and previous relieve of FCS (b).

5.2.4 Corrections

Five different types of corrections are applied to the total energy response in the calorimeter in the full AtlFast3: an energy resolution correction, energy-ϕ modulation correction, hadron total energy correction, a residual energy response correction and a simplified geometry shower shape correction. The energy resolution correlation is a weight-based change of the FCS energy parametrisation, where the weight applied is the proportion of the FCS and G4 PCA bins. Every RMS is calculated using at least 99% of the total events and in 3σ range around the mean. Figure 5.12 a shows an example for the pseudorapidity range of $0.4 < |\eta| < 0.45$, of how this correction worked, where the an initial FCS photon energy distribution is shown in blue, FCS distribution after the correction is shown in red and the Geant4 energy response in black. The agreement is significantly improved.



(a) $0.4 < |\eta| < 0.45$ (b) $1.65 < |\eta| < 1.7$, ϕ correction

Figure 5.12: Example impact of the energy response corrections applied to the FCS. Left: an example of energy resolution correction for a photon sample. Geant4 energy response is shown in black, FCS before correction in blue and FCS after correction in red [1]. Right: an example of energy-ϕ modulation correction. The Geant4 energy response with and without ϕ modulation is shown in grey and black respectively, FCS response is shown in red.

The energy-ϕ modulation correction is the correction applied to the energy response, due to the lack of functional dependence on ϕ, where $|\phi| = |(\phi_{calo}, \phi/512)|$, in FCS. FCS, has a set of simplifications, in comparison to Geant4, the lack of ϕ dependence considered, being one of them. Figure 5.12 b shows the difference between the original G4 energy of a photon sample in $1.65 < |\eta| < 1.7$ in grey and the G4 energy after removing the ϕ dependence. The FCS response is given in red and shows a very good agreement with the Geant4 energy prediction without ϕ dependence.

The hadron total energy correction is applied to correct for the use of pion energy response in the derivation of the parametrisation of all hadrons. It is a scaling of the energies involved, as follows:

$$E_{Total}^{corr,h} = \frac{\langle E_{G4}^{h} \rangle}{\langle E_{G4}^{\pi} \rangle} \times \frac{E_{kin}^{\pi,true}}{E_{kin}^{h,true}} \times E_{Total} \tag{5.1}$$

$E_{Total}^{corr,h}$ is the total hadron corrected energy, $\langle E_{G4}^{\pi} \rangle$ is the mean G4 simulated pion response, $\langle E_{G4}^{h} \rangle$ the mean G4 simulated hadron response, $E_{kin}^{\pi,true}$ is the true kinetic energy of the pion and $E_{kin}^{h,true}$ of the hadron and E_{Total} is the total energy before the hadron correction. This correction is largest at small kinetic energies and decreases with the increase of energy.

The residual energy response correction is the final correction of the energy response applied after all reconstruction and simulation ATLAS procedures have been finalised. The final corrected energy is given with:

$$E_{Total}^{corr,res}(p) = \langle E_{G4}(p) \rangle / \langle E_{AF3}(p) \rangle \times E_{Total}(p) \tag{5.2}$$

$\langle E_{G4}(p) \rangle$ is the mean energy of the particle p (electron, photon or pion) in G4 and $\langle E_{AF3}(p) \rangle$ in the AF3 part of the fast simulation. This residual energy correction is shown in Figure 5.13 for the three particles, where it is slightly higher overall for pions (green), than it is for the other two particles.

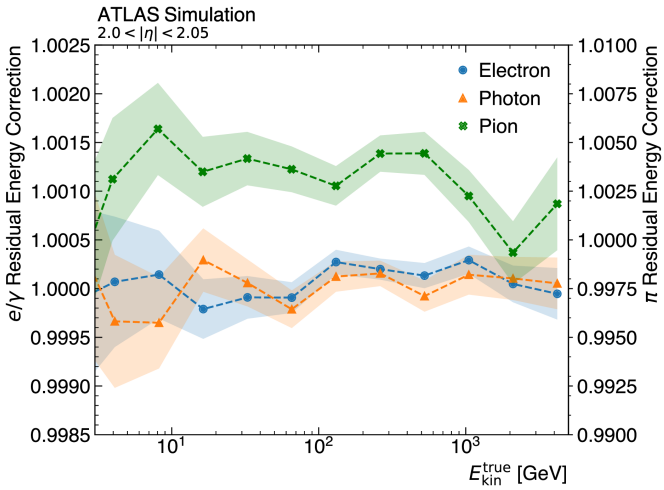


Figure 5.13: Residual energy correction for pions(green), photons(orange) and electrons(blue). The electron and photon scale is shown on the left and the pion's on the right y-axis [1].

The final correction, the simplified geometry shower shape correction is needed due to the simplified version of the calorimeter cells in FCS in comparison to G4. A significant number of hits are misplaced in ϕ hitting neighbouring to the

correct cells. This effect is shown in Figure 5.14 a), where there's a clear lack of consistency of the ratio of G4 and FCS number of hits between the cells. After correction, shown in Figure 5.14 b), FCS is in good agreement with G4.

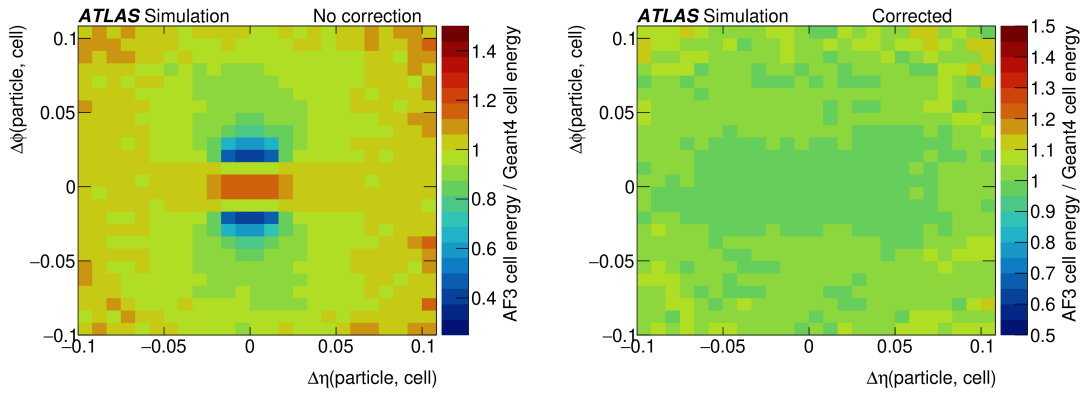


Figure 5.14: Ratio of calorimeter hits in G4 and FCS a) before and b) after the simplified geometry shower shape correction [1].

5.2.5 Reconstruction of Physical Objects

All comparisons of AtlFast3 with G4 and the previous iteration of AtlFast3 are made by comparing how the simulations perform at modelling and reconstructing different physical objects. The physical objects relevant for the work in this thesis are photons, electrons and pions, as they were used for the energy interpolation part of the simulation. The performance for those particles depends mainly on the calorimeter performance and on the performance of the tracking detectors.

Photons and electrons are reconstructed by considering energy depositions of the topological clusters in the calorimeter, where for electrons, tracks from the inner detector are matched to the clusters from the calorimeter. The electrons and photons used for the FCS studies in this thesis satisfy the "tight" isolation and identification requirements, which are set to be the strictest rules about the shower shape and particle's identification efficiency. For electrons, AtlFast3 results in a total difference of about 2% in the full variable phase space apart from the $30 < p_T < 300$ GeV, where it differs with about 5%. The situation for photons is similar, where G4 and AtlFast3 have negligible differences (Figure 5.15). a good agreements are shown at middle range and high range p_T and small differences in low p_T.

Pions often occur as daughter particles in τ hadronic decays. Taus are reconstructed by measuring the quantities of the daughter particles which range from one to three charged ot neutral pions [112, 113, 114].

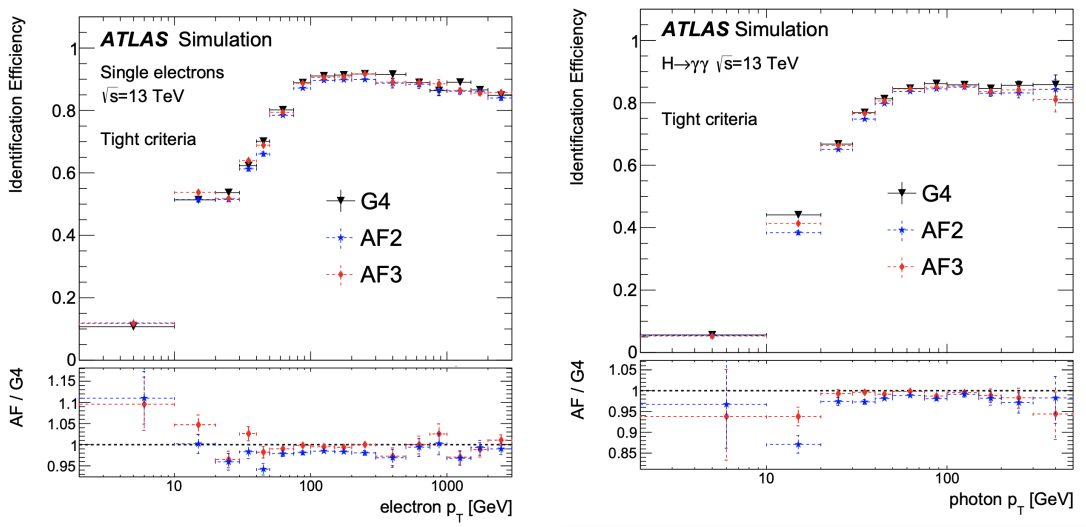


Figure 5.15: Comparison of current (AF3), previous (AF2) iteration of the FCS simulation and Geant4 (G4) reference samples for electrons and photons in p_T [1].

5.2.6 Performance of FCS

The average CPU time of the G4 ATLAS simulation for the simulation of a particle is about 200 times larger than the average CPU time for the current iteration of the simulation (AF3) for 8 GeV photons produced on the calorimeter surface and 600 times larger for 256 GeV photons. This demonstrates the efficiency of fast simulations in general and the high overall performance of FCS in this particular case (Figure 5.16). For the calorimeter simulation alone, G4 is 500 times slower than AF3. With the full simulation production chain, the required CPU time for completing one event is 5 times larger in G4, than it is in AF3. The conclusion made from the above is the need for a fast simulation for the tracking detector, which would improve the overall CPU performance even further. Personal performance studies completed in 2018 before the release of the last literation of AlFast3 [1] in 2021, and only on the FastCaloSim part of the fast simulation, can be found in Appendix D.

5.2.7 Physics List Range

The physics lists, as explained in Section 5.1.1 are used to describe the interactions of particles in Geant4. The main physics list used by ATLAS is the $FTFP$. The energy region 3-12 GeV for the incoming particle was investigated with the goal of improving the spline fit in that region, if necessary. Sixteen new samples with finer granularities were generated with energies between 1 and 16

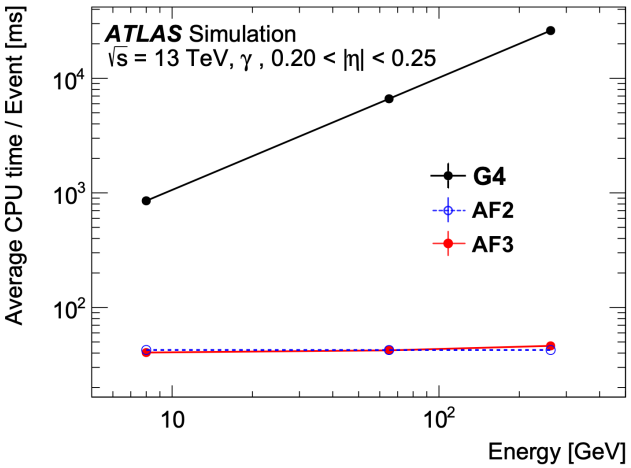


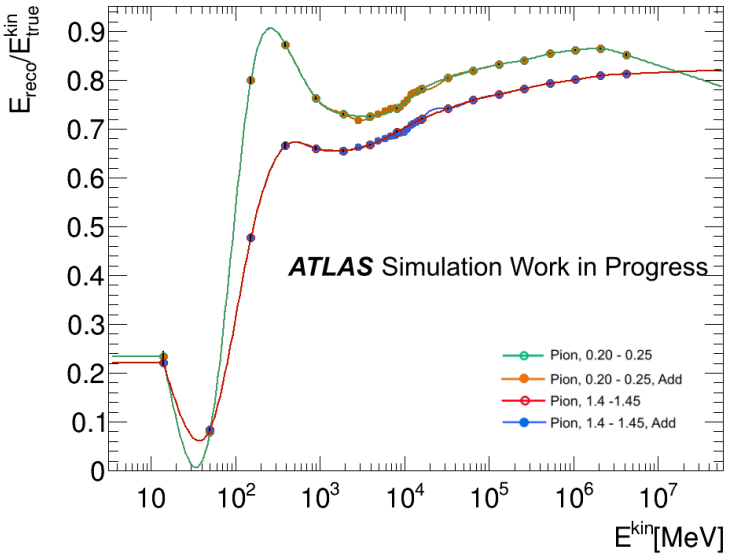
Figure 5.16: CPU performance of FCS in latest (AF3 red) and previous iteration (AF2 blue) in comparison to G4 black. Calorimeter simulation only. Example is for a photon sample in range $0.20 < |\eta| < 0.25$ [1].

GeV for two pseudorapidity regions $0.20 < \eta < 0.25$ (away from the crack and transition regions) and $1.40 < \eta < 1.45$ (transition region). The kinetic energy distributions were then added to the samples from the original production in order to create a new updated spline fit shape to be compared with the previous one. A final version of the plot is shown on Figure 5.17 alongside with a zoom on the desired part of the spectrum. The conclusion made was that the difference observed can be neglected as it is at most 10% and generally much less, therefore no further work for the improvement of that energy region was necessary.

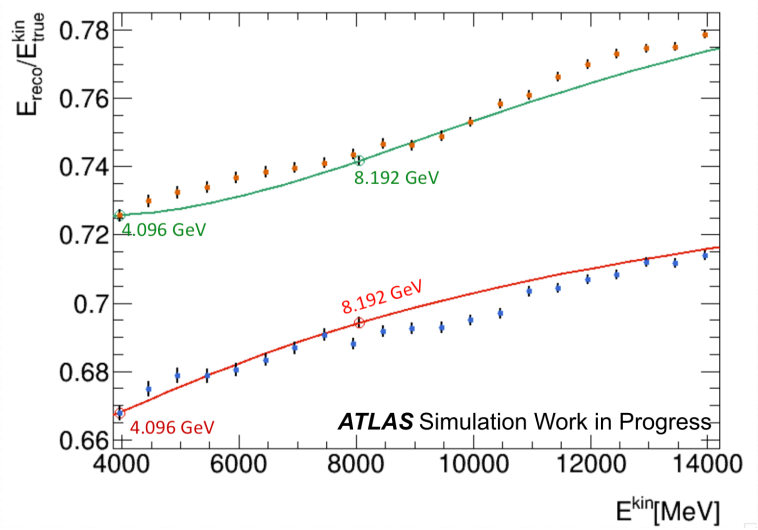
5.2.8 Conclusions of the AtlFast3 Studies

AtlFast3 is the new generation of the fast calorimeter simulation of the ATLAS calorimeter. It uses parametrised detector response, and the samples for this parametrisation used to only be simulated for discrete values of particle energies. A technique was developed and described in this thesis, which interpolates the detector response between these simulated energy points, and it was demonstrated that this interpolation is accurate. A piece-wise third order polynomial spline fit function was adopted, in order to interpolate to intermediate energies. Furthermore, linear extrapolation is used to reach energies beyond those of the simulated input samples. The spline fit interpolations are generated for each particle and each slice and are used to rescale the total energy response from the parameterization points. In addition to the interpolation of the total energy response, the other longitudinal and lateral shower shape properties also need to be interpolated. The shape interpolation is done by randomly selecting the parameterization from the nearest energy point with a probability linear in energy and fitted such that unit probability is reached for the grid energy points. For electrons and photons the spline fit for the energy response ranges down to 16 MeV, below which a linear extrapolation is used. For hadrons the energy response ranges down to a kinetic energy of 200 MeV, below which Geant4 is used for the simulation.

The interpolation enables the simulation of particles with any energy in AtlFast. It is used in AtlFast3, which significantly improves the agreement between the fast simulation with the full ATLAS simulation, most notably for jet substructure variables [1]. AtlFast3 is planned to be the default simulation for Run 3 physics production with the ATLAS detector. AtlFast3 significantly improves the modelling of reconstructed objects for physics analysis in comparison to the previous version of the fast calorimeter simulation. In the majority of studies, AtlFast3 agrees with the full Geant4 within a few percent. AtlFast3 requires only 20% of the CPU time required by Geant4 to simulate an event, AtlFast3 is currently being used to simulate 7 billion events with the Run 2 data.



(a) E_{reco}/E^{kin}_{true} Green and red: original spline fits; orange and blue: spline fit after added points



(b) E_{reco}/E^{kin}_{true} Close up in the region of interest.

Figure 5.17: Physics list investigation for checking the agreement after added points. The ratio of the reconstructed and true kinetic energy as a function of the true kinetic energy is shown. Comparison of π spline fit with and without additional points with energies 1 to 16 GeV.

Chapter 6

Machine Learning

"Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed."

Arthur Samuel (1959)

By its definition, the idea of machine learning could be seen as the programming equivalent of a very basic example for how humans evolve in a lifetime. We start by having some initial predispositions, which then change with respect to what we learn by observing numerous examples for how a certain task is done. We copy them, attempt them in our own lives and then we decide what to take from each experience (aka learning) before moving on to the next one, with a now better understanding for how to perform the task.

This thesis deals with a task (or problem) of how to separate events, in which we have interest, i.e. signal, from events which we do not need, i.e. background. The basic approach to solve the above is to look at certain regions of phase space for each variable in the data using what we call the cut-based approach. We "cut" on each variable's range to find an area of phase space with the largest signal/background ratio. Most real life problems tend to be too complex for the cut-based approach to be enough. If, for example, we have two variables x_1 and x_2 (Fig 6.1 a) and three classes of objects to be classified (blue, green and red), one way to perform the cut-based method would be to divide the 2D variable space into squares and then make a decision on which class each square belongs to by counting the number of objects with a certain colour in each square (Fig 6.1 b) [115]. The final goal of the classification in this case is being able to predict which class a certain data point belongs to by knowing its coordinates in the (x_1, x_2) space. An example point is marked with a black cross in Fig 6.1. Assuming equal distribution, using the square cut-based approach, would classify this data point as red with a good certainty (small chance of being green too)

because it falls within a red square.

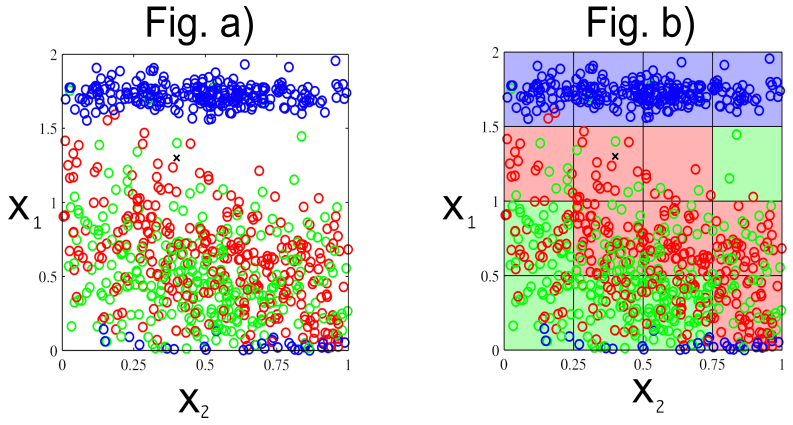


Figure 6.1: a) Three objects, represented with blue, green and red data points within a range (0.0) to (1,2) for two variables x_2 and x_1. b) Cut-based approach on a) [115].

For most real life problems, a simple cut-based approach would be limiting, due to the complex nature of the problems. The example described above deals with only two variables. If the number of variables is increased by even just one (Fig 6.2 c), the problem of classification becomes a lot more difficult and the difficulty grows exponentially with the increase of the number of variables. This exponential growth of the number of cells (or squares in 2D) is what is called the "curse of dimensionality" and is the reason for the development of more sophisticated ML techniques. More sophisticated techniques include using the most useful variables, as well as reducing the number of features.

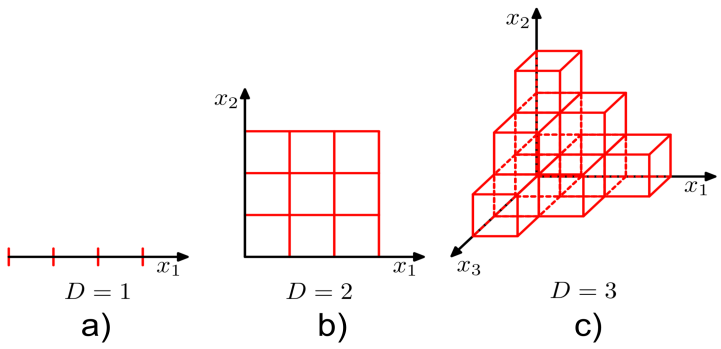


Figure 6.2: The curse of dimensionality illustrated for D = 1 (a), D = 2 (b) and D = 3 (c) variable dimensions. The number of points needed to sample the space spanned by the axes in D dimensions grown exponentially with D [115].

The tasks machine learning algorithms can perform include Projection Pursuit Regression (PPR), which predicts a real-value output, to give best estimate for a

certain output quantity, under some conditions, or logistic regression (otherwise called classification), which predicts a discrete-valued output, to separate certain components of the input data into categories.

6.1 Project Pursuit Regression

Linear regression (having a linear model fit for input data) is the most basic type of Projection Pursuit Regression (PPR). If, for example, we consider the most simplistic case of having two parameters: a and b and one input variable x, the hypothesis for the best fit for a training sample i is:

$$h(x) = a_i + b_i x \tag{6.1}$$

The idea of linear regression is to find a set of parameters (a, b), which minimises the shortest length between the data points y_i and their linear fit $h(x)$, in a number of training samples m:

$$\min_{a_i, b_i} J(a_i, b_i) = \frac{\partial}{\partial(a_i, b_i)} \frac{1}{2m} \sum_{i=1}^{m} (h(x)_i - y_i)^2 \tag{6.2}$$

$J(a_i, b_i)$ is known as the loss or cost function. The algorithm learning starts with a procedure called Batch Gradient Descent. The $\frac{1}{2}$ in Equation 6.2 is to cancel the factor of 2 from implicit differentiation. We set some initial values for the parameters a_i and b_i (the machine learning equivalent of predispositions in our human analogy), which then keep changing simultaneously. New values are assigned: $a_{i+1} = a_i - \alpha \frac{\partial}{\partial a_i} J(a_i, b_i)$ and $b_{i+1} = b_i - \alpha \frac{\partial}{\partial b_i} J(a_i, b_i)$ until a local minimum of the cost function is reached. Here α is called the learning rate. It controls the size of the step, taken from one estimated value of a parameter for a training sample i, to the next estimated value for a training sample $i + 1$. The learning rate needs to be carefully chosen so that it is not too small, which would make the gradient descent too slow, but also not too big, which will overshoot so as to miss the minimum.

A slightly more complicated scenario would be when we have multiple variables for the above described linear model. The work of the algorithm in this case is the same, only we need to make sure our variables are in comparable ranges and if not, use variable scaling as well as make sure we don't have redundant (highly correlated) variables and if we do, use variable regularization. Variable regularization is the process of determining which variables are most useful for the final goal, and which are least useful, and removing the later to save compu-

tational time and power, or to avoid over-fitting of the model [116]. If the number of weights is n, the loss function becomes:

$$J(\overrightarrow{\theta}) = J(\theta_0, \theta_1, ..., \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h(x)_i - y_i)^2. \tag{6.3}$$

6.2 Logistic Regression

Logistic regression refers to the part of a machine learning algorithm, which deals with problems requiring classification of the outputs in particular categories. The output is classified in a binary form of 0 or 1. In particle physics, classification is usually used to separate signal, denoted with $y = 1$, from background, denoted with $y = 0$. The algorithm would therefore output a number between 0 and 1, which will correspond to the probability for a certain event to be signal (closer to 1) or background (closer to 0). The hypothesis function corresponding to this case is:

$$h(\theta) = \frac{1}{1 + e^{-\overrightarrow{\theta}^T x}} \tag{6.4}$$

This is called the logistic or sigmoid function and is illustrated in Fig 6.3. $\overrightarrow{\theta}^T$ is the transpose of the parameter vector $\overrightarrow{\theta}$.

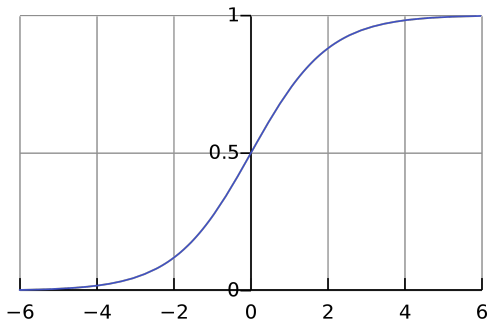


Figure 6.3: The sigmoid activation function $h(z) = \frac{1}{1+e^{-z}}$.

The loss function for logistic regression in two steps, is given by:

$$J(\overrightarrow{\theta}) = \begin{cases} -\log\left(\frac{1}{1+e^{-\overrightarrow{\theta}^T x}}\right) \text{ if } y = 1 \\ -\log\left(1 - \frac{1}{1+e^{-\overrightarrow{\theta}^T x}}\right) \text{ if } y = 0 \end{cases} \tag{6.5}$$

The logarithmic operation is used to ensure the local minima free nature of the function, which guarantees the convergence of the gradient descent to the

only minimum. To simplify, we can also have one equation for both cases and rewrite Equation 6.5 as:

$$J(\overrightarrow{\theta}) = \frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))] \tag{6.6}$$

However, as mentioned above, most real life problems require more complicated learning algorithms going beyond a simple linear and/or logistic regression, because we have multiple variables, as well as multiple parameters and the parameter space becomes a complex higher-dimensional area. As mentioned earlier, this problem is usually referred to as *the curse of dimensionality*. Therefore, a polynomial of higher order is usually more appropriate. Unfortunately, the higher the order, the higher the computational power needed, so the increased complexity of the problem, increases the need for a more architecturally complex but less computationally heavy algorithm.

6.2.1 Over-fitting and variable scaling

To make sure that the data is properly modelled, the possibilities for either over-fitting or under-fitting should be minimised. Over-fitting is the process of the model learning to fit the data so well, that it negatively biases the performance of the model on new data. In other words, the noise or random fluctuations in the data are picked up as a main part of the model and included in the fit. This results in any changes, however small they may be, to influence the model significantly and incorrectly. Examples of under-fitting, correct fitting and over-fitting can be seen in Figure 6.4, where the example given is of a classification problem (similar to the objective of the first neural network in these studies described later in this chapter). The linear function in Figure 6.4 left side is not enough to describe the division between signal and background in the 2D variable space, contrary to Figure 6.4 right side quartic function, which is too precise to the extent of damaging the flexibility of data used in the model. In the Figure 6.4 middle, we see an example of a good fit to this particular case, with a quadratic function.

In addition to dealing with over- and under-fitting, techniques such as feature scaling and splitting the data into smaller parts, to ensure the most unbiased solution to the problem is found, can be adopted. Variable scaling ensures the uniformity of the ranges of the different input variables. Features, which have a drastically different scale than all the others could lead to complications or complete instability of the final convergence when finding the global minimum. Feature scaling deals with that problem in the so called data pre-processed step

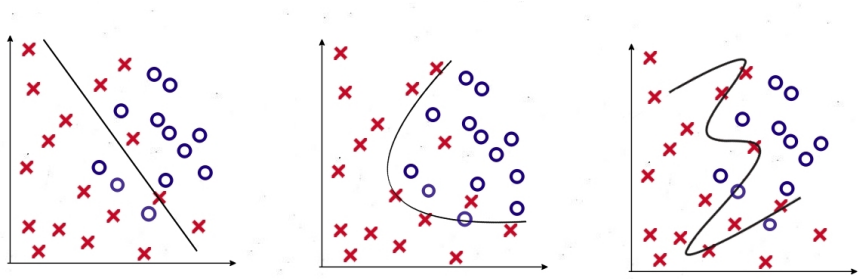


Figure 6.4: Examples of under-fitting (left), a good fit (middle) and over-fitting (right) of the background (blue circles) in a problem of classification of signal (red crosses) categories.

and each variable, x, and its mean, μ, and variance, σ, is scaled using:

$$\tilde{x_i} = \frac{(x_i - \mu_i)}{\sigma_i}. \tag{6.7}$$

6.3 k-fold Cross Validation

Once a model has been trained, it is critical to evaluate its generality and robustness on an independent dataset. The method of validation used in this thesis is called k-fold cross validation. It involves the separation of the total data set into k smaller parts and running k times in total. The idea behind it, is that all the results will be validated statistically, while we do not lose events, as each and every event is used in the training regardless. The procedure for this to be achieved is that after the data sets is divided in k parts, 1 of those parts is used for validation and the remaining $k-1$ for training. Then the procedure is repeated $k-1$ more times, where each time a different set of events aka fold is used for the training and all the rest for validation. When the process is finished, the mean and standard deviation for each calculated value of each loss function for the particular epoch aka pass though the training data are used as the final result and its error. If the training agrees with validation for each of the final resulting values, then it can be concluded that they are not over- or under-trained. Due to the computationally intensive procedures of the machine learning techniques used in this thesis, three folds were used, which proved to be sufficient as the values between the folds did not vary significantly. The usual number of folds in the current ML world is $3 < k < 10$ [117].

6.4 Neural Networks

Neural Networks were originally meant to be the programming equivalent of neurons from the human brain. They started being widely used in '80s and early '90s, after which their popularity diminished due to the insufficient computational power (time and capability) of the computers at that time. They re-emerged in recent years with an enormous success and are the state of the art technique for many fields, from economics and politics to science and technology.

The central nervous system is made up of two basic types of cells: neurons and glia. The neurons are the most diverse and important part of the brain. They carry information by transmitting electrical impulses (signals) and have three basic parts: a cell body, an axon and dendrites, see Figure 6.5. The dendrites receive information (input), the nucleus processes the received information and the axon sends the processed information to other neurons (output). In other words, the neuron can be called a computational unit - the connection between nodes in neural networks.

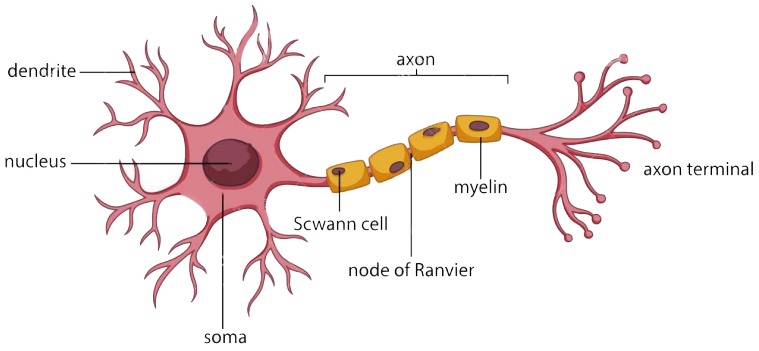


Figure 6.5: A basic representation of a neuron, as the simplified basic unit of the central nervous system. Picture: https://byjus.com.

A neural network in machine learning is a collection of units (neurons), which transmit and process information, see Figure 6.6. They consist of layers. The first layer is the input with units x_i, last layer the output and in the middle, we have what are called "hidden layers", with units $a_i^{(j)}$; for node i and layer j. When working with neural networks, the parameters are called weights and are, just as in the above ML examples, represented with the vector $\Theta^{(j)}$. In order to compute precisely the output, or the value, computed by the hypothesis function, the information received in each unit of each layer has to be calculated. For example, for unit $a_1^{(2)}$, this calculation would be:

$$a_1^{(2)} = g(\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 + \Theta_{13}^{(1)} x_3). \tag{6.8}$$

In the above equation, $g(x) = \frac{1}{1+e^{-\vec{\Theta}^T x}}$ is the sigmoid or logistic function and Θ^j is a matrix of weights controlling the function, which maps layer j to layer $j+1$. For example, $\Theta^(2)_1 4$ is the weight corresponding to the information transferred between node 4 of layer $j = 2$ and node 1 of layer $j + 1 = 3$.

A bias value x_0 allows you to shift the argument of the activation function (g) to the left or right, which may be critical for successful learning. A bias unit can therefore move the sigmoid (or any other activation function) curve to fit the prediction with the data better. Considering all the above and taking into account the bias unit (usually $x_0 = 1$) the hypothesis function, which calculates the output is now given by:

$$h_\Theta(x) = g(\Theta^{(2)}_{10} a^{(2)}_0 + \Theta^{(2)}_{11} a^{(2)}_1 + \Theta^{(2)}_{12} a^{(2)}_2 + \Theta^{(2)}_{13} a^{(2)}_3) \tag{6.9}$$

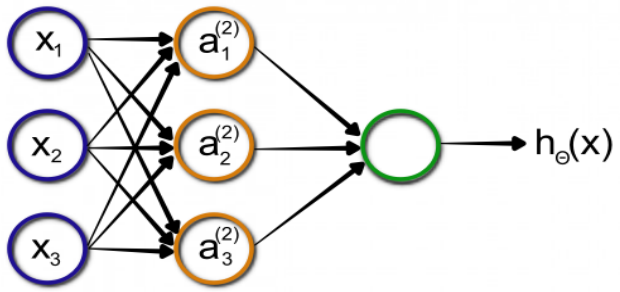


Figure 6.6: A basic representation of a neural network. x_1 - x_3 are the inputs layers (blue), a^2_1 - a^2_3 the hidden layers (orange) and $h_\Theta(x)$ is the output from the output layer (green).

The cost function used for neural networks is a generalisation of the one used for logistic regression:

$$J(\Theta) = -\frac{1}{m}[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} h_\Theta(x^{(i)}) + (1-y_k^{(i)}) \log(1-h_\Theta(x^{(i)}))] + \frac{r}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(\Theta_{ij}^{(l)})^2 \tag{6.10}$$

In the above Equation 6.10, m is the number of training samples, K is the total number of units, L the total number of layers, s_l the number of units (not counting bias unit) in layer l, and r is the regularisation parameter responsible for the penalisation of terms, which may contribute to over-fitting, but don't help the algorithm to learn.

Several expressions can be used for the activation function (g in Equation 6.9). Activation functions can be divided into two types: linear and non-linear. The former is only used for simple problems and the latter for problems, which have more variables and require more complex learning. The most common types of activation functions are: sigmoid (mentioned in Chapter 6), tanh, relu [118], softplus [119] and Gaussian, shown in Figure 6.7. For the purpose of the studies of this thesis the relu activation function was chosen.

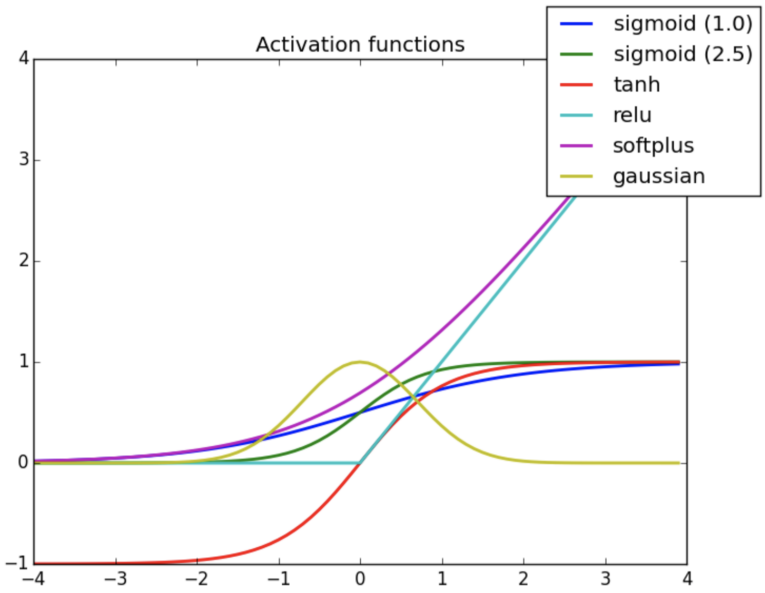


Figure 6.7: Most common activation functions used in ML. The value of the activation function ($g(x)$) as a function of the argument x is shown. Source: https://www.analyticsvidhya.com/.

Finally, the exact architecture of the NN depends on the particular problem to be solved. Initial decisions (before optimisation) need to be taken about the number of hidden layers, number of input units, number of units per hidden layer and weights. The code needs to include an implementation of the forward propagation (transmission of information towards the output layer) to get $h_\Theta(x^{(i)})$, computation of the loss function $J(\Theta)$ defined in Equation 6.10 and computation of the minimum by finding the derivative $\frac{\partial}{\partial \Theta_{jk}^{(l)}} J(\Theta)$. The neural networks use densely connected layers, which map every input variable to every output variable and also batch normalisation layers which standardise the variables from the preceding layer (taking the values from the previous densely connected layer and scaling them such that they have a mean of zero and a standard deviation of one). This is part of regularisation and is needed to cope with a potential situation of a single node giving a very different value than the average, therefore

shifting it significantly. Each node in a dense layer outputs a set of values that are dependent on the weights in that layer and the choice of activation function. The next layer is tasked with learning new features based on the activation values a of the previous one.

6.4.1 Neural Network (NN) Architecture

The configuration of hidden layers and nodes (as shown in 6.4) is referred to as a neural network architecture. The exact neural network architecture depends on the physical problem considered. Generally, as the number of nodes and layers increases, so does the power of the network. The reason for the increasing power is the increase of the number of variables (weights), which leads to a higher dimensional variable space computed by the NN, which can be used for more complex physical problem. Two neural networks are used for the analysis in this thesis. Both operate with the open-source software platform Keras [120], which provides a python based interface for the TensorFlow platform [121]. The former was constructed specifically for neural networks, while the later can be used on its own too for numerous ML related tasks [122] .

6.5 Adversarial Neural Networks (ANN)

6.5.1 The ANN Methodology

Adversarial neural networks [123] (classifier and adversary) are where two neural networks work with different objectives towards the same common goal. Important input variables for this analysis are: transverse momentum p_T, pseudorapidity η and azimuth angle ϕ of the two photons (labelled X in Fig 6.8). The classifier (first neural network) is developed as described in 6.4 and exploits non-linear combinations of the inputs in order to find the optimal cuts to reject background events. The classifier is tasked with predicting mass signal labels (0 or 1 based on the probability for a certain event to be signal or background). The adversary is tasked with inferring the values of the variables, from which the output of the classifier should be de-correlated. The output of the classifier is then treated as input to the adversary which is trained to decorrelate the variables from $M_{\gamma\gamma}$. These correlations lead to a sculpted background in $M_{\gamma\gamma}$, shown in Figure 6.9. After rejecting background with the classifier, the distribution of background events left, peaks approximately where the signal peaks. There is,

therefore, no way of knowing, exactly how many of the remaining events are signal events and how many background, as illustrated in Figure 6.9. This problem of bias is resolved with the adversary, which is trained to de-correlate the input variables with the invariant mass distribution $M_{\gamma\gamma}$ distribution and therefore to remove the sculpting.

The adversary is designed to create a probability density function (PDF) p_{adv} for $M_{\gamma\gamma}$ conditional on $J_{cls}(\overrightarrow{\theta})$ by using a Gaussian Mixture Model (GMM). A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions, which have parameters that keep being updated at each optimisation iteration of the gradient descent procedure for finding the local minimum for the loss function. Auxiliary inputs (aux) which are specifically useful for the decorrelation can be provided as additional input to the adversary, in a similar manner as the original inputs are provided to the classifier.

The final loss function, which includes the auxiliary variable aux (a variable introduced after classifier training not as an input to the classifier but rather an additional variable to the adversary) is then given by:

$$J_{adv}(\overrightarrow{\theta}_{adv}) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log p_{adv}(M_{\gamma\gamma}|\theta_{adv}, J_{cls}(\overrightarrow{\theta}_{cls}), aux)]. \tag{6.11}$$

The ANNs use both batch normalization and dense layers.

The adversarial training uses both neural networks, which have opposing goals, and therefore compete with each other to achieve the best possible balance between maximum background rejection and minimum background sculpting. The training of the two networks simultaneously is done using a gradient reversal technique. A gradient reversal layer brings both networks closer after performing some transformation and learns discriminative and invariant features, which gives the most optimal balance between J_{cls} and J_{adv}: J_{ANN}. For the analysis in this thesis, the logarithm of the variable with the highest correlation with $M_{\gamma\gamma}$, the transverse momentum of the leading photon, is used as an auxiliary variable. The logarithmic scale is used, so that the auxiliary variable's range can correspond to the range of the classifier output.

A binary function which depends on the data, the weights on the neural networks θ_{cls} and θ_{adv} and a parameter, which controls the loss function's parameter λ is used to monitor the balance between minimising the loss function J_{cls} (maximising background rejection) while maximising J_{adv} (minimising back-

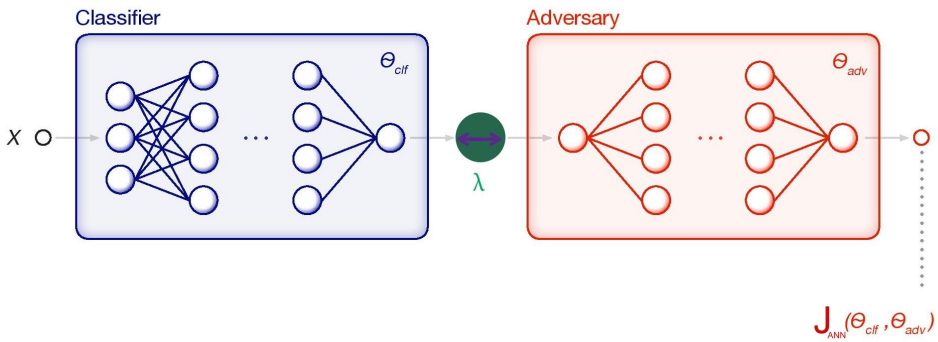


Figure 6.8: Adversarial neural networks schematics. Shown from left to right - X: data, θ_{cls} and θ_{adv}: weights on classifier and adversary, λ: parameter, which controls the loss function $J_{adv}(\theta_{cls},\theta_{adv})$. The classifier is tasked with predicting mass signal labels (0 or 1 based on the probability for a certain event to be signal or background). The adversary is tasked with inferring the values of the variables, from which the output of the classifier should be de-correlated by parametrising a PDF as a Gaussian mixture model. The training of the two networks simultaneously is done using a gradient reversal technique, which gives the most optimal balance between J_{cls} and J_{adv}: J_{ANN}.

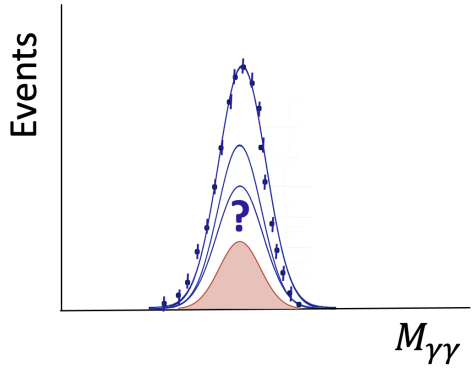


Figure 6.9: A graphical example of the problem of sculpting in the invariant mass distribution $M_{\gamma\gamma}$, where the red curve represents signal and the blue background. After rejecting background with the classifier, the distribution of background events left, peaks approximately where signal peaks.

ground sculpting). The loss function J_{cls} is aimed to be minimised and the loss function J_{adv} is aimed to be maximised.

$$\min_{\theta_{cls}}\max_{\theta_{adv}} J_{ANN} = J_{cls}(\theta_{cls}) - \lambda J_{adv}(\theta_{adv}\theta_{cls}) \tag{6.12}$$

In Equation 6.12, minimising the classifier loss function $J_{cls}(\theta_{cls})$ is a static problem, which means that there is a local minimum, which the network looks for by using the gradient descent procedure described in Section 6.1, where after ev-

ery optimisation step, the values of the weights of the classifier and the adversary networks are optimised simultaneously until a local minimum is reached. The final loss function J_{ANN} is calculated by including the loss function of the second neural network $\lambda J_{adv}(\theta_{adv}\theta_{cls})$. The losses are being updated simultaneously after each optimisation iteration. The classifier is trained stand-alone, which is why its loss only depends on its own weights. However, the adversary can only be trained in combination with the classifier, so the adversary's loss depends on both the weights of the classifier and the adversary. The minus sign in front of the adversary term in Equation 6.12 accounts for the classifier and the adversary's competing objectives or in other words, the classifier being trained to maximise the loss function of the adversary, which is trained to be minimised by the adversary itself. The parameter, which controls the trade-off between classifier and adversary λ can be set to 0 for a stand-alone classifier training (or just background rejection here) and to any value greater than 0, depending on the trade-off of level of importance in background rejection versus background sculpting minimisation.

6.5.2 ANN Architecture

The full ANN architecture with all its nodes, layers and inner connections can be found illustrated in Appendix A. The novelty challenge to the ML world, which ANNs deal with, is the joint optimisation of the networks, which ultimately work against one another. In this analysis, θ_{cls} and θ_{adv} undergo simultaneous optimisation by using gradient reversal [124], which deals with optimising with more than one objective at a time, by optimising the connection between the two neural networks and then updates the weights simultaneously and accordingly. The update of the weights happens at each node-node connection, within each layer of the neural networks. The gradient scaling operation is applied directly to the connection between the networks, instead of to each of them separately and consecutively.

During each epoch, the gradient, which propagates from classifier to adversary and then back to classifier is scaled by $-\lambda$. Due to the more complex nature of the adversary due to the fact that it depends on both adversarial and classifier weights and that it is meant to resolve a problem caused by the classifier, additional help by the user needs to be set to ensure the correct convergence to minima. Therefore, to ensure that the joint optimisation converges properly and the true minimum is found, the learning rate of the classifier (10^{-2}) is given to be smaller than that of the adversary 10^{-1}.

6.5.3 ANN Hyperparameters

Hyperparameters, which need to be optimised for the final set-up of the ANN architecture are the number of units, the activation function, the usage of batch normalisation, the number of layers, the number of epochs, the shuffling of events after every epoch, the learning rate of each neural network, the type of optimiser, the number of epochs used for pre-training, the parameter which controls the power of the adversary λ, the learning ratio between the two networks, the batch size and the loss function. Pre-training is the process before fine-tuning the network of starting with random weights and then training the network until the weights are optimised. Most of these parameters have already been discussed in Chapter 6. In addition, the learning ratio is the ratio between the learning rates of both networks and is important, because it is part of the balance between them, which decides on which network's objective to put more weight [125].

6.6 Boosted Decision Trees

Another machine learning technique very commonly used in particle physics is that of boosted decision trees (BDT, Figure 6.10). These were used in the analysis described in Section 7. It is very similar to a neural network, in the sense that it uses an N- dimensional hyperparameter space of input variables and it uses the information provided by their relationships to either classify or regress. The name of this technique comes from the idea of using segmented predictor space (as seen in Figure 6.10 top left for the segmented space and bottom right for the prediction space) of the input variables to make certain decisions by calculation of the mean/mode between the different segments (which could also be called branches). The decision trees are trained in sequence, enumerated by a boosting step n and similarly to neural networks, there is a weight a^n calculated. The final learning algorithms using input data X, is then given by:

$$BDT(X) = \sum_n a^n DT^n(X). \tag{6.13}$$

If, for example, there was a regression problem to be resolved, the procedure of using decision trees (shown in Figure 6.10) includes splitting the variable space into a number of main regions (which are usually called leaves of the tree or more technically terminal nodes) and the connections between them are known as branches. The importance of and relationships between the variables are then taken into consideration and a final prediction for the desired result is made. For

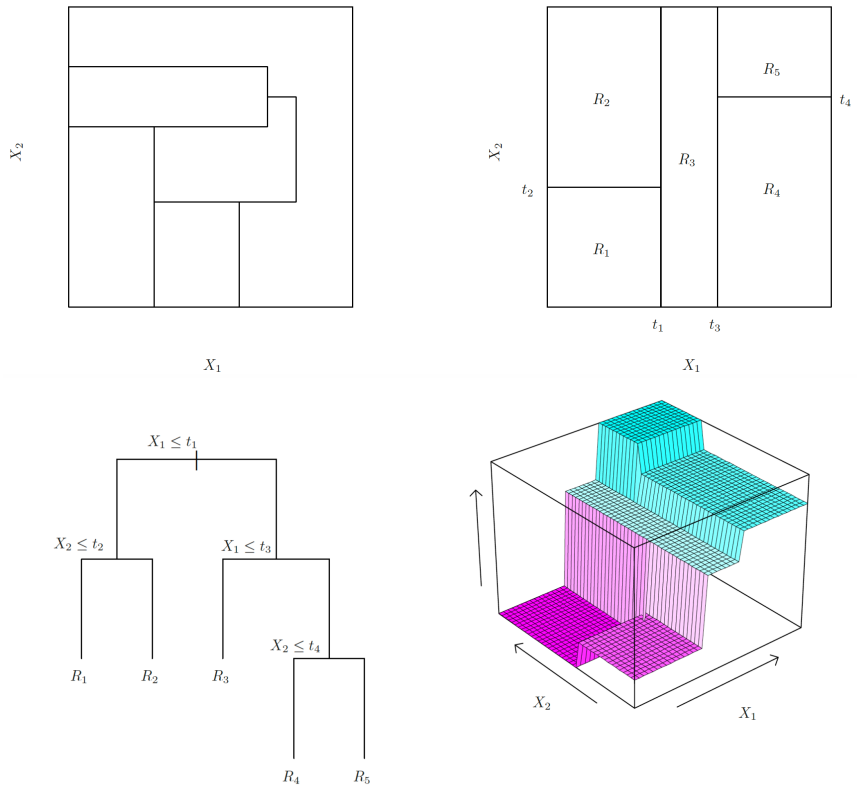


Figure 6.10: The steps in BDTs. Top left plot shows an example for a 2D segmented space with variables X_1 and X_2. Top right plot illustrates the splitting of regions. Bottom left shows the tree that is used to create the segmented space and bottom right illustrates the prediction space [117].

every observation that falls into the same region, the same prediction would be made which would be the mean of the response values for all the input observations in that specific region. Next, the process is repeated, looking for the best predictor and best cut-point in order to split the data further so as to minimize the residual sum of squares (Equation 6.14) within each of the resulting regions.

There are several ways for the main regions to be chosen and several shapes, which they can take. The most common shape is a 2D box and the most usual way similarly to how we do this with NNs, by minimising the distance between the observable y_i and the mean response of the corresponding box $\hat{y}_{R_j}$:

$$\sum_{j=1}^{J} \sum_{i} (y_i - \hat{y}_{R_j})^2, \tag{6.14}$$

where J is the number of samples in the box. A "greedy " approach is then applied by choosing the best possible split for each step of the process of starting from the top of the tree and going down until finished. Similar to any other

machine learning techniques, when working with decision trees, it is crucial to avoid over-fitting. A too complex tree with too many regions would be time consuming and may lead to over-fitting. A too simple tree may give the optimal minima and therefore may give precise enough predictions. The way around that problem when using neural network is to control the learning rate. A large tree with the sole aim to reach the optimal minima, on the other hand, is first built and then it is sized down.

Boosting is an additional technique, which deals with the production of multiple decision trees used towards the same goal. It is particularly useful in high energy physics (HEP) scenarios, where the decision function is not usually expected to be discontinuous like in the case described above. The input variables, for either the classification or regression problem is usually not expected to exhibit discontinuous jumps. Boosting is used to deal with the discontinuity of a decision tree. The idea is to combine a set of learners and use them together to construct a stronger final learner.

6.7 Jenson-Shannon Divergence

The Kullback-Leibler divergence (KLD) is a measure of relative entropy [126]. If it is equal to 0, this means that the two distributions in comparison are completely identical and if it is equal to 1, that they have nothing in common with one another. To measure the relative entropy of A with respect to B, we need to consider both the entropy of A: $H(A)$, and the cross-entropy between A and B,

$$KL(A||B) = H(A,B) - H(A) = -\sum_i A_i \log_n B_i + \sum_i A_i \log_n A_i \qquad (6.15)$$

The KL divergence for two distributions A and B with i bins each is given by:

$$KL(A||B) = \sum_i A_i \log_n \left(\frac{A_i}{B_i}\right). \qquad (6.16)$$

All entropy calculations in this thesis will use a logarithm with base 2, with the purpose of an easy to use range for the divergence of $[0, 1]$. In case of identical distributions, $H(A,B) = H(A)$, $KL = 0$. The bigger the difference between the two distributions, the bigger the cross-entropy, therefore the larger the KLD coefficient. As described above, KLD is prone to mathematical instabilities. An example for that is if one or more bins of a given distribution is bigger than 0,

while the same bin for the second distribution is exactly 0. In this case, Equation 6.15 will go to infinity due to the cross entropy factor. Another potential problem is the asymmetry of KLD with respect to its arguments. In this analysis, when classifying the events as signal or background, ideally, we should not have to deal with a scenario where the same event leads to different classification value, which would be the case if the relative cross-entropy is not symmetrical with respect to its arguments. The above leads to the need for a divergence similar to KLD but one, which can avoid all instabilities.

The Jenson-Shannon divergence (JSD) [127] is frequently used in statistics and machine learning to quantify the difference between the shapes of two distributions. In this thesis the JSD is used to determine how much the classification algorithms sculpt the $M_{\gamma\gamma}$ background. Visually, there can only be an insufficient knowledge gained for exactly how much sculpting there is in different cases and for different optimisation set-ups of the NNs. The JSD is a generalisation of the KLD divergence, which solves the mentioned problems. The JSD between two distributions A and B and $M = \frac{A+B}{2}$ is given by:

$$JSD(A||B) = H(M) - \frac{1}{2}\left(H(A) + H(B)\right) = \frac{1}{2}\left(KL(A||M) + KL(B||M)\right). \tag{6.17}$$

Similar to KLD, in the JSD, when the two distributions are identical, $H(A) = H(B) = H(M)$ and $JSD = 0$. When the two distributions are completely different and have no overlapping bins, $H(M) = \frac{H(A)+H(B)}{2} + \log_2(2)$ and $JSD = \frac{H(A)+H(B)}{2} + \log_2(2) - \frac{H(A)+H(B)}{2} = 1$ and $JSD = 1$.

Chapter 7

Measurements in $H \to \gamma\gamma$ decay channel

The event selection and signal extraction techniques of $H \to \gamma\gamma$ studied in this thesis follow the baseline ATLAS analysis, described in [43]. The signature of the Higgs boson in the di-photon decay channel is a narrow peak on top of the smoothly falling $M_{\gamma\gamma}$ distribution. The width of the peak is consistent with the resolution of the detector and is typically between 1 GeV and 2 GeV, depending on the kinematics of the event. The mass and event rate of the Higgs boson can be inferred from the first of the $M_{\gamma\gamma}$ distribution. Major Higgs boson production processes, including ttH were generated using *Powheg Box v2* [128]. All generated events for the processes are interfaced to *Pythia 8.2* [129] to model parton showering, hadronization and the underlying event. All events are generated with a Higgs boson mass of 125 GeV and an intrinsic width of 4.07 MeV. Prompt di-photon production is simulated with the *Sherpa 2.2.4* [130] generator. The production of $tt\gamma\gamma$ events is modelled using *MadGraph5_aMC@NLO 2.3.3* [131].

7.1 Event reconstruction and selection

7.1.1 Photons

Photon reconstruction includes separation of the photons from electrons and jets. The photons are divided into groups of the so called "converted" and "unconverted" candidates, where converted means that they decay to an electron-positron pair. Photon candidates are required to deposit the majority of their energy in the electromagnetic calorimeter, and to have a lateral shower shape

consistent with that expected from a single electromagnetic shower [19]. Photon candidates are separated from jet backgrounds using an identification criteria based on calorimeter shower shape variables. The full efficiency range after all cuts, which lead to *tight* photons with energy $E_\gamma > 25$ GeV, is 84% - 94% (85% to 98%) for unconverted (converted) photons. The final criteria for photons is that the energy in a radius $\Delta R = 0.2$ around the photon, excluding the energy containing the photon shower, has to be less than 6.5 % of the photon transverse momentum of each photon. Photons are considered *isolated* when each photon has a track isolation of less than 5% of the transverse energy. A track isolation is the scalar sum of the transverse momenta of all tracks with radius $\Delta R < 0.2$ around the photon candidate. In case of a candidate converted photon, the tracks associated to the conversion are excluded from the sum.

Additional requirements are set on the di-photon system [43]. The system is created by choosing the two highest p_T photons. The primary vertex is then reconstructed with a neural network. Each of the two photon candidates is then required to satisfy the *tight* criteria. The efficiency of the tight identification for unconverted (converted) photons ranges from about 84% (85%) at $p_T = 25$ GeV to 94% (98%) for $p_T > 100$ GeV. A final cut is applied on the leading and subleading candidates of $p_T/M_{\gamma\gamma} > 0.35$ and 0.25 respectively. Events, which fail the isolation or the identification criteria are used as an approximation to data or for modelling purposes.

7.1.2 Leptons

Electrons are reconstructed by matching EM clusters from the ECAL with the corresponding tracks seen in the ID part of the detector and are required to have $p_T > 10$ GeV and $|\eta| < 2.47$ excluding the transition region and to satisfy *medium* selection, based on shower shapes and track parameters [132]. Additionally, the longitudinal impact parameter z_0 (the distance of closest approach of the track to the collision point along the z axis) is set to $|z_0 \sin\theta| < 0.5$ mm and the transverse impact parameter divided by the uncertainty is set to be $|d_0|/\sigma_{d_0}$ <5.

Muons are reconstructed from a combination of tracks built in the inner detector and the muon spectrometer and must have $p_T > 10$ GeV and $|\eta| < 2.7$ and satisfy the *medium* identification requirement, described in [133]. The identification efficiency is 95-97% for muons with $p_T = [10 - 60]$ GeV and 99% for muons with $p_T > 60$ GeV. The tracks have to satisfy $|z_0 \sin\theta| < 0.5$ mm and $|d_0|/\sigma_{d_0} <$ 3.

7.1.3 Top quark

Top quark candidates are reconstructed and identified using a BDT discriminant. The BDT targets both leptonic and hadronic top quark signatures. The BDT is trained using the XGBoost package [134]. It is trained with the ttH sample with the idea to infer the three-jet combination appearing most like the hadronic decay products of a top quark. In the hadronic case for the decay of the W bosons, the triplet with the highest BDT score is taken as the primary top quark candidate and in the leptonic case for events containing only one lepton, a W boson candidate is first constructed from the lepton and the missing transverse momentum, followed by the reconstruction of the top by considering the highest BDT score jet, as well. After the highest score top quark is selected, if there are at least three additional jets, a second top quark candidate is reconstructed following the same procedure [135].

7.1.4 Jets

Jets are reconstructed using a particle flow [136] algorithm of topological clusters [137] of energy deposits in the calorimeter implemented in the anti-kT algorithm [138, 139] with a radius parameter R=0.4. The key feature of the anti-kT is that soft particles do not modify the shape of the jet, while hard particles do. I.e. the jet boundary in this algorithm is resilient with respect to soft radiation, but flexible with respect to hard radiation. Initial cuts are set on transverse momentum and η as follows: $p_T > 25$ GeV and $|\eta| < 4.4$. The constructed jet four-momenta are corrected for the signal losses due to noise threshold effects, energy losses in gap regions and pile-up. An additional jet-vertex-tagger discriminant is applied to jets with $p_T < 60$ GeV and pseudorapidity of $|\eta| < 2.4$ to suppress pile-up. b-jets are tagged with a separate algorithm [140] with four different efficiency working points: 60%, 70%, 77% and 85%. The tagging of b-jets is the process of identifying them against a large jet background containing c-hadrons or light-flavour jets. The jet-vertex-tagger [141] is constructed using a two-dimensional likelihood derived using simulated di-jet events and based on a k-nearest neighbour (kNN) algorithm [142]. A relative probability is calculated for each event to be of type signal by calculating the ratio between the number of hard-scatter jets and the number of hard-scatter plus pile-up jets found in a local neighbourhood with k neighbours.

7.1.5 Missing energy E_T^{miss}

The missing transverse energy is defined as the negative vector sum of the transverse energies of the selected photon, electron, muon and jets and also of other particles associated with the diphoton vertex, estimated using tracks matched to the diphoton primary vertex but not assigned to any of the selected objects [43] [143].

7.2 Event Categorisation

The events passing the event selection are classified into mutually exclusive event categories, targetting ttH production for different ranges of the transverse momentum of the Higgs boson pt^H. In each of these categories, the sensitivity is defined as [43]:

$$Z = \sqrt{2((S+B)\ln(1+(S/B)-S)}, \tag{7.1}$$

where S is the signal yield, B the background yield from the continuum diphoton distribution, which includes all non-ttH Higgs processes and $f = S/(S+B)$ is the purity. The total sensitivity, after splitting into different categories is the square of the quadratic sum of all.

The sensitivity in the full analysis in [43] are calculated using Equation 7.1 for eight selected and 1 unselected categories and shown on Table 7.1. The categories correspond to ranges of binary BDT (boosted decision tree with two possible outputs) values, which are chosen to maximise Z. If an event fails to enter a selected final category, it is placed in the un-selected category. A class is split into two categories if this leads to an improvement of more than 5% in the expected sensitivity, and into three categories if a further improvement of at least 5% relative to the two-category configuration can be achieved. The categories are referred to as high-purity, mid-purity and, in the case of a 3-category split, low-purity.

7.3 Published ttH measurement and uncertainty

When probing the Higgs boson production mechanism, the different production channels considered are ggF, VBF, WH, ZH and the combined tH and ttH channels. The measurement reported [43] is in terms of $(\sigma \times B_{\gamma\gamma})$, where σ is the fiducial cross-section and $B_{\gamma\gamma}$ is the branching ratio of the di-photon

Categories	S	B	f	Z
$p_T^H < 60$ GeV, High purity	3.2	5.0	0.39	1.3
$p_T^H < 60$ GeV, Mid-purity	3.5	15	0.18	0.8
$60 \geq p_T^H < 120$ GeV, High purity	5.1	4.3	0.54	2.1
$60 \geq p_T^H < 120$ GeV, Mid-purity	3.7	10	0.26	1.1
$120 \geq p_T^H < 200$ GeV, High-purity	6.1	3.8	0.62	2.6
$120 \geq p_T^H < 200$ GeV, Mid-purity	3.1	8.1	0.28	1.0
$200 \geq p_T^H < 300$	4.6	1.7	0.73	2.7
$p_T^H \geq 300$ GeV	3.6	1.0	0.78	2.6
Unselected (incl tH)	11	120	0.08	1.0

Table 7.1: The expected signal (S) and background (B) yields, purity (f) and sensitivity (Z) for analysis categories targetting the ttH production [43].

decay channel. The observed (expected) sensitivity values for the ttH+tH process is 4.7 (5.0)σ. The total uncertainties are decomposed into components for data statistics (Stat), and systematic uncertainties (Syst). The best-fit value and uncertainty for ($\sigma_{ttH+tH} \times B_{\gamma\gamma}$) are $1.2^{+0.4tot}_{-0.3tot}(\pm 0.3 \pm 0.1)$ fb, where ± 0.3 fb is the Stat uncertainty and ± 0.1 fb the Syst uncertainty. The SM prediction is 1.3 ± 0.1 fb. For the ttH process, the leading experimental uncertainty is related to the measurement of jets, and it can be as large as 6%. The dominant systematic uncertainties in the $ttH + tH$ are from the photon energy resolution: $\pm 4.9\%$, photon efficiency: $\pm 2.4\%$ and luminosity and trigger: $\pm 2.3\%$. The full list of contributors can be found in [43]. The statistical uncertainty is much larger than the systematic uncertainty in the ttH channel, so the systematic uncertainty is considered negligible in this thesis.

7.4 Signal and Background Modelling

The shapes of the background and signal distributions add up to create a probability density function (PDF) of all events left after full background rejection procedure is completed. They are both modelled with analytical functions of $M_{\gamma\gamma}$ to create a final analysis likelihood function. The likelihood function also includes all uncertainties, which are incorporated as nuisance parameters, each of which corresponds to a Gaussian shaped PDF. The Higgs boson cross sections are included as parameters to the final likelihood model and the Higgs mass is assumed to be 125.09 ± 0.24 GeV [144].

7.4.1 Signal Modelling

The shape of the invariant mass $M_{\gamma\gamma}$ distribution of the ttH signal is described by an analytical function. Te same as previous ATLAS analyses before the 2020 publication [43], the *double-sided Crystal Ball* (DSCB) function is used. It is a composite function with 6 parameters formed by a Gaussian core, which models the peak, and two power-law tails, given by Equation 7.2:

$$f_{\mathrm{DSCB}}(M_{\gamma\gamma}) = N \times \begin{cases} e^{-t^2/2} & \text{if } -\alpha_{low} \leq t \leq \alpha_{high} \\ \dfrac{e^{-\frac{1}{2}\alpha_{low}^2}}{\left[\frac{1}{R_{low}}\left(R_{low} - \alpha_{low} - t\right)\right]^{n_{low}}} & \text{if } t < -\alpha_{low} \\ \dfrac{e^{-\frac{1}{2}\alpha_{high}^2}}{\left[\frac{1}{R_{high}}\left(R_{high} - \alpha_{high} + t\right)\right]^{n_{high}}} & \text{if } t > \alpha_{high} \end{cases} \tag{7.2}$$

The N denotes a normalization factor and the six parameters are

- μ_{CB} and σ_{CB} are the mean and the width of the Gaussian core, and are combined in $t = \left(m_{\gamma\gamma} - \mu_{\mathrm{CB}}\right)/\sigma_{\mathrm{CB}}$;
- α_{low} and α_{high} are the positions of the transitions from the Gaussian core to power-law tails on the low and high mass sides respectively;
- n_{low} and n_{high} are the exponents of the low and high mass tails. With the α's, they define $R_{low} = \frac{n_{low}}{\alpha_{low}}$ and $R_{high} = \frac{n_{high}}{\alpha_{high}}$.

Signal modelling for three high purity categories in different Higgs p_T regions is shown in Figure 7.1.

7.4.2 Background modelling

The background modelling procedure [43] includes the following two main steps: first a background model template histogram (example template given in Figure 7.2) is constructed after applying all analysis cuts and running either a BDT [43] or, used for this thesis, the adversarial neural networks platform with the purpose of rejecting maximum number of background events, while having the highest signal efficiency. For the ttH production, the templates are obtained from the simulated ttH events. In the second step, that template is used and run what is called a spurious signal test, which choses the best function among several options to fit the continuos background and therefore model the background

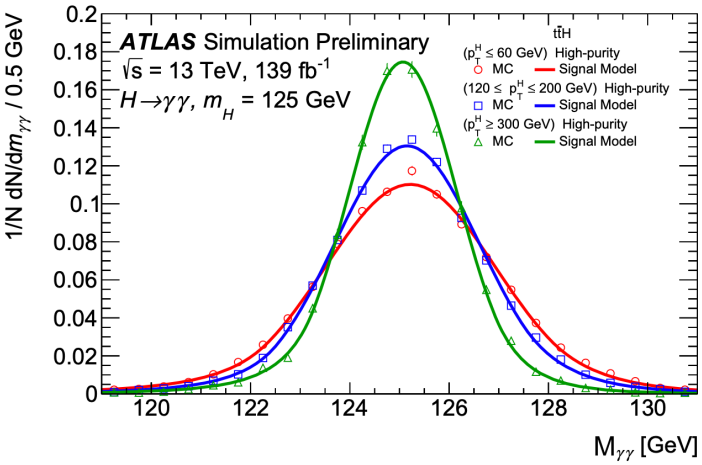


Figure 7.1: An example for signal modelling for three high purity ttH categories in different Higgs p_t regions [43].

distribution. The final goal is to choose an analytical function for the fit, which results in a small potential bias compared to the statistical uncertainty.

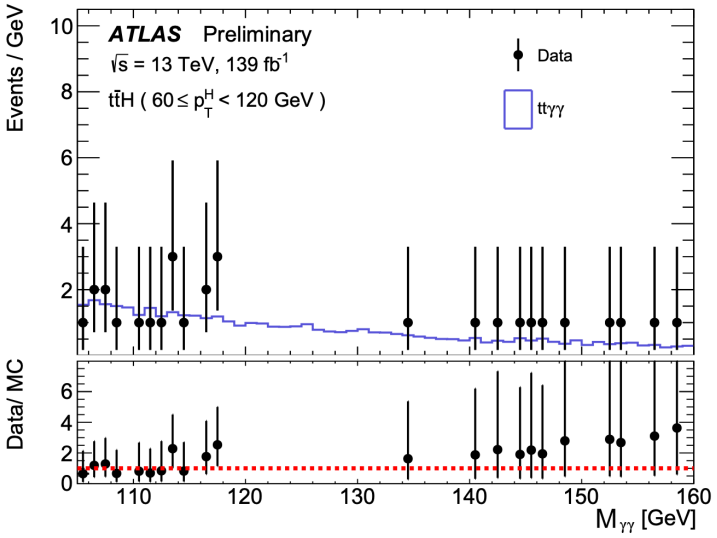


Figure 7.2: An example for constructing a background template. The data has been blinded in the signal region (120-130) GeV [43].

The background $M_{\gamma\gamma}$ shape is described using an analytic function, which is fitted to the $M_{\gamma\gamma}$ distribution in each analysis category. The considered functions, where the coefficients, c_i, are the free parameters used to define the function shape, are:

- Exponential function: $f(M_{\gamma\gamma}) = e^{c \cdot M_{\gamma\gamma}}$,

- Exponential function of 2^{nd} order polynomial called ExPoly2: $f(M_{\gamma\gamma}) = e^{c_1 \cdot M_{\gamma\gamma}^2 + c_2 \cdot M_{\gamma\gamma}}$,
- Exponential function of 3^{rd} order polynomial called ExPoly3: $f(M_{\gamma\gamma}) = e^{c_1 \cdot M_{\gamma\gamma}^3 + c_2 \cdot M_{\gamma\gamma}^2 + c_3 \cdot M_{\gamma\gamma}}$,
- Bernstein polynomial of order N: $B_N(M_{\gamma\gamma}) = \sum_{i=0}^{N} c_i \cdot b_{i,N}$ with $b_{i,N} = \binom{N}{i} m_{\gamma\gamma}^i (1 - M_{\gamma\gamma})^{N-i}$,
- First-order power law function called Pow: $f(m_{\gamma\gamma}) = M_{\gamma\gamma}^c$,
- Second-order power law function called Pow2: $f(M_{\gamma\gamma}) = M_{\gamma\gamma}^{c_1} + c_2 \cdot M_{\gamma\gamma}^{c_3}$.

The coefficients are assumed to be independent across categories, regardless of the functions chosen, and are always treated as free parameters in the fits to data.

Chapter 8

Results

The two main outcomes of this study are: (1) classifying ttH events with an Adversarial Neural Network performs comparably well as the classification techniques used by ATLAS so far; (2) The Adversarial Neural Network enables a balance between the efficient background rejection and the minimisation of the sculpting, which occurs in the process. This Chapter provides the results of a proof-of-principle analysis which answers this question. The ANN architecture developed for this thesis is provided in Section 8.1. Compared to the ATLAS analysis [43], this thesis uses a simplified set of input variables, described in Section 8.2. The loss functions for the two networks are shown in Section 8.3. The proposed Adversarial case is compared to the Scaled networks described in Chapter 8.4. They are first compared using performance metrics that are key to the $ttH(H \to \gamma\gamma)$ analysis, *i.e.*, classification (Section 8.5), decorrelation (Section 8.6) and a combined metric, simultaneously accounting for classification and sculpting performance (Section 8.7). Finally, results of the proof-of-principle $ttH(H \to \gamma\gamma)$ analysis are documented in Sections 8.8 and 8.10.

8.1 ANN Architecture

The hyperparameters used have been optimised for the Adversarial Neural Networks, used in [145] with the *Spearmint* library [146]. This uses Bayesian optimisation and scans the hyperparameter space for the optimal points, which correspond to the best balance between background sculpting minimisation and background rejection maximisation. The hyperparameters were chosen after a manual hyperparameter optimisation was performed for several values of the number of epochs, the adversarial network learning rate, the decay rate, and lastly the ratio between the classifier and adversarial learning rates. All the ANN

hyperparameter values required to reproduce the thesis results can be found in Table 8.1. The Adam optimiser [147] is an algorithm used for stochastic descent ML problems or for the calculation of exponential moving average of the gradient. The stochastic gradient descent maintains a single learning rate for all weight updates and the learning rate does not change during training. Adam was chosen due to its multiple advantages compared to other algorithms. It is computationally efficient, invariant to diagonal rescale of the gradients, works well with large data and it is also appropriate for problems with very noisy/or sparse gradients. The binary cross-entropy was chosen.

	Classifier	Adversary
Units	64	64
Activation Function	Relu	Relu
Architecture	3	1
Epochs	200	200
Batch Size	8192	8192
Loss Type	Binary Cross-Entropy	Binary Cross-Entropy
Learning Rate	1×10^{-2}	1×10^{-1}
Decay	1×10^{-3}	1×10^{-2}
Optimiser	Adam	Adam

Table 8.1: ANN hyperparameters. Units are the number of activation neurons, architecture corresponds to the number of layers. All hyperparameters are described in Chapter 6.

8.2 Input Variables

The input variables to the neural networks used for training were the energy E, transverse momentum p_T, pseudorapidity η and azimuthal angle ϕ of the leading (highest p_T) photon (γ_1) and the sub-leading (second highest p_T) photon (γ_2) and of the three leading jets (j_1, j_2, j_3), the difference between the pseudorapidities of the two photons $\Delta\eta$, the difference between the azimuthal angles of the two photons $\Delta\phi$ and the angular difference between the two photons ΔR. The two photons $\gamma\gamma$ come from the Higgs decay $H \to \gamma\gamma$ and the three leading jets from the decays of the W bosons. This information is used to differentiate between signal (MC $ttH, H \to \gamma\gamma$) and background and also to de-correlate the ANN discriminant from $M_{\gamma\gamma}$. Two background hypotheses are used: $tt\gamma\gamma$ MC and Non-Tight, Non-Isolated photon (NTNI) data. The idea behind NTNI data is trying to separate QCD jets (abundant, as they come from QCD processes)

from prompt photons (rare, as they come from electro-weak processes). The ATLAS calorimeter is good, but not 100% accurate when separating them. Isolated and tight cuts are therefore the cuts which are very likely to select photons rather then QCD jets, the non tight and non isolated are much more likely to select jets faking photons. Isolated refers to hadronic activity (tracks, calorimeter signals) around a photon. A QCD jet has a lot of hadronic activity and a prompt photon as little. Tight refers to identification requirement, which accounts for photon shape in the calorimeter. It means that the calorimeter assigns higher degree of confidence that this is a prompt photon. Loose, corresponds to a low degree of confidence. Therefore, selecting NTNI backgrounds gives a good impression of how the background looks like in the TI region.

The reason for using two alternatives is that the shape and the background composition to the $ttH(H \to \gamma\gamma)$ production is not predicted accurately. In an ideal scenario the choice of the classification working point used to extract the result in the data analysis would not depend on which background is assumed in the network training. The MC and the NTNI cases have notable differences in the input variable distributions and therefore enable us to probe how robust the network is.

Examples for the shapes of the variables used for training are shown in Figures 8.1 - 8.7 for the leading photon and jet in both MC and NTNI data. The figures shown in this section are obtained for the hadronic event selection (both W bosons decay hadronically). Similar conclusions are also obtained when considering the events passing the leptonic event selection (at least one W boson decays leptonically).

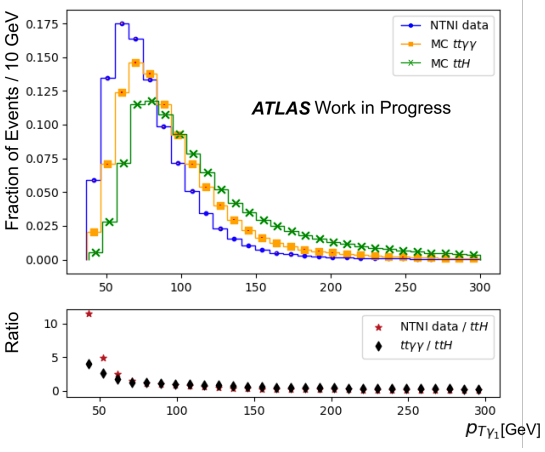


Figure 8.1: Transverse momentum distribution of the leading photon.

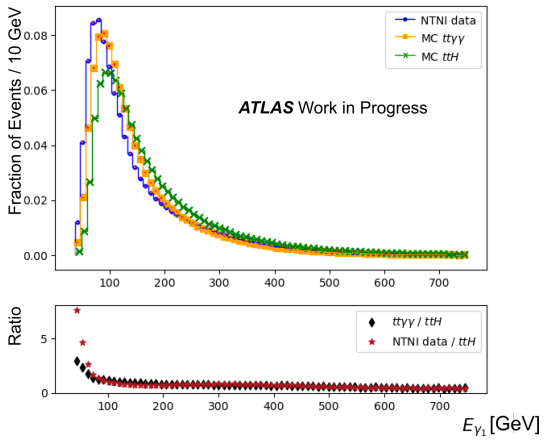


Figure 8.2: Energy distribution of the leading photon.

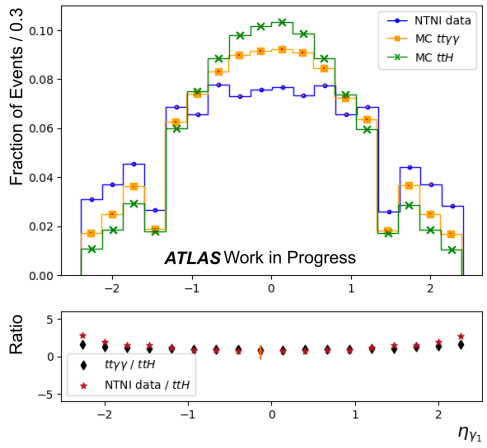


Figure 8.3: Pseudorapitidy distribution of the leading photon.

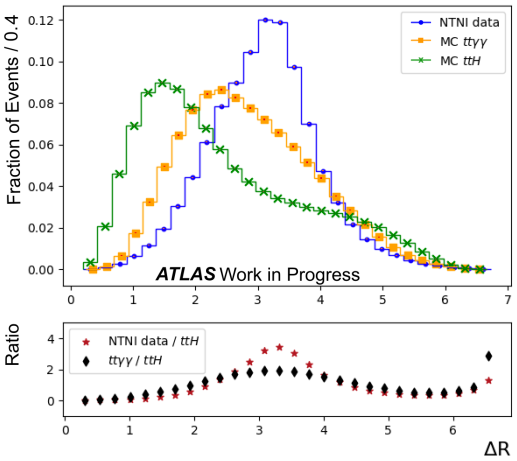


Figure 8.4: Angular difference distribution between the two photons.

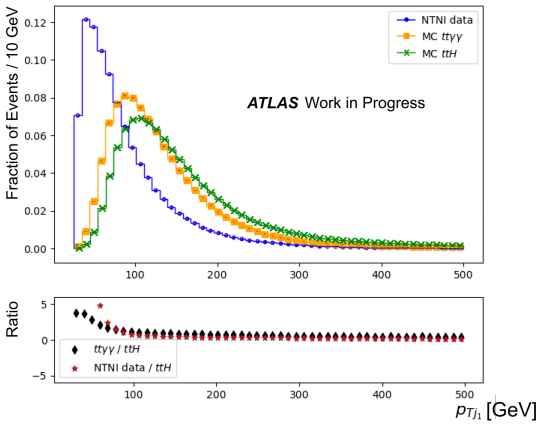


Figure 8.5: Transverse momentum distribution of the first jet.

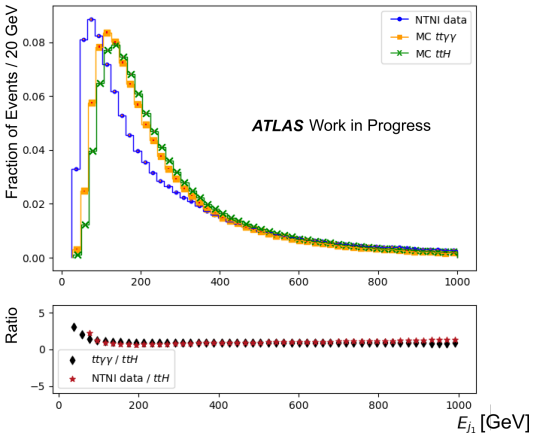


Figure 8.6: Energy distribution of the leading jet.

Figures 8.1, 8.2, 8.5 and 8.6 show the distributions of the transverse momenta and energies of the leading photon and jet. The transverse momentum and energy distributions tend to be the hardest for the ttH signal and softest for the NTNI data. The corresponding η distributions (Figure 8.3 and 8.7) show that the fraction of central events is the highest for signal and lowest for NTNI data. Both these observations can be explained by the fact that the ttH events are produced at the highest energy scales, since the final state contains three heavy particles $(t, \bar{t}, H)$. The NTNI background contains a large fraction of QCD background events with no massive final state particles, produced at low energy scales. The angular difference ΔR (Figure 8.4) shows that the distributions peak at higher

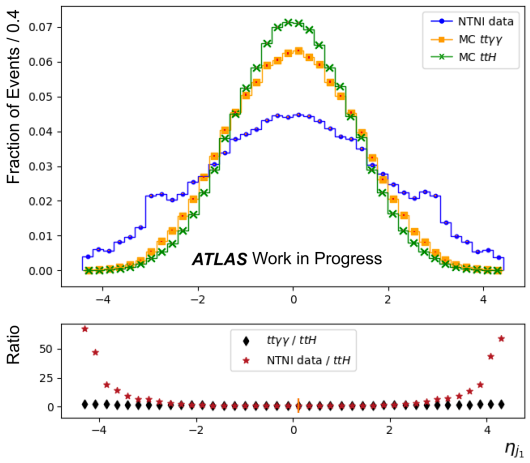


Figure 8.7: Pseudorapidity distribution and ratio of the leading jet.

ΔR values in $tt\gamma\gamma$ and NTNI data with respect to signal ttH and different shapes of the distributions. The signal ttH events, in which the photons are emitted from the Higgs boson, peak at smaller values compared to the backgrounds. The reason ttH has small ΔR is that the photons come from a boosted resonance, differently from NTNI data, which peaks at $\approx \pi$ because the photons are mostly the hardest objects in the events, thus back-to-back. The $tt\gamma\gamma$ case is somewhere in-between ttH and NTNI data in this aspect.

8.2.1 Correlations

The reason for the sculpting of the $M_{\gamma\gamma}$ distribution after background rejection are the correlations between the input variables and the $M_{\gamma\gamma}$ distribution. Some examples of correlations are shown in Figures 8.8 - 8.15 for both MC and NTNI background, where the colour scale (z-axis) represents the number of events. The correlation with a variable X, where n denotes all events in the sample, $\overline{X}$ and $\overline{M_{\gamma\gamma}}$ are the mean values and σ_X, $\sigma_{M_{\gamma\gamma}}$ the standard deviations of X and $M_{\gamma\gamma}$ respectively, is be given by:

$$c_{X,M_{\gamma\gamma}} = \frac{\sum_{i=1}^{n}(X - \overline{X})(M_{\gamma\gamma} - \overline{M_{\gamma\gamma}})}{\sigma_X \sigma_{M_{\gamma\gamma}}}. \tag{8.1}$$

Examples of variables with strong correlations with $M_{\gamma\gamma}$ are the transverse momentum and the energy of the leading photon (Figure 8.8 and 8.12).

The cuts used in this analysis include cuts on the fraction $p_T/M_{\gamma\gamma}$, in order to keep the general analysis' cuts. This introduces a further correlation, as it is another constraint on the relationship between the already most correlated

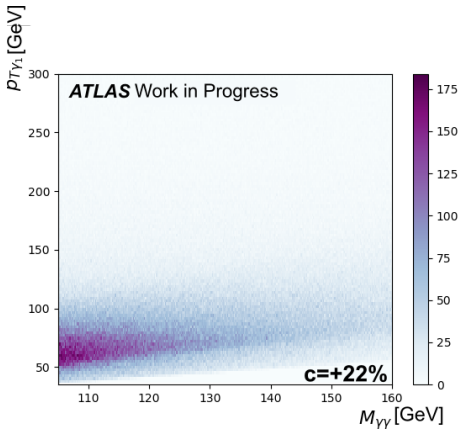


Figure 8.8: Correlation between the transverse momentum $p_{T\gamma_1}$ and $M_{\gamma\gamma}$ in simulated $tt\gamma\gamma$ background events.

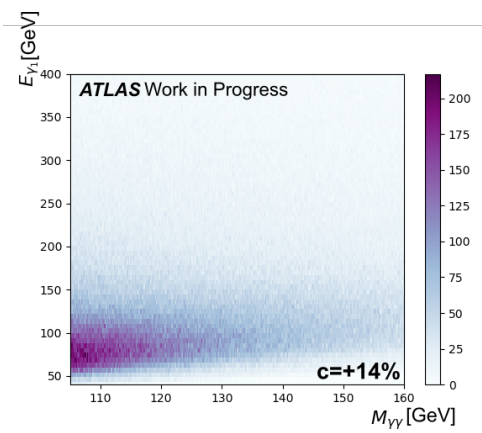


Figure 8.9: Correlation between the energy of the leading photon E_{γ_1} and $M_{\gamma\gamma}$ in simulated $tt\gamma\gamma$ background events.

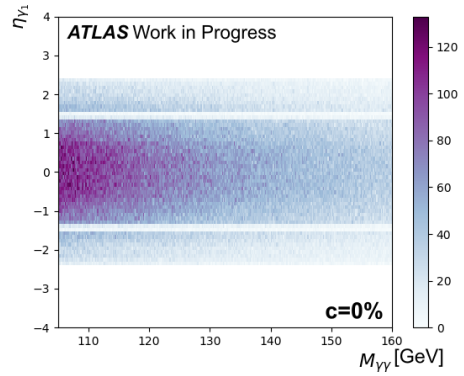


Figure 8.10: Correlation between the pseudorapidity of the leading photon η_{γ_1} and $M_{\gamma\gamma}$ in simulated $tt\gamma\gamma$ background events.

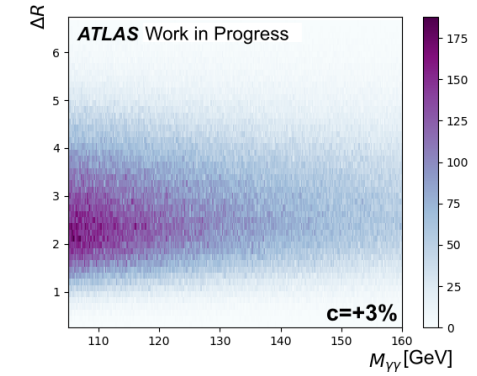


Figure 8.11: Correlation between the angular difference ΔR of the two photons and $M_{\gamma\gamma}$ in simulated $tt\gamma\gamma$ background events.

photon variable p_T and $M_{\gamma\gamma}$. Another example is the correlation between the $M_{\gamma\gamma}$ and the angular difference ΔR (Figures 8.11, 8.15). The correlations can be understood by considering the two-body decay of the Higgs boson to the massless photons. The energy-momentum conservation requires:

$$M_{\gamma\gamma}^2 = 2E_{\gamma_1}E_{\gamma_2}(1-\cos\theta), \tag{8.2}$$

where θ is the opening angle between the two photons.

Summaries of correlations between all input variables and $M_{\gamma\gamma}$ are shown in Figure 8.16 for MC $t\bar{t}\gamma\gamma$, Figure 8.17 for MC $t\bar{t}H$ and Figure 8.18 for NTNI data. On all correlation plots, the positive correlations are shown in red, the negative correlations in blue and the negligible to no correlations are shown in grey. On

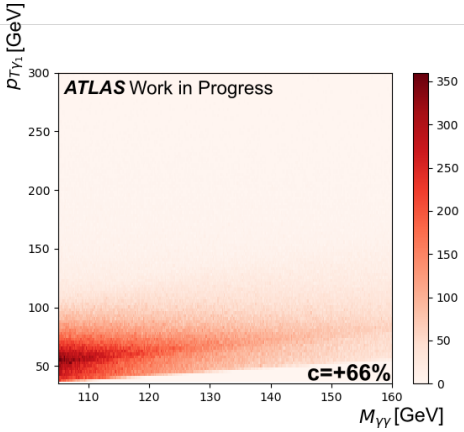


Figure 8.12: Correlation between the transverse momentum $p_{T\gamma_1}$ and $M_{\gamma\gamma}$ in the NTNI background events.

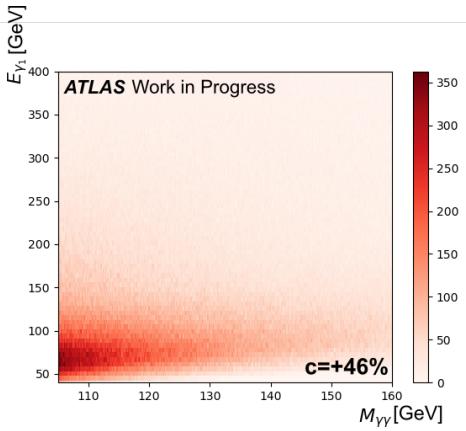


Figure 8.13: Correlation between the energy of the leading photon E_{γ_1} and $M_{\gamma\gamma}$ in the NTNI background events.

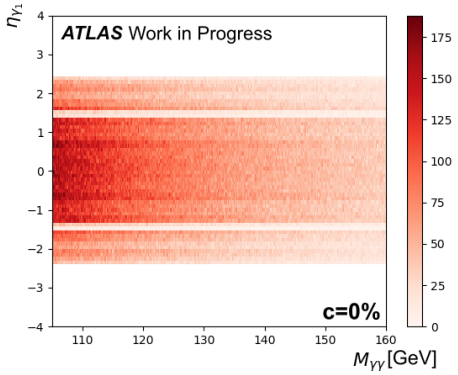


Figure 8.14: Correlation between the pseudorapidity of the leading photon η_{γ_1} and $M_{\gamma\gamma}$ in the NTNI background events.

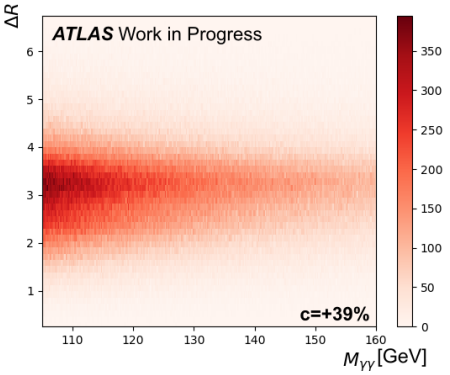


Figure 8.15: Correlation between the angular difference ΔR of the two photons and $M_{\gamma\gamma}$ in the NTNI background events.

top of each figure, there's a grid showing which variables are most correlated, where the ones most similar are always close neighbours.

In both NTNI data and MC $t\bar{t}\gamma\gamma$ (Figure 8.16 and 8.18), the most similar and highest correlated variables to the $M_{\gamma\gamma}$ distribution are the transverse momentum and energy of the leading and sub-leading photons and although the azimuthal angle ϕ and the pseudo-rapidity η themselves have negligible correlations with $M_{\gamma\gamma}$, while $\Delta\eta$, $\Delta\phi$ and ΔR have high correlations with $M_{\gamma\gamma}$.

The correlations between the different input variables in signal $t\bar{t}H$ (Figure 8.17) are similar to those in background, but with the important difference of them not being correlated with the $M_{\gamma\gamma}$ distribution, which is one of the main differences between signal and background that can be used for classification as well as decorrelation.

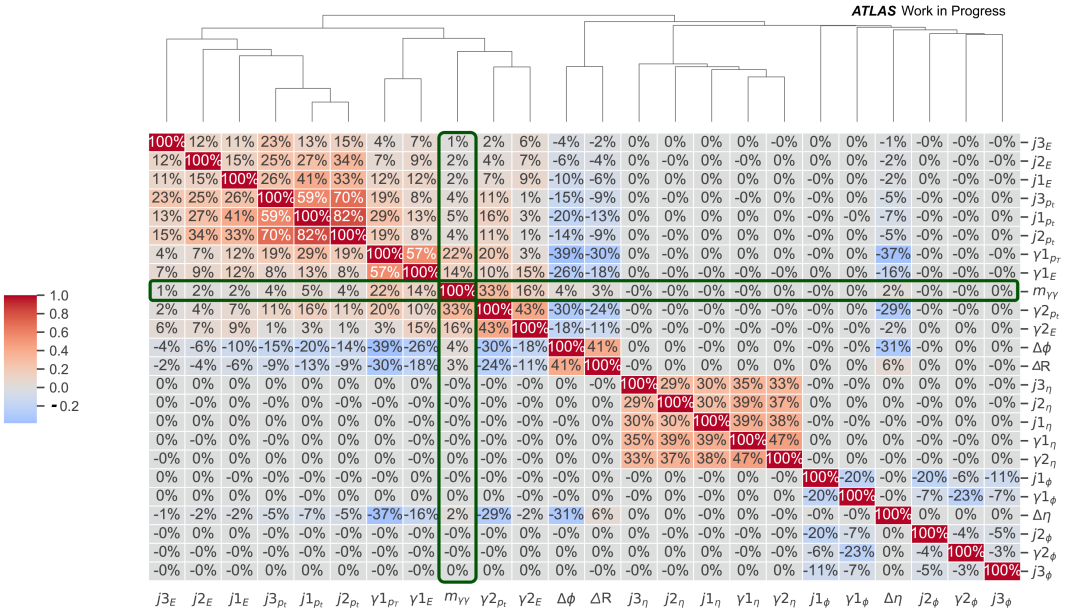


Figure 8.16: Correlations between the input variables in the $tt\gamma\gamma$ background events. The correlations with $M_{\gamma\gamma}$ are highlighted by the green box. The positive correlations are shown in red, the negative correlations in blue and the negligible to no correlations are shown in grey. The highest correlated variables to the $M_{\gamma\gamma}$ distribution are the transverse momentum and energy of the leading and sub-leading photons. The azimuthal angle ϕ and the pseudo-rapidity η have negligible correlations with $M_{\gamma\gamma}$ but $\Delta\eta$, $\Delta\phi$ but ΔR have high correlations with $M_{\gamma\gamma}$.

The relative importance of all used variables for training can be observed in Figure 8.19 for MC simulated events and in Figure 8.20 for NTNI data. for NTNI data. The ranking was obtained with ELI5 library, using classification accuracy as the figure of merit. The variables which contribute the most to the learning of the neural networks are, as expected, the ones that have the optimal balance between having the highest correlations with $M_{\gamma\gamma}$ and the highest importance for rejection of background events. Both in MC and NTNI data, $p_{T\gamma}$, η_γ and ϕ_γ are significantly stronger than the others, but regardless are not the only variables included. This is due to the way the ANN combines variables and creates relationships between them, which end up also contributing to the final result.

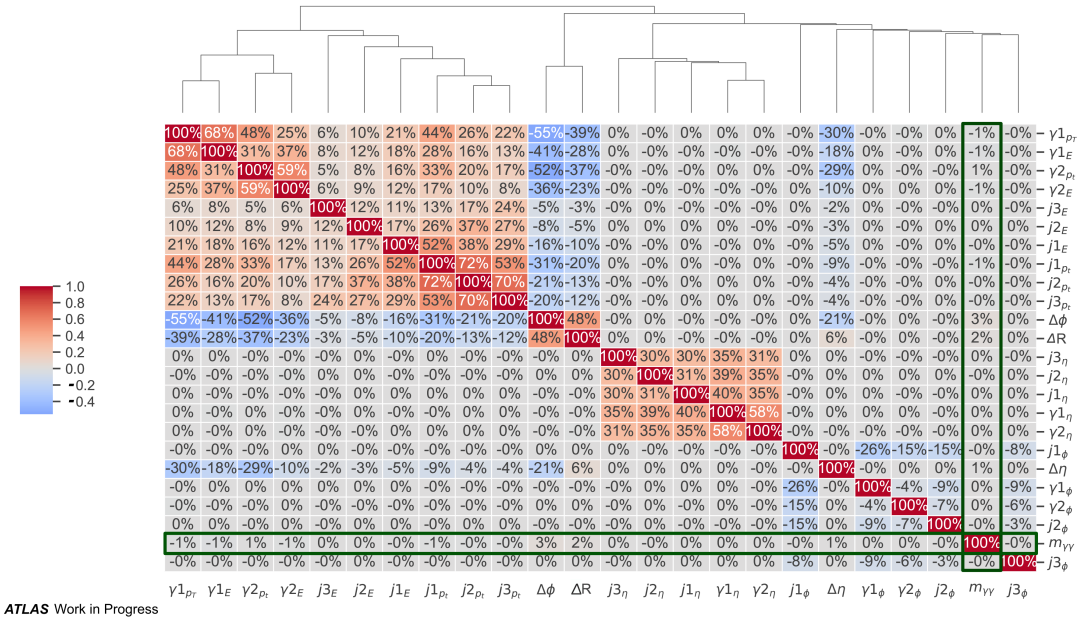


Figure 8.17: Correlations between the input variables in the $t\bar{t}H$ signal events. The correlations with $M_{\gamma\gamma}$ are highlighted by the green box. The positive correlations are shown in red, the negative correlations in blue and the negligible to no correlations are shown in grey.

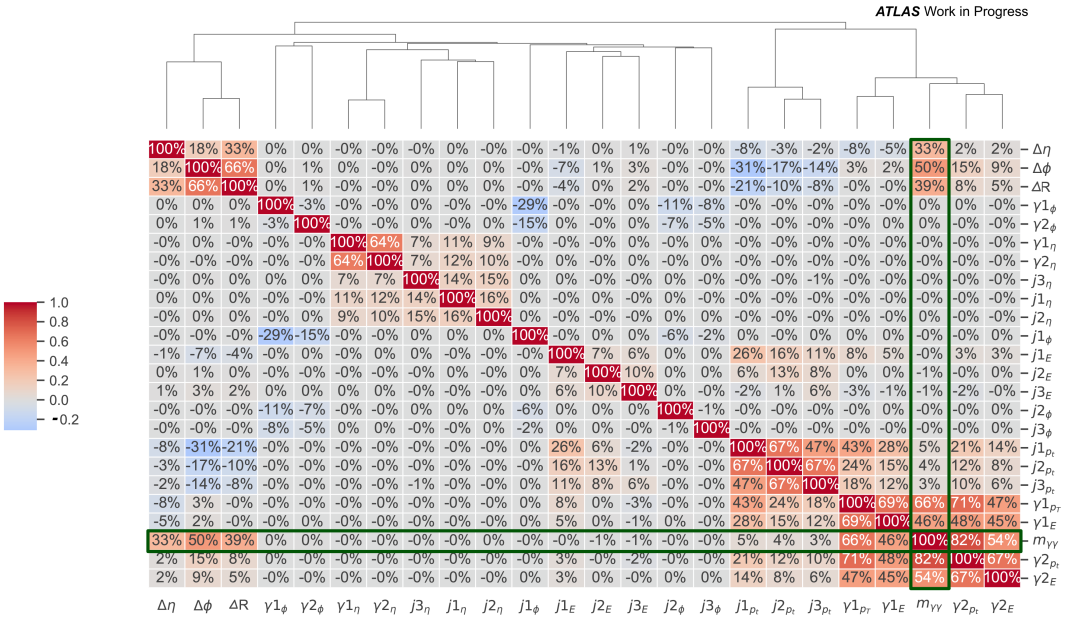


Figure 8.18: Correlations between the input variables in the NTNT data background events. The correlations with $M_{\gamma\gamma}$ are highlighted by the green box. The positive correlations are shown in red, the negative correlations in blue and the negligible to no correlations are shown in grey. The highest correlated variables to the $M_{\gamma\gamma}$ distribution are the transverse momentum and energy of the leading and sub-leading photons. The azimuthal angle ϕ and the pseudo-rapidity η have negligible correlations with $M_{\gamma\gamma}$ but $\Delta\eta$, $\Delta\phi$ and ΔR have high correlations with $M_{\gamma\gamma}$.

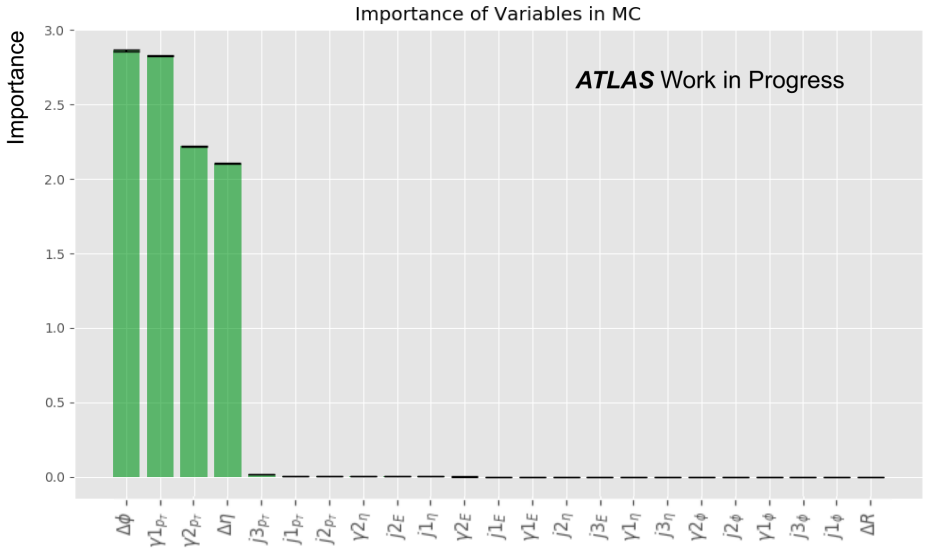


Figure 8.19: Ranking variables used for training in MC simulated ttH signal and $ttyy$ background events.

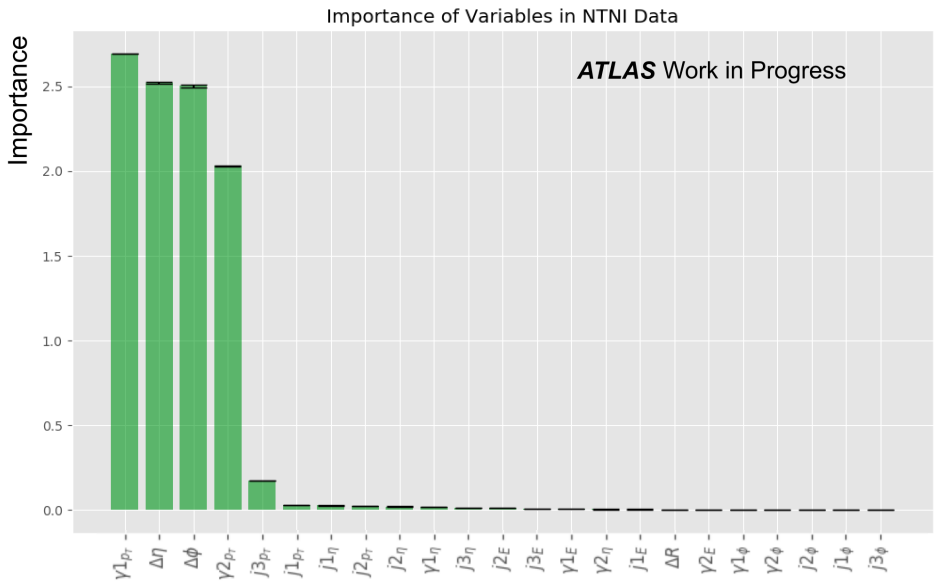


Figure 8.20: Ranking variables used for training in MC simulated ttH signal and NTNI data background events.

8.3 Loss Functions

The loss function is, as discussed in Section 6.1, the smallest difference between the fit and the data points, which is minimised (or maximised depending on the final goal) by finding the set of weights, which correspond to the absolute minima. The general loss function for a regression problem is given by equation 6.10 and the general loss function for a classification problem is given by equation 6.5. While the first neural network classifies the events as either signal or background, the second deals with the sculpting.

The neural networks are trained on 50% signal (ttH), 50% background (either $tt\gamma\gamma$ for MC or NTNI for real data) events. The total dataset is then randomised and split into two equal parts, one used for training and the other for calculating predictions. This ensures that the events, on which training is done, are different from the events used for a test, which is adopted for unbiased learning.

Figures 8.21 - 8.23 show the loss function of the classifier (J_{cls}), adversary (J_{adv}) and the combined network ($J_{ANN} = J_{cls} - \lambda J_{adv}$) respectively. The results are shown for the networks trained with the simulated $tt\gamma\gamma$ background events in the hadronic decay channel, but similar conclusions are obtained for the network trained with the NTNI background, and the networks trained in the leptonic decay channel. The results are shown for λ=25, and similar conclusions are obtained for other λ values. In all cases, the loss functions reach a plateau or an optimum (meaning, it no longer learns in further epochs). This means that training was sufficiently long for the neural network to reached a stage, where it no longer learns. The classifier in this analysis was pre-trained to an optimal configuration for classification, before the adversarial training commences.

The training was validated with k-fold cross-validation. All figures show a good agreement of the validation losses with the training losses, which confirms there is no over-fitting, i.e., differences between the training and validation.

8.4 Scaled Neural Network

A benchmark for evaluating the performance of the ANN is a Scaled neural network. In this Scaled NN the sculpting is reduced by a technique used in the ATLAS $ttH(H \to \gamma\gamma)$ analyses for the top quark Yukawa observation [43]. It is to scale the transverse momentum and the energy of the two photons by dividing them by $M_{\gamma\gamma}$, before using them as inputs to the machine learning classification algorithms.

The reason for the sculpting are the correlations between the input kinematic

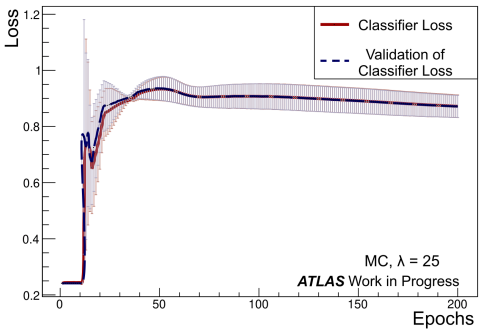


Figure 8.21: Loss function of the classifier network in training (solid) and in validation (dashed) samples. The number of epochs of training are given on the x-axis. First 10 epochs are for pre-training.

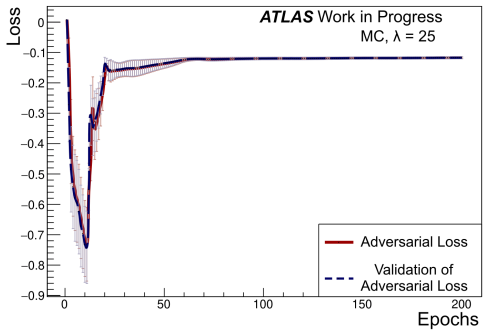


Figure 8.22: Loss function of the adversary network in training (solid) and in validation (dashed) samples. The number of epochs of training are given on the x-axis. First 10 epochs are for pre-training.

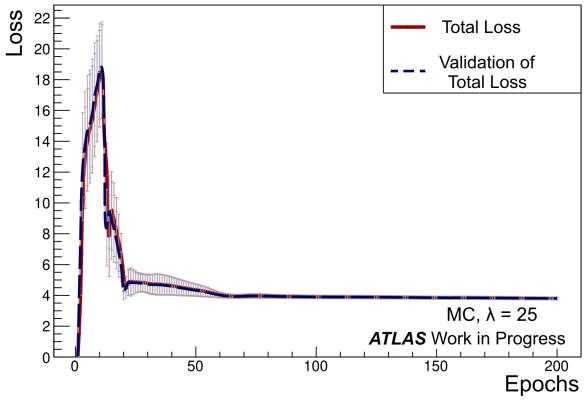


Figure 8.23: Total loss function $J_{ANN} = J_{cls} - \lambda J_{adv}$ for the ANN training and its corresponding validation loss. The number of epochs of training are given on the x-axis. First 10 epochs are for pre-training.

variables of the photons and $M_{\gamma\gamma}$. Dividing the p_T and E by $M_{\gamma\gamma}$ removes some of this dependence, but does not deal with the correlations between the photon angular variables (ΔR, $\Delta\eta$, $\Delta\phi$) and $M_{\gamma\gamma}$. Additionally, the division by $M_{\gamma\gamma}$ provides a single point in a space of potentially many solutions. The goal is to find the balance between having the highest possible sensitivity and having the lowest possible sculpting. While the division is simple, there is no way for the optimal solution to be selected among the different possible scenarios, and the probability for the single point in the solutions space to be the optimal one is negligible. The adversarial neural networks on the other hand, enable a much more flexible environment, where the network's hyperparameters and the input

variables can be varied until a true optimum is achieved for both sensitivity and sculpting.

8.5 Classification

8.5.1 Simulated events

The aim of classification is to use a discriminant cut to keep a high fraction of signal events, while rejecting most backgrounds. By rejecting the events, which do not contain the Higgs boson, originating from a pair of top quarks and decaying to a pair of photons, the classifier ensures, the final signal mass peak is clearer. In Figure 8.24, the discriminant distribution after background $M_{\gamma\gamma}$ distribution after classifier stand-alone training is shown for each of the ttH signal and $tt\gamma\gamma$ background. The discriminant here is a reduction technique that is commonly used for supervised classification problems. It is used for modelling differences in groups i.e. separating two or more classes; signal and background in this study.

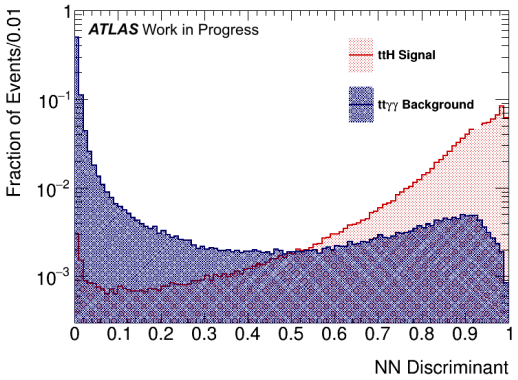


Figure 8.24: Signal ttH (in red) and background $tt\gamma\gamma$ (in blue) classifier NN discriminant shapes for MC hadronic events.

The performance of the network is calculated with what is called Receiver Operating Characteristic Curves (ROC). A ROC curve is a graph, which shows the balance between the False Positives (the rate of background misidentified as signal) and the True Positives (the rate of signal identified as signal). All ROC curves for the performance of the networks in each MC scenario discussed, are shown in Figure 8.25. The ANNs with a lower value of the parameter λ have a higher classification power, but are less effective in reducing the sculpting of the background $M_{\gamma\gamma}$ distribution. This is due to λ being the parameter, which controls how much more importance in the learning process is given to the adversary's job in comparison to the classifier's. The reduction of sculpting is discussed

in Section 8.6, where λ=20 is shown to be a good compromise. In this case a comparison between the Scaled NN and the ANNs (Figure 8.26) for MC leptonic and hadronic scenarios show values within 5% difference between the ROC curve areas for the Scaled NN and ANN cases.

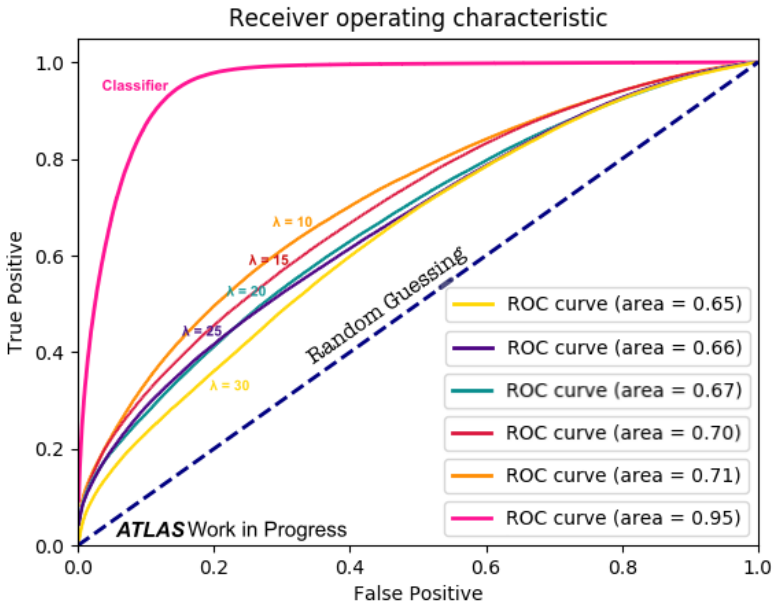


Figure 8.25: ROC curves and their corresponding areas under the curve for ANN training with MC hadronic ttH signal and $t\bar{t}\gamma\gamma$ background events. The pink curve shows the classifier's performance, the orange, red, cyan, purple and yellow curves show the adversaries' performance for the values of λ, the parameter controlling the loss function L_{adv}, of 10, 15, 20, 25 and 30 respectively.

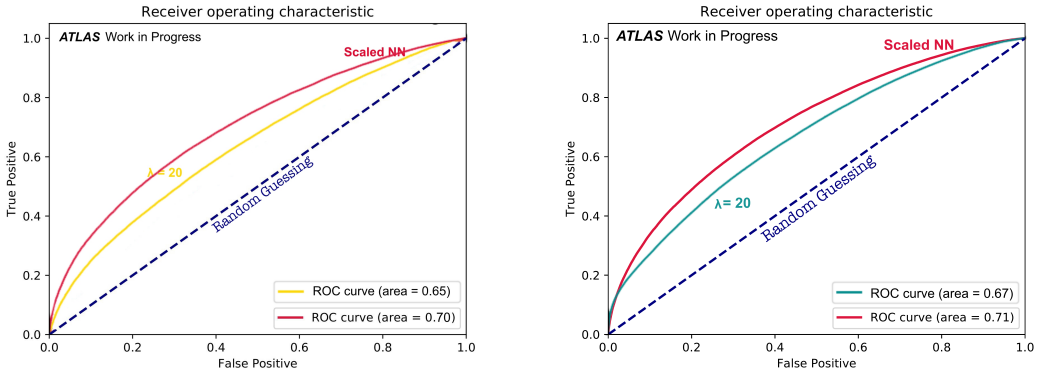


Figure 8.26: ROC curves and their corresponding areas under the curve for training with MC leptonic (left) and hadronic (right) events. The red curve shows the performance of a Scaled network, described in Section 7.4, the yellow and cyan the performance of the ANNs for $\lambda = 20$.

Network	Efficiency [%]		
	Signal	Rejection Background	Total
NN lep	87	85	86
NN had	95	85	90
ANN lep $\lambda = 20$	**67**	**62**	**65**
ANN had $\lambda = 10$	70	60	66
ANN had $\lambda = 15$	67	60	64
ANN had $\lambda = 20$	**67**	**55**	**61**
ANN had $\lambda = 25$	69	52	61
ANN had $\lambda = 30$	65	55	60
Scaled NN lep	**60**	**69**	**65**
Scaled NN had	**66**	**64**	**65**

Table 8.2: Overall efficiencies (in %) of the neural networks in signal and background and total (50 % signal and 50 % background) after training with MC leptonic (lep) and hadronic (had) events. NN is the classifier, ANN the combined classifier with adversary, and the Scaled NN is the benchmark network described in Section 8.4. Efficiency is the percentage of time the neural network learns correctly what it signal and what is background.

8.5.2 NTNI Data

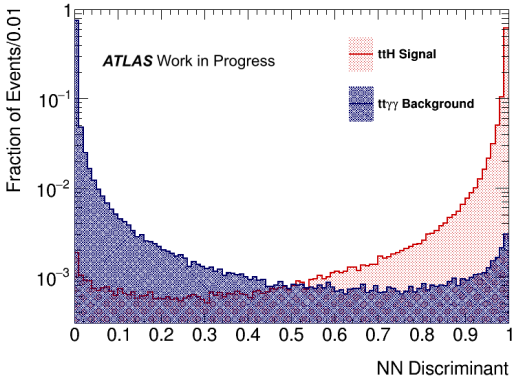


Figure 8.27: Distribution of signal ttH (in red) and background $tt\gamma\gamma$ (in blue) for NTNI data events. The discriminant is the probability for an event to be a signal event.

The ANNs were also used for the training with NTNI data as background. The signal and background distributions after classifier training with the NTNI data (Figure 8.27) are much better separated than what was observed in MC, due to the higher ANN performance. In this case, whatever discriminant is chosen between ≈ 0.05 and ≈ 0.9, the efficiencies of classification in both signal and background remain excellent (Table 8.3). After adversarial training, the efficien-

cies drop slightly, but remain excellent and vary between 84-95% in both signal and background for a discriminant cut of 0.5. This can also be seen from the high ROC areas values in Figure 8.28. Another tendency observed in NTNI data, in comparison to MC simulated events, is the higher parameter λ, required to achieve full sculpting minimisation.

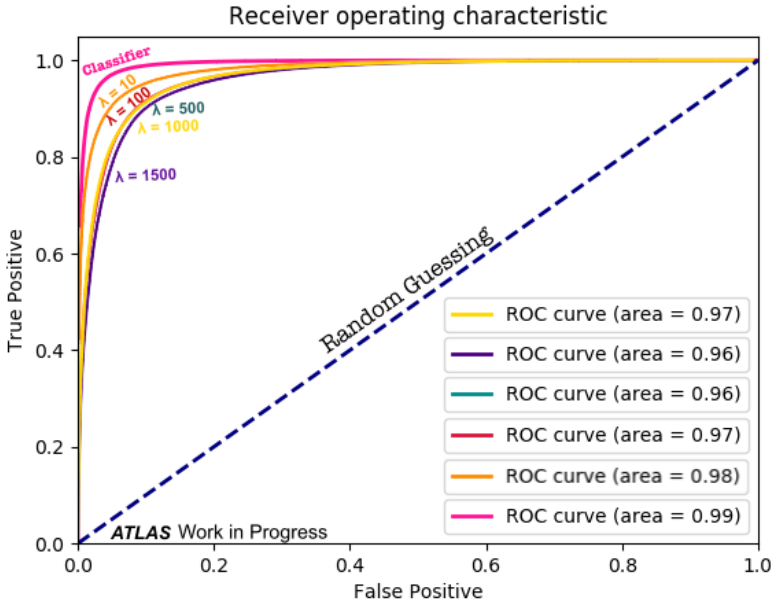


Figure 8.28: ROC curves and their corresponding areas for ANN training with NTNI hadronic data. The pink curve is the ROC shows the classifier's performance, the orange, red, cyan, purple and yellow curves show the adversaries' performance for the values of λ, the parameter controlling the loss function L_{adv}: 10, 100, 500, 1000 and 1500 correspondingly. Scaled network yeilds comparable results.

Network	Efficiency		
	Signal	Rejection Background	Total
NN lep	89	87	88
NN had	97	96	96
ANN lep $\lambda = 500$	**73**	**70**	**71**
ANN had $\lambda = 10$	94	92	93
ANN had $\lambda = 100$	94	87	91
ANN had $\lambda = 500$	**93**	**85**	**89**
ANN had $\lambda = 1000$	93	87	90
ANN had $\lambda = 1500$	95	84	90
Scaled NN lep	**63**	**53**	**58**
Scaled NN had	**93**	**90**	**92**

Table 8.3: Overall efficiencies (in %) of the neural networks in signal and background and total (50 % signal and 50 % background) after training with NTNI data leptonic (lep) and hadronic (had) events. NN is the classifier, ANN the combined classifier with adversary, and the Scaled NN is the benchmark network described in Section 8.4. Efficiency is the percentage of time the neural network learns correctly what it signal and what is background.

8.6 Decorrelation

8.6.1 Simulated MC Events

The second neural network, the adversary, deals with the decorrelation of the $M_{\gamma\gamma}$ distribution from the photon kinematic variables. Those correlations, as described in Section 8.2.1, are the reason for the sculpting of the background distribution after classification. In Figure 8.29, the $M_{\gamma\gamma}$ distributions can be seen in three phases: before classifier training in blue, after a cut on the classifier discriminant in red and after full adversarial training in green. Ideally, the shape of the distribution after the discriminant cut would have a consistent shape as the initial $M_{\gamma\gamma}$ distribution before any discriminant cuts. This would correspond to a fully solved sculpting problem. Figure 8.29 contains five plots for the hadronic case, which are different only in λ value, *i.e.*, the strength given to the adversary's task with respect to the classifier's task, which is one of the ANNs hyper-parameters controlled by the user. Figure 8.30 contains two plots of the training and predictions on MC leptonic events (left) and training on MC and predictions on NTNI leptonic events (right). The red curve is the same for all, as it is the $M_{\gamma\gamma}$ distribution after the stand-alone classifier. It shows strong sculpting and peaks at the mass of the Higgs boson. The adversary joins the classifier after the initial stand-alone classifier training to resolve that problem. The higher the value of the parameter λ, which controls the loss function, the lower the accuracy of the networks (Figure 8.25), but the higher the minimization of the sculpting (see Table 8.4). A balance is sought, which achieves both the goals of high performance and negligible to no sculpting simultaneously. In the case of $\lambda = 10$, the sculpting is significantly reduced, but not to an extent where a simple background modelling can be used for $M_{\gamma\gamma}$ after it has undergone adversarial training, similarly to before training. This is important to make sure that the spurious signal test (described in Section 8.10) passes with minimum additional complexity introduced. As λ increases, the sculpting diminishes, and on the plot in the far right, the optimal case for $\lambda = 20$ can be observed, which corresponds to a final sensitivity of Z = 3.3 (described in detail in 8.8). Apart from the slight change in slope, the initial and final distribution's shapes can be modelled with the same function. For Figure 8.29, a classifier discriminant cut D_{cls} was chosen for illustration purposes. The cut corresponds to the signal acceptance of 80%. The area under the ROC curve for this case is 0.67.

In the leptonic decay channel, the choice of $\lambda = 20$ yields optimal performance (metric explained in 8.6.3) for both the $tt\gamma\gamma$ and NTNI data background

hypotheses. The corresponding area under the ROC curve is 0.64 for $tt\gamma\gamma$ and 0.71 for the NTNI data respectively.

The combined ANN discriminant in $M_{\gamma\gamma}$ (Figure 8.31) varying with respect to λ in a discriminant range [0.2,1] shows the most evident regions of dependence. A case of full lack on dependence would have only fully horizontal white contour lines.

The same events used as input to the ANNs in both MC and NTNI data were used with scaled variables $\frac{p_T}{M_{\gamma\gamma}}$ and $\frac{E}{M_{\gamma\gamma}}$ for the scaled network as input to a stand-alone classifier and the sculpting and sensitivity were compared with the optimal case from the ANNs (Figures 8.32 and 8.33). In both cases, a change in slope from the original $M_{\gamma\gamma}$ distributions can be observed.

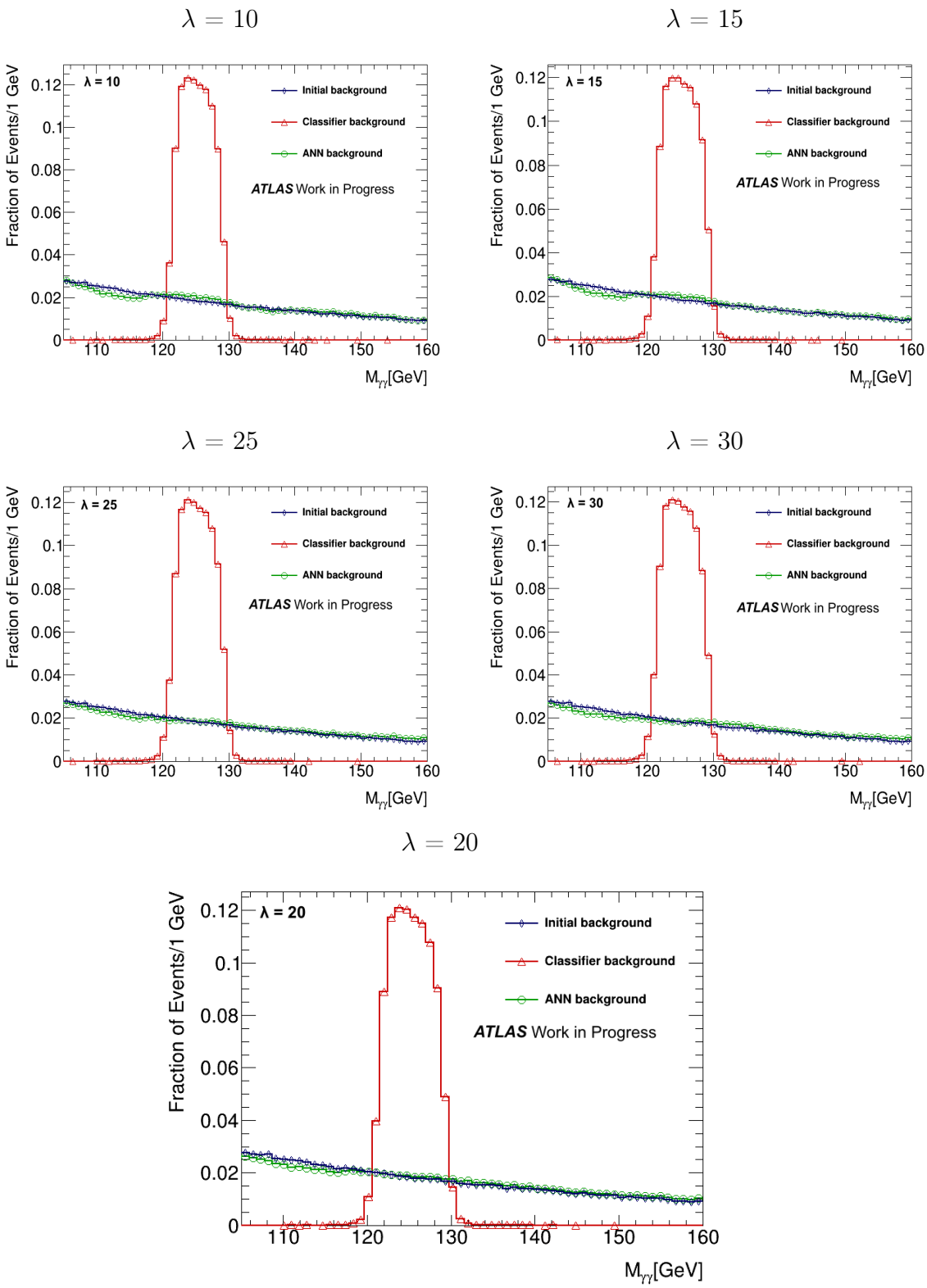


Figure 8.29: The shapes of the $M_{\gamma\gamma}$ distribution in the hadronic MC background $tt\gamma\gamma$ events. The initial background is shown in blue, the background after stand-alone classifier training in red and the background after adversarial training with both networks in green. The stand-alone classifier distribution does not depend on λ, and is therefore the same in all figures. Signal efficiency = 80%.

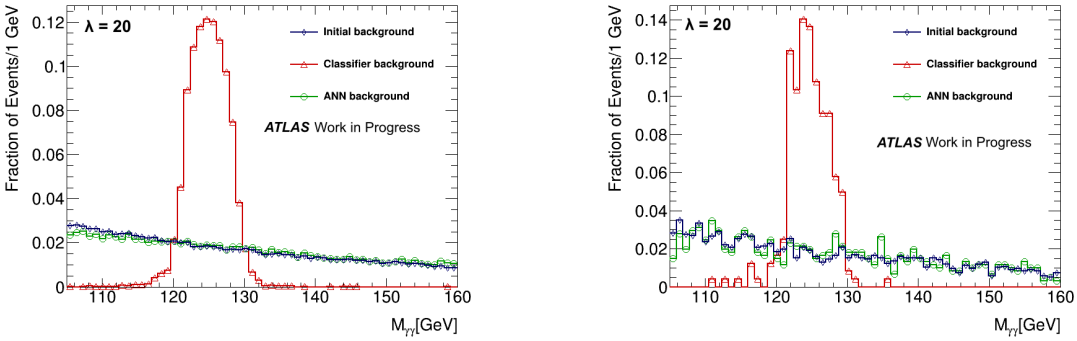


Figure 8.30: MC semi-leptonic and di-leptonic $M_{\gamma\gamma}$ background shapes for the three main steps of ANN training. The distributions are normalized to unit area. The initial background is shown in blue, the background after stand-alone classifier training in red and the background after adversarial training with both networks in green. Left plot: training and predictions on MC leptonic events. Right plot: training on MC leptonic events, predictions on NTNI leptonic events. Signal efficiency = 80%.

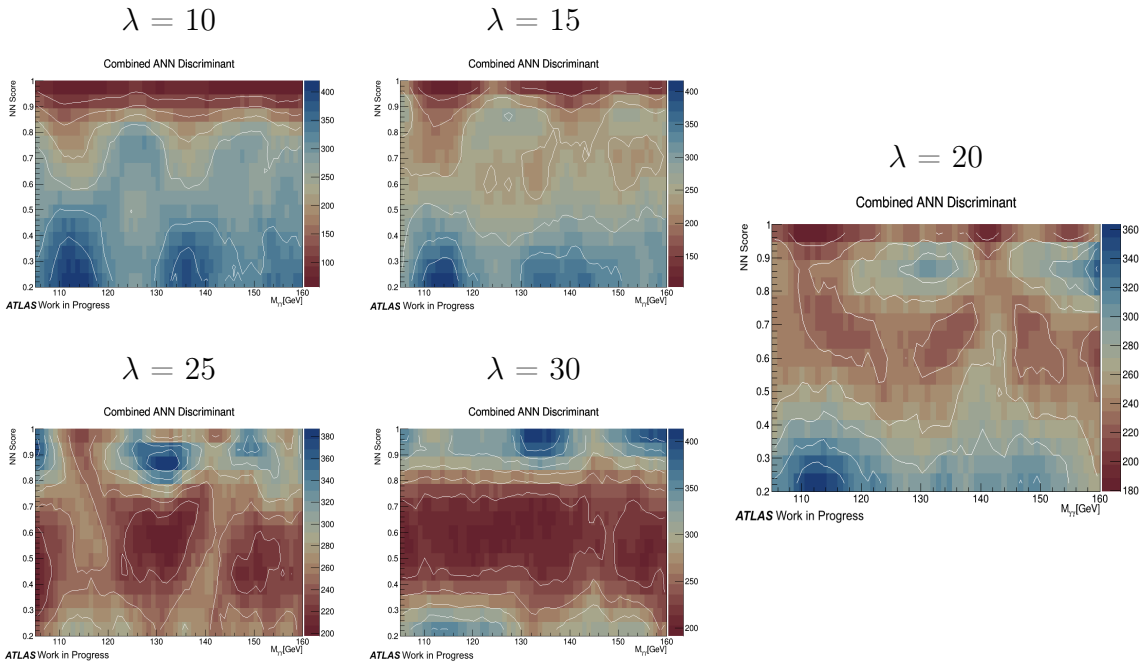


Figure 8.31: 2D Plots of the combined ANN discriminant with respect to $M_{\gamma\gamma}$ for various values of the regularization parameter λ. The white lines show the contours connecting the bins with the same number of events.

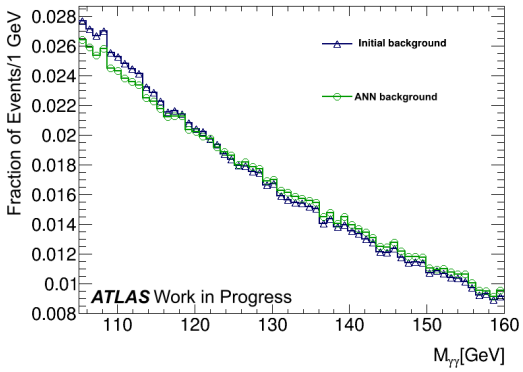


Figure 8.32: Background $M_{\gamma\gamma}$ distribution after ANN training of MC $tt\gamma\gamma$ background and ttH signal events with un-scaled photon kinematic variables. $\lambda = 20$.

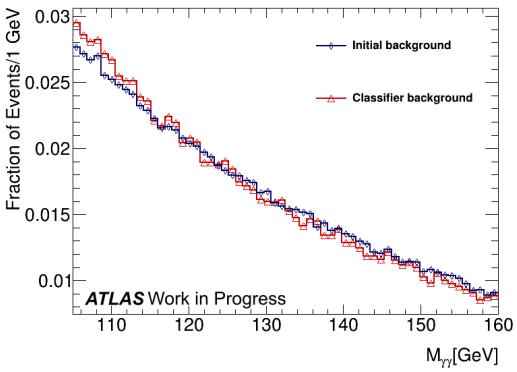


Figure 8.33: Background $M_{\gamma\gamma}$ distribution after classifier stand-alone training of MC $tt\gamma\gamma$ background and ttH signal events with scaled photon kinematic input variables.

8.6.2 NTNI data background

The sculpting minimisation achieved is shown in Figure 8.34 in the data driven background NTNI, for five different λ values, where the optimal case was chosen to be $\lambda = 500$. Just like in MC, a small change in slope can be observed as a difference between the initial and final $M_{\gamma\gamma}$ background distributions. The dependence of $M_{\gamma\gamma}$ distribution on the discriminant is significantly smaller than the observed in MC background. This is shown through the very nearly horizontal contour lines in Figure 8.35.

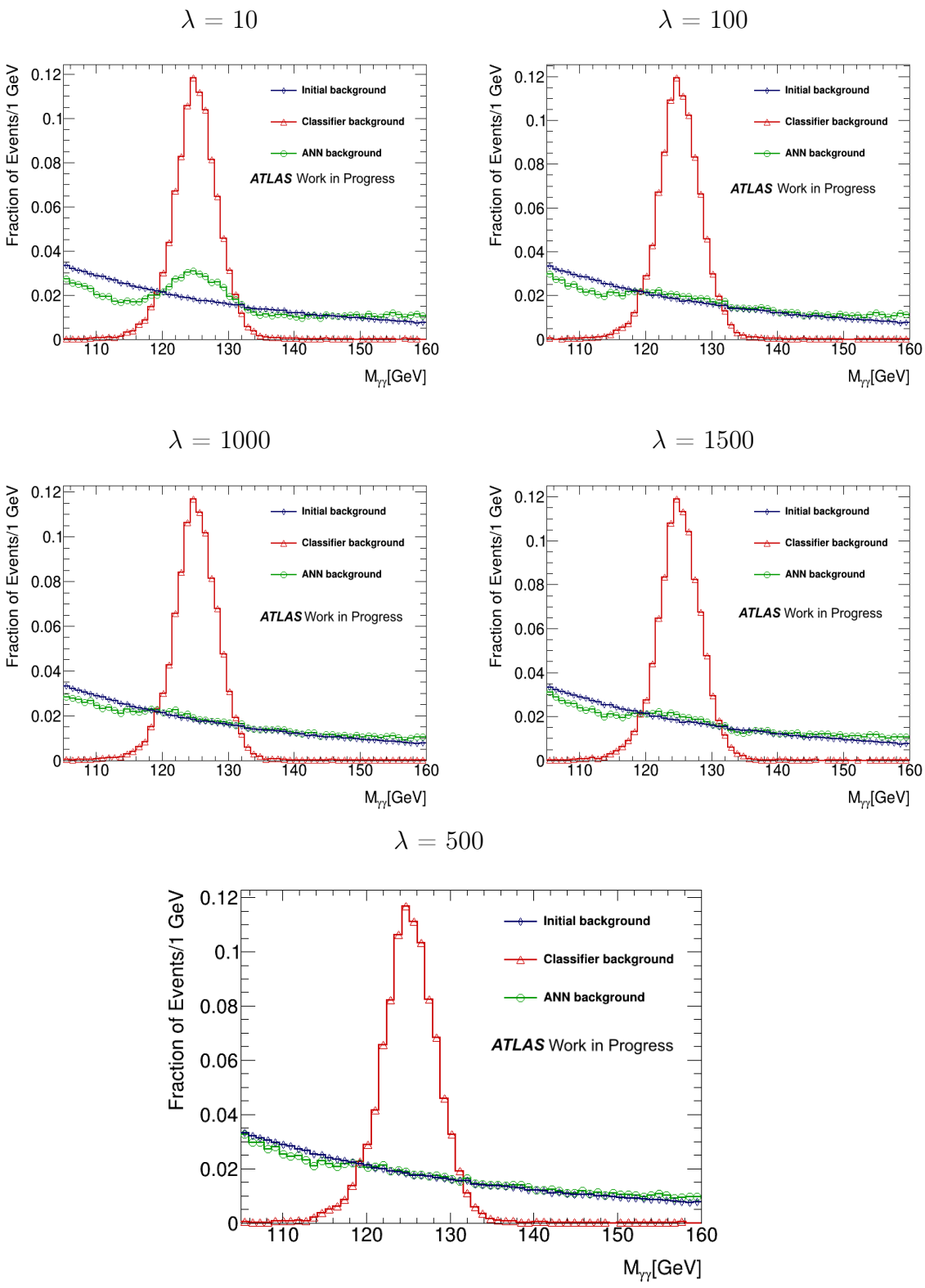


Figure 8.34: NTNI data hadronic $M_{\gamma\gamma}$ background distribution's integrated area shapes after ANN training. The initial background is shown in blue, the background after stand-alone classifier training in red and the background after adversarial training with both networks in green. Signal acceptance = 80%.

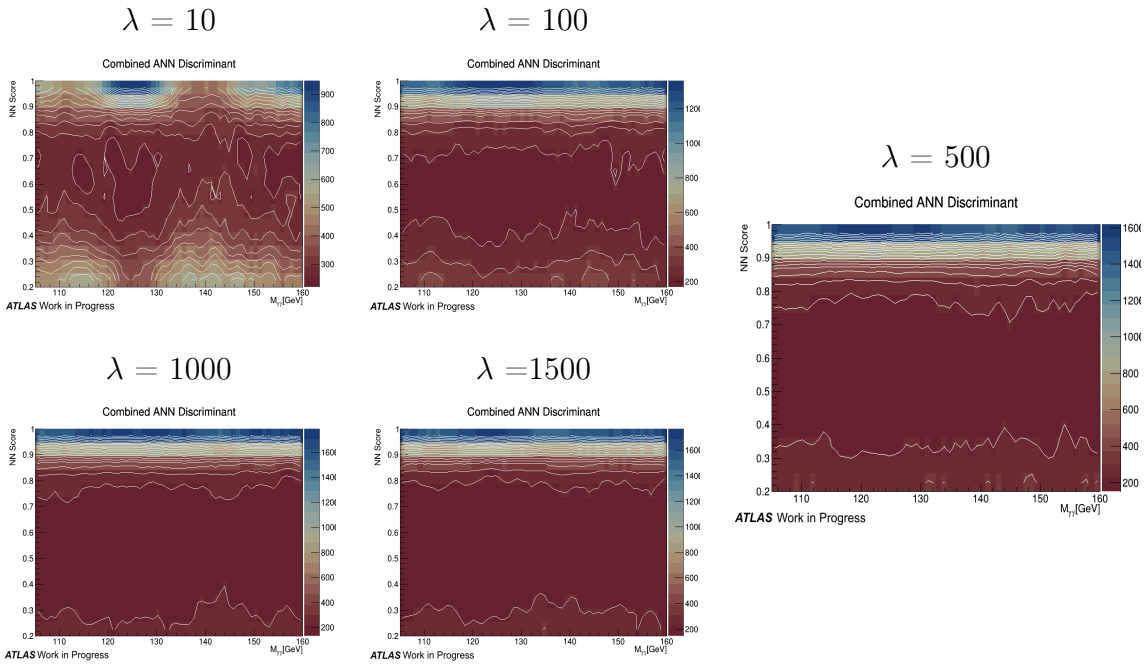


Figure 8.35: 2D Plots for NTNI data of the combined ANN discriminant (y axis) with respect to $M_{\gamma\gamma}$ (x axis) for various values of the regularization parameter λ. The white lines show the contours connecting the bins with the same number of events.

Differently from MC, the real data shows no change in slope in the scaled case (Figure 8.37) but has larger statistical fluctuations than the unscaled (Figure 8.36) at (120-140) GeV, which includes the mass of the Higgs.

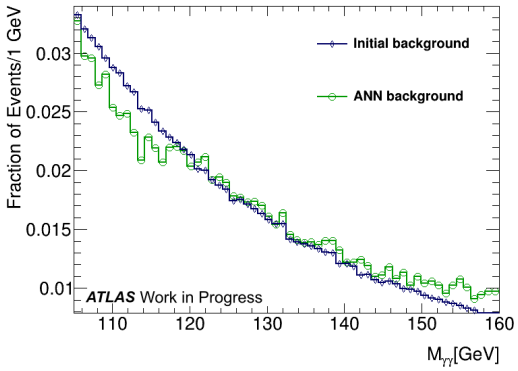


Figure 8.36: Optimal background $M_{\gamma\gamma}$ distribution after ANN training of NTNI Run 2 background data and ttH signal events with un-scaled photon kinematic variables. $\lambda = 500$.

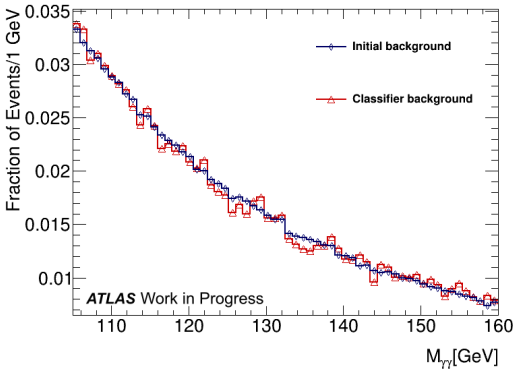


Figure 8.37: Classifier stand-alone training of NTNI real Run 2 background data and ttH signal events with scaled photon kinematic input variables.

8.6.3 Stability test of results

To quantify the sculpting effects, the Jenson Shannon Divergence (JSD), defined in Section 6.7, is calculated for the difference between the initial $M_{\gamma\gamma}$ background distribution and the $M_{\gamma\gamma}$ background distribution after cutting on the neural network discriminant.

The statistical uncertainty of the JSD is estimated from the statistical uncertainty of the underlying histograms by using pseudo-experiments. In each pseudo-experiment the histogram bins are randomly fluctuated according to their statistical uncertainty, and the JSD of the fluctuated histogram pair is evaluated. The spread of the JSD values over the pseudo-experiments is taken as the statistical uncertainty on the JSD. To validate the procedure randomly sampled histogram pairs from an identical underlying distribution are generated. On average, the resulting JSD is found to be unbiased, and the spread of the example histogram's JSD values is consistent with the assigned JSD uncertainty.

The JSD values for MC simulated events can be seen in Table 8.4. Two features of the differences between the $M_{\gamma\gamma}$ distributions before and after the discriminant cut are important, when quantifying sculpting: the full $M_{\gamma\gamma}$ range, as used in the $H \to \gamma\gamma$ analyses: 105 GeV $< M_{\gamma\gamma} <$ 160 GeV, and the range around the Higgs boson mass peak: 120 GeV $< M_{\gamma\gamma} <$ 130 GeV.

The sculpting in the full range was calculated to fall until $\lambda = 20$, where it reaches its minima. In the Higgs mass peak range, the values show a different pattern, where the sculpting keeps falling after $\lambda = 20$, but considering the negligible values, the error nearly equates to the value itself, so this does not go against the conclusion of an optimum at $\lambda = 20$. The first two rows (NN lep and NN had) show the JSD values obtained when running a NN classifier with no scaling of the photon p_T and E. As expected, this results in large sculpting and therefore large JSD values. A JSD value is accepted as negligible in this study, if JSD $< 10^{-3}$.

Network	JSD	
	(105-160)GeV	(120-130)GeV
Classifier NN lep	$(53.519 \pm 0.104) \times 10^{-2}$	$(71.634 \pm 0.213) \times 10^{-2}$
Classifier NN had	$(58.973 \pm 0.022) \times 10^{-2}$	$(76.345 \pm 0.341) \times 10^{-2}$
ANN lep $\lambda = 20$	**$(1.22 \pm 0.52) \times 10^{-3}$**	**$(0.99 \pm 0.45) \times 10^{-3}$**
ANN had $\lambda = 10$	$(1.02 \pm 0.11) \times 10^{-3}$	$(2.63 \pm 0.52) \times 10^{-3}$
ANN had $\lambda = 15$	$(0.85 \pm 0.11) \times 10^{-3}$	$(1.62 \pm 0.51) \times 10^{-3}$
ANN had $\lambda = 20$	**$(0.68 \pm 0.11) \times 10^{-3}$**	**$(0.73 \pm 0.48) \times 10^{-3}$**
ANN had $\lambda = 25$	$(0.81 \pm 0.11) \times 10^{-3}$	$(0.43 \pm 0.46) \times 10^{-3}$
ANN had $\lambda = 30$	$(1.16 \pm 0.11) \times 10^{-3}$	$(0.43 \pm 0.49) \times 10^{-3}$
Scaled NN lep	**$(0.20 \pm 0.48) \times 10^{-3}$**	**$(0.11 \pm 0.61) \times 10^{-3}$**
Scaled NN had	**$(0.32 \pm 0.36) \times 10^{-3}$**	**$(0.82 \pm 0.92) \times 10^{-3}$**

Table 8.4: Lowest JSD values for the full $M_{\gamma\gamma}$ distribution analysis range $(105 - 160)$ GeV and for the peak part of the distribution, at the mass of the Higgs boson $(120 - 130)$ GeV. See text for further explanations.

8.7 Combined Metric

A simultaneous study of the background rejection and the background sculpting is necessary to determine the optimal choice of the regularization parameter λ. The background rejection is defined as:

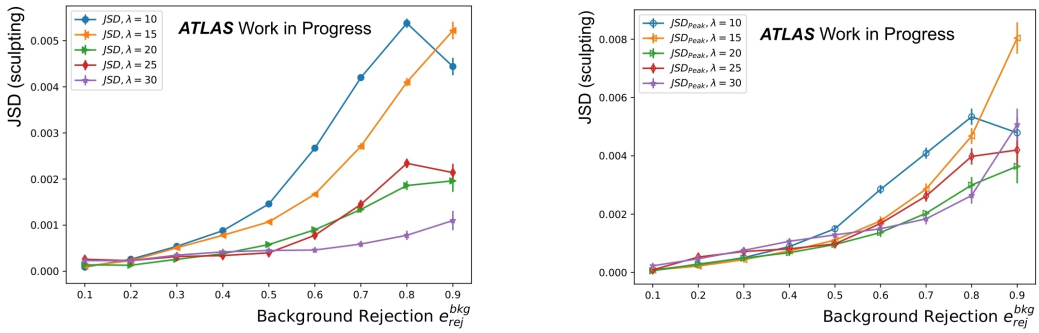
$$e_{rej}^{bkg} = 1 - e_{eff}^{bkg} = 1 - \frac{N_{after}^{bkg}}{N_{before}^{bkg}} \tag{8.3}$$

The ANN efficiency for background events is e_{eff}^{bkg}, the number of background events, before the ANN discriminant cut is N_{before}^{bkg} and the number of background events after the ANN discriminant cut as N_{after}^{bkg}. With the higher background rejection, also the JSD factor rises, or the more background events are rejected,

the higher the final sculpting. The connection between the two is observed to diminish with the raising values of λ (Figure 8.38 and Figure 8.39), but so does the efficiency of the networks in both signal and background, observed in Figures 8.25 and 8.28.

8.7.1 Metric in Simulated MC Events

A good compromise between minimising sculpting and maximising background rejection is again observed for $\lambda = 20$ using MC $tt\gamma\gamma$ background events, as shown in Figure 8.38. Both the full analysis range (105 - 160 GeV) and the Higgs mass peak range (120 - 130 GeV) are used. For this value of λ, any background rejection $e_{rej}^{bkg} < 0.6$ shows sculpting of $JSD < 10^{-3} \approx 0$ in both the full analysis and the Higgs mass peak ranges.



(a) Full analysis range, 105 GeV $< M_{\gamma\gamma} <$ 160 GeV

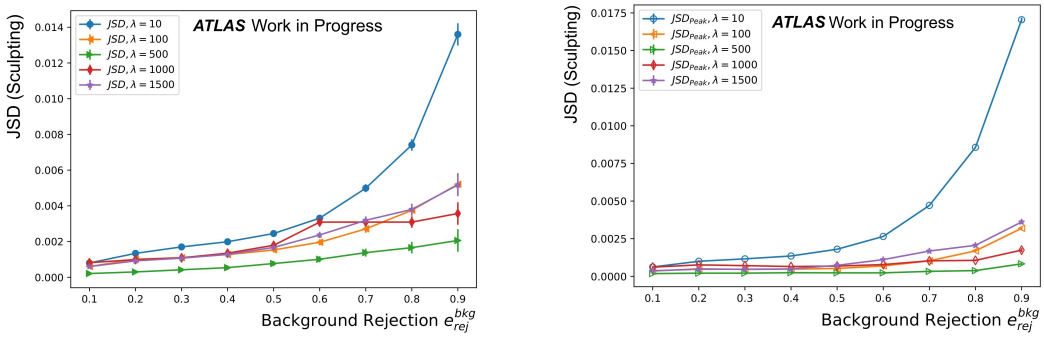
(b) Higgs mass peak range, 120 GeV $< M_{\gamma\gamma} <$ 130 GeV

Figure 8.38: Jenson Shannon Divergence for different ANN background rejection efficiencies in MC simulated data. Lines correspond to $\lambda = 10$, 15, 20, 25 and 30 and points correspond to ANN background rejection efficiency of (90 - 10)% in increments of 10%.

8.7.2 Metric in NTNI Data

NTNI data background (Figure 8.39) shows similar trends as the MC $tt\gamma\gamma$ background. The overall sculpting in the full range, as well as in the Higgs mass peak range, is bigger by ≈ 3 and ≈ 20 times respectively, due to the larger change in slope after the ANN training. Another significant difference is the optimality achieved with $\lambda = 500$ does not only show very high efficiencies of both networks (Table 8.3), but also the absolute optimal case for sculpting minimisation, where increasing λ more only gave higher sculpting with a higher dependence on the background rejection. Just like with $tt\gamma\gamma$ background, any background rejection

$e_{rej}^{bkg} < 0.6$ shows sculpting of $JSD < 10^{-3} \approx 0$ in both the full analysis and the Higgs mass peak ranges.



(a) Full analysis range, 105 GeV $< M_{\gamma\gamma} <$ 160 GeV

(b) Higgs mass peak range, 120 GeV $< M_{\gamma\gamma}$ $<$ 130 GeV

Figure 8.39: Jenson Shannon Divergence for different ANN background rejection efficiencies in NTNI real data. Lines correspond to λ = 10, 100, 500, 1000 and 1500 and points correspond to ANN background rejection efficiency of (90 - 10) % in spaces of 10%.

8.8 Signal Results

The most important performance metric of the classification network is the sensitivity to the signal ttH production. In this section the sensitivity achievable with the adversarial neural networks is presented. The results are compared to the network in which the photon momenta are scaled by $M_{\gamma\gamma}$. This scaled network is shown to be good approximation of the classification used in the ATLAS analyses [43].

To estimate the sensitivity to the ttH production, a cut on the classification discriminant is used to design categories enriched in signal ttH events. Two scenarios are studied:

- *1-category* scenario in which a single signal category is created, using a discriminant cut for which such selection yields the highest sensitivity.
- *2-category* scenario in which two signal categories are selected, such that their combined sensitivity is maximised.

The sensitivity is estimated using Equation 7.1, by counting the numbers of expected signal (S) and background (B) events in the region: 121 GeV$<$ $M_{\gamma\gamma}$ $<$129 GeV. For the 2-category scenario the two optimal boundary positions

are selected by scanning them all, and for each calculating the total sensitivity, which is the sensitivity of the 1st and 2nd categories summed in quadrature.

In these sensitivity estimates, which are used to select the discriminant cuts, only statistical uncertainty is considered. This is well motivated for a ttH, $H \to \gamma\gamma$ analysis, since the statistical uncertainty dominates the measurement precision in Run 2 [40].

The sensitivity estimate requires the knowledge of the absolute numbers of signal and background events, which are obtained by normalizing the corresponding yields to the preliminary integrated luminosity of the Run 2 dataset; $\mathcal{L} = 139$ fb^{-1}. The signal is normalized to the cross-section calculated at NLO QCD and NLO EW accuracy and the $H \to \gamma\gamma$ branching ratio as reported in Section 4.1.

Prior to the classification network discriminant cuts, 26.5 and 8.0 signal events are expected in the hadronic and leptonic decay channels respectively. The background events are normalized such that the expected number of events with 90 GeV$< M_{\gamma\gamma} <$105 GeV matches the corresponding tight isolated data yields. This side-band region used for the normalization is orthogonal to the analysis region. In the analysis the region of 105 GeV$< M_{\gamma\gamma} <$160 GeV is used to determine the signal yield, and in this region 916 and 76 background events are expected in the hadronic and leptonic channel prior to the discriminant cuts.

When selecting the discriminant cut which maximises the sensitivity, an additional requirement that the categories should have at least 20 expected background events is made. This ensures sufficient number of events to fit the background from the $M_{\gamma\gamma}$ spectrum.

Figures 8.40 and 8.41 show the sensitivity as a function of the 1-category discriminant cut as obtained in the leptonic decay channel ($N_{lep} >$0). The 1-category and 2-category discriminant cuts which maximise the sensitivity are denoted by vertical lines. In Figure 8.40 simulated $tt\gamma\gamma$ background events are used, whereas Figure 8.41 uses NTNI data events. The results obtained with the adversarial neural network are compared to the network in which the photon momenta are scaled by $M_{\gamma\gamma}$ (Scaled Network).

The sensitivities obtained in the leptonic decay channel are listed in Table 8.5. In each of the 1-category and 2-category scenarios, the sensitivities obtained with the ANN are comparable with the Scaled Network, with the Scaled Network yielding about 10% higher sensitivity. The 2-category scenario reaches substantially higher sensitivity and is therefore more interesting than the 1-category scenario. Comparing the two background hypotheses, the simulated $tt\gamma\gamma$ and NTNI backgrounds yield similar sensitivities with the Adversarial Network.

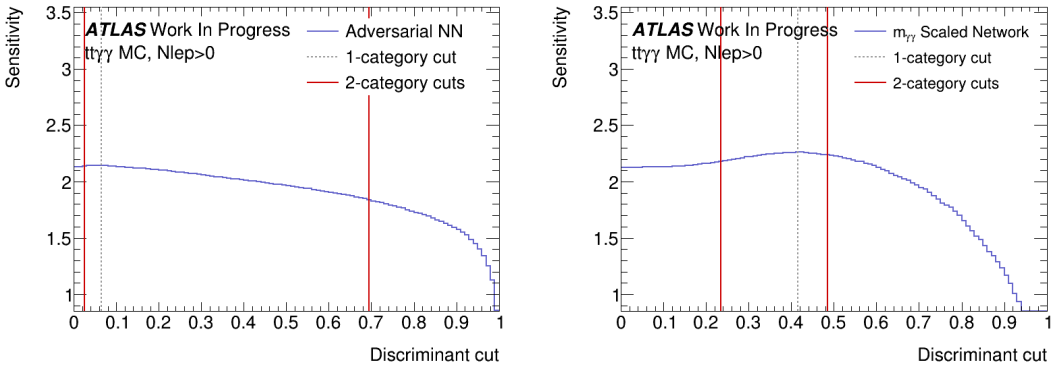


Figure 8.40: Sensitivity (Significance Z) to the ttH production in the leptonic decay channel as a function of the neural network discriminant cut, as obtained with the adversarial neural network (left) and the network in which the photon momenta are scaled by $M_{\gamma\gamma}$ (right). The sensitivity corresponds to creating a single signal-like category. The vertical lines show the discriminant cuts used to create one (dashed) or two (full line) signal-like categories. Simulated $tt\gamma\gamma$ background events are used.

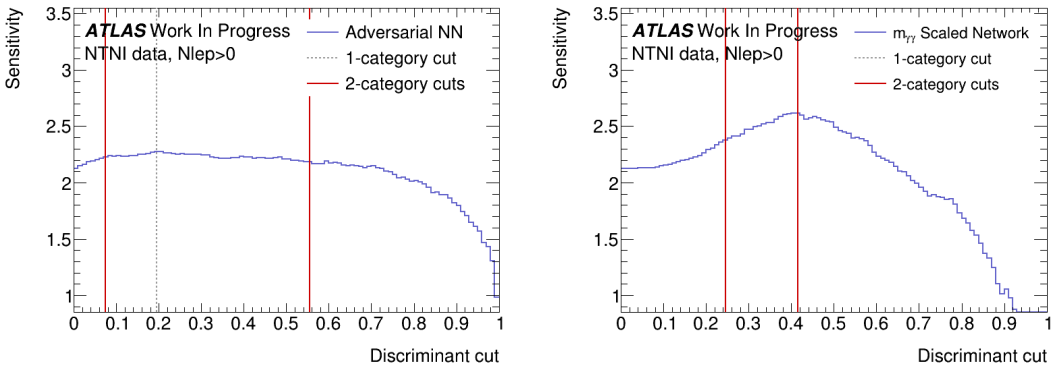


Figure 8.41: Sensitivity to the ttH production in the leptonic decay channel as a function of the neural network discriminant cut as obtained with the adversarial neural network (left) and the network in which the photon momenta are scaled by $M_{\gamma\gamma}$ (right). NTNI data background events are used. In the right figure, the 1-category discriminant cut overlaps with the 2-category cut of 0.415, and its corresponding dashed line is not visible.

As can be seen by comparing Figures 8.40 and Figure 8.41, the $tt\gamma\gamma$ and NTNI data also yield comparable ANN discriminant cut choices; if the 1-category cut optimised for the NTNI data was applied to the $tt\gamma\gamma$ background hypothesis, the sensitivity would only decrease by 1.8%. The Adversarial Network classification is therefore stable to the variations in the assumed background shape, which is important, because the background shape is not known precisely. Comparing

the two background hypotheses, the simulated $tt\gamma\gamma$ and NTNI backgrounds yield similar sensitivities with the Adversarial Network. The ratio of sensitivities is shown in Figure 8.42, the largest difference across the full discriminant range is about 14%.

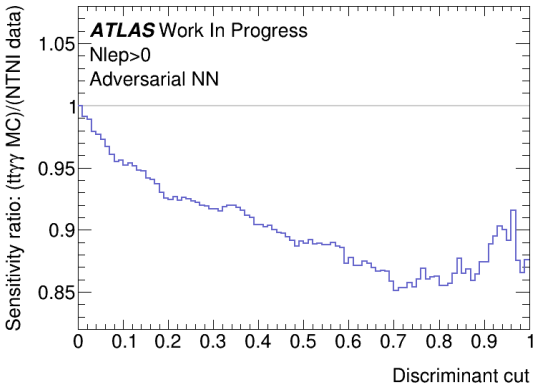


Figure 8.42: Sensitivity (Significance Z) ratio of $tt\gamma\gamma$ MC over the sensitivity of NTNI data in dependence of the discriminant cut.

	Z(1-category)	**Z(2-category)** {Z(low-D.),Z(high-D.)}
	Background: simulated $tt\gamma\gamma$:	
Adversarial Network	2.1	**2.3** {1.3,1.8}
$M_{\gamma\gamma}$ Scaled Network	2.3	**2.5** {1.0,2.2}
	Background: non-tight or non-isolated (NTNI) data:	
Adversarial Network	2.3	**2.5** {1.1,2.2}
$M_{\gamma\gamma}$ Scaled Network	2.6	**2.7** {0.8,2.6}

Table 8.5: Sensitivity (Significance Z) in the leptonic decay channel obtained with a single signal category (Z(1-D)) and two signal categories Z(2-D). In the 2-D case the total sensitivity, as well as the sensitivity of the individual categories, corresponding to low and high discriminant cuts {low-D.,high-D.}, is listed. Results are shown for two background hypotheses: simulated $tt\gamma\gamma$ and NTNI data events.

Figure 8.43 shows the sensitivity obtained in the hadronic decay channel, using simulated $tt\gamma\gamma$ background events. The obtained sensitivities are listed in Table 8.6. Similar to the leptonic decay channel, the Scaled Network yields about 10% higher sensitivity to the ttH production compared to the Adversarial Network.

Combining the lepton and hadron decay channel results by adding them in quadrature, the classification performed with the Adversarial network yields a sensitivity of $\mathbf{Z = 3.6}$, while the Scaled network yields a sensitivity of Z=4.1 assuming $tt\gamma\gamma$ background events. This is comparable to the published ATLAS

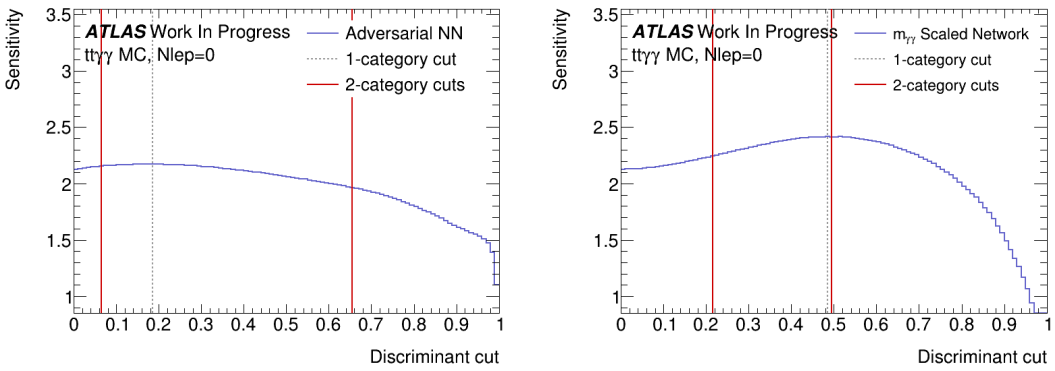


Figure 8.43: Sensitivity (Significance Z) to the ttH production in the hadronic decay channel as a function of the neural network discriminant cut, as obtained with the adversarial neural network (left) and the network in which the photon momenta are scaled by $M_{\gamma\gamma}$ (right). Simulated $tt\gamma\gamma$ background events are used.

	Z(1-category)	**Z(2-category)** {Z(low-D.),Z(high-D.)}]
Adversarial Network	2.3	**2.7** {2.0,1.8}
$M_{\gamma\gamma}$ Scaled Network	2.7	**3.2** {2.1,2.5}

Table 8.6: Sensitivity (Significance Z) in the hadronic decay channel obtained with a single signal category (Z(1-D)) and two signal categories Z(2-D). In the 2-category case the total sensitivity, as well as the sensitivity of the individual categories, corresponding to low and high discriminant cuts {low-D.,high-D.}, is listed.

ttH,$H \to \gamma\gamma$ analyses [43] with a sensitivity of Z = 4.4. The overall sensitivity obtained when applying the Adversarial Neural Network method is lower than the 5.0 σ sensitivity reached in the state of the art ATLAS analysis in [43]. This is due to several factors, such as more extensive categorisation of 11 categories, as well as feature engineering [148]. This is beyond the scope of this thesis, which aims to establish a proof of principle that ANNs can be used to handle the $M_{\gamma\gamma}$ sculpting. It is advisable for any future work, aiming to improve the sensitivity achievable with an ANN, to include a full hyperparameter optimisation of the networks by looking at the broader parameter hyperspace. In particular, the ANN Units and Architecture discussed in Section 8.1 could be optimised further.

In summary the results of this Section establish that the Adversarial network achieves an overall comparable sensitivity to the Scaled network. The Scaled network sensitivity is about 10% higher, and this difference is likely to be reduced after hyperparameter optimization. The sensitivity calculation described so far only accounted for the statistical uncertainty. While this uncertainty dominates the Run 2 ttH, $H \to \gamma\gamma$ measurement uncertainty, it is important to consider

whether the use of the novel adversarial neural network may exacerbate the background modelling uncertainty. To evaluate this, the Spurious Signal has to be evaluated.

8.9 Signal modelling

The fit of the DCSB to the ttH,$H \to \gamma\gamma$ signal events is shown in Figure 8.44.

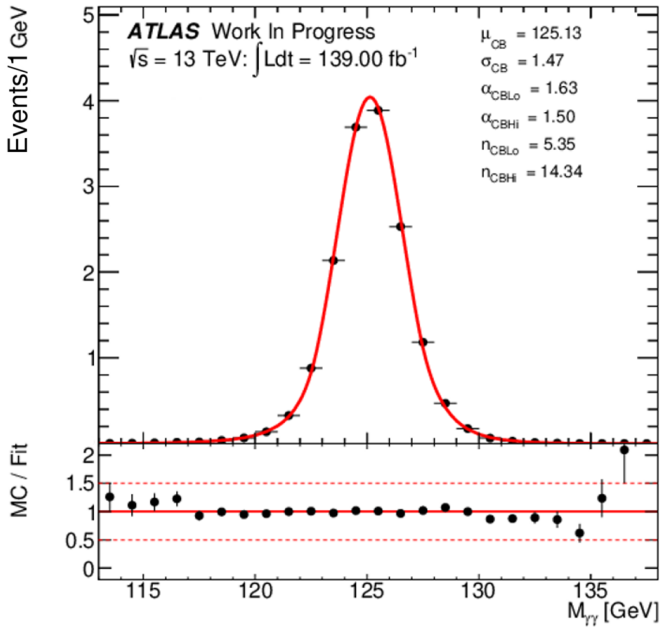


Figure 8.44: Double-sided Crystal Ball function fit to the ttH,$H \to \gamma\gamma$ signal events. The fitted parameters are described in Equation 7.2.

The fitted parameter values are kept fixed during the signal+background fits used to extract the spurious signal, and the same parameter values will be used in all categories. The reason for this is that the fits in the leptonic and hadronic decay channel yield very similar parameter values. Likewise, there are only small differences between the events passing high or low discriminant cuts of the classification networks. Therefore, the DSCB parameters are extracted only once, using all ttH events having at least three jets. The signal yield is not fitted as it is determined in each fitted category separately from the yields of the simulated ttH events normalized to the theory predictions detailed in Section 4.1.

8.10 Spurious Signal Evaluation

8.10.1 Spurious Signal Fit

Spurious signal (SS) is the bias in the signal yield, or the number of events, which can be classified as fake signal. For this, the background template is fitted with a signal+background model to determine both the signal and background event yields. The potential bias from differences between the actual background distribution and the fitted one is estimated from the fitted signal yield (spurious signal). The SS is performed on templates with a 1 GeV step in the $M_{\gamma\gamma}$ range of 120 GeV - 130 GeV in order to avoid accidentally small bias values at the nominal Higgs boson mass. The fitted number of spurious signal events is allowed to be either of positive or negative.

Categories pass the spurious signal test in the general analysis described in [43] if, they have a minimum of 100 events in their sidebands and $|N_{sp}|$ satisfies the following criteria:

- $|N_{sp}|$ is smaller than 10% of the expected number of total Higgs boson signal events
- $|N_{sp}|$ is smaller than 20 % of the statistical uncertainty of the fitted signal yield, σ_{exp}

If more than one analytic function passes the spurious signal test, then the function with the fewest parameters is selected.

The spurious signal fit used in this thesis closely follows the conventions used in ATLAS $H \to \gamma\gamma$ analyses [43] and is performed with the software developed by the ATLAS $H \to \gamma\gamma$ group. The signal + background model is fitted to background-only events in the range of $105 \leq M_{\gamma\gamma} \leq 160$ GeV. The number of fitted signal events as a function of the Higgs mass is computed in the range of $120 \leq M_{\gamma\gamma} \leq 130$ GeV in steps of 0.5 GeV. The number of spurious signal events N_{sp} corresponds to the maximum of the absolute value of the fitted number of signal events. In this thesis, the following criteria are required for the functional form to pass the spurious signal (SS) test:

- The goodness of fit: the χ^2 per number of degrees of freedom is required to be consistent with 1 with a probability higher than 1%.
- The number of spurious signal events should be less than 50% of the expected signal statistical uncertainty $\left(\frac{N_{sp}}{\Delta S} < 50\%\right)$.

The motivation for choosing 50% is that at this value, the spurious signal would be the dominant systematic uncertainty of the ttH measurement, and it would substantially degrade the analysis sensitivity. If the test passes at this value, it means, it passes at all expected signal thresholds.

Out of all functions passing the required criteria, the best function to use in the fit is selected as follows:

- If several functions pass the spurious signal test, the function with the lowest number of parameters is selected.
- In case of several functions with the same number of parameters, the one with the lowest $|N_{sp}|$ is chosen.

The reason for the preference of the low number of parameters is that most of the categories in this analysis contain a low number (<50) of expected background events. A fit of a function with many parameters to such low number of events would result in large uncertainties on the fitted parameters.

These criteria are somewhat different from the spurious signal criteria used in ATLAS $H \to \gamma\gamma$ analyses [43], which are intended for all Higgs boson production modes. In contrast to the ttH production, production modes such as the ggF use categories with high numbers of background events and the spurious signal dominates the measurement uncertainty. The spurious signal criteria for such production modes therefore allow for higher-order functions if this lowers the spurious signal yields. The analysis [43], attempts to construct uniform SS criteria, usable across all production modes and 100+ categories. These criteria are overly complicated and overly strict for the stat-dominated ttH channel.

8.10.2 Spurious Signal in the Leptonic Decay Channel

In the ttH final states with at least one lepton, categorising the events into two signal categories results in the sensitivity of Z=2.3 for the Adversarial network and Z=2.5 for the Scaled network, using $tt\gamma\gamma$ background events. The corresponding discriminant (D) cuts and background yields (N_b) within $105 \leq M_{\gamma\gamma} \leq 160$ GeV are:

Adversarial network:

- high-cut category: $D > 0.695$, N_b=24.7
- low-cut category: $0.025 < D < 0.695$, N_b=50.0

Scaled network:

- high-cut category: $D > 0.49$, N_b=24.7
- low-cut category: $0.24 < D < 0.49$, N_b=44.9

The $M_{\gamma\gamma}$ background spectra of these two categories are shown in Figure 8.45.

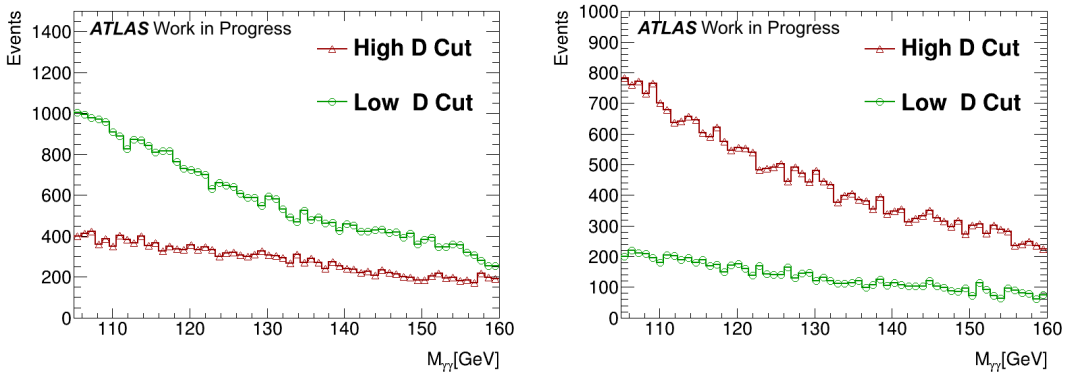


Figure 8.45: Distribution of $M_{\gamma\gamma}$ leptonic background for the Adversarial (left) vs distribution of $M_{\gamma\gamma}$ leptonic Scaled (right) NN scenarios in their best low discriminant (green) and high discriminant (red) cut categories.

Tables 8.7 and 8.8 show the results of the spurious signal fits in the high-cut and low-cut categories respectively. the SS fit allows both the lack and excess of signal, as both would bias the signal estimate. This leads to both negative (-) and positive (+) values for the spurious signal. The fitted function name and number of the parameters is shown, alongside the number of spurious signal events (N_{sp}), fraction of the signal uncertainty $\frac{N_{sp}}{\Delta S}$, the χ^2 per number of degrees of freedom, and the result of passing the spurious signal test requirements described in the previous section (result=pass or fail). For the function selected by the spurious signal procedure, the values are listed in **bold**. The information is shown for each of the Adversarial and Scaled network, and the last column provides a comparison of their performance. The listed value is the relative difference between the number of the spurious signal events obtained by the Adversarial network and the Scaled network, defined as:

$$\text{ANN} - \text{Scaled} = \frac{|N_{sp}(\text{Adversarial})| - |N_{sp}(\text{Scaled})|}{0.5(|N_{sp}(\text{Adversarial})| + |N_{sp}(\text{Scaled})|)}. \tag{8.4}$$

The trend to be mindful of is whether the value of $(\text{ANN} - \text{Scaled}) > 0$ is obtained frequently. This would mean that the Adversarial network exacerbates the spurious signal compared to the Scaled network.

		Adversarial Network				Scaled Network				
Name	Npar	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	diff [%]
Pow	1	1.08	21.5	1.86	fail	**-0.84**	**-15.2**	**0.97**	**pass**	24
Expo	1	**0.74**	**14.7**	**1.43**	**pass**	-1.41	-25.4	1.11	pass	-62
ExpPoly2	2	0.21	3.97	1.23	pass	-1.13	-19.2	1.0	pass	-136
ExpPoly3	3	-0.43	-7.56	1.08	pass	-1.2	-19.1	1.03	pass	-94
Pow2	3	0.32	6.3	1.36	pass	-1.03	-18.0	1.05	pass	-104
Bern3	3	-0.68	-11.9	0.96	pass	-1.08	-17.4	1.01	pass	-45
Bern4	4	-0.67	-11.8	0.76	pass	-1.17	-18.6	1.01	pass	-54
Bern5	5	-0.83	-13.5	0.79	pass	-1.79	-26.9	0.97	pass	-72

Table 8.7: Results of the spurious signal test for the leptonic high-cut category using $tt\gamma\gamma$ background events. The results of the functions selected by the spurious signal procedure are highlighted in bold. The columns are described in the text, where diff = ANN - Scaled (see equation 8.4). Explanation of functions can be found in Chapter 7.4.2

From the high-cut category spurious signal test shown Table 8.7, we see that a first-order function is selected for each of the Adversarial network (Exponential) and the Scaled network (Power Law). In the case of the Adversarial network, the power law fails the test due to a poor fit, which has a χ^2 probability of only 0.4%. The test results in two important positive outcomes:

- For the same functional form, the Adversarial network typically yields less spurious signal. With Expo fitted, the number of spurious events N_{sp} is: 0.74 (ANN) and 1.41 (Scaled), with ExpPoly2 = 0.21 (ANN) and 1.13 (Scaled) etc.

- For all functional forms, the spurious signal uncertainty is much smaller than the statistical uncertainty on the number of signal events. With Expo fitted, the spurious signal uncertainty $\frac{N_{sp}}{\Delta S}$ = 14.7% (ANN) and 25.4% (Scaled), with ExpoPoly2, it is = 3.97% (ANN) and 19.2% (Scaled) etc.

Table 8.8 shows the test results in the low-cut category. The Exponential function is selected for the Adversarial network and the Power Law for the Scaled network. All functions pass the test and yield small spurious signals.

Similar conclusions as for the 2-category classification with $tt\gamma\gamma$ background hold for the 1-category classification and also in case of the NTNI data background. We therefore conclude that the performance of the Adversarial network is satisfactory and comparable to that of the Scaled network in the leptonic decay channel.

		Adversarial Network				Scaled Network				
Name	Npar	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	diff [%]
Pow	1	0.89	15.4	1.66	pass	**0.39**	**7.26**	**1.08**	**pass**	77
Expo	1	**-0.29**	**5.05**	**0.83**	**pass**	-0.68	-12.1	1.04	pass	-79
ExpPoly2	2	-0.57	-9.04	0.78	pass	-0.47	-7.91	0.98	pass	19
ExpPoly3	3	-0.83	-12.3	0.8	pass	-0.52	-8.21	1.01	pass	46
Pow2	3	-0.84	-12.1	0.77	pass	0.47	9.84	1.01	pass	56
Bern3	3	-0.69	-10.2	0.81	pass	-0.4	-6.39	0.97	pass	52
Bern4	4	-0.81	-12.0	0.68	pass	-0.53	-8.29	0.9	pass	42
Bern5	5	-0.92	-12.9	0.71	pass	-0.79	-11.7	0.89	pass	15

Table 8.8: Results of the spurious signal fit for the leptonic low-cut category using $tt\gamma\gamma$ background events. diff = ANN-Scaled (see equation 8.4). Explanation of functions can be found in 7.4.2.

8.10.3 Spurious Signal in the Hadronic Decay Channel

In the hadronic decay channel, the spurious signal test turns out to be very challenging for the categories obtained with the Adversarial network, especially for the high-cut category in the categorization with two signal categories. The discriminant (D) cuts and background yields (N_b) within $105 \leq M_{\gamma\gamma} \leq 160$ GeV are:
Adversarial network:

- high-cut category: $D > 0.955$, N_b=28.9
- low-cut category: $0.115 < D < 0.955$, N_b=787

Scaled network:

- high-cut category: $D > 0.795$, N_b=28.7
- low-cut category: $0.415 < D < 0.795$, N_b=407

The corresponding $M_{\gamma\gamma}$ distributions are shown in Figure 8.46. It is evident that the high-cut category spectrum obtained with the Adversarial network can only be fitted with a higher-order function.

Tables 8.9 and 8.10 show the results of the spurious signal test in the high-cut and low-cut categories respectively.

The results for the high-cut category in Table 8.9 confirm what is expected from Figure 8.46: when the Adversarial network is used, only 5th order Bernstein Polynomial passes the spurious signal test. All other functions fail the goodness of fit requirement (χ^2 probability $> 1\%$) and some also yield $\frac{N_{sp}}{\Delta S} > 50\%$. The Scaled network shows superior performance, with all functions passing the test and a Power Law function selected.

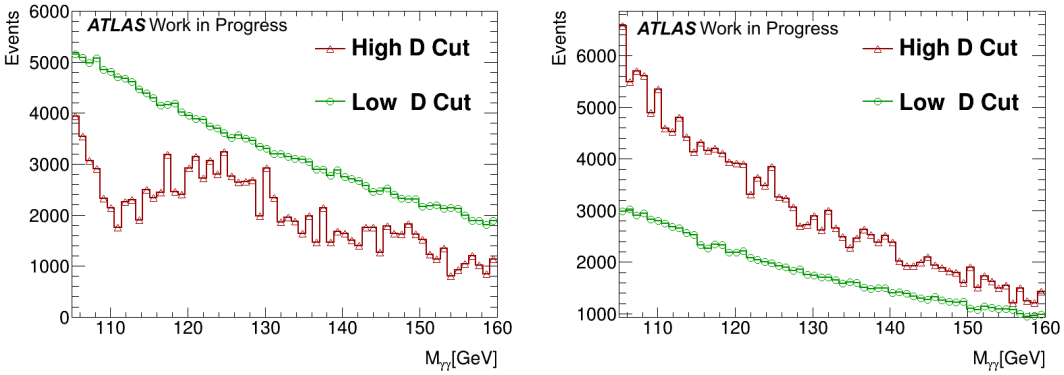


Figure 8.46: Distribution of $M_{\gamma\gamma}$ leptonic background for the Adversarial (left) vs distribution of $M_{\gamma\gamma}$ leptonic Scaled (right) scenarios in their best low (green) and high (red) cut categories. In both figures, the high-cut category yield has been scaled up by a factor of 20 for visibility.

		Adversarial Network				Scaled Network				
Name	Npar	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	diff [%]
Pow	1	3.83	56.8	9.85	fail	**0.33**	**6.68**	**0.78**	**pass**	168
Expo	1	3.23	47.8	7.96	fail	-0.59	-11.5	1.1	pass	137
ExpPoly2	2	2.81	38.9	7.27	fail	-0.23	-4.25	0.77	pass	169
ExpPoly3	3	2.86	38.9	7.54	fail	0.23	4.17	0.75	pass	170
Pow2	3	3.83	70.8	10.6	fail	0.22	4.03	0.79	pass	178
Bern3	3	2.64	35.6	7.27	fail	0.0	0.0	0.81	pass	200
Bern4	4	2.06	27.4	4.87	fail	0.3	5.39	0.74	pass	149
Bern5	5	**0.17**	**2.15**	**1.0**	**pass**	0.39	6.29	0.79	pass	-75

Table 8.9: Spurious signal test for the hadronic high-cut category.The results of the functions selected by the spurious signal procedure are highlighted in bold. ANN requires a 5th order function in the high-cut category, therefore, it is less performant, than the scaled in this case. Explanation of functions can be found in 7.4.2. diff = ANN-Scaled (see equation 8.4).

		Adversarial Network				Scaled Network				
Name	Npar	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	diff [%]
Pow	1	2.92	28.5	4.94	fail	1.73	16.4	2.56	fail	51
Expo	1	**1.14**	**11.1**	**1.11**	**pass**	**-0.93**	**-8.61**	**0.84**	**pass**	19
ExpPoly2	2	-0.42	-3.8	0.65	pass	-0.78	-6.76	0.81	pass	-59
ExpPoly3	3	0.27	2.37	0.59	pass	-0.47	-3.91	0.79	pass	-54
Pow2	3	2.41	23.6	3.81	fail	-0.89	-7.56	1.0	pass	92
Bern3	3	0.37	3.24	0.59	pass	-0.43	-3.55	0.74	pass	-14
Bern4	4	-0.59	-5.03	0.64	pass	-0.35	-2.86	0.76	pass	51
Bern5	5	0.55	4.38	0.65	pass	0.36	2.76	0.82	pass	42

Table 8.10: Spurious signal test for the hadronic low-cut category. Explanation of functions can be found in 7.4.2. diff = ANN-Scaled (see equation 8.4). The ANN shows comparable performance to the Scaled.

The results for the low-cut category in Table 8.10 show that in this category the Adversarial network and the Scaled network perform comparably, with Exponential function being selected in the test.

Overall, the Adversarial network is not adequate for the 2-category classification in the hadronic decay channel, since it requires a 5th order function in the high-cut category. As this category only contains 28.9 expected background events, such function would not be fitted accurately in the data. Given the failure of the 2-category classification, Table 8.11 shows the spurious signal test for the 1-category classification in the hadronic decay channel. The 1-category discriminant cut and the expected number of background events are:

- Adversarial network: $D > 0.215$, N_b=693
- Scaled network: $D > 0.575$, N_b=189

		Adversarial Network				Scaled Network				
Name	Npar	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	N_{sp}	$\frac{N_{sp}}{\Delta S}$[%]	χ^2	result	ANN-Scaled [%]
Pow	1	3.32	34.8	5.82	fail	1.63	16.4	2.22	fail	68
Expo	1	1.85	19.4	2.1	fail	**-0.72**	**-7.05**	**0.85**	**pass**	87
ExpPoly2	2	**0.75**	**7.37**	**1.29**	**pass**	-0.46	-4.18	0.8	pass	49
ExpPoly3	3	1.04	9.99	1.26	pass	-0.34	-2.99	0.76	pass	101
Pow2	3	1.32	13.2	1.36	pass	1.45	29.0	0.78	pass	-9
Bern3	3	1.04	9.84	1.3	pass	0.48	4.32	0.71	pass	73
Bern4	4	0.74	6.95	1.12	pass	0.41	3.69	0.75	pass	57
Bern5	5	-0.52	-4.36	0.81	pass	0.74	6.03	0.83	pass	-35

Table 8.11: Spurious signal test for the 1-category classification in the hadronic decay channel. Explanation of functions can be found in Section 7.4.2

For the 1-category classification the Adversarial network performance is found to be adequate. While the 1-parameter functions fail the spurious signal test, the functions with two or more parameters pass the test. Such functions could be fitted accurately, as the category has 693 expected background events. The spurious signals yields obtained with the Adversarial network are comparable to the ones obtained with the Scaled network, and much smaller than the statistical uncertainty on the signal.

8.10.4 Conclusions of the Spurious Signal test

In this Section it has been demonstrated that the performance of the Adversarial network for the spurious signal test matches that of the Scaled network in the leptonic decay channel.

In the hadronic decay channel, the Adversarial network also performed comparably well to the Scaled network, except for the categories with high ANN discriminant cuts. In these categories the background $M_{\gamma\gamma}$ spectrum was sculpted. It could therefore only be fitted with functions with many (5 or more) free parameters. However, these categories are expected to contain only about 30 background events, which is too few to constrain many parameters in a fit. Thus, categories with high ANN discriminant cuts could not be used. The Scaled network required no such restrictions on high discriminant categories. Due to this, the hadronic channel sensitivity achievable with the Scaled network (Z=3.2) exceeded the sensitivity achievable with the Adversarial network (Z=2.3).

What are the possible ways to improve the Adversarial network performance in the hadronic decay channel? The Adversarial network is a system of two networks (classifier and adversary), and therefore has more degrees of freedom (node weights) compared to the Scaled network (classifier). This larger number of degrees of freedom implies the need for more training data. A study with about 5 times more training data than available for this thesis was conducted in Ref. [149]. The additional training data was obtained from two sources: (1) larger signal and background samples, (2) a larger fraction of data allocated to the training sample, and a smaller fraction to the validation and prediction ones. In addition, Ref. [149] used a smaller number of classifier nodes and Gaussian Model Mixture components. With more training data and fewer degrees of freedom, the sculpting in the high ANN discriminant categories was alleviated. To match the training sample statistics of Ref. [149], ATLAS should produce about 12 million simulated signal ttH and 12M background $tt\gamma\gamma$ events. This is a factor of about a factor 2.5 more compared to the samples available to Run 2 analyses.

The Adversarial network setup developed in this thesis is thus applicable to the ttH, $H \to \gamma\gamma$ classification in most categories. Because of the sculpting in the high ANN discriminant categories of the hadronic decay channel, the sensitivity to ttH production achievable with the Adversarial network classification is lower compared to the Scaled network. Combining the leptonic and hadronic decay channels, the sensitivity with the Adversarial network is $\mathbf{Z}$=3.3, compared to the Scaled network $\mathbf{Z}$=4.1. In future ATLAS analyses, the performance of the Adversarial network could be improved by larger simulated samples and hyperparameter optimization.

Chapter 9

Conclusion and Outlook

The work performed for this thesis consists of two parts. The first part was key to the development of the new generation of AtlFast, the fast simulation of the ATLAS calorimeter. AtlFast uses a parametrised detector response, and the samples for this parametrisation can only be simulated for discrete values of particle energies. A technique which interpolates the detector response between these simulated energy points, has been developed, and demonstrated that this interpolation is accurate. The interpolation enables the simulation of particles with any energy in AtlFast. It is used by the collaboration in AtlFast3, which significantly improves between the fast simulation with the full ATLAS simulation which is now excellent.

In the second and main part of this thesis the use of adversarial neural networks for the classification of $ttH(H\gamma\gamma)$ events was investigated. The goal of the approach is to prevent sculpting of the background $M_{\gamma\gamma}$ distribution after background rejection, while retaining high background rejection and signal selection efficiencies. The new adversarial technique developed in this work is proven to perform comparably to the classification techniques currently used by ATLAS, with the benefit that the background sculpting was automatically handled by the adversary network, and required no approximate, manual approach. However, the sensitivity to ttH production achievable with the Adversarial network classification ($\mathbf{Z}$=3.3) is still lower compared to the ATLAS classification techniques ($\mathbf{Z}$=4.1). This is primarily due to the sculpting observed in high ANN discriminant categories of the hadronic decay channel. Means to improve the adversarial networks were identified based on Ref. [149]: more training data and hyper-parameter optimisation. If ATLAS produced at least 12 million simulated ttH signal and 12 million $tt\gamma\gamma$ background events, the sensitivity could be further enhanced.

For the imminent ATLAS Run3 analyses, AtlFast3 will be the baseline simulator, which was made possible by the improvements described in the first part of the thesis. AtlFast3 will speed-up the detector simulation by about a factor of five compared to the full simulation with Geant4 [1]. With this speed-up, it could be feasible to produce 2.5 times larger ttH signal and ttyy background samples, as required for the adversarial neural network training.

At the future HL-LHC the $ttH(H\gamma\gamma)$ measurements are projected to have a statistical uncertainty of 4.2% and a systematic uncertainty of 4.0% [47]. Hence, reduction of the systematic uncertainty would substantially enhance the measurement sensitivity. In this scenario, adversarial networks could be used to provide a classification discriminants independent of $M_{\gamma\gamma}$, and robust to systematic uncertainties. A proof-of-principle of such uncertainty-aware classification has been provided by phenomenology studies, including Ref [150]. A substantial speed up of the simulated sample production is planned for the HL-LHC, and is projected to enable a production of much larger simulated signal and background samples. With these, there are several directions in which the adversarial neural networks may improve the current ATLAS classification solutions.

Appendices

Appendix A

Adversarial Neural Networks Structure

The number of layers, nodes and and the full architecture of the ANNs is shown on Figures A.1, A.2 and A.3.

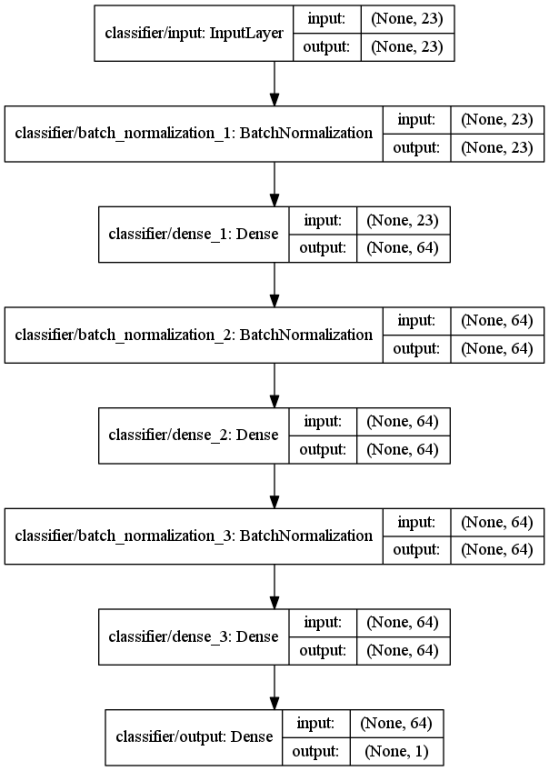


Figure A.1: Classifier Neural Network set-up, described in Section 8.1. The layer dimensions are displayed in the format: (batch size, number of nodes), where the batch size is a hyper-parameter, and is therefore listed as 'None'. The classifier's InputLayer takes 23 features, and feeds them to three hidden layers. Each of these hidden layers uses batch normalisation. The classifier outputs the probability that the even is signal-like.

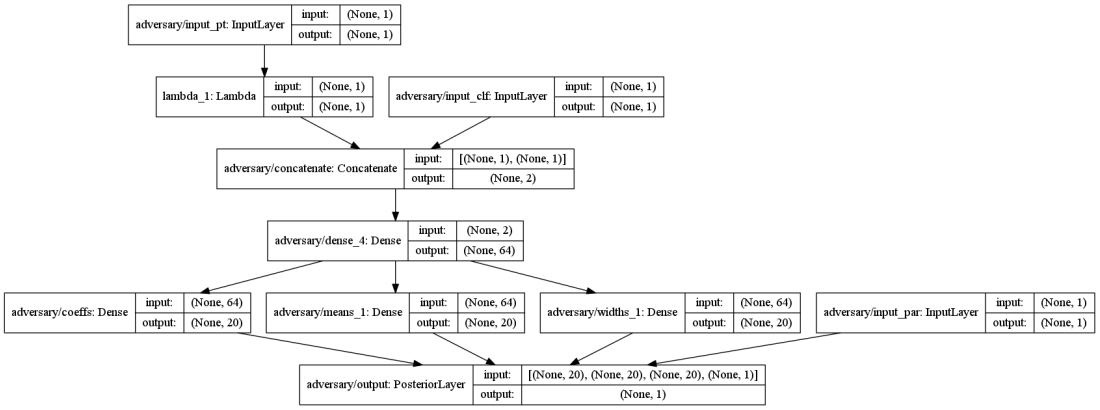


Figure A.2: Adversary Neural Network set-up, described in Section 8.1. The layer dimensions are displayed in the format: (batch size, number of nodes), where the batch size is a hyper-parameter, and is therefore listed as 'None'. The adversary concatenates two inputs: the auxiliary variable and the classifier output. It feeds these to one hidden layer, and subsequently to a Gaussian Mixture Model of 20 Gaussian distributions. The adversary determines the means, widths, and normalisation coefficients of these Gaussians, and compares the Gaussian Mixture Model distribution with the actual $M_{\gamma\gamma}$ distribution.

Appendix B

ROC curves

A receiver operating characteristic curve (ROC) is a graphical illustration of the performance ability of a machine learning tool to diagnose in a certain task, as its discrimination threshold is varied. The curve consists of a plot of the true positive rate (TP, the sensitivity of detection) vs. the false positive (FP, probability or false alarm) of the output of a machine learning algorithm. It can also be seen as a graphics, which represents the Type 1 Error in statistics. Type 1 Error is the mistaken rejection of a null hypothesis (Figure B.1) [151]. Type 2 Error is the false negative probability (FN). In every ROC plot, the diagonal line from the left bottom of the plot to the right top, represents the guessing line, which means beyond that line, all values obtained as output are random guessing and the machine learning algorithm has not worked at all. The rule for overall performance rates is to stay as far from the guessing line corresponding to 50% as possible. A ROC area of $\approx 70\%$ is usually considered acceptable.

In mathematical terms, the ROC curves plots $TP(D) = \int_D^{\inf} f_0(x)dx$ versus $FP(D) = \int_D^{\inf} f_1(x)dx$, where D is the discriminant and $f(x)$ are the probability density functions.

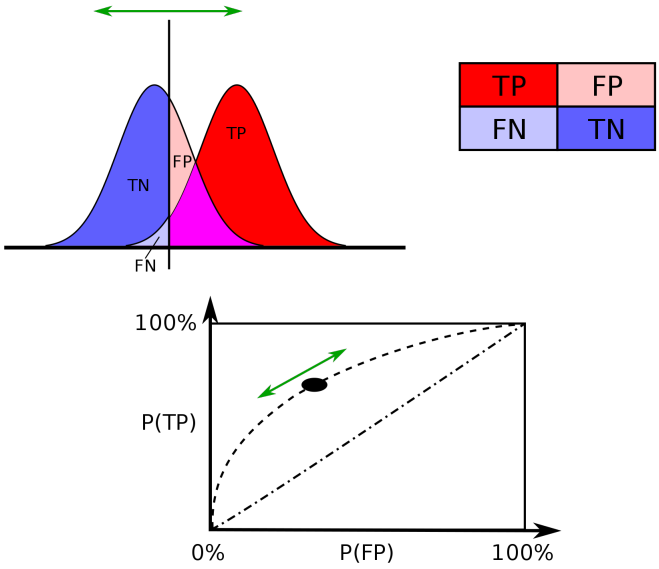


Figure B.1: An illustration of the true positive (TP), false positive (FP), false negative (FN), true negative (TN) at top and a ROC curve at the bottom of the plot. The green arrow shows the possibility for varying the discriminant, the two bell curves represent the two possible hypothesis.

Appendix C

Manual Background Modelling

Four functions were considered for the modelling of the final $M_{\gamma\gamma}$ distribution after ANN training (Figure C.1): linear, Chebyshev first order exponential and second order exponential, explained in Section 7.4.2. Those functions have been used in the analysis categories, due to their simplicity, which does not lead to complications with the spurious signal.

C.0.1 Simulated MC Events

The best fitting one was found to be the first order exponential, which gave promising values of χ^2/ndf for $\lambda = 20$. Further background fitting was performed after spurious signal studies, which can be found in Section **??**.

C.0.2 Real Run 2 Data Events

The NTNI $M_{\gamma\gamma}$ distribution before ANN training already appeared more complex than the MC $tt\gamma\gamma$, which could also explain the need for an order of magnitude bigger value for the parameter, which controls the adversary's loss: λ (Figure C.2). With manual optimisation, from the functions used, the exponential was the closest match, which is why it was used for a preliminary comparison between the initial and final distributions, but is not the function, which was used for the final modelling of the background. In the optimal case of $\lambda = 500$, the sculpting is fully minimised and the final χ^2 shows good approximation to the original value, which corresponds to no additional complexity added by the ANN training.

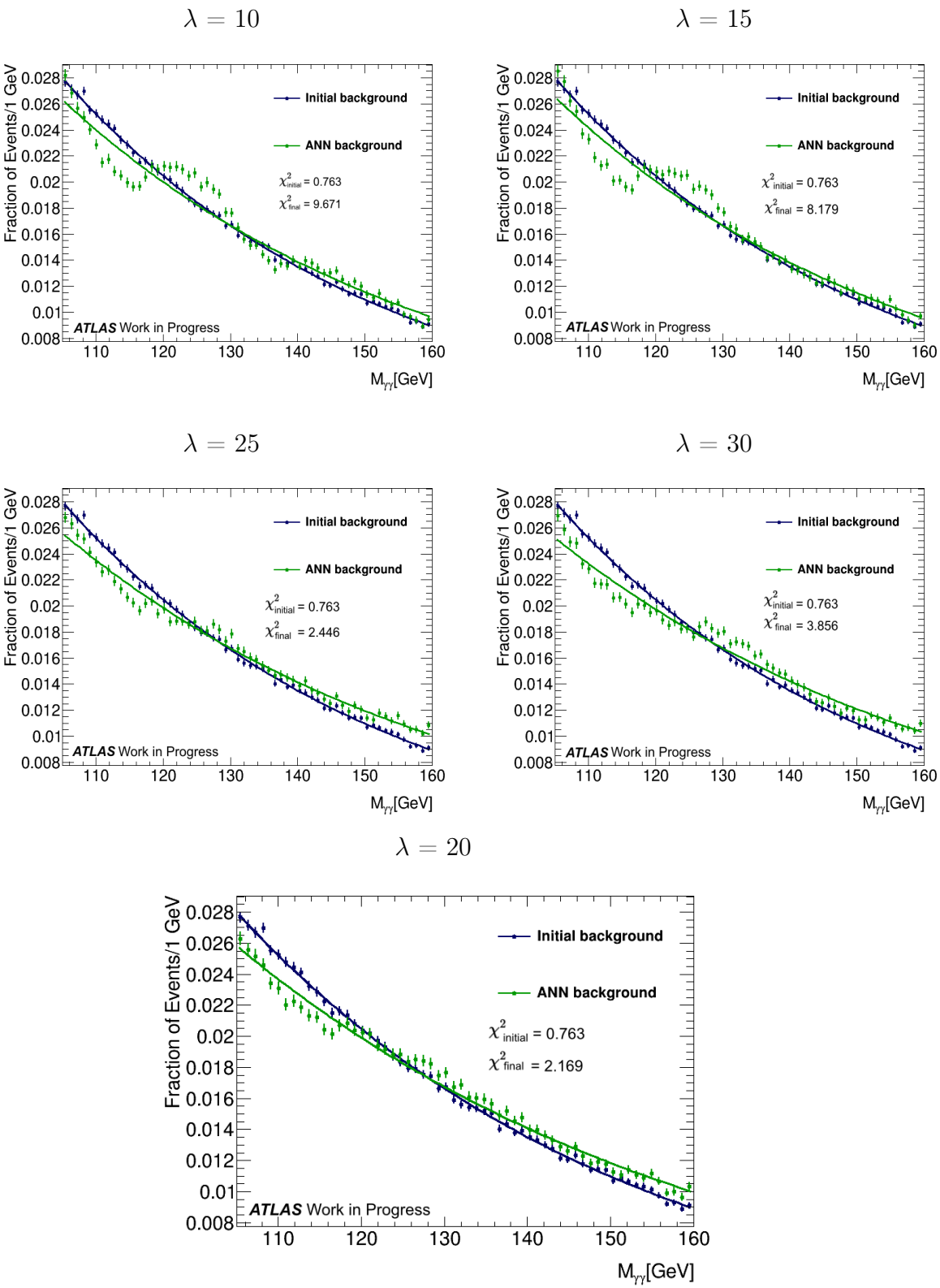


Figure C.1: Manual modelling of the $tt\gamma\gamma$ MC background distribution before ANN training (blue) and after (green) with a first-order exponential function for λ = 10, 15, 20, 25 and 30. where ndf is the number of degrees of freedom. Each χ^2 value corresponds to the fraction $\frac{\chi^2}{ndf}$, where ndf is the number of degrees of freedom.

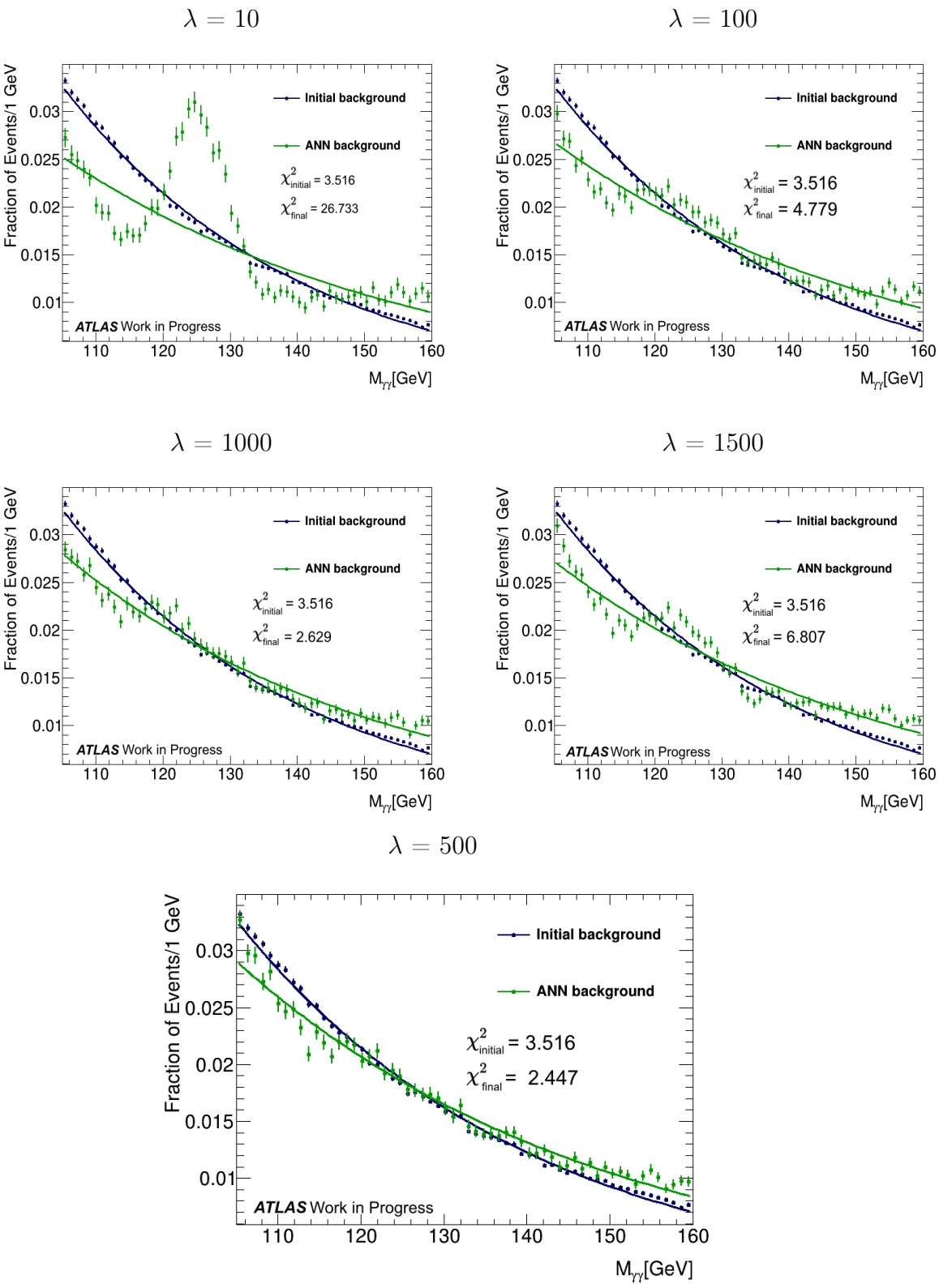


Figure C.2: Manual modelling of the NTNI real data background distribution before ANN training (blue) and after (green) with a first-order exponential function for $\lambda = 10$, 100, 500, 1000 and 1500. where ndf is the number of degrees of freedom. Each χ^2 value corresponds to the fraction $\frac{\chi^2}{ndf}$, where ndf is the number of degrees of freedom.

Appendix D

Simulation Profiling

CPU time reduction is the main reason for using fast simulations. Sometimes a part of the software can do the job but take a very long time, so optimization of code is a crucial part of particle physics simulations. Three profilers were used to study the performance of the StepinfoSD package, part of the new FastCaloSim (improved with the exclusion of the ParamAlg algorithm). Ten single photon simulation samples were generated for four different pseudorapidity ranges and two different energies. The first profiler investigated was Perfmon [152]. The final outputs for time spent per event can be seen in Table D.1 and part of the output plots on Figure D.1. A few minutes per event was already a significant improvement from the previous version of the simulation which had CPU with an order of magnitude bigger.

E \ η	0.2 - 0.25	1.00 - 1.05	2.00 - 2.05	3.00 - 3.05	4.00 - 4.05
1 TeV	2.6 min	3.0 min	11.2 min	2.9 min	0.7 min
4 TeV	11.3 min	14.0 min	14.8 min	12.1 min	8.6 min

Table D.1: Perfmon CPU time per event for photon generated samples. First line represents the pseudorapidities η of the generated samples used and first column their energies E.

Next, the StepinfoSD package was investigated in more detail in order to see whether a certain function or an algorithm takes a sufficiently more time than the rest of the code. Two profilers were tested for the purpose - Callgrind [153] and GPerfTools [154] [155].

Callgrind is a tool which uses the number of functions executed, the relationships between them, the caller/callee function relationship and the number of such calls. An example of the output for one of FastCaloSim's packages is

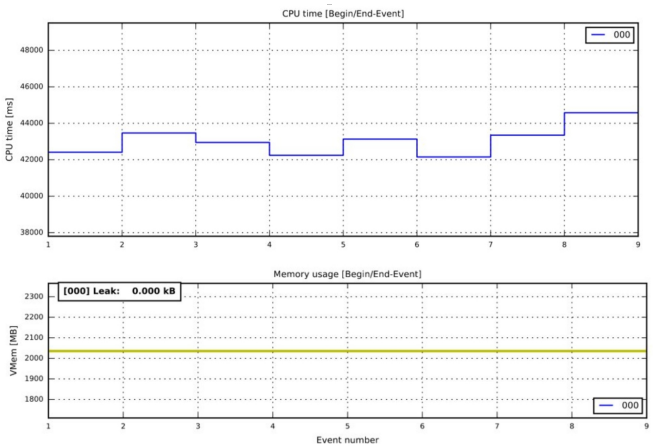


Figure D.1: Perfmon output for 1TeV photon, pseudorapidity 4 – 4.05. CPU time per event (top) and memory leakage (bottom).

shown on Figure D.2. The first percentage in each box represents the fraction of time this particular library took with respect to the whole software package in one athena (ATLAS' software framework) event and the second the fraction with respect to its caller. It took more than three hours for only one event and the profiling of only a certain part of the code appeared to be challenging for the profiler. Considering that the output from event to event can vary especially in the first few events due to library loading, this profiler was considered useful only if there is interest in the specific libraries and the time or memory they consume.

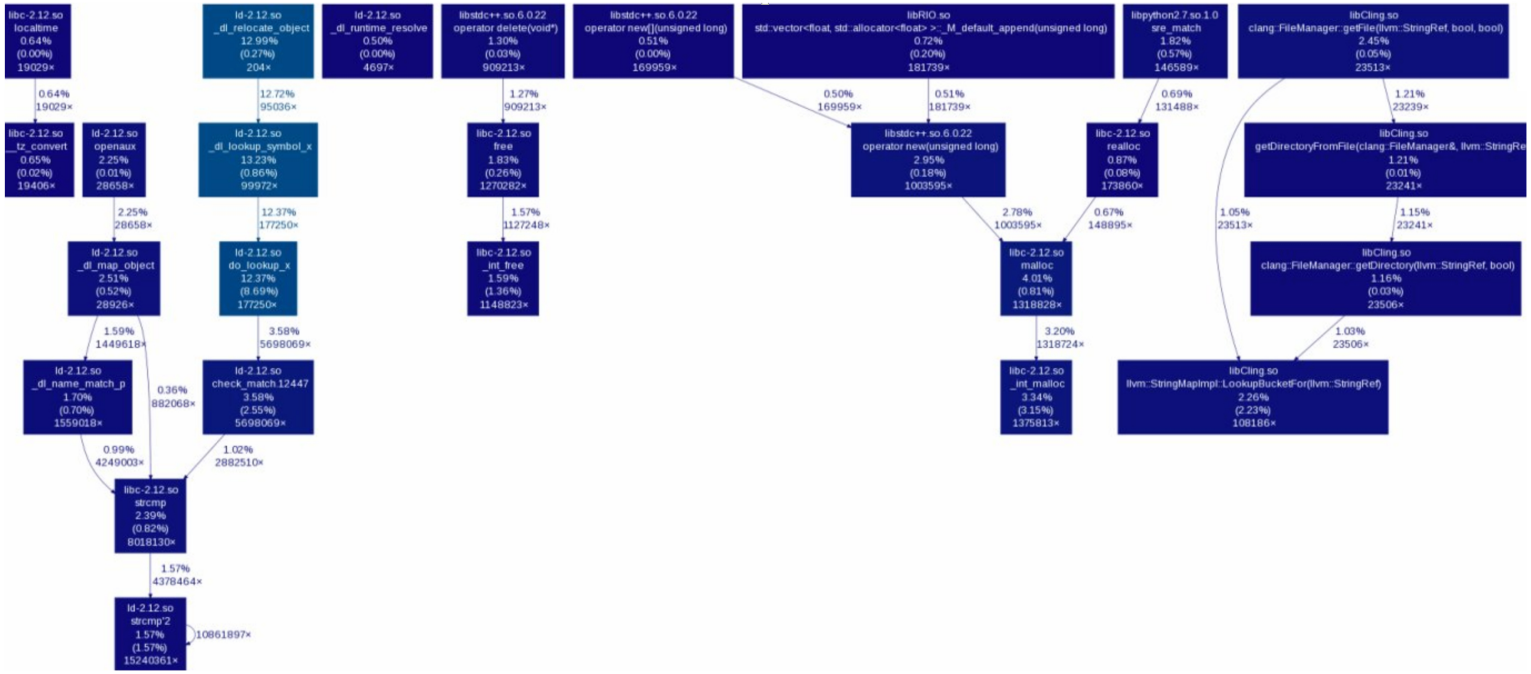


Figure D.2: Example for Callgrind Profiling.

The last and considered most useful profiler used was GPerfTools. Due to the possible worse performance of the first fifteen events, twenty events were run for each sample. An example of the output can be seen in Table D.2. The last column of the table represents the profiling samples in a particular function and its branches. There is a one to one correspondence of percentage to seconds. One profiling sample in a given function would correspond to spending ten ms in that function. It was noted that in all tested samples, the FCS-SteInfoSD::update-map class took the most time and a suggestion was made to the group for further optimization.

Finally, detailed instructions for how to run, use and understand the output of the three profilers were presented to the FastCalSim group for possible future investigations.

	1	2	3	4	5	6
1TeV, η: 3-3.05	FCS-SteInfoSD::update-map	3766	3.9%	30.4%	3820	4.0%
	LArFCS-StepInfoSD::ProcessHit	280	0.3%	67.2%	524	0.5%
	LArFCS-StepInfoSD::ConvertID	186	0.2%	75.4%	186	0.2%
1 TeV, η: 1-1.05	FCS-SteInfoSD::update-map	23927	10.7%	10.7%	24160	10.8%
	LArFCS-StepInfoSD::ProcessHit	1411	0.6%	38.5%	11974	5.4%
	LArFCS-StepInfoSD::ConvertID	1326	0.6%	39.7%	1326	0.6%

Table D.2: Example for GPerfTools output for two of the generated samples, where 1 is the name of the function/class, 2 - Number of profiling samples in this function, 3 - percentage of profiling samples in this function, 4 - percentage of profiling samples in the functions printed so far, 5 - number of profiling samples in this function and its branches and 6 - percentage of profiling samples in this function and its branches

Appendix E

Comparison in Signal/Background

In terms of classification power, the differences are obvious in the shape of the signal and background with respect to the classifier discriminant. The events, which are classified as signal or background with a high certainty lie around 0 or 1 and the discrimination is easier in the un-scaled case (Figure E.1), as compared to the scaled case (Figure E.2).

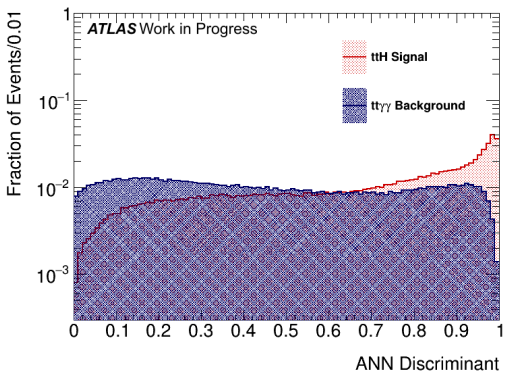


Figure E.1: Signal and Background with respect to the final ANN discriminant in MC $tt\gamma\gamma$ background and ttH signal events with un-scaled photon kinematic variables. $\lambda = 20$.

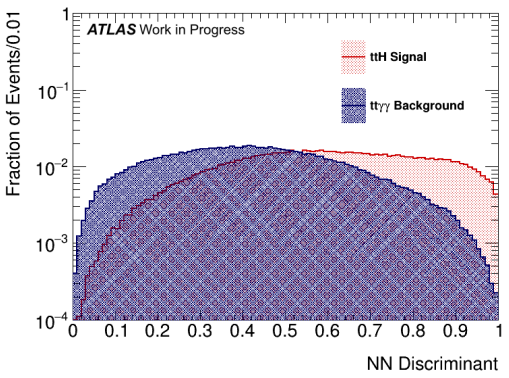


Figure E.2: Signal and Background with respect to the classifier discriminant in MC $tt\gamma\gamma$ background and ttH signal events with scaled photon kinematic input variables.

The separation between signal and background is slightly better in the un-scaled but overall similar to the scaled. (Figures E.3 and E.4).

The ANN discriminant distributions for signal and $tt\gamma\gamma$ background events are shown in Figure E.5, using $\lambda = 20$.

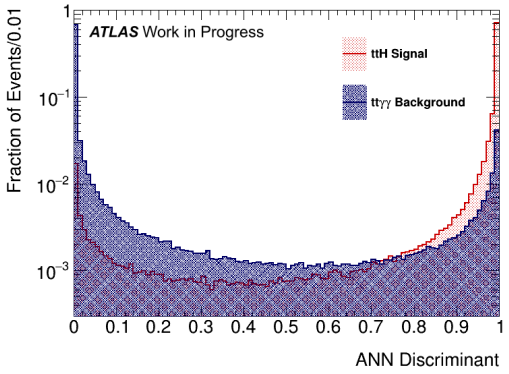


Figure E.3: Signal and Background with respect to the final ANN discriminant in NTNI real run 2 background and ttH signal events with un-scaled photon kinematic variables. $\lambda = 500$.

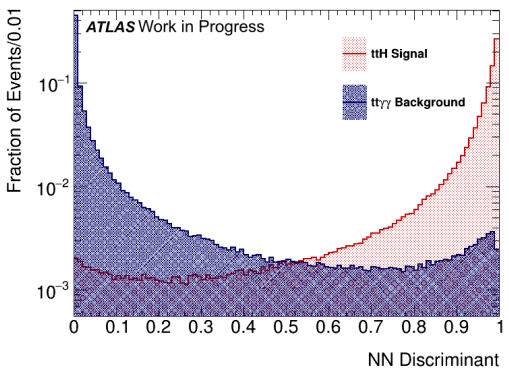
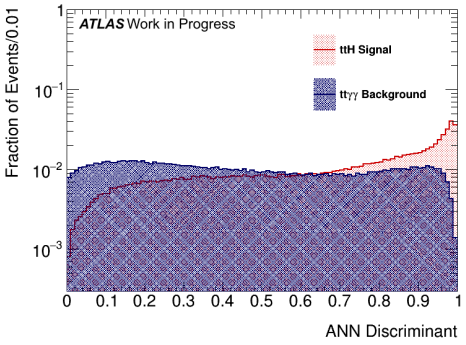


Figure E.4: Signal and Background with respect to the classifier discriminant in NTNI real run 2 background and ttH signal events with scaled photon kinematic input variables.



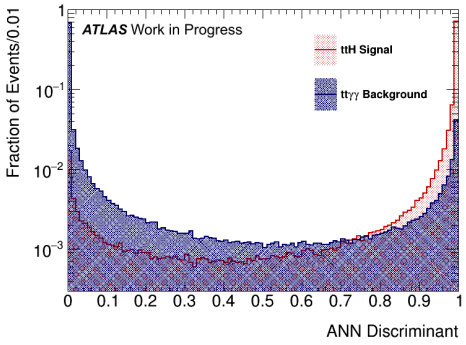


Figure E.5: ANN discriminant distributions for signal and background events in hadronic events. Background hypothesis: $tt\gamma\gamma$ MC (left), NTNI data (right). The regularization parameter is set to $\lambda = 20$ and $\lambda = 500$ in the left and right figures respectively.

Bibliography

[1] ATLAS Collaboration. AtlFast3: The next generation of fast simulation in ATLAS. *Comput. Softw. Big Sci.*, 6:7, 2022.

[2] Chase Shimmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Søgaard. Decorrelated Jet Substructure Tagging using Adversarial Neural Networks. *Phys. Rev. D*, 96(7):074034, 2017.

[3] Georges Aad et al. Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 81(6):537, 2021.

[4] G. L Kane. *Modern elementary particle physics.* Addison-Wesley, 1987.

[5] J. Sapirstein. Quantum electrodynamics. *Springer Handbook of Atomic, Molecular, and Optical Physics*, 2006.

[6] W. Marciano and H. Pagels. Quantum chromodynamics. *Physics Reports*, 36(3):137, 1978.

[7] M. Thomson. *Modern particle physics.* Cambridge University Press, 2013.

[8] Andreas Vogt, S Moch, and JAM1109 Vermaseren. The three-loop splitting functions in qcd: the singlet case. *Nuclear Physics B*, 691(1-2):129–181, 2004.

[9] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt. LHAPDF6: Parton density access in the LHC precision era. *Eur. Phys. J. C*, 75:132, 2015.

[10] J. C. Collins, D. E. Soper, and G. Sterman. Factorization of hard processes in QCD. In *Perturbative QCD*, page 1. World Scientific, 1989.

[11] T. Aaltonen et al. High-precision measurement of the W boson mass with the CDF II detector. *Science*, 376(6589):170–176, 2022.

[12] S. L. Glashow. The renormalizability of vector meson interactions. *Nucl. Phys.*, 10:107, 1959.

[13] A. Salam. Weak and electromagnetic interactions. In *Selected papers of Abdus Salam: (With commentary)*, page 244. World Scientific, 1994.

[14] D. J. Gross and R. Jackiw. Effect of anomalies on quasi–renormalizable theories. *Phys. Rev. D*, 6:477, 1972.

[15] ATLAS Collaboration. Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[16] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.

[17] I. van Vulpen. The standard model Higgs boson: Part of the lecture particle physics II. *UvA Particle Physics Master*, 2012, 2011.

[18] G. Bhattacharyya, D. Das, and P. B. Pal. Modified Higgs couplings and unitarity violation. *Phys. Rev. D*, 87:011702, 2013.

[19] ATLAS Collaboration. Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 98:052005, 2018.

[20] D. de Florian et al. Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector. *arXiv:1610.07922*.

[21] F. Halzen and A. D Martin. Quarks and leptons, 1984.

[22] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard model Higgs–boson branching ratios with uncertainties. *Eur. Phys. J. C*, 71:1753, 2011.

[23] M. Spira, A. Djouadi, D. Graudenz, and P. M. Zerwas. Higgs boson production at the LHC. *Nucl. Phys. B*, 453:17, 1995.

[24] M. Spira. Higgs boson production and decay at hadron colliders. *Prog. Part. Nucl. Phys.*, 95:98, 2017.

[25] J. M. Campbell and R. K. Ellis. Higgs constraints from vector boson fusion and scattering. *JHEP*, 04:030, 2015.

[26] S. Dittmaier et al. Handbook of LHC Higgs cross sections: 2. Differential distributions. *arXiv:1201.3084*.

[27] J. R. Andersen et al. Handbook of LHC Higgs cross sections: 3. Higgs properties. *arXiv:1307.1347*.

[28] CMS Collaboration. Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 79(5):421, 2019.

[29] ATLAS Collaboration. A combination of measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. ATLAS-CONF-2020-027.

[30] U. J. Saldaña Salazar. A principle for the Yukawa couplings. *J. Phys. Conf. Ser.*, 761(1):012064, 2016.

[31] U. J. Saldaña Salazar. The flavor–blind principle: A symmetrical approach to the Gatto–Sartori–Tonin relation. *Phys. Rev. D*, 93(1):013002, 2016.

[32] Y. Koide. Fermion–boson two–body model of quarks and leptons and Cabibbo mixing. *Lett. Nuovo Cim.*, 34:201, 1982.

[33] C. M. Becchi and G. Ridolfi. Breaking of accidental symmetries. In *An introduction to relativistic processes and the standard model of electroweak interactions*, page 95. Springer, 2006.

[34] C Patrignani, Particle Data Group, et al. Review of particle physics. *Chin. Phys. C*, 40(10):100001, 2016.

[35] ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett. B*, 784:173, 2018.

[36] CMS Collaboration. Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{\text{s}} = 13$ TeV. *JHEP*, 07:027, 2021.

[37] A. Montalbano. Measurement of higgs boson properties using the atlas detector. Technical Report ATL-COM-PHYS-2020-607.

[38] CMS Collaboration. Measurements of $t\bar{t}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel. *Phys. Rev. Lett.*, 125(6):061801, 2020.

[39] ATLAS Collaboration. Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 97(7):072016, 2018.

[40] ATLAS Collaboration. Measurement of the properties of higgs boson production at 13 tev in the channel using 139 fb^{-1} of collision data with the atlas experiment. tech. rep. ATLAS-CONF-2020-026, CERN, 2020,: https://cds. cern. ch/record.

[41] ATLAS Collaboration. Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4\ell$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 03:095, 2018.

[42] Morad Aaboud et al. Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 97(7):072016, 2018.

[43] ATLAS Collaboration. Measurement of the properties of Higgs boson production at $\sqrt{s} = 13$ TeV in the $H \to \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment. *arXiv:2207.00348*.

[44] CMS Collaboration. Measurement of $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel in 41.5 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV. CMS-PAS-HIG-18-030.

[45] ATLAS collaboration. Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector. In *12th International Workshop on Top Quark Physics, Beijing, China*, page 102, 2019.

[46] CMS Collaboration. Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81(4):378, 2021.

[47] M. Cepeda et al. Report from Working Group 2: Higgs Physics at the HL–LHC and HE–LHC. *CERN Yellow Rep. Monogr.*, 7:221, 2019.

[48] D. J. Fixsen. The temperature of the cosmic microwave background. *Astrophys. J.*, 707:916, 2009.

[49] R. Agnese et al. Search for low–mass weakly interacting massive particles with SuperCDMS. *Phys. Rev. Lett.*, 112(24):241302, 2014.

[50] L. D. Duffy and K. van Bibber. Axions as dark matter particles. *New J. Phys.*, 11:105008, 2009.

[51] G. Bertone and T. M. P. Tait. A new era in the search for dark matter. *Nature*, 562(7725):51, 2018.

[52] A. de Gouvea, D. Hernandez, and T. M. P. Tait. Criteria for natural hierarchies. *Phys. Rev. D*, 89(11):115005, 2014.

[53] E. J. Copeland, M. Sami, and S. Tsujikawa. Dynamics of dark energy. *Int. J. Mod. Phys. D*, 15:1753, 2006.

[54] R Aaij et al. First observation of CP violation in the decays of B_s^0 mesons. *Phys. Rev. Lett.*, 110(22):221601, 2013.

[55] Roel Aaij et al. Observation of CP Violation in Charm Decays. *Phys. Rev. Lett.*, 122(21):211803, 2019.

[56] H.-Y. Cheng and C.-W. Chiang. Revisiting CP violation in $D \to PP$ and VP decays. *Phys. Rev. D*, 100(9):093002, 2019.

[57] G. C. Branco, L. Lavoura, and J. P. Silva. *CP violation*. Number 103. Oxford University Press, 1999.

[58] I. I. Bigi and A. I. Sanda. *CP violation*, volume 9. Cambridge University Press, 9 2009.

[59] LHCb Collaboration. Test of lepton universality in beauty–quark decays. *Nature Phys.*, 18(3):277, 2022.

[60] L. Evans. The Large Hadron Collider (LHC). *New J. Phys.*, 9:335, 2007.

[61] ATLAS Collaboration. Operation of the ATLAS trigger system in Run 2. *JINST*, 15(10):P10004, 2020.

[62] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[63] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[64] LHCb Collaboration. The LHCb detector at the LHC. *JINST*, 3:S08005, 2008.

[65] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[66] C. Wiesner et al. LHC beam dump performance in view of the high luminosity upgrade. In *8th International Particle Accelerator Conference*, 5 2017.

[67] ATLAS Collaboration. Alignment of the ATLAS inner detector and its performance in 2012. ATLAS-CONF-2014-047.

[68] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV. *Eur. Phys. J. C*, 76(5):292, 2016.

[69] ATLAS Collaboration. The ATLAS simulation infrastructure. *Eur. Phys. J. C*, 70:823, 2010.

[70] F. Hugging. The ATLAS pixel detector. *IEEE Trans. Nucl. Sci.*, 53(3):1732, 2006.

[71] ATLAS Collaboration. Technical design report for the ATLAS inner tracker pixel detector. Technical Report CERN-LHCC-2017-021, ATLAS-TDR-030.

[72] B. Abbott et al. Production and integration of the ATLAS insertable B–layer. *JINST*, 13(05):T05008, 2018.

[73] *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. CERN, Geneva, 1997.

[74] B. Mindur. ATLAS transition radiation tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider. Technical Report ATL-INDET-PROC-2016-001.

[75] R. Wigmans. *Calorimetry: Energy measurement in particle physics*, volume 107. Oxford University Press, 2000.

[76] C. W. Fabjan and F. Gianotti. Calorimetry for particle physics. *Rev. Mod. Phys.*, 75:1243, 2003.

[77] M. Aharrouche et al. Response uniformity of the ATLAS liquid argon electromagnetic calorimeter. *Nucl. Instrum. Meth. A*, 582:429, 2007.

[78] ATLAS Collaboration. Readiness of the ATLAS tile calorimeter for LHC collisions. *Eur. Phys. J. C*, 70:1193, 2010.

[79] ATLAS Collaboration. Readiness of the ATLAS liquid argon calorimeter for LHC collisions. *Eur. Phys. J. C*, 70:723, 2010.

[80] W. Van Roosbroeck. Theory of the yield and Fano factor of electron–hole pairs generated in semiconductors by high–energy particles. *Phys. Rev.*, 139(5A):A1702, 1965.

[81] M Aleksa and M Diemoz. Discussion on the electromagnetic calorimeters of ATLAS and CMS. Technical report, CERN, Geneva, 2013.

[82] E. Diehl. Calibration and performance of the ATLAS muon spectrometer. In *Meeting of the APS Division of Particles and Fields*, 9 2011.

[83] G. Avoni et al. The new LUCID–2 detector for luminosity measurement and monitoring in ATLAS. *JINST*, 13(07):P07017, 2018.

[84] S. Ask et al. Luminosity measurement at ATLAS: Development, construction and test of scintillating fibre prototype detectors. *Nucl. Instrum. Meth. A*, 568:588, 2006.

[85] P. Jenni, M. Nessi, and M. Nordberg. Zero degree calorimeters for ATLAS. Technical Report CM-P00072881.

[86] P. Jenni, M. Nordberg, M. Nessi, and K. Jon-And. ATLAS forward detectors for measurement of elastic scattering and luminosity. Technical Report ATLAS-TDR-018.

[87] ATLAS Collaboration. Performance of the ATLAS trigger system in 2015. *Eur. Phys. J. C*, 77(5):317, 2017.

[88] D. et al. Adams. The ATLAS computing model. Technical Report ATL-SOFT-2004-007, ATL-COM-SOFT-2004-009, CERN-ATL-COM-SOFT-2004-009, CERN-LHCC-2004-037-G-085.

[89] C. Ay et al. Monte Carlo generators in ATLAS software. *J. Phys. Conf. Ser.*, 219:032001, 2010.

[90] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX. *JHEP*, 06:043, 2010.

[91] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with parton shower simulations: The POWHEG method. *JHEP*, 11:070, 2007.

[92] P. Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[93] R. D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[94] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015.

[95] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. Technical Report ATL-PHYS-PUB-2014-021.

[96] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152, 2001.

[97] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas. NLO QCD corrections to $t\bar{t}H$ production in hadron collisions. *Nucl. Phys. B*, 653:151, 2003.

[98] J. Alwall et al. The automated computation of tree–level and next–to–leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[99] ATLAS Collaboration. The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model. Technical Report ATL-PHYS-PUB-2016-017.

[100] Georges Aad et al. ATLAS data quality operations and performance for 2015–2018 data-taking. *JINST*, 15(04):P04003, 2020.

[101] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. Technical report, CERN, Geneva, 2019. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-021.

[102] P. Calafiura et al. The Athena control framework in production, new developments and lessons learned. In *14th International Conference on Computing in High-Energy and Nuclear Physics*, page 456, 2005.

[103] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[104] J. Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

[105] ATLAS Collaboration. ATLAS computing: Technical design report. Technical Report CERN-LHCC-2005-022, ATLAS-TRD-017.

[106] M. Dobbs and J. B. Hansen. The HepMC C++ Monte Carlo event record for high energy physics. *Comput. Phys. Commun.*, 134:41, 2001.

[107] A et al. Ribon. Status of geant4 hadronic physics for the simulation of lhc experiments at the start of lhc physics program. technical report cern-lcgapp-2010-02. *CERN-LCGAPP*, 2:2010, 2010.

[108] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist. A model for low p_T hadronic reactions, with generalizations to hadron–nucleus and nucleus–nucleus collisions. *Nucl. Phys. B*, 281:289, 1987.

[109] B. Andersson, A. Tai, and B.-H. Sa. Final state interactions in the (nuclear) FRITIOF string interaction scenario. *Z. Phys. C*, 70:499, 1996.

[110] B. Ganhuyag and V. Uzhinsky. Modified FRITIOF code: Negative charged particle production in high energy nucleus nucleus interactions. *Czech. J. Phys.*, 47:913, 1997.

[111] ATLAS Collaboration. The new Fast Calorimeter Simulation in ATLAS. Technical Report ATL-SOFT-PUB-2018-002.

[112] ATLAS Collaboration. Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment. *Eur. Phys. J. C*, 76(5):295, 2016.

[113] ATLAS Collaboration. Reconstruction, energy calibration, and identification of hadronically decaying tau leptons. ATLAS-CONF-2011-077.

[114] M. Hübner. Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using *pp* collisions at 13 TeV. Technical Report ATL-COM-PHYS-2018-1121.

[115] C. M. Bishop. *Pattern recognition and machine learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[116] Tanin Sirimongkolkasem and Reza Drikvandi. On regularisation methods for analysis of high dimensional data. *Annals of Data Science*, 6(4):737–763, 2019.

[117] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[118] A. F. Agarap. Deep learning using rectified linear units (ReLU). *arXiv:1803.08375*.

[119] H. Zheng et al. Improving deep neural networks using softplus units. In *2015 International Joint Conference on Neural Networks (IJCNN)*, page 1. IEEE, 2015.

[120] Keras library. `https://github.com/keras-team/keras`.

[121] M. Abadi et al. TensorFlow: Large–scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467*.

[122] G. Louppe, M. Kagan, and K. Cranmer. Learning to pivot with adversarial networks. *arXiv:1611.01046*.

[123] H. Ajakan et al. Domain–adversarial neural networks. *arXiv:1412.4446*.

[124] Y. Ganin et al. Domain–adversarial training of neural networks. 2016.

[125] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[126] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79, 03 1951.

[127] B. Fuglede and F. Topsoe. Jensen–Shannon divergence and Hilbert space embedding. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, page 31. IEEE, 2004.

[128] Carlo Oleari. The POWHEG-BOX. *Nucl. Phys. B Proc. Suppl.*, 205-206:36–41, 2010.

[129] 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. *Comput. Phys. Commun.*, 191:159–177, 2015.

[130] E. Bothmann et al. Event generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[131] J. Alwall and M. Herquet. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):1–40, 2011.

[132] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *JINST*, 14(12):P12006, 2019.

[133] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV. *Eur. Phys. J. C*, 76(5):292, 2016.

[134] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[135] ATLAS Collaboration. CP properties of Higgs boson interactions with top quarks in the $t\bar{t}H$ and tH processes using $H \to \gamma\gamma$ with the ATLAS detector. *Phys. Rev. Lett.*, 125(6):061802, 2020.

[136] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS detector. *Eur. Phys. J. C*, 77(7):466, 2017.

[137] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017.

[138] M. Cacciari, G. P. Salam, and G. Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[139] M. Cacciari, G. P. Salam, and G. Soyez. The anti–k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[140] ATLAS Collaboration. ATLAS b–jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019.

[141] ATLAS Collaboration. Performance of pile–up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 76(11):581, 2016.

[142] A. Hoecker et al. TMVA–toolkit for multivariate data analysis. *arXiv: physics/0703039*, 2007.

[143] ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV. *Eur. Phys. J. C*, 78(11):903, 2018.

[144] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[145] ATLAS Collaboration. Performance of mass–decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS. Technical Report ATL-PHYS-PUB-2018-014.

[146] L. Yang and A. Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295, 2020.

[147] D P Kingma and J Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[148] C Reid Turner, Alfonso Fuggetta, Luigi Lavazza, and Alexander L Wolf. A conceptual basis for feature engineering. *Journal of Systems and Software*, 49(1):3–15, 1999.

[149] Reducing the biases in machine learning algorithms for Higgs physics, Masters thesis, University of Edinburgh, 2022. ATLAS-CONF-2020-027.

[150] C. Englert, P. Galler, P. Harris, and M. Spannowsky. Machine learning uncertainties with adversarial neural networks. *Eur. Phys. J. C*, 79(1):4, 2019.

[151] P. Kaur, J. Stoltzfus, et al. Type i, ii, and iii statistical errors: A brief overview. *Int. J. Acad. Med.*, 3(2):268, 2017.

[152] R. Enbody. Perfmon: Performance monitoring tool. perfmon providing a functionally and logically consistent set of capabilities with a consistent hardware interface. *http://www.cps.msu.edu/ enbody/perfmon.html*.

[153] J. Weidendorfer. Sequential performance analysis with callgrind and kcachegrind. In *Tools for high performance computing*, page 93. Springer, 2008.

[154] ATLAS Collaboration. ATLAS offline software performance monitoring and optimization. *J. Phys. Conf. Ser.*, 513:052022, 2014.

[155] GPerfTools. http://code.google.com/p/gperftools/.